

Accepted Manuscript

A review of substances found positive in 1 of 3 in vitro tests for skin sensitization

Susanne N. Kolle, Andreas Natsch, G. Frank Gerberick, Robert Landsiedel



PII: S0273-2300(19)30144-8

DOI: <https://doi.org/10.1016/j.yrtph.2019.05.016>

Reference: YRTPH 4391

To appear in: *Regulatory Toxicology and Pharmacology*

Received Date: 19 February 2019

Revised Date: 15 April 2019

Accepted Date: 16 May 2019

Please cite this article as: Kolle, S.N., Natsch, A., Gerberick, G.F., Landsiedel, R., A review of substances found positive in 1 of 3 in vitro tests for skin sensitization, *Regulatory Toxicology and Pharmacology* (2019), doi: <https://doi.org/10.1016/j.yrtph.2019.05.016>.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

A Review of Substances Found Positive in 1 of 3 In Vitro Tests for Skin Sensitization

Susanne N. Kolle¹, Andreas Natsch², G. Frank Gerberick³ and Robert Landsiedel¹

¹BASF SE, Experimental Toxicology and Ecology, Ludwigshafen, Germany

²Givaudan Schweiz AG, Dübendorf, Switzerland

³GF3 Consultancy, LLC, West Chester, USA

Keywords: Skin sensitization; allergic contact dermatitis; defined approaches; hazard assessment; human reference data; LLNA reference data; adverse outcome pathway; 2 out of 3 approach; applicability domain; concordance

Abbreviations:

ACD, allergic contact dermatitis;

AOP, adverse outcome pathway;

3-AP, 3-Aminophenol;

ARE, antioxidant response element;

BA, Benzoic acid;

BB, Benzyl benzoate;

BBP, Butylbenzylphthalate;

BC, Benzyl cinnamate;

- 21 BP, Benzoyl peroxide;
- 22 BS, Benzyl salicylate;
- 23 BZA, Benzyl alcohol;
- 24 DAs, defined approaches;
- 25 DBP, Dibutyl phthalate;
- 26 DEP, Diethyl phthalate;
- 27 DPRA, direct peptide binding assay;
- 28 DT, N,N-Diethyl-m-toluamide;
- 29 EBA, Ethyl benzoylacetate;
- 30 EV, Ethyl vanillin;
- 31 EC3, estimated concentration that produces a stimulation index of 3 in the LLNA;
- 32 h-CLAT, human Cell Line Activation Test;
- 33 HMT, human maximization test;
- 34 HRIPT, human repeat insult patch test;
- 35 HS, Hexyl Salicylate;
- 36 ITS, integrated testing strategy
- 37 LLNA, local lymph node assay;
- 38 LOEL, lowest observed effect level;
- 39 6-MC, 6-Methylcoumarin;
- 40 MIO, α -iso-Methylionone;
- 41 MMP, 2-Methoxy-4-methylphenol;
- 42 mMUSST, modified myeloid U937 skin sensitization test;
- 43 NESIL, no expected sensitization induction level;

44 NOEL, no observed effect level; LOEL, lowest observed effect level;

45 OECD, Organization for Economic Cooperation and Development;

46 PA, Phthalic anhydride;

47 SI, stimulation index;

48 10-UDE, Undec-10-enal;

49 WoE, Weight of Evidence,

50

51

Abstract

There has been significant progress in recent years in the development and application of alternative methods for assessing the skin sensitization potential of chemicals. The pathways involved in skin sensitization have been described in an OECD adverse outcome pathway (AOP). To date, a single non-animal test method is not sufficient to address this AOP so numerous approaches involving the use of 2 or more assays are being evaluated for their performance. The 2 out of 3 approach (DPRA, KerationSensTM/LuSens and h-CLAT) is a simple approach that has demonstrated very good sensitivity, specificity and overall accuracy numbers for predicting the skin sensitization potential of chemicals. Chemicals with at least two positive results in tests addressing Key events 1 - 3 are predicted sensitizers, while chemicals with none or only one positive outcome are predicted non-sensitizers. In this report we have thoroughly reviewed the discordant results of 29 chemicals with 1 out of 3 positive results to understand better what led to the results observed and how this information might impact our hazard assessments of these chemicals. We initially categorized each chemical using a weight of evidence approach as positive, negative or indeterminate based on review of available human and animal data as well as what skin sensitization alerts were triggered using two versions of OECD Toolbox and DEREK Nexus. We determined that 4 of the 29 chemicals should be classified as indeterminate and not included in analysis of method performance based on insufficient, borderline and/or conflicting data to confidently categorized the chemicals as allergens or non-allergens. Of the 29 chemicals included in this analysis, 17 were classified as negative and would be correctly identified using a 2 out of 3 approach while 8 chemicals were classified as positive in vivo and would be false-negative with this approach. For some of these chemicals, the outcomes observed can be explained by *in vitro* borderline results (13 chemicals)

or in some instances there is mechanistic understanding of why a chemical is positive or negative in a particular assay (9 chemicals). Thus, when comparing the performance of different defined approaches, one should attempt to only include chemicals which demonstrate clear evidence to be categorized as allergens or non-allergens. Finally, when interpreting the results obtained for an individual unknown chemical it is critical that the *in vitro* skin sensitization data is reviewed critically and there is a good understanding of the variance and applicability domain limitations for each assay being used.

1. Introduction

Allergic contact dermatitis (ACD) continues to be an important clinically relevant outcome with an estimated 15-20% of the general population becoming sensitized over the course of their lives (Thyssen et al, 2007). Therefore, it is critical that there are robust test methods and defined approaches available to address this important toxicological endpoint. Over the past few decades, there has been significant progress in the development of alternative methods for assessing the skin sensitization potential of chemicals. There has also been progress in understanding what key events are being addressed with an individual assay. The key events involved in skin sensitization have been described by the Organization for Economic Co-operation and Development (OECD) in a document titled “The Adverse Outcome Pathway for Skin Sensitization Initiated by Covalent Binding to Proteins” which has been helpful in facilitating the development of methods and approaches (OECD, 2012). The four key mechanistic events covered in the adverse outcome pathway (AOP) include: (1) binding of haptens to endogenous proteins in the skin, (2) keratinocyte activation, (3) dendritic cell activation, and (4) proliferation

of antigen specific T cells. *In chemico* and *in vitro* methods addressing the first three mechanistic events of the skin sensitization AOP have been adopted by the OECD including the Direct Peptide Reactivity Assay (DPRA, OECD TG 442C) (OECD 2015), the ARE-Nrf2 Luciferase Test Method (KeratinoSensTM and LuSens, OECD TG 442D) (OECD 2018a) and the human Cell Line Activation Test (h-CLAT, OECD TG 442E, OECD 2018b). It is proposed that these methods, and others should be considered in the context of defined approaches (DA) to testing and assessment and in combination with other relevant complementary information if available, e.g., physical–chemical properties, information on other key events of the skin sensitization AOP as well as non-testing methods, including read-across from chemical analogues (Casati et al, 2018) in Integrated Approaches to Testing and Assessment (IATA).

Since none of the currently accepted *in vitro* methods are approved as stand-alone methods, test strategies for skin sensitization based on the AOP are preferred (Mehling et al, 2012; Rovida et al, 2015; Basketter et al, 2013; Casati et al, 2018; Kleinstreuer et al, 2018). There are numerous integrated approaches being developed and evaluated. A selection of DAs was submitted to the OECD (OECD 2016a, 2016b). The Cosmetics Europe skin tolerance task force and the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods partnered to produce a review of DAs in a transparent manner and evaluated them using a data set of 128 substances chosen based on the availability of both animal and human skin sensitization information (Hoffmann et al. 2018; Kleinstreuer et al, 2018) and an OECD expert working group has been established to set up a guideline to eventually facilitate the mutual acceptance of data of the DAs (OECD 2018c).

One simple approach has been to use a 2 out of 3 approach using OECD approved methods: DPRA, KeratinoSensTM or LuSens and h-CLAT, U937-CD86 and mMUSST (Bauch et al, 2012; Natsch et al, 2013; van der Veen et al, 2014; Urbisch et al, 2015). Placed in the context of the AOP, the DPRA evaluates key event 1 – the protein/peptide reactivity of a substance, the KeratinoSensTM and LuSens assays represent key event 2 and give a measure of keratinocyte activation, and the h-CLAT, U937-CD86 and mMUSST describe key event 3, namely dendritic cell activation. Thus, when used together, they cover the first three of the four key events of the sensitization AOP, thus being of mechanistic relevance and supporting the scientific rationale for using a combination of these methods in an AOP-based integrated testing strategy. Chemicals with at least two positive results in tests addressing Key events 1 - 3 are rated positive, while chemicals with none or only one positive outcome are rated negative.

Not surprising some chemicals are not compatible with specific assays so it becomes important to understand, as well as possible, how a chemical is behaving in each assay. Obviously, this becomes critical when one obtains discordant results between assays or there are results not supporting what is known of the how the chemicals behave, from a skin sensitization point of view, in human and/or animal models. Investigators have reported challenges due to solubility issues with materials that have high LogP values; molecules reacting with water in aqueous incubation media; surfactants that can disrupt membranes in cell-based assays and due to the need for metabolic activation/biotransformation of pre/prohaptens. However, using a LogP cut-off is no longer part of the 442D guideline and there is discussion to have it refined to be less restrictive in the 442E guideline. It is also important to note, regarding metabolic activity of cell-based assays, that it was shown that there is some metabolic activity in these systems (Fabian et al, 2013). Pre- and pro-haptens are also thought not to be a significant problem as theory suggest

for reactivity-based assays (Patlewicz et al, 2016; Urbisch et al, 2016). No doubt the future development of multi-cell 3D models will address some of these issues but there will still be a need with any future model to know its strengths and limitations for testing a diversity of chemistries. Thus, it is critical to develop a deep understanding of how chemicals perform in each individual assay so one can have confidence in accepting the results and using them for hazard identification and risk assessment purposes. Finally, it is important to understand that human categorization of a chemical's skin sensitization potency can be exceptionally challenging, especially for chemicals where there is limited sensitization data and/or exposure data. Efforts have been undertaken to categorize human sensitizers to help facilitate evaluation of DAs (Basketter et al. 2014; Api et al. 2017). The proposed categories for evaluating DAs included: category 1, extreme; category 2, strong; category 3, moderate; category 4, weak; and category 5 and 6, negative. The approach taken by Basketter *et al.* integrated human sensitization information from both predictive human tests (Human repeat insult patch test, HRIPT and Human Maximization test, HMT) and clinical data on frequency of positive patch tests (taking exposure into account). The latter is influenced not only by sensitizer potency but very much also by exposure. The approach described in Api *et al.* is mainly targeted at risk assessment to define a NESIL (No Expected Sensitization Induction Level). This is determined from a No Observed Effect Level (NOEL) in human predictive tests. Since many human tests have led to no cases of sensitization, they only report a NOEL and no LOEL (lowest observed effect level) which limits usefulness to determine sensitizer potency, although this approach is highly valuable for determining a NESIL in risk assessment. These notes of caution are important when evaluating alternative methods for their appropriateness to predict human sensitization potential and will be discussed here at the level of the individual chemicals.

Generally, performance statistics are used for evaluation of how well a particular defined approach is performing, especially when comparing one approach to another. This, of course, makes perfect sense but it is our view that it is also important to vet carefully the individual results to see if there are obvious scientific reasons or possibilities why a particular outcome was achieved. In this report, discordant predictions by *in vitro* tests in published data (Urbisch et al., 2015) are evaluated to see if on an individual chemical basis one can better understand what led to the discordant results and how it might impact the 2 out of 3 approach for hazard assessments which is based on the first three key events of the AOP of skin sensitization. Specifically, this manuscript reviews the data on 29 compounds with one positive and two negative results in the DPRA, KeratinoSens™/LuSens or h-CLAT/ U937-CD86 /mMUSST, i.e. chemicals rated negative by the 2 out of 3 approach in the data set (n=213) published by Urbisch et al., 2015. In a number of instances having 2 out of 3 negative results gave the correct prediction based on the expert review of the available *in vivo* data. Even in these cases where a correct prediction was obtained, the discordant results were evaluated to determine if it could be understood why it may have occurred. Where the expert review concluded the chemical is a true sensitizer and hence should be positive in the *in vitro* assays, the two discordant results were evaluated to understand why they gave negative results. Finally, of the 29 compounds, 5 were predicted to be indeterminate compounds based on the available *in vivo* data, for these compounds having borderline *in vitro* results might not be unexpected.

2. Materials and methods

2.1. Data set

2.1.1. The data set consist of 29 materials previously evaluated according 442C, 442D and 442E guidelines and yielding 1 out of 3 positive results in tests addressing key events 1, 2 and 3 in the AOP for skin sensitization (Urbisch et al, 2015), they are a subset of the 213 chemicals in the Urbisch et al. 2015 database and present 14% of all chemicals assessed therein. For each substance, the proposed reaction mechanisms/alerts (OECD Toolbox profilers based on OECD and OASIS algorithms and Derek Nexus; Lhasa Limited; Leeds UK) and test data (compiled from Urbisch et al., 2015) are listed in Tables 1 and 2, respectively.

2.2. Non-animal test methods

2.2.1. *Direct peptide reactivity assay.* The Direct Peptide Reactivity Assay (DPRA) was developed by (Gerberick et al., 2004; Gerberick et al., 2007). The DPRA has been formally validated and the OECD Testing Guideline TG 442C (OECD, 2015) was adopted in 2015. Briefly, the DPRA measures the depletions of two model peptides containing a cysteine- or lysine- residue as a reactive nucleophilic center after they are incubated with a test substance. The classification threshold for the DPRA is the mean depletion of 6.38% of the two peptides compared to the depletion in the reference controls (OECD, 2015). If the mean lysine- and cysteine- peptide depletion is above this threshold, a test substance is considered to be peptide reactive. Depending on the regulatory framework a positive result of the DPRA can serve as standalone information for classifying substances into Category 1 for skin sensitization. However, as also emphasized in the ECHA Guidance on information

requirements and Chemical Safety Assessment Chapter R.4a (ECHA 2016) the DPRA should not be used in isolation for identifying a skin sensitizer or non-sensitizer. In the DPRA a test substance was considered borderline if the mean peptide depletion observed was between 4.86 and 7.90% (Leontaridou et al., 2017).

2.2.2. *ARE-Nrf2 luciferase method.* The ARE-Nrf2 luciferase method utilizes the gene induction regulated by the antioxidant response element (ARE) in transgenic human keratinocyte cell lines.) The ARE-Nrf2 luciferase method is covered by KeratinoSensTM (Natsch et al., 2011) and LuSens (Ramirez et al., 2014). In ARE-Nrf2 luciferase methods the keratinocyte activating potential is determined by measuring Nrf2-induced luciferase activity after treatment with a test substance treatment relative to concurrent vehicle controls (OECD Test Guideline TG 442D, 2018a). A statistically significant fold induction (FI) of the luciferase activity above 1.50 is considered to indicate a keratinocyte activating potential of a test substance. The classification threshold for both assays is FI = 1.50, above which a substance is considered to have a keratinocyte activating potential. Similar to the DPRA, the ARE-Nrf2 luciferase method result in isolation is not considered suitable for classifying substances as skin sensitizers or non-sensitizers if used in isolation (ECHA 2016). In the LuSens a test substance was considered borderline if the FI observed was between 1.27-1.73 (Leontaridou et al., 2017).

