Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2021)11/ANN6

Unclassified

English - Or. English

16 July 2021

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

ANNEX 6: ANALYSIS OF LLNA REFERENCE DATA TO CONCLUDE ON PREDICTIVITY OF ALTERNATIVE METHODS FOR SKIN SENSITISATION FOR LIPOPHILIC CHEMICALS

Series on Testing and Assessment,
No. 336

JT03479532

OECD Expert Group on Defined Approaches for Skin Sensitisation

Annex 6: Analysis of LLNA reference data to conclude on predictivity of alternative methods for skin sensitisation for lipophilic chemicals

Authors:

Andreas Natsch, Switzerland; David Asturiol, European Commission-Joint Research Centre

IOMC
INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS
A cooperative agreement among FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD

Environment Directorate
ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT
Paris 2021

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 36 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in eleven different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides**; **Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents;** and **Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

Abstract

It had been observed, that the sensitivity of h-CLAT, KeratinoSens™ and DPRA predictions against LLNA reference classifications is lower for lipophilic chemicals with a LogP ≥3.5. This is not due to a lack of exposure by these poorly soluble chemicals, as cytotoxicity in most cases is reached in cell-based assays. An initial analysis of the human reference data for these chemicals indicated instead, that most of the false-positive LLNA results cluster in this particular physiochemical range. In this meta-analysis, this observation was investigated in detail using different reference data: (i) The human data curated by the DASS expert group, (ii) the expert judgment on human sensitisation data published by Cosmetics Europe and (iii) a weight-of-evidence analysis based on DASS data and clinical data. All three analyses indicate, that in this physicochemical range, the specificity of the LLNA as reported in the DASS reference database is reduced, which may explain the apparent reduced sensitivity of *in vitro* data when evaluated against the LLNA only. Further support comes from the analysis of studies identifying false-positives in the LLNA vs. guinea pig data and *in silico* structural alerts. Analysis of the reference data from the LLNA validation study indicates that this could not have been recognized at the time of LLNA validation due to a lack of comparative data in this particular physicochemical range in the validation set. This analysis indicates that for negative calls in *in vitro* assays in this physicochemical range resorting to the LLNA is not necessarily a better option.

1. Background: Rationale to analyse LLNA reference data for lipophilic chemicals

While analysing predictivity of the *in vitro* tests as stand-alone methods vs. LLNA data and only based on the curated DASS reference database, a reduced sensitivity of all three tests in OECD TG 442C (DPRA), 442D (KeratinoSens™) and 442E (h-CLAT) for chemicals with a LogP ≥ 3.5 was observed [see Annex V]. In addition a reduced sensitivity of the h-CLAT as stand-alone method vs. LLNA data had been reported before (Takenouchi et al., 2013) and was included as a limitation in OECD TG442E (OECD, 2018). The concomitant reduced sensitivity for all three assays was hypothesized by subsequent analysis to be caused by a potentially limited exposure by poorly soluble chemicals when tested in the aqueous (KeratinoSens™, h-CLAT) or partly aqueous (DPRA) incubation media (hypothesis presented by DK to DASS expert group; 5.7.2020).

Nevertheless, a more in-depth evaluation of the data generated with the cell-based assays, shows that the vast majority of the test chemicals with a LogP ≥ 3.5 producing false negative results compared to LLNA classifications, while not inducing the markers for positivity (luciferase induction or CD86 / CD54 expression), still lead to cytotoxicity in KeratinoSens™ and/or h-CLAT tests[1]. These data prove that in the given exposure range, these chemicals with a LogP≥3.5 provide sufficient cell exposure. Furthermore, for the DPRA, a recent study (Yamamoto et al., 2019) investigated which of the 82 chemicals initially tested by Gerberick *et al.* (Gerberick et al., 2007) to assess DPRA predictivity lead to visible precipitation in the DPRA. Precipitation would indicate that the solution is oversaturated and that the dissolved exposure concentration is below the nominal concentration. Analysing the data in that study, the predictivity of the DPRA as stand-alone method is still at 80% (20/25) for those partly dissolved chemicals. This indicates that the reaction can proceed with the amount of the test chemical being in solution, and additional test chemical can get dissolved as the dissolved test chemical has reacted, in case the reaction rate is slower than the dissolution rate. These observations indicate that a limited exposure for more hydrophobic chemicals due to insolubility is not a general reason for not detecting the biological activity (in cell assays) or reactivity (in *in chemico*) assays.

[1] For h-CLAT 19/26 (73%) of all false-negatives vs. LLNA have a reported CV75 for cytotoxicity; for 4/26 a value of 1000 µg/ml, i.e. below max test concentration of h-CLAT is reported. For the chemicals with LogP≥ 3.5, this amounts to 10/12 (83%) with two chemicals only tested up to 1000 µg/ml; for KeratinoSens™, with a lower maximal test concentration, for 22/39 (56%) of all the false-negatives vs. LLNA, 50% cytotoxicity was reached. In the range of LogP≥3.5, 13/13 (100%) of the chemicals caused at least 50% cytotoxicity in the tested concentration range.

Given that the observation of reduced sensitivity as stand-alone methods was made for all three assays, and only for evaluations against LLNA data, there exists the theoretical possibility that, actually, the LLNA produces more false-positive in the physiochemical range of LogP≥ 3.5 than for other ranges. A higher number of false-positives for the LLNA reference data in the high LogP range would automatically lead to an apparent reduced sensitivity of all the alternative methods and defined approaches (DA) in which such methods are integrated. The question then is not "is the reference data (LLNA) wrong in general in that physicochemical range" – rather the question is: "Are there indications of a *reduced specificity* of the LLNA in that physicochemical range?" –.

This analysis is of particular importance, since during the curation of the DASS reference database and based on the evaluation criteria applied to review the LLNA data for inclusion as positives or negatives into the DASS reference database, a number of observations with regard to the specificity of the LLNA reference data were made by the expert group:

- Of the five negative reference substances reported in the LLNA performance standards (OECD TG429), one (salicylic acid, LogP= 2.26) is positive based on the data review, for two (chlorobenzene (LogP= 2.84) and methyl salicylate LogP= 2.55) no conclusive LLNA outcome could be made, for one (lactic acid, LogP= -0.72) a negative attribution could only be made by an expert assessment and taking irritation data from other information sources (irritation not measured in the LLNA) into account and only one chemical (isopropanol. LogP= 0.05) was considered to be clearly negative.
- One of the guidelines recommended LLNA vehicles (DMSO LogP= -1.35) is positive in the LLNA in the DASS reference database.
- The specificity of the LLNA vs. human is only 22.5% (only 2 out of 9 human non-sensitisers are correctly predicted by the LLNA). While this particularly low specificity may partly be due to low number of chemicals, the specificity would still be limited in case *e.g.* 10 extra human negative chemicals with concordant LLNA negative classifications were considered in the evaluation, which is rather unlikely given the amount of LLNA false-positives in the first 9 evaluated chemicals.
- For 2 more of human negative chemicals, the LLNA was non-conclusive based on the data review.

These observations indicate that the specificity of the LLNA as evaluated based on the DASS reference database may be different from that calculated during the validation of the LLNA. This may be partly linked to the fact that the evaluation criteria applied today are largely based on a required higher test dose in OECD TG429 as compared to the typical maximal test concentrations used during the LLNA validation (Kolle et al., 2020).

2. Analysis of the DASS reference database: Predictivity of LLNA vs. human data in the two physicochemical ranges

Composition of the DASS DB

The DASS LLNA and human database (HDSG) was constructed starting from a database collated previously at the United States National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). The chemical set was later expanded based on additional chemicals with LLNA and *in* vitro data as reported in Urbisch *et al* (Urbisch et al., 2015). This dataset was later complemented with all publicly available reference data (LLNA and human) as well as with further data provided by the Research Institute for Fragrance Materials (RIFM). The dataset was subject to a strict curation procedure which is described in detail in the Annex 4 and 5. Some of the most relevant criteria which explain the final composition of the DASS DB are listed below:

HDSG

- Only substances with data from HRIPT (human repeat insult patch test) or HMT (Human maximisation test) were included; clinical data and use data were excluded from the evidence contrary to the compilation by Basketter et al.(Basketter et al., 2014)
- Only substances that have been tested at ≥25% and have not caused sensitisation are considered NC

LLNA

- In the LLNA reference database, study results were accepted as negative, if for all test concentrations SI values were < 3 and if the test substance was tested up to a highest concentration tested (HCT) of at least 50% unless a justification was available that the tested concentration was the highest achievable (for technical or toxicological reasons).

The DASS DB contains 168 chemicals with unambiguous LLNA hazard classifications as agreed by the EG, from now on this will be referred to as DASS DB-168 (see Table 1).