2.2.3. *human Cell Line Activation Test.* The human Cell Line Activation Test (h-CLAT) (Ashikaga et al., 2006; Ashikaga et al., 2010; Sakaguchi et al., 2006; Sakaguchi et al., 2010) determines the dendritic cell activating potential by measuring the induction of the expression of the cell surface markers CD54 and CD86 after

treatment with a test substance relative to concurrent vehicle controls in sensitization human monocytic leukemia THP-1 cells as a surrogate of DCs. As indicated in the OECD testing guideline TG 442E (OECD, 2018b) a two-fold induction of the CD54 expression and/or 1.50-fold induction of CD86 expression at relative cell viabilities of at least 50% is considered to indicate a dendritic cell activating potential of a test substance. The classification thresholds for the h-CLAT are CD54 FI = 1.50 and CD86 FI = 2.00. Like for the DPRA, KeratinoSensTM and LuSens, the method should not be used in isolation for classifying skin sensitization potential (ECHA 2016). In the h-CLAT a test substance was considered borderline if the fold induction of CD54 expression observed was between 1.81-2.19 and or the fold induction for CD86 expression was between 1.24- 1.76(Leontaridou et al., 2017).

2.2.4. *U937-CD86 and mMUSST*. The U937-CD86 was conducted as described in Natsch et al., 2013. Briefly U937 cells were obtained from the American Type Culture Collection (ATCC, Rockville, MD, USA) and were cultured and treated in a manner similar to that described by Ade et al. (2006). Substances that induced CD86 expression more than 1.5-fold compared to the vehicle controls in the U937-CD86 assay were predicted to have a DC activating potential.

The modified MUSST (mMUSST) was conducted as described in Bauch et al., 2012. For the mMUSST U937 cells were obtained from the German Resource Center for Biological Material (DMSZ, Braunschweig, Germany). Substances that induced CD86 expression more than 1.2-fold compared to the vehicle controls in mMUSST assay were predicted to have a DC activating potential.

2.2.5. *The 2-out-of-3 approach for sensitization skin sensitization potential.* The 2-out-of-3 approach (Bauch et al., 2012; Urbisch et al., 2015; Sauer et al. 2016; OECD 2016 a and b) is an integrated testing strategy for the assessment of skin sensitization potential. According to this approach, 2 out of 3 concordant test results using the DPRA, ARE-NrF2 luciferase method, and the h-CLAT determine the prediction. The ARE-NrF2 luciferase method can be covered by LuSens or KeratinoSensTM. The 2-out-of-3 approach addresses the first three consecutive key events of the AOP for skin sensitization and it is a selected case study for integrated approaches to testing and assessment (IATA) (Bauch et al., 2012 and Urbisch et al., 2015).

2.2.6. *Structural alerts.* In silico analysis performed with the OECD QSAR Toolbox version 3.2 was published in Urbisch et al., 2015. Here we update that analysis by the use of a more recent OECD QSAR Toolbox version 4.2. Further, structural alerts were also generated with Derek Nexus: 6.0.1 Nexus: 2.2.1 (Lhasa Limited).

2.3. *Approach for Evaluation of Discordant Results.* Each material was carefully vetted to understand its *in vivo* sensitization potential and performance in each *in vitro* assay. Expert review of available *in vivo* and clinical data was used to determine the correct outcome expected from *in vitro* skin sensitization assays: positive (8), negative (16) or indeterminate (5). For the indeterminate compounds, it was our expert opinion that these compounds are borderline in their potential to cause ACD and therefore mixed results would be expected using different *in vitro* skin sensitization test methods. The *in vitro* data for each compound was reviewed to determine if the observed responses were borderline or clear-cut positive/negative. A determination was also made on whether

there was a discordant result within the same key event in case multiple results were available. Finally, determinations were made on whether sufficient data existed to support a discordant compound being out of applicability domain for a technical of mechanistic reason.

3. Results

3.1. *DPRA+/KeratinoSens-/h-CLAT- (n=4)*. The AOP for skin sensitization is initiated by protein binding. Therefore, any chemical which is a true positive in the DPRA should also be active in the cell-based assays. There are however a few mechanisms known by which the DPRA can give a “false positive” result. Examples are formation of peptide dimers (Natsch et al, 2007), oxidation of the cysteine group to a sulfoxide or possibly non-covalent peptide binding (absorption). Also, a false-positive prediction is suspected in the DPRA if a substance lacks a structural alert associated with protein binding reactivity. In contrast, cellular detoxification may explain why an effect was only observed in the cell-free test. With both cellular responses missing, the test substance would not be expected to drive the AOP towards a skin sensitization response.

3.1.1. *Phthalic anhydride* - The skin sensitization potential of phthalic anhydride (PA) in humans is not clear cut. It is, however, clearly established as a respiratory sensitizer (Dearman et al, 2013). Basketter and Kimber (2016) report that there is a lack of evidence to substantiate PA as a contact allergen that can cause allergic contact dermatitis in humans. The key question to ask is if this view is skewed because there is very limited diagnostic patch test data available (evidence of

absence or absence of evidence). It is certainly accepted that other respiratory allergens, like glutaraldehyde, can cause skin allergy (Dearman et al, 2013). It is well known that PA and other respiratory allergens elicit positive responses in skin test methods including mice and guinea pigs (Basketter and Kimber, 2016). It is interesting to speculate whether PA would induce and elicit skin sensitization in humans if tested using a standard repeat insult patch test protocol or human maximization protocol. Despite limited human data, from the available evidence, one would expect PA to be identified as positive using *in vitro* skin sensitization test methods. In support of this view, PA generates an acylation alert for binding to proteins which is a common mechanism by which skin sensitizers cause allergy (Roberts et al, 2007; OECD QSAR Toolbox) and generates a cyclic acid anhydride, azlactone or analogue alert in Derek Nexus (Table 1).

PA was positive in the DPRA but negative in h-CLAT, U-SENSTM, LuSens and KeratinoSensTM (Urbisch et al., 2015, Piroird et al., 2015). If one considers PA to be a human skin allergen, then the 2 of 3 approach is a misclassification. One explanation could be that in the aqueous culture systems the acid anhydrides undergo rapid hydrolysis prior to reaching cellular molecular targets which could lead to false negatives (Narita et al, 2017). The KeratinoSensTM/LuSens, like the DPRA, are methods that are based upon measurement of the inherent electrophilic activity of chemicals but acylating agents are difficult to detect in the ARE-Nrf2 reporter assays (probably because acylating agents primarily target lysine residues, as observed by predominant Lys-reactivity in the DPRA). It is also important to note

that other acid anhydrides are positive in the h-CLAT while PA was not (Nukada et al, 2013) but has been described to be difficult to dissolve in both DMSO and culture medium and hence maybe could not be tested at a sufficiently high concentration (Takenouchi et al., 2013). PA is Natsch et al (2015) reported that PA led to complete peptide depletion and adduct formation at both Lys-residues of the Cor1-420C peptide and reported PA to be positive in a U937 cell-based assay. Thus, using the 2 out of 3 approach, PA would be classified incorrectly as a non-sensitizer based on the negative KeratinoSens and h-CLAT results (Urbisch et al, 2015). Thus, the KeratinoSens/LuSens reactivity key event results are discordant with the DPRA and there appears to be a technical applicability issue with the h-CLAT and KeratinoSens yielding negative results for PA (Table 2).

3.1.2. *Benzoyl peroxide (BP)* – BP is a well-known skin allergen in humans and is assigned Category 3, a common cause of contact allergy (Basketter et al, 2014). The categorization is supported by clinical diagnostic patch test data which is often associated with acne medicine. There are multiple positive LLNA studies with BP giving EC3 values of 0.0044-0.3% (references available in Hoffman et al, 2018).

BP generates an acylation alert for binding to proteins and gives an acyl peroxide radical generation and acylation alerts with Derek Nexus (Takenouchi et al, 2015; OECD QSAR Toolbox). Although positive in the DPRA, BP is negative in the KeratinoSens, LuSens (negative but close to the cut-off value) and h-CLAT assays (Urbisch et al, 2015) and has been described to rapidly hydrolyze (Nacht et al.,

1981). Besides the potential shortcoming of the ARE-Nrf2 assays in detection of acylating agents, another explanation is that the albumin in the fetal bovine serum used in these assays might be acting as a peroxidase to reduce BP's activity (Pirisino et al, 1988). However, the positive results from the U937-CD86 and U-SENSTM assays are not consistent with this explanation (Natsch et al. 2013, and Piroird et al., 2015). BP gives a positive (moderate) response in the SENS-IS and SensCeeTox assays but is non-sensitizing in the IL-18 epiCS[®] and IL-18 EpiDerm assays (Hoffman et al, 2018 and Mehling et al., 2019). Noteworthy in the IL-18 EpiDerm assay, BP was correctly identified when tested from a fresh non-blinded stock was assessed supporting the hydrolysis hypothesis by Nacht et al., 1981. Using the 2 out of 3 approach, BP would be classified incorrectly as a non-sensitizer based on the negative KeratinoSens and h-CLAT results (Urbisch et al, 2015). It is important to note that the KeratinoSens and LuSens give a borderline response and there exist the possibility of a mechanistic applicability issue with the both the KeratinoSens (Table 2).

3.1.3. *Benzoic acid (BA)* - BA is a non-sensitizer based on the review of available animal and human data. It is negative in the LLNA and does not generate any skin sensitization reactivity alerts using OECD QSAR Toolbox and Derek Nexus (Urbisch et al, 2015). Benzoic acid is a well-known contact urticariant and skin irritant but there exists very little data supporting BA as a contact allergen (Frosch, 2006). Although not included in Basketter's human categorization paper (2014), the available data would support it being included in Category 6 or at worst Category 5

based on only a few rare clinical reports of contact allergy. Thus, it would be expected that BA should generate negative responses in *in vitro* skin sensitization assays.

Since BA does not contain any structural elements associated with reactivity, it should not cause peptide depletion in the DPRA (Table 1). Thus, the BA response in the DPRA should be considered to be a clear false positive. BA was negative in the KeratinoSens/LuSens and h-CLAT and U-SENSTM assays and thus gave the correct prediction using the 2 out of 3 approach (Urbisch et al, 2015, Piroird et al., 2015). The positive response with the DPRA is considered to be a technical applicability issue (Table 2).

3.1.4. *N,N*-Diethyl-*m*-toluamide (DT) - DT is placed in Category 6 based on lack of evidence associating the material as a skin allergen (Basketter et al, 2014). DT is widely used as an insect repellent with little to no associations with causing allergic contact dermatitis in dermatology clinics. DT is negative in the LLNA when tested up to 60% (Kern et al, 2010). It is reported to have no alerts (Urbisch et al, 2015; OECD QSAR Toolbox) and to be negative using Derek Nexus (Takenouchi et al, 2015). Based on the totality of the available data one would expect DT to be negative in *in vitro* skin sensitization assays.

Since DT does not contain any structural elements associated with reactivity, it should not cause any peptide depletion in the DPRA. Thus, the DT response in the

DPRA (mean depletion of 6.4%) should be considered to be a borderline false positive. Interestingly, DT is reported as being negative in the DPRA in an independent study (Takenouchi et al, 2015) and was borderline in Urbisch et al (2015) paper. Using the 2 out of 3 approach, DT would be classified correctly as a non-sensitizer based on the negative KeratinoSens and h-CLAT results (Urbisch et al, 2015) (Table 2).

3.2. *DPRA-/KeratinoSens+/h-CLAT-* ($n=12$) - The activation of the ARE-Nrf2 pathway in the KeratinoSens/LuSens assay by this group of chemicals may be unrelated to the skin sensitization process (Nguyen et al., 2009), or may not be sufficient by itself. For most chemicals reviewed here, there is no indication that the predictions of the DPRA and the h-CLAT assays might be false negative ones. For some of the compounds, the lack of peptide reactivity is consistent with the lack of structural alerts. With antigen formation and dendritic cell activation missing, the test substance would not be expected to drive the AOP towards a skin sensitization response. Of course, this assumes that no critical metabolites are being generated in KeratinoSens/LuSens assays that are not formed in the h-CLAT.

3.2.1. *4-Methoxyacetophenone (Acetanisole)* - Acetanisole is considered to be a non-sensitizer. In humans it has been tested up to 4138 $\mu\text{g}/\text{cm}^2$ in the human maximization protocol without any responses (Gerberick et al, 2001). It is reported to be negative in the LLNA when tested up to 50% (Gerberick et al, 2005). Acetanisole does not generate any skin sensitization alerts and is thought to be non-

protein binding (Patlewicz et al, 2008; Urbisch et al, 2015; OECD QSAR Toolbox; Derek Nexus). However, Takenouchi (2015) lists acetanisol as positive using Derek Nexus analysis but they do not classify it as sensitizer based on their ITS approach. Based on the totality of the available data one would expect acetanisol to be negative in *in vitro* skin sensitization assays.

It is not clear why acetanisol is clearly positive in both the KeratinoSens and LuSens. It has also been reported to be positive in the U937 assay (Natsch et al, 2013) but negative in the mMUSST (Bauch, et al, 2012) so it seems less likely that the test material is interfering with different cell-based assays. Importantly, the correct prediction was obtained using the 2 out of 3 approach with negative DPRA and h-CLAT data (Urbisch et al, 2015). Thus, there appears to be a technical applicability issue with the KeratinoSens (Table 2).

3.2.2. *6-Methylcoumarin* (6-MC) - There is limited data to robustly categorize 6-MC as a human sensitizer or non-sensitizer (Urbisch et al, 2015). Structurally related coumarin is listed in Category 3 which is defined as a common cause of contact allergy, perhaps due to higher exposures. 6-Methyl coumarin is known to be a photoallergen in humans. However, there are few positive diagnostic patch test data reported for 6-MC. LLNA data supports 6-MC, when tested up 25%, as being a non-sensitizer (Gerberick et al, 2005). 6-MC generates an alpha, beta-unsaturated ester or precursor skin sensitization alert with Derek Nexus while giving an acylating alert and no alert using OECD algorithms and OASIS algorithms (OECD

441 QSAR Toolbox), respectively (Table 1). Interestingly, only a single peptide was
442 modified by 6-MC in keratinocyte cell lysates and no 6-MC modifications were
443 observed in the skin lysates (Parkinson et al, 2018). Overall, the data on 6-MC is
444 mixed and it would be a borderline compound with leaning towards being a weak
445 human skin allergen. Thus, one could expect positive or negative findings using *in*
446 *vitro* skin sensitization assays.

447
448 6-MC is negative in h-CLAT, U937-CD86, mMUSST and U-SENSTM and DPRA
449 but clearly positive in KeratinoSens and LuSens (despite the acylating alert from the
450 computational tools) (Natsch et al, 2013; Bauch et al, 2012; Nukada et al, 2012;
451 Urbisch et al, 2015, Piroird et al., 2015). Using the 2 out of 3 approach, 6-MC
452 would be classified as a non-sensitizer based on the negative DPRA and h-CLAT
453 results (Urbisch et al, 2015). There does not appear to be any chemical or biological
454 reasons why 6-MC is giving different results in the various *in vitro* assays other than
455 6-MC being a borderline skin sensitizer, thus getting mixed results would not be
456 unexpected (Table 2).

457
458 3.2.3. *Tween 80* - Tween 80 is considered to be a non-sensitizer. It is listed in Category
459 6, a substance essentially absent of convincing evidence of contact allergy
460 (Basketter et al, 2014). In support, Tween 80 does not generate any skin
461 sensitization alerts (OECD QSAR Toolbox; Derek Nexus) (Table 2).

Tween 80 is negative in the h-CLAT and borderline negative DPRA (mean peptide depletion 5.07%) but clearly positive in KeratinoSens and LuSens (Nukada et al, 2012; Urbisch et al, 2015) but reported positive (mean peptide depletion 31.0%) in the DPRA in Hoffmann et al., 2018 and the U-SENSTM (Piroird et al., 2015). Tween 80 is non-sensitizing in SENS-IS, the IL-18 epiCS[®] and IL-18 EpiDerm but an extreme sensitizer in the SensCeeTox (Hoffman et al, 2018 and Mehling et al, 2019). The sensitizer-induced luciferase signal in the KeratinoSensTM cells is completely dependent on Nrf2. The same holds true for the luciferase induction observed for Tween 80, indicating that the false positive result is not due to recruitment of an alternative transcription factor (Emter et al, 2013). The authors hypothesize two explanations. First, Tween 80 could be activating the formation of endogenous Nrf2-inducing compounds, such as 4-hydroxynonenal formed by lipid peroxidation or the messenger H₂O₂. Or, it may trigger a signaling cascade that results in Nrf2-activation, which is not directly related to electrophilic stress. However, using the 2 out of 3 approach, Tween 80 would be classified correctly as a non-sensitizer based on the negative DPRA and h-CLAT results (Urbisch et al, 2015). Thus, Tween 80 gives a borderline negative DPRA result and mechanistic applicability domain issue for the KeratinoSens and LuSens assays (Table 2).

3.2.4. *Coumarin* - Coumarin is reported to be a known human allergen and is listed as Category 3, a common cause of contact allergy (Basketter et al, 2014). It is one of the 26 fragrances currently requiring individual labeling (SCCNFP Opinion, 2012). It has a reported human HRIPT NOEL of 3543 µg/cm² and HMT NOEL of 5517

486 $\mu\text{g}/\text{cm}^2$. A LOEL of 8858 $\mu\text{g}/\text{cm}^2$ is also reported. Positive patch results are rarely
487 (0.3% of 1855 ACD patients) reported for coumarin (Frosch et al, 2006). In
488 contrast, 10% pure coumarin was very well tolerated in humans (Vocansons et al.,
489 2006). However, there is a negative LLNA result (Gerberick et al, 2005) as well as
490 positive LLNA results using commercial grade coumarins due to contaminants
491 while pure coumarin was reported negative (Vocanson et al, 2006). Coumarin
492 generates an alpha, beta-unsaturated ester or precursor skin sensitization alert with
493 Derek Nexus while giving an acylating alert and no alert using OECD algorithms
494 and OASIS algorithms (OECD QSAR Toolbox), respectively (Urbisch et al, 2015).
495 However, Takenouchi et al (2015) lists a negative entry for DEREK. One reason the
496 for the mixed results on coumarin might be due to the reported contaminants that
497 have been found in coumarin samples and thought to be responsible for the positive
498 results observed in mice and humans (Vocanson et al, 2006). Thus, it is difficult to
499 determine with confidence how one would expect coumarin to behave in *in vitro*
500 skin sensitization assays.

501
502 Coumarin is negative in h-CLAT, U937-CD86, and DPRA but clearly positive in
503 the KeratinoSens (Natsch et al, 2013; Nukada et al, 2012; Urbisch et al, 2015)
504 (despite the acylating alert from the computational tools). Coumarin gives a positive
505 response in the U-SENSTM (Piroird et al., 2015) and (weak) response in the SENS-
506 IS assay (Hoffman et al, 2018). As mentioned, not knowing the purity of samples
507 tested in the different assays, it is difficult to know if KeratinoSens is giving the
508 correct or incorrect response (Table 2).

3.2.5. *Ethyl vanillin (EV)* – There is no category given for EV in Basketter's (2014) human categorization paper but the structurally related vanillin is placed in Category 5, a rare cause of contact allergy. EV is negative in the LLNA when tested up to 50% (Basketter et al, 2001). The SCCNFP opinion on fragrance materials (2012) states that the human evidence of EV being a contact allergen is limited. In support of EV being a weak allergen is that it generates a Schiff base/quinone precursor alert and is listed as a pre/pro hapten (Urbisch et al, 2015). EV is also listed as giving a positive result in Derek (Nukada et al, 2013). However, EV does not generate any alerts using the OECD QSAR Toolbox or with Derek Nexus. Based on the totality of the available data one would expect EV to be borderline in *in vitro* skin sensitization assays.

Ethyl vanillin is negative in h-CLAT, U937-CD86, U-SENSTM and borderline negative in the DPRA but positive in the KeratinoSens (borderline negative in Natsch et al, 2013; positive in Bauch et al., 2012; Nukada et al, 2012; borderline positive in LuSens Urbisch et al, 2015, Piroird et al., 2015). Takenouchi (2015) reports EV as positive in the DPRA and classifies it as a weak allergen using their ITS approach. EV was not tested in SENS-IS but vanillin is non-sensitizing in the assay (Hoffman et al, 2015). If one accepts EV being a weak human allergen, then the most likely explanation for its misclassification using the 2 out of 3 approach is that EV was not biotransformed from a pre/pro-hapten into a reactive chemical in the one DPRA and cell-based assays, h-CLAT and U937-CD86 assays. Of course,

if one assumes EV is not a significant human allergen, then the 2 out of 3 approach result is correct in its prediction of non-sensitizing. Although EV gives borderline response in various *in vitro* assays, the *in vivo* data suggest it should be considered a borderline compound (Table 2).

3.2.6. *Ethyl benzoylacetate* - Ethyl benzoylacetate (EBA) is negative in the LLNA (Gerberick et al, 2005) and there is little to no human patch test data available to assess its skin allergy potential. EBA does not trigger any skin sensitization alerts (Urbisch et al, 2015; Takenouchi et al, 2015; Nukada et al, 2013; OECD QSAR Toolbox; Derek Nexus) (Table 1). Although limited information is available for EBA, it should be classified as a non-sensitizer in humans.

Ethyl benzoylacetate is negative in h-CLAT, mMUSST, U937-CD86, U-SENSTM, borderline negative in the DPRA and borderline positive in the KeratinoSens (Natsch et al, 2013; Nukada et al, 2012; Urbisch et al, 2015, Piroird et al., 2015). Since EBA does not contain any structural elements associated with reactivity, it should not be positive in the reactivity-based assay. There is no apparent chemical or biological reason of why EBA would give a false positive response in the KeratinoSens assay. However, using the 2 out of 3 approach, EBA would be classified correctly as a non-sensitizer based on the borderline negative DPRA and h-CLAT results (Urbisch et al, 2015). Thus, the overall classification of EBA is dependent on the borderline results observed (Table 2).

3.2.7. *Dibutyl phthalate (DBP)* - DBP is not evaluated in Basketter's (2014) human categorization paper but the structurally related diethyl phthalate is, and it is placed in Category 6, a substance that is essentially absent of convincing evidence of contact allergy. DBP is negative in the LLNA when tested up to 50% (Kern et al, 2010). In support, DBP does not trigger any skin sensitization alerts (Urbisch et al, 2015; OECD QSAR Toolbox; Derek Nexus) and there are very few reported cases of patch test positive results referenced in the literature (Frosch, 2006). Although limited information is available for DBP, it should be classified as a non-sensitizer in humans.

Dibutyl phthalate is negative in h-CLAT and DPRA but clearly positive in the KeratinoSens (Urbisch et al, 2015; Takenouchi et al, 2015). Natsch and Emter (2008) reported that diethyl phthalate was repeatedly positive in the ARE-based assay but only at very high concentrations. DBP was not tested in the SENS-IS assay but diethyl phthalate was categorized as non-sensitizing in the assay (Hoffman et al, 2015). Since DBP does not contain any structural elements associated with reactivity, it should not be positive in the KeratinoSens reactivity-based assay. There is no apparent chemical or biological reason of why DBP would give a false positive response in the KeratinoSens assay other than a possible mechanistic applicability domain issue (Table 2). However, using the 2 out of 3 approach, DBP would be classified correctly as a non-sensitizer based on the negative DPRA and h-CLAT results (Urbisch et al, 2015).

3.2.8. *Butylbenzylphthalate (BBP)* - BBP is negative in the LLNA when tested up to 50% (Kern et al, 2010). There is little to no human patch test data available to assess its skin allergy potential. However, BBP does trigger a skin sensitization reactivity alert, a substance reacting by nucleophilic substitution by SN2 mechanism using OECD algorithms in the OECD QSAR Toolbox, but no alert when using OASIS algorithms (Urbisch et al, 2015; Takenouchi et al, 2015). The SN2 alert is also generated with Derek Nexus (Table 1). Although limited information is available, BBP should be classified as a non-sensitizer and give negative results when using *in vitro* assays.

Butylbenzylphthalate is negative in h-CLAT and DPRA but clearly positive in the KeratinoSens (Urbisch et al, 2015; Takenouchi et al, 2015). Since BBP does contain structural elements associated with reactivity, it is not surprising that it could be positive in the KeratinoSens reactivity-based assay. Using the 2 out of 3 approach, BBP would be classified correctly as a non-sensitizer based on the negative DPRA and h-CLAT results (Urbisch et al, 2015). Thus, there appears to be a mechanistic applicability domain issue with the KeratinoSens (Table 2).

3.2.9. *Benzyl cinnamate (BC)* - BC is assigned Category 4 (weak) based on a HRIPT NOEL of 4720 $\mu\text{g}/\text{cm}^2$ and HMT NOEL of 5517 $\mu\text{g}/\text{cm}^2$ as well as moderately high tonnage usage (Api et al, 2017). The SCCNFP opinion on fragrance materials (2012) list BC as a contact allergen in humans that is less frequently reported and thus less documented compared to other consumer allergens. It is one of the 26

601 fragrances currently requiring individual labeling. BC is positive in the LLNA with
602 an EC3 value of 18.4% (Kern et al, 2010). BC generates a Michael acceptor and
603 SN2 nucleophilic substitution alerts for protein binding (OECD QSAR Toolbox;
604 Derek Nexus; Urbisch et al, 2015; Kern et al, 2010) (Table 1). Based on the totality
605 of the available data one would expect BC to be positive in *in vitro* skin
606 sensitization assays.

607
608 Benzyl cinnamate is negative in h-CLAT and borderline negative in DPRA but
609 clearly positive in the KeratinoSens and U-SENSTM (Urbisch et al, 2015;
610 Takenouchi et al, 2015; Natsch and Emter, 2008, Piroird et al., 2015). It does form a
611 direct peptide adduct with the DPRA-Cys peptide and the Cor1-C420 peptide
612 (Natsch and Gfeller, 2008; Roberts and Natsch, 2009), but depletion is below the
613 DPRA classification threshold indicating a weak peptide reactivity. It has also been
614 reported to cause precipitation in the h-CLAT (Takenouchi et al, 2015). Thus, and
615 maybe not surprising, BC was capable of giving a weak positive response in the 3-D
616 SENS-IS assay (Hoffman et al, 2015). However, using the 2 out of 3 approach, BC
617 would be classified incorrectly as a non-sensitizer based on the negative DPRA and
618 h-CLAT results (Urbisch et al, 2015). Thus, overall assessment for BC is potentially
619 explained by a borderline result due to weak reactivity in the DPRA and a possible
620 technical applicability issue with the h-CLAT (Table 2).

621
622 3.2.10. *Benzyl benzoate (BB)* – BB is assigned Category 5, a rare cause of contact allergy,
623 based on a very high HRIPT NOEL of 20690 $\mu\text{g}/\text{cm}^2$ (Basketter et al, 2014).

However, the SCCNFP opinion on fragrance materials (2012) list BB as a contact allergen in humans that is less frequently reported and thus less documented compared to other consumer allergens. It is one of the 26 fragrances currently requiring individual labeling, but 0 positive results in 1870 tested patients and 1 in 2003 patients was reported by Schnuch *et al.* (2007, 2015) indicating extremely rare sensitization in humans. BB is positive in the LLNA with an EC3 value of 17% (Gerberick *et al.*, 2005) but negative in an additional LLNA study (referred to in Hoffmann *et al.*, 2018). BB triggers a skin sensitization reactivity alert, substance reacting by nucleophilic substitution, SN2 (Urbisch *et al.*, 2015; Takenouchi *et al.*, 2015; OECD QSAR Toolbox; Derek Nexus) but this reaction mechanism could not be verified experimentally by peptide adduct formation (Natsch and Gfeller, 2008). Based on the totality of the available human data one would expect BB to be negative in *in vitro* skin sensitization assays.

Benzyl benzoate is clearly negative in h-CLAT, U937-CD86 and DPRA but clearly positive in the KeratinoSens (Urbisch *et al.*, 2015; Takenouchi *et al.*, 2015; Natsch *et al.*, 2013; Nukada *et al.*, 2013) and positive according to the “6 rule strategy” only in the U-SENSTM (Piroird *et al.*, 2015). One explanation for it being negative in the h-CLAT might be related to limited solubility in cell-based assays with oil droplets observed (Takenouchi *et al.*, 2015). It is reported that the DPRA yields false negative results relative to the LLNA with benzylic esters with common structural alerts, including benzyl benzoate, benzyl salicylate, butylbenzyl phthalate and benzyl cinnamate (Urbisch *et al.*, 2015). Finally, BB produces a weak positive

response in the 3-D SENS-IS assay (Hoffman et al, 2015). Using the 2 out of 3 approach, BB would be classified correctly as a non-sensitizer based on the negative DPRA and h-CLAT results (Urbisch et al, 2015).

3.2.11. *Undec-10-enal (10-UDE)* - There is very limited human sensitization information on 10-UDE. However, it is reported to have an EC3 of 6.8% in the LLNA (Patlewicz et al, 2002). 10-UDE triggers a skin sensitization reactivity alert, a substance reacting by Schiff base formation (Urbisch et al, 2015; Takenouchi et al, 2015; OECD QSAR Toolbox; Derek Nexus). Based on the available data, although limited, one would expect 10-UDE to be a positive in *in vitro* skin sensitization assays.