These 168 chemicals correspond to 135 LLNA positive (sensitisers) (80%) and 33 LLNA negative (non-sensitisers) (20%). Of these 168 chemicals, only 56 have human HDSG data considered unambiguous by the EG, and these chemicals correspond to 47 human positives (84%) (also named human sensitisers throughout the text) and 9 human negatives (16%) (also named human non-sensitisers throughout the text). All statements regarding human sensitisers based on the HDSG analysis refer to human predictive patch test data, while all reference to human sensitisers in the Basketter database discussed below refer to an expert WoE assessment made on human sensitisation potential.

It is obvious that the strict curation procedure applied by the EG yielded a high-quality database. This database, however, is also highly biased towards sensitising substances (>80%) and contains a low number of human non-sensitisers (n=9). The distribution of substances in the database is not ideal to try to understand the performance of the methods in a real scenario. A method with systematic skin sensitisation overprediction would look better in a database with a higher number of sensitisers, as most of the sensitisers would be correctly predicted, while many of the few non-sensitisers would be incorrectly predicted as non-sensitisers. As a consequence, the low performance in predicting non-sensitisers would be masked because it would be difficult to understand whether the reason for the low performance was the systematic overprediction or the low number of chemicals. It is estimated that only ~20% of the substances placed in the market are skin sensitisers (Experts from DK, personal communication based on QSAR predictions). With such numbers of market prevalence, a method that would correctly predict 80% of sensitisers and only 50% of non-sensitisers would, in reality, incorrectly predict 45% of the chemicals to be placed in the market. However, such a biased distribution can be an advantage if one wants to make sure to not miss any sensitising substance.

Kathon CG did not have an assigned LogP value in the DASS DB-168. The active components of Kathon CG are Methylchloroisothiazolinone (MCI) and Methylisothiazolinone (MI) with a proportion between them of about 76% MCI/24% MI. Both components of the mixture have calculated LogP<3.5. According to the Kathon CG data safety sheet, the measured LogP is 0.401. Consequently, this mixture will be considered in the group of substances of LogP<3.5.

Table 1. Composition of the DASS DB-168 in terms of LLNA and HDSG results, and their distribution based on LogP values.

Subset	Total	Positive	Negative
LLNA	168	135 (*80%*)	33 (*20%*)
LogP≥3.5	39 (23%)	33 (24%)	6 (18%)
LogP<3.5	129 (77%)	102 (76%)	27 (82%)
HDSG	56	47 (*84%*)	9 (*16%*)
LogP≥3.5	12 (21%)	6 (13%)	6 (66%)
LogP<3.5	44 (79%)	41 (87%)	3 (33%)

Values in parenthesis correspond to the proportion that each value represents with respect to the total of that row or column. The percentages of the LLNA and HDSG rows (shown in italics) are with respect to the total of the row (*e.g.* for the LLNA Positives=135 (80%), the percentage of 80% is with respect to LLNA Total, 135/168=80%), while the percentages of the LogP rows are with respect to the total of the column (*e.g.* for LLNA LogP≥3.5 Positives =33(24%), the percentage of 24% is with respect to the LLNA Positives, 33/135=24%).

Predictivity of LLNA vs HDSG

In order to confirm whether a) DPRA, KeratinoSens™ and h-CLAT suffer from an increase rate of false-negatives for substances with high LogP, *i.e.* LogP≥3.5, when compared to the LLNA, or b) if instead the LLNA has an increased rate of false-positives compared to human data for substances with high LogP, the predictivity of LLNA with respect to HDSG for the DASS DB has been analysed. Performance statistics have been calculated for the overall dataset, and for the subsets of substances with LogP<3.5 and LogP≥3.5.

Table 2. Predictivity of the LLNA for chemicals of the DASS DB which have LLNA and HDSG data available (n=56). Three subsets have been calculated: all (n=56), LogP<3.5 (n=44), and LogP≥3.5 (n=12)

Set	FN	FP	TN	TP	n	Acc.	Bal-Acc	Sens.	Spec.	PPV	NPV	Prev.
All	3	7	2	44	56	0.82	0.58	0.94	0.22[1]	0.86	0.4	0.8
LogP<3.5	3	2	1	38	44	0.89	0.63	0.93	0.33	0.95	0.25	0.93
LogP≥3.5	0	5	1	6	12	0.58	0.58	1	0.17	0.55	1	0.5

[1)] The specificity values are based on a relatively low number of chemicals. Clopper Pearson 95% confidence intervals were thus calculated: The CI for the specificity of 0.22 is [0.02, 0.60].