Undec-10-enal is clearly negative in h-CLAT and borderline negative DPRA but clearly positive in the KeratinoSens and U-SENS™ (Urbisch et al, 2015; Takenouchi et al, 2015, Piroird et al., 2015). It is difficult to judge if the 2 out 3 is correct or not since there is little to no human data available on the material so one is not able to know if LLNA result is correct but the aldehyde alert would indicate so. Using the 2 out of 3 approach, 10-UDE would be classified incorrectly as a non-sensitizer based on the negative DPRA and h-CLAT results (Urbisch et al, 2015). The 10-UDE results are potentially explained by a borderline result in the DPRA.

3.2.12. *Benzyl salicylate (BS)* - Benzyl salicylate (BS) is assigned Category 5, a rare cause of contact allergy, based on a very high HRIPT NOEL of 17717 $\mu\text{g}/\text{cm}^2$ and

no induction up to 20690 $\mu\text{g}/\text{cm}^2$ in the HMT (Basketter et al, 2014). Only 2 positive reactions among 2041 patients tested were reported (Schnuch et al, 2007). However, the SCCNFP opinion on fragrance materials (2012) lists BS as a contact allergen in humans that is less frequently reported and thus less documented compared to other consumer allergens. It is one of the 26 fragrances currently requiring individual labeling. BS is positive in the LLNA with an EC3 value of 2.9% (Gerberick et al, 2005). The benzyl ester of BS triggers a nucleophilic substitution alert by SN2 using OECD algorithm in OECD QSAR Toolbox whereas the OASIS algorithm gives an acylating alert (OECD QSAR Toolbox; Urbisch et al, 2015; Takenouchi et al, 2015). Derek Nexus gives a quinone Michael acceptor, free radical generator alert (Table 1). Finally, it is reported that the substituted phenol of BS could be a pro/prehaptens producing a Michael acceptor or phenolic radical (Payne and Walsh, 1994). Based on the totality of the available data one would expect BS to be a borderline compound in *in vitro* skin sensitization assays with the human data indicating no significant sensitization risk.

Benzyl salicylate is negative in h-CLAT and borderline negative in the DPRA but positive in the KeratinoSens and U-SENSTM (Urbisch et al, 2015; Takenouchi et al, 2015; Natsch and Emter, 2008, Piroird et al., 2015). It is reported that the DPRA yields false negative results with benzylic esters with common structural alerts, including benzyl benzoate, benzyl salicylate, butylbenzyl phthalate and benzyl cinnamate (Urbisch et al, 2015) when compared to LLNA data. Finally, BS produces a weak positive response in the 3-D SENS-IS assay (Hoffman et al, 2015).

With borderline compounds like BS it is not unexpected to get mixed results in the different *in vitro* skin sensitization assays (Table 2). Using the 2 out of 3 approach, BS would be classified as a non-sensitizer based on the negative DPRA and h-CLAT results (Urbisch et al, 2015).

3.3. DPRA-/KeratinoSens-/h-CLAT+ ($n=13$) – For this group of chemicals, the activation of the dendritic cells in the h-CLAT may be unrelated to the skin sensitization process (Gallucci et al., 1999). In some instances, there is no indication that the predictions of the DPRA and the KeratinoSens/LuSens assays might be false negative ones. For example, the lack of peptide reactivity is consistent with the lack of structural alerts for a number of these compounds. If no critical metabolites are generated and antigen formation and keratinocyte activation is missing, the test substances would not be expected to drive the AOP towards a skin sensitization.

3.3.1. Diethyl phthalate (DEP) - Diethyl phthalate (DEP) is considered to be a non-sensitizer. It is listed in Category 6, essentially absent of convincing evidence of contact allergy (Basketter et al, 2014). In support, DEP does not generate any skin sensitization alerts (OECD QSAR Toolbox; Derek Nexus; Urbisch et al, 2015; Takenouchi et al, 2015) and there are very few reported cases of patch test positive results referenced in the literature (Frosch, 2006). RIFM routinely uses a 75%DEP/25% ethanol vehicle for their human patch testing without issues of sensitization. DEP is negative in the LLNA when tested up to 100% (Gerberick et al, 2005). Based on the totality of the available data one would expect DEP to be negative in *in vitro* skin sensitization assays.

Diethyl phthalate is negative in KeratinoSens, U937-CD86 and DPRA but clearly CD54 positive in the h-CLAT, U-SENSTM and LuSens (Urbisch et al, 2015; Natsch et al, 2013; Nukada et al, 2012; Takenouchi et al, 2015, Piroird et al., 2015). However, Natsch and Emter (2008) reported that diethyl phthalate was repeatedly positive in a non-standard ARE-based assay but only at very high concentrations. Finally, DEP is non-sensitizing in the 3-D SENS-IS assay (Hoffman et al, 2015). There is no apparent chemical or biological reason for why DEP would give a false positive response in the h-CLAT assay with CD54, but not CD86. DEP does give the correct prediction using the 2 out of 3 approach using the KeratinoSens data (Urbisch et al, 2015).

3.3.2. *Pyridine* - Pyridine is assigned Category 5, a rare cause of contact allergy, based on a very high HRIPT NOEL of 41051 $\mu\text{g}/\text{cm}^2$ (Basketter et al, 2014). It is also reported as a weak sensitizer in a small human maximization test with one of 24 subjects responding at a 50% induction concentration and 10% challenge concentration (Kligman, 1966). Pyridine does not generate any skin sensitization alerts (OECD QSAR Toolbox; Derek Nexus; Urbisch et al, 2015; Takenouchi et al, 2015) and there are very few reported cases of patch test positive results referenced in the literature (Frosch, 2006). However, pyridine is positive in the LLNA with a high EC3 value of 71.2% which would categorize it as a weak sensitizer (Gerberick et al, 2005). Overall, with the data on pyridine we conclude it to be a non-sensitizer. Thus, one could expect negative findings using *in vitro* skin sensitization assays.

Pyridine is negative in KeratinoSens (borderline negative), LuSens (borderline negative), mMUSST, U-SENSTM and DPRA but borderline CD86 positive in the h-CLAT (Urbisch et al, 2015; Bauch et al, 2012; Natsch and Emter, 2008; Nukada et al, 2012, Piroird et al., 2015). There is no apparent chemical or biological reason for why pyridine would give a false positive response in the h-CLAT assay with CD86, but not CD54, being detected. One explanation for discordant result may be related to the h-CLAT result being borderline at a very high test substance concentration of 29 mM (Table 2). Using the 2 out of 3 approach, pyridine would be classified as a non-sensitizer based on the negative DPRA and KeratinoSens results (Urbisch et al, 2015).

3.3.3. *Resorcinol* - Resorcinol is assigned Category 4, a substance that is rarely an important clinical allergen because it requires considerable/prolonged exposure to higher dose levels to produce sensitization (Basketter et al, 2014). Resorcinol was judged to be in this category based on its low frequency of human sensitization reports, despite widespread use at high levels in hair dyes and topical skin preparations, including pharmaceuticals. Diagnostic patch test clinical reports show overall low incidence of sensitization for resorcinol (Frosch, 2006; DeGroot, 2017). (Darsis and Goosens, 2016). Tested in the human maximization test at 15% for induction and 5% for challenge, no subjects were sensitized (Kligman, 1966). Resorcinol triggers a quinone precursor/radical substitution alert (Urbisch et al, 2015). It is considered to be a prohaptens with a pro-Michael acceptor binding

mechanism (Patlewicz et al, 2016). No alerts were generated using the OECD QSAR Toolbox while a plausible prohapten free radical alert is generated with Derek Nexus. The LLNA results are mixed with positive and negative results (ICCVAM). EC3 values of 5.5 and 6.3 are reported for resorcinol (Basketter et al, 2007). Based on the totality of the available data one would expect resorcinol to be positive in *in vitro* skin sensitization assays.

Resorcinol is negative in KeratinoSens, LuSens, and DPRA but positive in the h-CLAT (CD54), mMUSST/ U937-CD86 and U-SENS™ (Urbisch et al, 2015; Natsch et al, 2013; Nukada et al, 2013; Takenouchi et al, 2015, Piroird et al., 2015). Finally, resorcinol is moderate in the 3-D SENS-IS assay, week in the SensCeeTox and sensitizing in the IL-18 epiCS® and IL-18 EpiDerm assays (Hoffman et al, 2015 and Mehling et al., 2019). The reason that resorcinol, a putative prohapten, is not being detected may be due to the lack of a metabolic system in the DPRA and limited metabolic capacity in the cell-based assays. However, Nakata (2013) reports positive DPRA results for resorcinol but the value is a borderline positive. Thus, the discordant results are likely due to mechanistic applicability domain issue with KeratinoSens and borderline DPRA results (Table 2). Using the 2 out of 3 approach, resorcinol would be classified as a false negative based on the negative DPRA and KeratinoSens results (Urbisch et al, 2015). It is classified positive in another 2 out of 3 ITS, however (Takenouchi et al, 2015; Nukata et al, 2013).

3.3.4. *Benzyl alcohol (BZA)* - BZA is assigned Category 4 (weak) based on a HRIPT NOEL of 5906 $\mu\text{g}/\text{cm}^2$ and HMT NOEL of 6897 $\mu\text{g}/\text{cm}^2$ as well as moderately high tonnage usage and a LOEL of 8858 $\mu\text{g}/\text{cm}^2$ (Api et al, 2017). The SCCNFP opinion on fragrance materials (2012) list BZA as a contact allergen in humans that is less frequently reported and thus less documented compared to other consumer allergens. It is one of the 26 fragrances currently requiring individual labeling. BZA does not trigger any skin sensitization alerts and is negative in the LLNA when tested up to 50% (OECD QSAR Toolbox; Derek Nexus; Kern et al, 2010). Based on the totality of the available data one would expect BZA to be a borderline compound in *in vitro* skin sensitization assays, with the human data indicating a weak sensitization potential.

Benzyl alcohol is borderline negative in KeratinoSens and borderline negative (5.06% mean peptide depletion) in the DPRA, negative in the U-SENSTM but positive in the h-CLAT (CD86) and clearly positive in the LuSens (Urbisch et al, 2015, Piroird et al., 2015). Finally, BZA is weak in the 3-D SENS-IS assay (Hoffman et al, 2015). The reason BZA is being detected positive in some assays and not others may be due to the sensitivity of the individual assays, especially noting the borderline results that are observed (Table 2). It is important to note, however, that BZA is not detected in the LLNA and does not trigger any skin sensitization alerts. Using the 2 out of 3 approach, BZA would be classified as a non-sensitizer based on the negative DPRA and KeratinoSens results (Urbisch et al, 2015).

3.3.5. *Octanoic acid (Caprylic acid)* - Octanoic acid is considered to be a non-sensitizer.

It is listed in Category 6, a compound that is essentially absent of convincing evidence of contact allergy with only a few positive patch test results reported (Basketter et al, 2014). In support, octanoic acid does not generate any skin sensitization alerts (OECD QSAR Toolbox; Derek Nexus; Urbisch et al, 2015; Takenouchi et al, 2015) and is non-sensitizing in the LLNA when tested up to 50% (Gerberick et al, 2005). Based on the totality of the available data one would expect octanoic acid to be classified as a negative compound in *in vitro* skin sensitization assays.

Octanoic acid is clearly negative in KeratinoSens and DPRA but positive in the h-CLAT (CD54), U937-CD86, U-SENS™ (Urbisch et al, 2015; Natsch et al, 2013; Nukata et al, 2012; Takenouchi et al, 2015, Piroird et al., 2015). Finally, octanoic acid is categorized as non-sensitizing in the 3-D SENS-IS assay (Hoffman et al, 2015). There is no apparent chemical or biological reason for why octanoic acid would give a discordant false positive response in the h-CLAT assay with CD54, but not CD86, being detected (Table 2). One explanation could be that some irritants like octanoic acid can give a false positive response in the h-CLAT by activating specific cytokines like CD54. Using the 2 out of 3 approach, octanoic acid would be classified correctly as a non-sensitizer based on the negative DPRA and KeratinoSens results (Urbisch et al, 2015).

3.3.6. *Nonanoic acid* - Although not included in Basketter's human categorization paper (2014), one would expect nonanoic acid, a well-known human irritant, to be included in Category 6 based on its lack of human skin sensitization data (Frosch, 2006). In support, nonanoic acid does not generate any skin sensitization alerts (OECD QSAR Toolbox; Derek Nexus; Urbisch et al, 2015; Takenouchi et al, 2015). Nonanoic acid is reported negative in the LLNA at a concentration up to 60% (Natsch et al, 2013). However, Adenuga et al (2012) reported an EC3 of 18.7% and Takenouchi (2015) reports an EC3 of 21%. Based on the totality of the available data one would expect nonanoic acid to be classified as a negative compound in *in vitro* skin sensitization assays.

Nonanoic acid is negative in KeratinoSens and DPRA but positive in the h-CLAT (CD86) and mMUSST / U937-CD86 (Urbisch et al, 2015; Natsch et al, 2013; Takenouchi et al, 2015). There is no apparent chemical or biological reason for why nonanoic acid would give a false positive response in the h-CLAT assay with CD86, but not CD54, being detected. One explanation could be that some irritants like nonanoic acid could give a false positive response in the h-CLAT by activating specific cytokines like CD86. Using the 2 out of 3 approach, nonanoic acid would be classified correctly as a non-sensitizer based on the negative DPRA and KeratinoSens results (Urbisch et al, 2015).

3.3.7. *Clofibrate (Ethyl (2-(4-chlorophenoxy)-2-methylpropanoate)* - There is limited human data available to categorize the potency of Clofibrate. It is a lipid-lowering

agent used to control high cholesterol and triacylglyceride levels in the blood. It has also been used in the treatment of atopic dermatitis for its anti-inflammatory and anti-itching activities. Clofibrate does not trigger any skin sensitization alerts (OECD QSAR Toolbox; Derek Nexus; Urbisch et al, 2015; Takenouchi et al, 2015) and has been reported to be (borderline) negative in the LLNA when tested up to 50% (Kern et al, 2011; Natsch et al, 2013). Based on the available data, although limited, one would expect Clofibrate to be classified as a negative compound in *in vitro* skin sensitization assays.

Clofibrate is negative in KeratinoSens, U937-CD86 and DPRA but CD54 positive in the h-CLAT (Urbisch et al, 2015; Natsch et al, 2013; Takenouchi et al, 2015). There is no apparent chemical or biological reason for why Clofibrate would give a discordant false positive response in the h-CLAT assay with CD54, but not CD86, being detected other than there is technical applicability domain issue. Using the 2 out of 3 approach, Clofibrate would be classified correctly as a non-sensitizer based on the negative DPRA and KeratinoSens results (Urbisch et al, 2015).

3.3.8. *Chlorobenzene* - Although not included in Basketter's human categorization paper (2014), one would expect chlorobenzene to be included in Category 6 or at worst, Category 5, based on the lack of human skin sensitization data (Frosch, 2006). In support, chlorobenzene does not generate any skin sensitization alerts (OECD QSAR Toolbox; Derek Nexus; Urbisch et al, 2015; Takenouchi et al, 2015) and is reported negative in the LLNA at concentrations up to 25% (Gerberick et al,

2005). Furthermore, chlorobenzene is positive when tested at higher concentrations (Kolle et al, 2013). Based on the totality of the available data one would probably expect chlorobenzene to be classified as a negative compound in *in vitro* skin sensitization assays.

Chlorobenzene is negative in KeratinoSens, U937-CD86, U-SENSTM and DPRA but CD86 positive in the h-CLAT (Urbisch et al, 2015; Natsch et al, 2013; Nukata et al, 2012; Takenouchi et al, 2015 Piroird et al., 2015). There is no apparent chemical or biological reason for why chlorobenzene would give a false positive response in the h-CLAT assay with CD86, but not CD54, being detected. However, Bauch (2012) excluded chlorobenzene from the study due to it demonstrating solubility issues for cell-based assays. Using the 2 out of 3 approach, chlorobenzene would be classified correctly as a non-sensitizer based on the negative DPRA and KeratinoSens results (Urbisch et al, 2015).

3.3.9. *Linalool* - Linalool is assigned Category 4, a substance that demonstrates infrequent association with causing contact allergy in relation to its level of exposure (Basketter et al, 2014). Linalool has an HRIPT NOEL of 13793 $\mu\text{g}/\text{cm}^2$ (Basketter et al, 2014; Gerberick et al, 2001) and HMT NOEL of 55176 $\mu\text{g}/\text{cm}^2$ (Greif, 1967) with no cases of sensitization reported in both tests conducted at high concentration. The SCCNFP opinion on fragrance materials (2012) lists patch test data on 5423 subjects with only 18 reported cases globally (0.3%), despite widespread use of linalool (present in > 95% of cosmetic products investigated).

SCCNFP opinion lists Linalool as a well-established contact allergen in humans, however this assessment is largely based on data on Linalool put under forced oxidation for several months (Christensson et al, 2010), which is not relevant for the regulatory assessment of the parent compound and also appears not to represent commercial use (Kern et al. 2014). It is one of the 26 fragrances currently requiring individual labeling. Linalool is known to be a prehapten that contains an alert for autooxidation to hydroperoxide formation (SCCNFP, 2012; Patlewicz et al, 2016). It is given an equivocal terpenoid alert using Derek Nexus. Linalool is positive in the LLNA with reported EC3 values of 30% and 46% (Gerberick et al, 2005). Based on the totality of the available human data one would expect linalool to be negative in *in vitro* skin sensitization assays.

Linalool is negative in KeratinoSens and DPRA but positive in the h-CLAT (CD54), U937-CD86, and U-SENS™ (Urbisch et al, 2015; Natsch and Emter, 2008; Natsch et al, 2013; Nukata et al, 2013; Takenouchi et al, 2015, Piroird et al., 2015). Finally, linalool is categorized as weak in the 3-D SENS-IS assay (Hoffman et al, 2015). Given the low frequency of positive patch test reactions in spite of high usage, a negative classification based on *in vitro* data may be acceptable or even preferable as compared to the LLNA classification (Natsch et al, 2013). Using the 2 out of 3 approach, linalool would be classified as a non-sensitizer based on the negative DPRA and KeratinoSens results (Urbisch et al, 2015).

3.3.10. *α*-iso-Methylionone (MIO) - MIO is assigned Category 5, a rare contact allergen except in selected patients with eczema. It has a reported HRIPT NOEL of 70866 $\mu\text{g}/\text{cm}^2$ (no case of sensitization at this very high test substance concentration) along with known high usage. MIO generates a Michael acceptor alert for protein binding (OECD QSAR Toolbox; Derek Nexus; Urbisch et al, 2015). It is positive in the LLNA with a high EC3 value of 21.8% (Kern et al, 2010). Based on the available human data, one would expect MIO to be negative in *in vitro* skin sensitization assays.

α-iso-Methylionone is negative in KeratinoSens and DPRA but CD54 positive in the h-CLAT and U-SENSTM (Urbisch et al, 2015; Natsch and Emter, 2008; Takenouchi et al, 2015, Piroird et al., 2015). Complete lack of reactivity despite a Michael acceptor alert was shown by Natsch and Emter (2017). Finally, MIO is categorized as weak in the 3-D SENS-IS assay (Hoffman et al, 2015). Given the very high HRIPT NOEL and the very low frequency of positive patch test results (2 cases in 1870 patients tested and 1 case in 2004 tested despite wide-spread use (Schnuch et al., 2007, 2016)), a negative classification based on *in vitro* data may be preferable as compared to the LLNA classification (Table 2). Using the 2 out of 3 approach, MIO would be classified as a non-sensitizer based on the negative DPRA and KeratinoSens results (Urbisch et al, 2015).

3.3.11. 2-methoxy-4-methylphenol (MMP) - MMP is assigned Category 2, a strong allergen (Api et al, 2017). However, there is limited human data to support a strong

category with a reported HRIPT NOEL of 118 $\mu\text{g}/\text{cm}^2$ along with no diagnostic patch test data (Api et al, 2017; SCCNFP, 2012). MMP is a pre/pro-hapten that generates a quinone precursor/radical substitution alert for protein binding (Urbisch et al, 2016). No alerts are generated with OECD QSAR Toolbox and with Derek Nexus a probable pro Michael acceptor/free radical alert is generated. It is positive in the LLNA with an EC3 value of 5.8% (Estrada et al, 2003). Based on the available data, one would expect MMP to be positive in *in vitro* skin sensitization assays.

MMP is negative in KeratinoSens and borderline negative in the DPRA (mean depletion 5.7%) but positive in the h-CLAT (CD 54 and CD 86), U937 and U-SENSTM (Urbisch et al, 2015; Natsch et al, 2013; Nukada et al, 2012; Piroird et al., 2015). The reason that MMP, a putative prohaptent, is not being detected may be due to the lack of a metabolic system in the DPRA and limited metabolic capacity in the cell-based assays. A demethylation of the methoxy group or the formation of a quinone methide might not occur under the conditions of the DPRA for a compound like MMP whereas the KeratinoSens has shown a positive result in a structurally similar compound, dihydroeugenol, which indicates no general lack in detecting chemicals with similar reaction mechanisms as MMP. Thus, the discordant results are likely due to and the borderline negative DPRA results (Table 2). Using the 2 out of 3 approach, MMP would be classified as a false negative based on the negative DPRA and KeratinoSens results (Urbisch et al, 2015).

3.3.12. 3-Aminophenol (3-AP) - There is limited human sensitization data available for 3-Aminophenol (3-AP). Some diagnostic patch data for 3-AP exists but it is sometimes difficult to interpret the positive responses based on the possibility of cross-reactivity to other hair dye ingredients (Frosch, 2006; DeGroot, 2017). Structurally related 2-Aminophenol is placed in Category 2, a frequent cause of contact allergy (Basketter et al, 2014). 3-AP is a pre/pro-hapten and generates a quinone precursor/radicalic substitution alert for protein binding (Urbisch et al, 2015). Patlewicz (2016) list 3-AP as prohaptens with a Michael acceptor alert. No alerts are generated using OECD QSAR Toolbox while a pre-Schiff Base, pro-electrophile is generated using Derek Nexus. It is positive in the LLNA with reported EC3 values of 0.24 and 3.2% (Gerberick et al, 2005; SCCP, 2007). Based on the available data, one would expect 3-AP to be positive in *in vitro* skin sensitization assays.

3-AP is negative in DPRA, borderline negative in the KeratinoSens and but positive in the h-CLAT (CD54), U937 and U-SENS™ (Urbisch et al, 2015; Natsch et al, 2013; Nukada et al, 2012, Piroird et al., 2015). Interestingly, 3-AP is listed as positive for induction of ARE-regulated luciferase activity (Natsch and Emter, 2008). The reason that 3AP, a putative prohaptens, is not being detected may be due to the lack of a metabolic system in the DPRA and limited metabolic capacity in the cell-based assays. Thus, the discordant results are likely due to mechanistic applicability domain issue with DPRA and borderline KeratinoSens results (Table 2). In support, 3-AP is detected positive in the PPRA with the peroxide/peroxidase

activation system (Troutman et al, 2011). Using the 2 out of 3 approach, 3-AP would be classified as a false negative based on the negative DPRA and KeratinoSens results (Urbisch et al, 2015).

3.3.13. Hexyl Salicylate (HS) - HS is assigned Category 4, a substance that is rarely an important clinical allergen because it requires considerable/prolonged exposure to higher dose levels to produce sensitization (Basketter et al, 2014). HS is reported to have an HRIPT NOEL of 35433 $\mu\text{g}/\text{cm}^2$ and HMT NOEL of 20654 $\mu\text{g}/\text{cm}^2$ (Api et al, 2017), with no sensitization observed in both studies. The SCCNFP (2012) reports HS lacks evidence to be categorized as a human sensitizer. HS does not generate any skin sensitization alerts (OECD QSAR Toolbox; Urbisch et al, 2015; Takenouchi et al, 2015). An equivocal pre/pro Michael acceptor and pre/pro free radical generator alert is generated with Derek Nexus. A very low EC3 of 0.18% is reported for this material (ECHA, 2006). Based on the totality of the data one would expect HS to be classified as a negative compound based on the available human data.

Hexyl salicylate is negative in KeratinoSens (borderline negative) and DPRA but CD54 positive in the h-CLAT and positive in the U-SENSTM (Urbisch et al, 2015; Takenouchi et al, 2015, Piroird et al., 2015). In addition, HS is categorized as non-sensitizing in the 3-D SENS-IS assay (Hoffman et al, 2015). There is no apparent chemical or biological reason for why HS would give a false positive response in the h-CLAT assay with CD54, but not CD86, being detected. One explanation

could be that some irritants like HS could give a false positive response in the h-CLAT by activating specific cytokines like CD54. Again, HS reveals no obvious alerts for peptide reactivity, so it would not be expected to be detected in the DPRA or KeratinoSens assays. Moreover, it is known that irritation is a confounding factor in the LLNA and is believed to lead to false positive responses. It is also possible that the low EC3 value reported for HS could be due to a contaminant (Urbisch et al, 2015). Using the 2 out of 3 approach, HS would be correctly classified as a non-sensitizer based on the negative DPRA and KeratinoSens results (Urbisch et al, 2015).

Table 1: Protein binding alerts^a

Test substance		OECD QSAR Toolbox v.3.2 ^b protein binding alters		OECD QSAR Toolbox v.4.2 protein binding alters		Derek Nexus: 6.0.1, Nexus: 2.2.1 ^c
Name	Cas #	OASIS algorithm	OECD algorithm	OASIS algorithm	OECD algorithm	
Diethyl phthalate	84-66-2	no alert	no alert	no alert	no alert	non-sensitizer
Octanoic acid (Caprylic acid)	124-07-2	no alert	no alert	no alert	no alert	non-sensitizer
4-Methoxyacetophenone (Acetanisole)	100-06-1	no alert	no alert	no alert	no alert	non-sensitizer
Benzoic acid	65-85-0	no alert	no alert	no alert	no alert	non-sensitizer
Clofibrate (Ethyl (2-(4-chlorophenoxy)-2-methylpropanoate)	637-07-0	no alert	no alert	no alert	no alert	non-sensitizer
Ethyl benzoylacetate	94-02-0	NA	no alert	NA	no alert	non-sensitizer
Tween 80	9005-65-6	no alert	SN2	no alert	no alert	non-sensitizer

Dibutyl phthalate	84-74-2	no alert	no alert	no alert	no alert	non-sensitizer
Butylbenzylphthalate	85-68-7	SN2	SN2	no alert	SN2	equivocal (SN2)
N,N-Diethyl-<i>m</i>-toluanimide	134-62-3	no alert	no alert	no alert	no alert	non-sensitizer
(+/-) Linalool	78-70-6	no alert	no alert	no alert	no alert	equivocal (terpenoid)
α-iso-Methylionone	127-51-5	MA	MA	MA	MA	plausible (MA)
Benzyl benzoate	120-51-4	SN2	SN2	SN2	SN2	plausible (SN2)
Chlorobenzene	108-90-7	no alert	no alert	no alert	no alert	non-sensitizer
Nonanoic acid	112-05-0	no alert	no alert	no alert	no alert	non-sensitizer
Hexyl salicylate	6259-76-3	no alert	no alert	no alert	no alert	equivocal (pre/pro MA, pre/pro free radical generator)
Benzyl cinnamate	103-41-3	MA / SN2	MA / SN2	MA	MA	plausible (MA)
Undec-10-enal	112-45-8	SB	SB	SB	SB	plausible (SB)
2-methoxy-4-methylphenol	93-51-6	no alert	no alert	no alert	no alert	probable (alert: pro MA, pre/pro free radical generator)
Resorcinol	108-46-3	no alert	no alert	no alert	no alert	plausible (alert: pre/pro free radical generator)

						radical generator)
3-Aminophenol	591-27-5	no alert	no alert	no alert	no alert	plausible (pre-SB, pro-electrophile)
Phthalic anhydride	85-44-9	Ac	Ac	Ac	Ac	probable (cyclic acid anhydride, azlactone or analogue)
Benzoyl peroxide	94-36-0	Ac	Ac	Ac	Ac	probable (acyl peroxide radical generation, Ac)
Benzyl alcohol	100-51-6	no alert	no alert	no alert	no alert	non-sensitizer
Coumarin	91-64-5	no alert	Ac / MA	no alert	Ac	plausible (alpha,beta-Unsaturated ester or precursor Ac)
6-Methylcoumarin	92-48-8	no alert	Ac / MA	no alert	Ac	probable (photoallergen alpha,beta-Unsaturated ester or precursor Ac)

Ethyl vanillin	121-32-4	no alert	no alert	no alert	no alert	non-sensitizer
Pyridine	110-86-1	no alert	no alert	no alert	no alert	non-sensitizer
Benzyl salicylate	118-58-1	Ac / SN2	SN2	Ac	SN2	equivocal (pre-/pro MA, pre/pro free radical generator)

^aAc, acylating agent; MA, Michael acceptor; NA, nucleophilic addition; SB, Schiff base formation; SN2, S_N2 substitution

^bTaken from Urbisch et al., 2015.

^cCertain, there is proof that the proposition is true; probable, there is at least one strong argument that the proposition is true and there are no arguments against it; plausible, the weight of evidence supports the proposition; equivocal, there is an equal weight of evidence for and against the proposition; doubted, the weight of evidence opposes the proposition; improbable, there is at least one strong argument that the proposition is false and there are no arguments that it is true; impossible, there is proof that the proposition is false; open, there is no evidence that supports or opposes the proposition; contradicted, there is proof that the proposition is both true and false.

Table 2: Substances with a negative prediction based on the 2 out of 3 approach along with rationale for the observed outcome and sorted by the expert prediction.