Table 2 shows that out of a total of 56 substances with LLNA and HDSG data, 10 have discordant LLNA and human data (18%). Out of these 10 substances, 7 correspond to FPs and 3 to FNs. In terms of LogP<3.5 and LogP≥3.5 subsets, 5 of the discrepancies correspond to LogP<3.5 substances and the other 5 to LogP≥3.5 substances. While the number of substances with discordant results is the same for both subsets, their proportion is higher for the LogP≥3.5 subset, as this subset contains a total of 12 substances (5/12=42%) while the other contains a total of 44 substances (5/44=11%).

The table above shows that the LLNA is very good at predicting human sensitisers, since 44 out of the 47 human sensitisers are correctly predicted by the LLNA. This is reflected in a sensitivity of 0.94. This predictivity for human sensitisers seems not to be dependent on LogP as the sensitivity for the two sets, LogP<3.5 and LogP≥3.5, is 0.93 and 1.0, respectively. However, the table also shows that the LLNA is not good at predicting human non-sensitisers. Only 2 out of the 9 human non-sensitisers are correctly predicted by the LLNA. This is reflected in a specificity of 0.22. This value is highly uncertain due to low number of chemicals, but the 95% confidence interval indicates that it actually is below 0.6. Specificity changes significantly for LogP<3.5 and LogP≥3.5 as it is of 0.33 (1 correctly predicted HDSG non-sensitiser out of 3) and 0.17 (1 correctly predicted out of 6), respectively. These observations cannot be considered conclusive as it is difficult to establish a cause-effect relationship with this low number of substances. Further analysis and data shown below, however, will support this option.

Interpretation of the results
The table above clearly shows poor performance of the LLNA when predicting HDSG non-sensitisers. The overall specificity of 0.22, although with a rather low number of chemicals (*i.e.* 9 HDSG non-sensitisers), is an indication that the LLNA is prone to provide FP results. Therefore, when comparing methods against the LLNA, the sensitivity and number of FNs must be analysed and interpreted with care as there are high chances (7 out of 9 if based on the DASS DB) that what is considered a FN when compared to the LLNA, is in reality a TN when compared to human data.
Due to the low number of LogP≥3.5 chemicals with HDSG data in the DASS DB (n=12), and especially in the subgroup of LogP≥3.5 HDSG non-sensitisers (n=6), it is difficult to draw conclusions on whether the LLNA is more prone to produce FPs for the LogP≥3.5 than for the LogP<3.5. The prevalence of HDSG sensitisers/non-sensitisers is 0.93 for Log<3.5 and 0.5 for LogP≥3.5, which means that the ratio of HDSG non-sensitisers is higher in the subgroup LogP≥3.5, but it is also closer to the estimated "reality" of 20% sensitisers in the chemical universe. This difference in prevalence between the high and low LogP substances can make the statistics of the LogP≥3.5 subset look worse than what they really are. However, the results for the LLNA specificity are so poor that even if 10 human non-sensitisers with concordant LLNA negative results were added to the LogP≥3.5 group there would still be 5 out of 16 human non-sensitisers predicted as FP by the LLNA.

Another possibility to look at the data is to count the FP vs. human in the LLNA as a fraction of the total LLNA positives in order to understand how many of the LLNA positive predictions are real sensitisers in humans, *i.e.* determine the so-called False Discovery Rate (FDR) =FP/(FP+TP). There are 40 chemicals in the DASS DB with LogP<3.5, positive LLNA results, and human data. Of these, only 2 are FP in the LLNA, therefore the FDR for the LLNA in the range of low LogP substances is 2/40 (5%). Similarly, the FDR of LLNA in the high LogP range is 5/11 (45%). Due to the higher FDR of LLNA in the LogP≥ 3.5 range, when evaluating a method with equal predictivity for low and high LogP substances using the LLNA as reference, this method would show a lower sensitivity in the high LogP range.

In order to shed light on the increase FDR rate of LLNA for high LogP substances, one can further explore Basketter *et al.* (Basketter et al., 2014) human data compilation, as it significantly enlarges the number of substances of the DASS DB-168 that have human data from 56 to 96.

3. Analysis of the DASS LLNA reference database vs. human data collected by Basketter et al. in various physicochemical ranges

The compilation made by Basketter *et al.* (Basketter et al., 2014) took human evidence from different sources, namely clinical data, history of safe, widespread use, HMT and HRIPT into account. This expert judgment, which, however, lacked documentation on the sources evaluated and used in the overall assessment of the clinical data and the way in which they were selected, does not follow the strict rules applied by the DASS data review, but it has the advantage of presenting a much larger dataset. The HDSG required a negative predictive test to be conducted at ≥25% to accept the negative result, this high test concentration was not used for many chemicals, which led to the low number of human NC.
Compared to the DASS reference database, the LLNA has clearly a higher predictivity for human data when evaluated with the Basketter *et al.* dataset, which comprises 68 sensitisers and 28 non-sensitisers.