Test substance		Reference data (<i>in vivo</i>)			<i>in vitro</i>				
		Sensitization Status Expert	In vivo results ^a	Uncertainty	2 out of 3 result (DPRA/ KeratinoSens/ h-CLAT) ^a	Uncertainty (OECD TG 442x)		Out of Applicability domain (OECD TG 442x)	
Name	Cas #	(Authors' Judgement)		borderline or discordant		borderline	discordant within same KE (OECD TG 442x)	technical	mechanistic
Diethyl phthalate	84-66-2	Negative	human 0 LLNA 0		0/0/1		D ^b , E		
Octanoic acid (Caprylic acid)	124-07-2	Negative	human 0 LLNA 0		0/0/1			E	
4-Methoxyacetophenone (Acetanisole)	100-06-1	Negative	human 0 LLNA 0		0/1/0		E	D	
Benzoic acid	65-85-0	Negative	human 0 LLNA 0		1/0/0			C	

Clofibrate (Ethyl (2-(4-chlorophenoxy)-2-methylpropanoate)	637-07-0	Negative	human 0? LLNA 0		0/0/1		E		
Ethyl benzoylacetate	94-02-0	Negative	human ? LLNA 0		(0)/(1)/0	C, D			
Tween 80	9005-65-6	Negative	human 0 LLNA 0		(0)/1/0	C	E		D
Dibutyl phthalate	84-74-2	Negative	human 0? LLNA 0		0/1/0				D
Butylbenzylphthalate	85-68-7	Negative	human ? LLNA 0		0/1/0				D
N,N-Diethyl-<i>m</i>-toluanimide	134-62-3	Negative	human 0 LLNA 0		(1)/0/0	C			
(+/-) Linalool	78-70-6	Negative	human 0 LLNA 1	YES	0/0/1				C, D
α-iso-Methylionone	127-51-5	Negative	human 0 LLNA 1	YES	0/0/1				

Benzyl benzoate	120-51-4	Negative	human 0 LLNA 0/1	YES	0/1/0		E	E	
Chlorobenzene	108-90-7	Negative	human 0 LLNA 0/1	YES	0/0/1		E		
Nonanoic acid	112-05-0	Negative	human 0 LLNA 1?	YES	0/0/1			E	
Hexyl salicylate	6259-76-3	Negative	human 0 LLNA 1	YES	0/(0)/1	D			
Pyridine	110-86-1	Negative	human 0 LLNA 1	YES	0/(0)/(1)	D, E	E		
Benzyl cinnamate	103-41-3	Positive	human 1 LLNA 1		(0)/1/0	C	E	E	
Undec-10-enal	112-45-8	Positive	human ? LLNA 1	YES	(0)/1/0	C	E		
2-methoxy-4-methylphenol	93-51-6	Positive	human ? LLNA 1		(0)/0/1	C			
Resorcinol	108-46-3	Positive	human 1 LLNA 0/1	YES	(0)/0/1	C			D
3-Aminophenol	591-27-5	Positive	human ?		0/(0)/1	D			C

			LLNA 1						
Phthalic anhydride^c	85-44-9	Positive	human 0? LLNA 1	YES	1/0/0		E		D
Benzoyl peroxide^d	94-36-0	Positive	human 1 LLNA 1		1/(0)/0	D	E		D
Benzyl alcohol	100-51-6	Positive	human 1 LLNA 0	YES	(0)/(0)/1	C, D	D, E		
Coumarin^e	91-64-5	Indeterminate	human 0/1 LLNA 0/1	YES (contaminants in LLNA test commercial grade test material)	0/1/0		E		D
6-Methylcoumarin^f	92-48-8	Indeterminate	human (0) LLNA 0	YES (photosensitizer)	0/1/0				D
Ethyl vanillin	121-32-4	Indeterminate	human 0?	YES	(0)/(1)/0	C, D			

			LLNA 0						
Benzyl salicylate	118-58-1	Indeterminate	human 0	YES	(0)/1/0	C	E		
			LLNA 1						

^a0, negative; 1, positive; 0?;absence of clear evidence; 1?, unclear evidence; 0/1, equivocal; (0), borderline negative; (1), borderline positive; ?, lack of data

^bKolle 2019

^cPhthalic anhydride is a respiratory sensitizer (Dearman et al., 2013) was positive in a U937-CD86 assay but negative in the h-CLAT (reviewed in Urbisch et al., 2015).

^dBenzoyl peroxide was positive in a U937-CD86 assay but negative in the h-CLAT (reviewed in Urbisch et al., 2015).

^eDifferent LLNA results reported for Coumarin due to contaminants (Vocanson et al., 2006).

^f6-Methylcoumarin is known to be a photoallergen in humans.

4. Discussion

The 2 out of 3 approach (DPRA, KerationSensTM/LuSens and h-CLAT) has been shown to give very good sensitivity, specificity and overall accuracy numbers for predicting the skin sensitization potential of chemicals (Bauch et al, 2012; Natsch et al, 2013; Urbisch et al, 2015). However, it is clear that the overall performance of any defined approach is critically dependent on the data set used for analysis. In the early development of individual assays or defined approaches they are often compared against ‘gold-standard’ lists of chemicals where there is substantial evidence to support a chemical’s categorization as a skin sensitizer or non-sensitizer (Hoffmann et al., 2008, Kolle et al, 2013; ECETOC Technical Report #77; ECVAM, 2008; ICCVAM, 2003, 2009). The variability of the local lymph node assay has been discussed by several groups Kolle et al. (2013), Hoffmann (2015), Dumont et al., (2016), and Dimitrov et al. (2016) and has to be considered when evaluating the reproducibility and predictive performance of non-animal methods (Paparella et al. 2013). When these datasets get expanded to larger numbers the possibility exist that some of the chemicals may not have been vetted sufficiently to establish their use in such analysis. For example, there can be inconsistencies between the available human and LLNA data or lack of clear evidence to support human categorization as positive or negative (Urbisch et al, 2015). It is equally important to review the *in vitro* data to see if the results are robust, not borderline, and whether any discordant results can be explained based on technical or mechanistic reasons. In this study, we examined the data of 29 substances with at least one discordant result to see if there was information available to inform us on the outcomes that were observed with these materials and how it relates to a WoE *in vivo* sensitization classification.

Sensitization Status Expert Prediction – Prior to running any *in vitro* skin sensitization assays and analyzing the data generated, it is critical to establish the predicted outcome for each substance before experimentation. This is especially true when wanting to know how a particular approach can predict a compound's hazard or potency based on the available human and animal data. It is generally accepted that human data are supposed to overrule the outcome of non-animal test method results as well as *in vivo* animal data such as the LLNA. Efforts have been made in recent years to permit categorization of the human potency of substances by reviewing data available to support their categorization (Basketter et al, 2014; Api et al, 2017). Typical data that inform on such an exercise can include prevalence data from the general population and/or dermatology clinics. For hazard identification, these human categories were dichotomized by considering categories 1–4 as sensitizer and categories 5 and 6 as non-sensitizers. However, in separate analyses categories 1–5 were designated as sensitizers (4 and 5 grouped together as weak) and category 6 designated as non-sensitizer. The determination of using one approach versus the other is based on expert opinion and maybe the desire to add additional conservatism to the analyses. Of course, the intrinsic potency of any substance for humans can only be defined accurately if there is an understanding of the exposure to that substance. Api (2017) added annual use volume data to their human potency categorization paper. Although this information can be helpful in understanding human sensitization outcomes in the marketplace to a particular ingredient, it may fall short for some materials. For example, it could be possible that a material that is manufactured and distributed at high tonnage levels could be based on its use in laundry products versus its use in products such as fine fragrances. The consumer exposures to a

fragrance ingredient in a detergent would be significantly lower than one used in a fine fragrance.

No doubt there are some substances where it is quite easy to assign a hazard and, in some instances, potency but it is clear there are potential limitations of available information for this purpose (Basketter et al, 2014). For example, there are substances that are believed strongly to be non-sensitizers (e.g. petrolatum, sodium lauryl sulfate and glycerol) even though one can find occasional patch test results in the literature (Basketter et al, 2014). There exist thousands of chemicals that have been used in patch testing even though many of them are not considered significant skin sensitizers (DeGroot, 2008). For example, positive patch test results are described in the literature for parabens even though their frequent use in general corroborates this class to be non-sensitizers (Schnuch et al, 2011; Basketter et al, 2006). Using a weight of evidence approach, we assigned each of 29 compounds to one of three groups: positive, negative or indeterminate for how they should be predicted by *in vitro* test methods.

All available data was used to make a weight of evidence assessment including review of *in vivo* data (animal and human) and clinical diagnostic patch data as well as whether a chemical had any structural alerts known to be associated with causing skin sensitization (Table 1). We assigned 17 chemicals as negative. Five of them, diethyl phthalate, octanoic acid, Tween 80, α -iso-Methylionone and N,N-diethyl-m-toluamide had been assigned Human Category 6, essentially absent of systemic evidence of contact allergy while pyridine was placed in Category 5 (substances with a very low intrinsic ability to cause skin sensitization, Basketter et al, 2014). Only three of the 17 compounds placed in the negative *in vitro* prediction category generated a

protein binding alert, butylbenzylphthalate, α -iso-methylionone and benzyl benzoate, when analyzed using the OECD QSAR Toolbox or Derek Nexus. We assigned 8 chemicals as *in vivo* positive. All these chemicals revealed clear *in vivo* evidence for sensitization and all of them generated protein binding alerts. However, three of them (3-methoxy-4-methylphenol, resorcinol and 3-aminophenol) generated alerts as pre/pro-haptens.

Finally, we assigned 4 compounds as indeterminate. There was not sufficient evidence to make an assessment whether these substances would be expected to yield positive or negative results using *in vitro* skin sensitization assays. Benzyl salicylate was placed in Category 5 which are described as substances that have very low intrinsic ability to cause skin sensitization (Basketter et al, 2014). It is our expert opinion that the mixed or insufficient data for these materials should exclude them from inclusion into a dataset for analysis of an individual assay or a defined approach. No doubt, assigning substances to any category is based on expert judgement which of course can be challenged by other experts. The quality of the data available is not always easy to establish, especially for human data. It is important that if robust human exposure data was available for these substances, it may be possible to move them into a positive or negative category. Unfortunately, this critical exposure data for consumers or workers is very difficult to obtain. Thus, in the absence of evidence it is best not to include these compounds in datasets being used to evaluate new test methods or defined approaches.

In vivo borderline – It has been shown that when retesting 22 LLNA performance standards in the standard LLNA protocol, a reproducibility of only 77% was found for the investigated substances (Kolle et al, 2013). In the present dataset of 29 chemicals, resorcinol, coumarin,

chlorobenzene and nonanoic acid have mixed and/or borderline LLNA results. No doubt having borderline or mixed LLNA results can be problematic for assigning a compound as positive or negative. However, using a WoE approach that includes review human data and the chemical's reactivity potential, we were able to assign chlorobenzene and nonanoic acid as negative and resorcinol as positive. Coumarin, however, was determined to be indeterminate based on possibility of contamination issues with some data that was reviewed (Vocanson et al, 2006).

In vitro borderline section - The borderline range depicts the variance of the individual test methods, including technical and biological variability (Leontaridou et al., 2017). It addresses the uncertainty of the three assays around their respective classification thresholds and represents a range in which the likelihood to obtain a positive or negative result just below or above the classification threshold is equal. This range was determined for each assay statistically (pooled standard deviations), using historical BASF data (Leontaridou et al., 2017). It is useful especially for assays for which no individual statistical analysis is possible due to low number of replicates per treatment (e.g. h-CLAT and DPRA). This evaluation is an amendment to the evaluation given in the respective OECD Test Guidelines. In this analysis, there are 13 compounds with borderline *in vitro* results for one or two assays. Understanding the variance of each individual test method can be important in the interpretation of 2 out of 3 approach results. The influence of the individual methods' uncertainty on the predictive capacity of the individual methods as well as the 2 out of 3 approach has recently been analysed (Leontaridou et al., 2019).

Structural alerts – in silico tools - Structural alerts can be extremely useful for assessing a chemicals skin sensitization potential in a WoE approach as used here. However, there are limitations that need to be considered when using structural alerts generated from *in silico* tools.

Electron-withdrawing or donating effects enhancing or reducing reactivity or steric effects are not, or only partially taken into account and the presence of a structural alert should not be taken as sufficient evidence that a chemical indeed will react with a particular reaction mechanism. Reactivity experiments (e.g. a DPRA with LC-MS based determination of peptide adducts or more sophisticated NMR experiments) may further substantiate whether a molecule indeed reacts by that particular mechanism. Thus, *in silico* tools indicate potential for acyl-transfer or S_N2 substitution at an acyl group for a number of chemicals discussed here, also for chemicals with no or very weak human sensitization potential (PA, coumarin, BB, BS, BBP, and BC). These reactions may not in all cases occur under dilute aqueous conditions. Thus LC-MS investigations did confirm this reaction mechanism only for PA among these six chemicals (Natsch and Gfeller, 2008; Natsch et al., 2015). A direct Michael acceptor alert is given for BC and MIO, however, experimental evidence by peptide reactivity monitored by LC-MS could confirm this mode of action only for BC (Natsch et al., 2015).

The 2 out of 3 approach is able to detect most pre- and pro-haptens and a negative prediction is considered acceptable. *In vitro* investigations (Urbisch et al., 2016 and Patlewicz et al., 2016) using compounds requiring molecular transformation to attain a sensitizing potential have shown that pre-haptens can readily be detected in the DPRA, many of which involve autoxidation processes. Moreover, many pro-haptens are also activated by non-enzymatic oxidation (and therefore are pre- and pro-haptens). The cellular models h-CLAT and LuSens have been shown to detect pro-haptens more efficiently; respective enzyme activities were detected in the cell lines (Fabian et al., 2013 and Patlewicz et al., 2016). Thus, within the 2 out of 3 approach, potentially relevant molecular transformations have been considered.

Applicability domain – The high overall accuracy of the 2 out of 3 approach indicates that in many cases positives or negatives in single assays are actually FP or FN, respectively. False-positives in the different assays might be due to different mechanisms; thus, in KeratinoSensTM unspecific activation of the antioxidant response due to other mechanisms than covalent modification of Keap1 is possible, while false-positives in DPRA may be generated by unspecific peptide oxidation. In the h-CLAT, non-sensitizing irritants such as octanoic acid may also lead to surface marker expression. False negatives results might occur due to solubility issue or limited metabolic or oxidative activation. In this dataset of 29 substances, 10 compounds were determined to indicate mechanistic reasons for the *in vitro* results observed in specific assays. However, in the absence of a mechanistic understanding for a particular outcome, there is likely an unknown technical reason. It is common for all toxicology assays to have applicability domain issues with specific compounds and it is critical to understand these as best possible so the strengths and limitations of each individual assay being used is known.

False-negative predictions in relation to sensitization risk

From a regulatory view-point, false-negative predictions made by a testing strategy deserve most scrutiny. Here we report out of the dataset on 213 chemicals published by Urbisch *et al.* (2015), 8 bona-fide false-negatives against WoE *in vivo* data when applying the rule that a 1 out of 3 result classifies chemicals as non-sensitizers. An important question is: does this group include chemicals which pose a serious sensitization risk to the human population? Based on the detailed analysis of this group of chemicals, the answer is yes probably only for Benzoyl peroxide, while for the other seven chemicals the human data summarized above indicate, as far as available, a

rather weak sensitization potency (high LOEL and/or infrequent positive patch tests). *Vice versa*, among the frequently reported sensitizers from the clinic (e.g. frequent fragrance allergens, sensitizing preservatives, acrylates, hair dyes, glove allergens, etc.) or among the chemicals with known low LOEL in human predictive tests, none is contained in the group of these eight chemicals. Thus, the human sensitizers missed by the 2 out of 3 approach tend to be rather weak allergens. Of course, this observation may need further testing by testing more clinically relevant or occupationally important allergens.

5. Conclusions

It has been demonstrated clearly that the use of five non-animal test methods (DPRA, KerationSensTM/LuSens, h-CLAT and (m)MUSST) are beneficial in predicting the respective AOP key events with high accuracy. Specifically, the 2 out of 3 approach has been shown to give very good sensitivity, specificity and overall accuracy numbers for predicting the skin sensitization potential of chemicals. However, it is clear that the overall performance of any defined approach is critically dependent on the data set used for analysis. In this report, the existing human, animal and chemical reactivity information for 29 chemicals was thoroughly reviewed so each chemical could be categorized as positive, negative or indeterminate. Five chemicals were identified to have insufficient information to be categorized as positive or negative which makes them of little use for evaluating a defined approach.

The 17 negative and 8 positive chemicals with one discordant *in vitro* test result were evaluated on an individual basis to understand better what led to the discordant result. It is clear that for some chemicals the discordant result might be explained by an observed borderline result.

Understanding the variance of each individual test method is critical for interpreting the statistical performance of a defined approach. It is also evident that some discordant results can be explained from how a chemical may interact mechanistically with the chemical or biological assay system being used. Although structural alerts can be extremely useful for assessing a chemicals skin sensitization potential in a WoE approach, it is important to understand there can be limitations when using alerts since there can be effects from neighboring groups enhancing or reducing reactivity that can impact a particular reaction mechanism.

6. Outlook

Tremendous progress has been made in recent years in the development of *in vitro* methods and defined approaches for predicting the skin sensitization potential of chemicals. There has also been recent progress in developing human datasets to evaluate how well these new approaches are working. Of course, it can be extremely challenging categorizing chemicals which have limited human data and/or exposure data. Critical review of human patch test data is specifically important since the relevance of the data is not always clear when there are limited reports available. Going forward it will be important for individuals building these datasets to review all available data thoroughly and, if needed, obtain relevant exposure data so a clear assessment of a chemical's sensitization potential can be determined.

Another need is for test developers and regulators to work continually to understand and share the strengths and limitations of each assay being used for hazard and risk assessment purposes. This includes knowing well the variance of each assay so borderline response can be put into perspective. As discussed in this paper, there can be chemical and/or biological explanations for

why a chemical gives a particular response in an assay. In addition, the identification of skin sensitization alerts may need additional understanding around the electronic effects of enhancing or reducing reactivity from neighboring groups or steric effects. No doubt in the near future we will see the emergence of new approaches for skin sensitization testing that may include the use of sophisticated 3D skin models or novel *in silico* approaches. Where ever the new science leads us, it will still be important to understand the strengths and limitations of each method or new defined approach along with using caution in what human or animal datasets are used for determining their performance.

Acknowledgement

The authors would like to thank Daniel Urbisch for conducting DEREK Nexus.

References

- Ade, N., Martinozzi-Teissier, S., Pallardy, M., Rousset, F. 2006. Activation of U937 cells by contact sensitizers: CD86 expression is independent of apoptosis. *J. Immunotoxicol.* 3, 189–197.
- Adenuga, D., Woolhiser, M.R., Gollapudi, B.B., Boverhof, D.R. 2012. Differential gene expression responses distinguish contact and respiratory sensitizers and nonsensitizing irritants in the local lymph node assay. *Toxicol. Sci.* 126(2), 413-425.
- Api, A.M., Parakhia, R., O'Brian, M.S., Basketter, D.A. 2017. Fragrances categorized according to their relative human skin sensitization potency. *Dermatitis* 28, 1-9.
- Ashikaga, T., Yoshida, Y., Hirota, M., Yoneyama, K., Itagaki, H., Sakaguchi, H., Miyazawa, M., Ito, Y., Suzuki, H., Toyoda, H. 2006. Development of an in vitro skin sensitization test using human cell lines: the human cell line activation test (h-CLAT) I. Optimization of the h-CLAT protocol. *Toxicol. In Vitro* 20, 767–773.
- Ashikaga T., Sakaguchi ,H., Sono, S., Kosaka, N., Ishikawa, M., Nukada, Y., Miyazawa, M., Ito, Y., Nishiyama, N., Itagaki, H. 2010. A comparative evaluation of in vitro skin sensitisation tests: the human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA). *Altern. Lab. Anim.* 38(4), 275-284.
- Basketter, D.A., Wright, Z.M., Warbrick, E.V., Dearman, R.J., Kimber, I., Ryan, C.A., Gerberick, G.F., White, I.R. 2001. Human potency predictions for aldehydes using the local lymph node assay. *Contact Dermatitis* 45(2), 89-94.

- Basketter, D.A., McFadden, J., Evans, P., Andersen, K.E., Jowsey, I., 2006. Identification and classification of skin sensitizers: identifying false positives and false negatives. *Contact Dermatitis* 55, 268–273.
- Basketter, D.A., Sanders, D., Jowsey I.R. 2007. The skin sensitization potential of resorcinol: experience with the local lymph node assay. *Cont. Derm.* 56(4), 196-200.
- Basketter, DA., Clewell, H., Kimber,I., Rossi, A., Blaauboer, B., Burrier, R., Daneshian, M., Eskes, C., Goldberg, A., Hasiwa, N., Hoffmann, S., Jaworska, J., Knudsen, T.B., Landsiedel, R., Leist, M., Locke, P., Maxwell, G., McKim, J., McVey, E.A., Ouédraogo, G., Patlewicz, G., Pelkonen, O., Roggen ,E., Rovida, C., Ruhdel, I., Schwarz, M., Schepky, A., Schoeters, G., Skinner, N., Trentz, K., Turner, M., Vanparys, P., Yager, J., Zurlo, J., Hartung, T. 2012. A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing. *ALTEX*. 29(1), 3-91.
- Basketter, D.A., Alepee, N., Ashikaga, T., Barroso, J., Gilmour, N., Goebel, C., Hibatallah, J., Hoffmann, S., Kern, P., Martinozzi-Teissier, S., Maxwell, G., Reisinger, K., Sakaguchi, H., Schepky, A., Tailhardat, M., Templier, M., 2014. Categorization of chemicals according to their relative human skin sensitizing potency. *Dermatitis* 25, 11–21.
- Basketter, D.A., Kimber, I. 2016. Phthalic anhydride: Illustrating a conundrum in chemical allergy. *J. Immunotox.* 13, 767-769.

- Bauch, C., Kolle, S.N., Ramirez, T., Eltze, T., Fabian, E., Mehling, A., Teubner, W., van Ravenzwaay, B., Landsiedel, R. 2012. Putting the parts together: Combining in vitro methods to test for skin sensitizing potentials. *Reg. Tox. Pharm.* 63, 489-504.
- Casati, S., Aschberger, K., Barroso, J., Casey, W., Delgado, I., Kim, T.S., Kleinstreuer, N., Kojima, H., Lee, J.K., Lowit, A., Park, H.K., Regimbald-Krnel, M.J., Strickland, J., Whelan, M., Yang, Y., Zuang, V. 2018. Standardisation of defined approaches for skin sensitisation testing to support regulatory use and international adoption: position of the International Cooperation on Alternative Test Methods. *Arch. Toxicol.* 92(2), 611-617
- Christensson, J.B., Matura, M., Gruvberger, B., Bruze, M., Karlberg, A.T. 2010. Linalool - a significant contact sensitizer after air exposure. *Contact Dermatitis*, 2010. 62(1), 32-41.
- Darcis, J., Goossens, A. 2016. Resorcinol: a strong sensitizer but a rare contact allergen in the clinic. *Contact Dermatitis*, 74(5), 310-312.
- Dearman, R., Basketter, D., Kimber, I., 2013. Inter-relationships between different classes of chemical allergens. *J. Appl. Toxicol.* 33, 558-565.
- de Groot, A.C. 2017. *Monographs in Contact Allergy, Volume 1: Non-Fragrance Allergens in Cosmetics (Part 1 and Part 2)*. CRC Press, Taylor & Francis Group, Boca Raton, FL.

- Dumont, C., Barroso, J., Matys, I., Worth, A., Casati, S., 2016. Analysis of the Local Lymph Node Assay (LLNA) variability for assessing the prediction of skin sensitisation potential and potency of chemicals with non-animal approaches. *Toxicology in Vitro* 34, 220-228.
- ECETOC Technical Report No. 77. Skin and Respiratory Sensitisers: Reference Chemicals Data Bank (August 1999)
- ECVAM, 2008. Murine local lymph node assay performance standards. European Centre for the Validation of Alternative Methods. October 2008. Last accessed 30.07.12. Available at: [http://ecvam.jrc.ec.europa.eu/ft_doc/ECVAM%20LLNA %20Performance%20Standards-Sec.pdf](http://ecvam.jrc.ec.europa.eu/ft_doc/ECVAM%20LLNA%20Performance%20Standards-Sec.pdf).
- ECHA – European Chemicals Agency (2016). Guidance on information requirements and Chemical Safety Assessment. Chapter R.7a: Endpoint specific guidance. Draft version 5.0, June 2016.
- Emter, R., van der Veen, J.W., Adamson, G., Ezendam, J., van Loveren, H., Natsch, A. 2013. Gene expression changes induced by skin sensitizers in the KeratinoSens™ cell line: Discriminating Nrf2-dependent and Nrf2-independent elements. *Toxicol. In Vitro*. 27(8), 2225-2232.
- Estrada, E., Patlewicz, G., Chamberlain, M., Basketter, D., Larbey, S. 2003. Computer-aided knowledge generation for understanding skin sensitization mechanisms: the TOPS-MODE approach. *Chem. Res. Tox.* 16, 1226-1235.

- Fabian, E., Vogel, D., Blatz, V., Ramirez, T., Kolle, S., Eltze, T., van Ravaenzwaay, B. Oesch, F., Landsiedel, R. 2013 Xenobiotic metabolizing enzyme activities in cells used for testing skin sensitization in vitro. *Arch. Toxicol.* 87(9), 1683-96.
- Frosch, P.J., Menné, T., Lepoittevin, J.P. 2006. *Contact Dermatitis* 4th Edition. Springer, Berlin
- Gerberick, G.F, Robinson, M.K., Ryan, C.A., Dearman, R.J., Kimber, I., Basketter, D.A., Wright, Z., Marks, J.G.. 2001. Contact allergenic potency: correlation of human and local lymph node assay data. *Am. J. Cont. Derm.* 12, 156-161.
- Gerberick, G.F., Vassallo, J.D., Bailey, R.E., Chaney, J.G., Morrall, S.W., and Lepoittevin, J.P. 2004. Development of a peptide reactivity assay for screening contact allergens. *Toxicol. Sci.* 81, 332-343.
- Gerberick, G.F., Vassallo, J.D., Foertsch, L.M., Price, B.B., Chaney, J.G., and Lepoittevin, J.P. 2007. Quantification of chemical peptide reactivity for screening contact allergens: A classification tree model approach. *Toxicol. Sci.* 97, 417-427.
- Greif, N. 1967. Cutaneous safety of fragrance material as measured by the maximization test. *Am. Perf. Cosm.* 82, 54-57.
- Gallucci, S., Lolkema, M., Matzinger, P. 1999. Natural adjuvants: endogenous activators of dendritic cells. *Nat. Med.* 5(11), 1249-55.

326
327 Hoffmann S, Edler L, Gardner I, Gribaldo L, Hartung T, Klein C, Liebsch M, Sauerland S,
328 Schechtman L, Stamatii A and Nikolaidis E. 2008. Points of reference in validation - the report
329 and recommendations of ECVAM Workshop. *Altern. Lab. Anim.* 2008, 36:343-352.
330
331 Hoffmann, S., 2015. LLNA variability: An essential ingredient for a comprehensive assessment
332 of non-animal skin sensitization test methods and strategies, *Altex*, pp. 379-383.
333
334 Hoffman, S., Kleinstreuer, N., Alépée, N., Allen, D., Api, A.M., Ashikaga, T., Clouet, E.,
335 Desprez, B., Gellatly, N., Goebel, C., Kern, P.S., Klaric, M., Kühnl, J., Lalko, J.F., Martinozzi-
336 Teissier, S., Mewes, K., Miyazawa, M., Parakhia, R., van Vliet, E., Zang, Q., Petersohn, D.
337 2018. Non-animal methods to predict skin sensitization (I); The Cosmetics Europe database.
338 *Crit. Rev. Tox.* 48, 359-374

339 ICCVAM, 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and
340 Alternative Test Methods. NIH Publication No. 03-4508. National Institute of Environmental
341 Health Sciences, Research Triangle Park, NC, Last accessed 30.07.12. Available at:
342 <http://iccvam.niehs.nih.gov/SuppDocs/submission.htm>.

343 ICCVAM, 2009. Recommended Performance Standards: Murine Local Lymph Node Assay.
344 NIH Publication 09-7357. National Institute of Environmental Health Sciences, Research
345 Triangle Park, NC, Last accessed 30.07.12. Available at:
346 http://iccvam.niehs.nih.gov/methods/immunotox/llna_PerfStds.htm.

Kern, P.S., Gerberick, G.F., Ryan, C.A., Kimber, I., Aptula, N. and Basketter, D.A. 2010. Local lymph node data for the evaluation of skin sensitization alternatives: a second compilation. *Dermatitis*, 21:8-32.

Kern, S., Dkhil, H., Hendarsa, P., Ellis, G., Natsch, A. 2014. Detection of potentially skin sensitizing hydroperoxides of linalool in fragranced products. *Anal. Bioanal. Chem.* 406(25), 6165-78.

Kleinstreuer, N.C., Hoffman, S., Alépée, N., Allen, D., Ashikaga, T., Casey, W., Clouet, E., Cluzel, M., Desprez, B., Gellatly, N., Goebel, C., Kern, P.S., Klaric, M., Kühnl, J., Martinozzi-Teissier, S., Mewes, K., Miyazawa, M., Strickland, J., van Vliet, E., Zang, Q., Petershohn, D. 2018. Non-animal methods to predict skin sensitization (II): an assessment of defined approaches. *Crit. Rev. Tox.* 48(5), 359-374.

Kolle, S.N., Basketter, D.A., Casati, S., Strickland, J., van Ravenzwaay, B., Vohr, H.W., Landsiedel, R. 2013. Performance standards and alternative assays: practical insights from skin sensitization. *Reg. Toxicol. Pharm.* 65, 278-285.

Kolle, S.N., Corrigendum to "LuSens: a keratinocyte based ARE reporter gene assay for use in integrated testing strategies for skin sensitization hazard identification." by Ramirez et al. [*Toxicol In Vitro*. 2014 Dec; 28(8):1482-97. doi: 10.1016/j.tiv.2014.08.002]. *Toxicol In Vitro*. 2019 Feb 11. pii: S0887-2333(19)30091-8. doi: 10.1016/j.tiv.2019.02.008. [Epub ahead of print]

- Kligman, A.M. 1966. The identification of contact allergens by human assay. III. The maximization test: a procedure for screening and rating contact sensitizers. *J. Invest. Dermatol.* 47(5), 393-409.
- Leontaridou, M., Urbisch, D., Kolle, S.N., Ott, K., Mulliner, D.S., Gabbert, S., Landsiedel, R. 2017. The borderline range of toxicological methods: Quantification and implications for evaluating precision. *ALTEX* 34, 525-538.
- Leontaridou, M., Gabbert, S., Landsiedel, R. 2019. The impact of precision uncertainty on predictive accuracy metrics of non-animal testing methods. Epub ahead of print. doi:10.14573/altex.1810111
- Mehling, A., Eriksson, T., Eltze, T., Kolle, S., Ramirez, T., Teubner, W., et al., 2012. Non-animal test methods for predicting skin sensitization potentials. *Arch. of Toxicol.* 86 (8), 1273–1295.
- Mehling, A., Adriaens, E., Casati, S., Hubesch, B., Irizar, A., Klaric, M., Manou, I., Müller, B.P., Roggen, E., van Vliet, E., Letasiova, S., Engelking, O., Basketter, D., In Vitro RHE Skin Sensitisation Assays: Applicability to Challenging Substances. *Manuscript in Preparation*
- Nacht, S., Yeung, D., Beasley, J.N., Anjo, M.D., Maibach, H.I. 1981. Benzoyl peroxide: percutaneous penetration and metabolic disposition. *J. Am. Dermatol.* 4(1), 31-37.

- Natsch, A., and Emter, R. 2008. Skin sensitizers induce antioxidant response element dependent genes: Application to the *in vitro* testing of the sensitization potential of chemicals. *Toxicol. Sci.* 102, 110-119.
- Narita, K., Vo, P.T.H., Yamamoto, K., Kohima, H. Itagaki, H. 2017 Preventing false-negatives in the *in vitro* skin sensitization testing of acid anhydrides using interleukin-8 release assays. *Toxicol. In Vitro.* 42: 69-75.
- Natsch, A. and Gfeller, H. 2008 LC-MS-based characterization of the peptide reactivity of chemicals to improve the *in vitro* prediction of the skin sensitization potential. *Toxicol. Sci.* 106(2), 464-478.
- Natsch, A., Bauch, C., Foertsch, L.M., Gerberick, G.F., Norman, K., Hilberer, H., Inglis, H., Landsiedel, R., Onken, S., Reuter, R., 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. *Toxicol. In Vitro* 25, 733–744.
- Natsch, A., Ryan, C.A., Foertsch, L., Emter, R., Jaworska, J., Gerberick, F., Kern, P. 2013. A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. *J. Appl. Tox.* 33, 1337-1352.
- Natsch, A. Emter, R., Gfeller, H., Haupt, T. and Ellis, G. 2015. Predicting skin sensitizer potency based on *in vitro* data from KeratinoSens and kinetic peptide binding: global versus domain-based assessment. *Tox. Sci.* 143, 319-332.

- Natsch, A. and Emter, R. 2017. Reaction chemistry to characterize the molecular initiating event in skin sensitization: A journey to be continued. *Chem. Res. Toxicol.* 30(1), 315-331.
- Natsch, A., Gfeller, H., Rothaupt, M., Ellis, G. 2017. Utility and limitations of a peptide reactivity assay to predict fragrance allergens in vitro. *Toxicol. in Vitro.* 21(7), 1220-1226.
- Nguyen, T, Nioi P, Pickett CB. 2009. The Nrf2-antioxidant response element signaling pathway and its activation by oxidative stress. *J. Biol. Chem.* 284(20), 13291-13295.
- Nukada, Y., Ashikaga, T., Miyazawa, M., Hirota, M., Sakaguchi, H., Sasa, H., Nishiyama, N. 2012. Prediction of skin sensitization potency of chemicals by human Cell Line Activation Test (h-CLAT) and an attempt at classifying skin sensitization potency. *Toxicol. In Vitro.* 26(7), 1150-1160.
- Nukada, Y. Miyazawa, M., Kazutoshi, S., Sakaguchi, H., Nishiyama, N. 2013. Data integration of non-animal tests for the development of a test battery to predict the skin sensitizing potential and potency of chemicals. *Toxicol. in Vitro.* 27, 609-618.
- OECD. 2012. OECD Series on Testing and Assessment No. 168. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Assessment. Paris: OECD Publishing.
- OECD. 2014a. The OECD QSAR Toolbox. Paris: OECD Publishing. <http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm>.

- 428 OECD. 2014b. Test No. 428. Skin absorption: in vitro method. In: OECD Guidelines for the
429 Testing of Chemicals, Section 4: Health Effects. Paris: OECD Publishing.
- 430 OECD. 2015. Test No. 442C. In Chemico skin sensitization: direct peptide reactivity assay
431 (DPRA). In: OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects. Paris:
432 OECD Publishing.
- 433 OECD. 2016a. Guidance document on the reporting of defined approaches to be used within
434 integrated approaches to testing and assessment. Series on Testing and Assessment No. 255.
435 Paris: OECD Publishing.
- 436 OECD. 2016b. Guidance Document on the Reporting of Defined Approaches to be used within
437 Integrated Approaches to Testing and Assessment. Series on Testing and Assessment No. 256.
438 Paris: OECD Publishing.
- 439 OECD. 2018a. Test No. 442D. In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method.
440 In: OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects. Paris: OECD
441 Publishing.
- 442 OECD. 2018b. Test No. 442E. In vitro skin sensitisation: assays addressing the key event on
443 activation of dendritic cells on the adverse outcome pathway for skin sensitisation. In: OECD
444 Guidelines for the Testing of Chemicals, Section 4: Health Effects. Paris: OECD Publishing.
- 445 OECD 2018c Draft Guideline on Defined Approaches for Skin Sensitisation available at
446 http://www.oecd.org/env/ehs/testing/latestdocuments/GL%20DASS_8Oct2018_Clean.pdf
447

- Paparella, M., Daneshian, M., Hornek-Gausterer, R., Kinzl, M., Mauritz, I., Mühleegger, S. (2013): Food for Thought...Uncertainty of testing Methods – What Do We (Want To) Know? *ALTEX* 30, 131-143
- Parkinson, E., Aleksic, M. 2018. Determination of Protein Haptenation by Chemical Sensitizers Within the Complexity of the Human Skin Proteome. *Toxicol. Sci.* 162, 429-438.
- Patlewicz, G.Y., Wright, Z.M., Basketter, D.A., Pease, C.K., Lepoittevin, J.P., Arnau, E.G. 2002 Structure-activity relationships for selected fragrance allergens. *Contact Dermatitis* 47(4), 219-26.
- Patlewicz, G., Roberts, D.W., Uriarte, E., 2008. A comparison of reactivity schemes for the prediction skin sensitization potential. *Chem. Res. Toxicol.* 21 (2), 521– 541.
- Patlewicz, G., Casati, S., Basketter, D.A., Asturiol, D., Roberts, D.W., Lepoittevin, J-P., Worth, A.P., Aschberger, K. 2016. Can currently available non-animal methods detect pre and pro-haptens relevant for skin sensitization. *Reg. Tox. Pharm.* 82:147-155.
- Payne, M.P., Walsh, P.T. 1994. Structure-activity relationships for skin sensitization potential: development of structural alerts for use in knowledge-based toxicity prediction systems. *J. Chem. Inf. Comput. Sci.* 34(1), 154-61.

- Pirisino, R., Di Simplicio, P., Ignesti, G., Bianchi, G., Barbera, P. 1988. Sulfhydryl groups and peroxidase-like activity of albumin as scavenger of organic peroxides. *Pharmacol. Res. Com.* 20, 546-552.
- Piroird, C., Ovigne, J.M., Rousset, F., Martinozzi-Teissier, S., Gomes, C., Cotovio, J., Alépée, N. 2015. The Myeloid U937 Skin Sensitization Test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome pathway for skin sensitization. *Toxicol In Vitro.* 29(5):901-16.
- Ramirez, T., Mehling, A., Kolle, S.N., Wruck, C.J., Teubner, W., Eltze, T., Aumann, A., Urbisch, D., van Ravenzwaay, B., Landsiedel, R. 2014. LuSens: A keratinocyte based ARE reporter gene assay for use in integrated testing strategies for skin sensitization hazard identification. *Toxicol. in Vitro.* 28, 1482-1497.
- Roberts, D.W., Aptula, A.O., Patlewicz, G., 2007. Electrophilic chemistry related to skin sensitization. Reaction mechanistic applicability domain classification for a published dataset of 106 chemicals tested in the mouse local lymph node assay. *Chem. Res. Toxicol.* 20, 44–60.
- Roberts, D.W. and Natsch, A. 2009. High throughput kinetic profiling approach for covalent binding to peptides: Application to skin sensitization potency of michael acceptor electrophiles. *Chem. Res. Toxicol.* 22(3), 592-603.
- Rovida, C., Alépée, N., Api, A.M., Basketter, D.A., Bois, F.Y., Caloni, F., et al., 2014. Integrated Testing Strategies (ITS) for safety assessment. *ALTEX.* 32, 25-40.

- Sakaguchi, H., Ashikaga, T., Miyazawa, M. Yoshida, Y., Ito, Y., Yoneyama, K., Hirota, H., Toyoda, H., Suzuki, H. 2006. Development of an in vitro skin sensitization test using human cell lines; human Cell Line Activation Test (h-CLAT) II. An inter laboratory study of the h-CLAT. *Toxicol in Vitro* 20, 767–773.
- Sakaguchi, H., Ryan, C.A., Ovigne, J.M., Schroeder, K.R., Ashikaga, T. 2010. Predicting skin sensitization potential and inter-laboratory reproducibility of a human Cell Line Activation Test (h-CLAT) in the European Cosmetics Association (COLIPA) ring trials. *Toxicol. in Vitro* 24, 1810-1820.
- Sauer, U.G., Hill, E.H., Curren, R.D., Raabe, H.A., Kolle, S.N., Teubner, W., Mehling, A., Landsiedel, R. 2016. Local tolerance testing under REACH: Accepted non-animal methods are not on equal footing with animal tests. *Altern. Lab. Animal.* 44, 281-299.
- SCCP. 2007. Memorandum on hair dye substances and their skin sensitizing properties. Brussels, Belgium:European Commission. Available: http://ec.europa.eu/health/ph_risk/committees/04_sccp/docs/sccp_s_05.pdf [accessed 17 December 2010].
- Scientific Committee on Consumer Safety, 2012. Opinion on fragrance allergens in cosmetic products. Adopted at the 15th Plenary Meeting, 26Y27 June, 2012. Available at: http://ec.europa.eu/health/sites/health/files/scientific_committees/consumer_safety/docs/scs_o_102.pdf. Accessed December 21, 2016.

Schnuch, A., Uter, W., Geier, J., Lessmann, H., Frosch, P.J. 2007. Sensitization to 26 fragrances to be labelled according to current European regulation. Results of the IVDK and review of the literature. *Contact Dermatitis*. 57(1), 1-10.

Schnuch, A., Mildau, G., Kratz, E.M., Uter, W., 2011. Risk of sensitization to preservatives estimated on the basis of patch test data and exposure, according to a sample of 3541 leave-on products. *Contact Dermatitis* 65, 167–174.

Schnuch, A., Wolfgang, U. Lessmann, H., Geier, J. 2015. Risk of sensitization to fragrances estimated on the basis of patch test data and exposure, according to volume used and a sample of 5451 cosmetic products. *Flavour Frag. J.*, 30(3), 208-217.

Strickland, J., Zang, Q., Paris, M., Lehmann, D.M., Allen, D., Choksi, N., Matheson, J., Jacobs, A., Casey, W., Kleinstreuer, N. 2017. Multivariate models for prediction of human skin sensitization. *J. Appl. Toxicol.* 37, 347-360.

Schneider, K., Akkan, Z., 2004. Quantitative relationship between the local lymph node assay and human skin sensitization assays. *Regul. Toxicol. Pharmacol.* 39, 245–255.

Takenouchi, O., Fukui, S., Okamoto, K., Kurotani, S., Imai, N., Fujishiro, M., Kyotani, D., Kato, Y., Kasahara, T., Fujita, M., Toyoda, A., Skiya, D., Watanabe, S., Seto, H., Hirota, M., Ashikaga, T., Miyazawa, M. 2015. Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin sensitizing potential and potency of chemicals. *J. Appl. Tox.* 35, 1318-1332.

- Thyssen, J.P., Linneberg, A., Johansen, J.D., 2007. The epidemiology of contact allergy in the general population – prevalence and main findings. *Contact Dermatitis* 57, 287–299.
- Troutman, J.A, Foertsch, L.M., Kern, P.S., Dai, H.J., Quijano, M., Dobson, R.L., Lalko, J., Lepoittevin, J-P. and Gerberick G.F. 2011. The incorporation of lysine into the peroxidase peptide reactivity assays for skin sensitization assessments. *Toxicol. Sci.* 122, 422-436.
- Urbisch, D., Becker, M., Honarvar, N., Kolle, S.N., Mehling, A., Teubner, B., Landsiedel, R. 2016. Assessment of Pre- and Pro-haptens using nonanimal test methods for skin sensitization. *Chem. Res. Toxicol.*, 29, 901-913.
- Urbisch, D., Mehling, A., Guth, K., Ramirez, T., Honarvar, N., Kolle, S., Landsiedel, R., Jaworska, J., Kern, P.S., 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. *Reg. Toxicol. Pharm.* 71, 337-351.
- van der Veen, J.W., Rorije, E., Emter, R., Natsch, A., van, L.H., Ezendam, J., 2014. Evaluating the performance of integrated approaches for hazard identification of skin sensitizing chemicals. *Regul. Toxicol. Pharmacol.* 69 (3), 371–379.
- Vocanson, M., Goujon, C., Chabeau, G., Castelain, M., Valeyrie, M., Floc’h, F., Maliverney, C., Gard, A., Nicolas, J.F. 2006. The skin allergenic properties of chemicals many depend on contaminants - evidence from studies on coumarin. *Int. Arch. Allergy Immunol.* 140, 231-238.

1 **Highlights**

- 2 • The 2 out of 3 approach is a simple approach that has demonstrated very good
- 3 predictivity for the skin sensitization hazard
- 4 • Chemicals with none or only one positive outcome are predicted non-sensitizers some of
- 5 which may be false predictions
- 6 • In this report we have thoroughly reviewed the results of 29 chemicals with 1 out of 3
- 7 positive results
- 8 • 8 chemicals were classified as positive in vivo and would be false-negative with the 2 out
- 9 of 3 approach
- 10 • Of those 8, 7 chemicals were borderline in at least one assay and the 8th missed substance
- 11 was probably out of domain

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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

GFG reports personal fees from BASF SE, during the conduct of the study; and GFG was involved in the development of the DPRA. AN is employee of Givaudan. Givaudan developed the KeratinoSens assay and is using the 2 out of 3 approach to register fragrance molecules. RL and SNK are employees of BASF, a chemical company which developed the LuSens assay and the 2 out of 3 approach and is using the 2 out of 3 approach to develop and register commercial products.