Table 3. Predictivity of the LLNA for chemicals of the DASS DB which have LLNA and human data from Basketter et al. (n=96). Three subsets have been calculated: all (n=96), LogP<3.5 (n=71), and LogP≥3.5 (n=25)

Set	FN	FP	TN	TP	n	Acc.	Bal-Acc	Sens.	Spec.	PPV	NPV	Prev.
All	1	17	11	67	96	0.81	0.69	0.99	0.39[1]	0.8	0.92	0.71
LogP<3.5	1	8	10	52	71	0.87	0.77	0.98	0.56	0.87	0.91	0.75
LogP≥3.5	0	9	1	15	25	0.64	0.55	1	0.10	0.62	1	0.60

[1)] The specificity values are based on a relatively low number of chemicals. Clopper Pearson 95% confidence intervals were thus calculated: The CI for the specificity of 0.39 is [0.21, 0.59].

Table 3 shows that out of a total of 96 substances, 18 have discordant LLNA and human data (Basketter et al., 2014) (19%). Of these 18 substances, 17 correspond to FPs and only 1 to FNs. Nine of the discrepancies correspond to LogP<3.5 and the other 9 to LogP≥3.5 substances. While the number of substances with discordant results is the same for both subsets, their proportion is higher for the LogP≥3.5 subset, as this subset contains a total of 25 substances (9/25=36%) while the LogP<3.5 subset contains a total of 71 (9/71=13%). Despite the difference in absolute numbers, the percentage of discrepancies and their distribution within the subsets is almost identical to that of the LLNA-HDSG comparison.

Table 3 shows that the LLNA is very good at predicting Basketter human sensitisers, 67 out of the 68 human sensitisers are correctly predicted by the LLNA. This is reflected in a sensitivity of 0.99. This predictivity for human sensitisers seems not to be dependent on the LogP as the sensitivity for the two subsets, LogP<3.5 and LogP≥3.5, is 0.98 and 1.0, respectively. The same excellent sensitivity was observed in the LLNA-HDSG comparison. Table 3 also shows that the LLNA is not good at predicting Basketter human non-sensitisers (Basketter et al., 2014) as only 11 out of the 28 human non-sensitisers are correctly predicted by the LLNA. This is reflected in a specificity of 0.39 which, although being poor, almost doubles that of the LLNA-HDSG comparison, which was 0.22. Again, this value may be influenced by the low number of chemicals, but similar to the analysis above, the 95% confidence interval has an upper limit of 0.59, indicating limited specificity, even if the exact value may be uncertain. Specificity also changes significantly for LogP<3.5 and LogP≥3.5 subsets as it is 0.56 (10 Basketter non-sensitisers correctly predicted out of 18) and 0.10 (1 Basketter non-sensitiser correctly predicted out of 10), respectively. Comparing with the LLNA-HDSG LogP≥3.5 subgroup, the LLNA-Basketter comparison contains 4 human non-sensitisers more (n=6 vs. n=10) and all 4 are FPs in the LLNA when compared to human data (9 FP/10NS in the LLNA-Basketter vs 5FP/6NS in the LLNA-HDSG).

Interpretation of the results

The comparison of the DASS DB-168 substances with LLNA data and human data compiled by Basketter et al. (Basketter et al., 2014) shows a very similar picture to that provided by the LLNA-HDSG comparison. In both comparisons the LLNA provides discordant classifications in ~20% of the cases, the proportion being higher for LogP≥3.5 substances (~40%) than LogP<3.5 (~10%). In both comparisons, the predictivity of the LLNA for human sensitisers are excellent, with sensitivity values ≥90% irrespective of the LogP value, and being close to 1.0 for the LLNA-Basketter comparison. The specificities of the two LogP subsets in the LLNA-Basketter comparison are significantly different, being 0.56 for LogP<3.5 and 0.1 for LogP≥3.5 substances. The LLNA-Basketter LogP<3.5 subset contains 15 more human non-sensitisers than the LLNA-HDSG LogP<3.5, of which 9 were correctly predicted and 6 were FP, resulting in a total of 10 out of 18 human non-sensitisers correctly predicted by the LLNA. For the LogP≥3.5 subgroup, the 4 extra human non-sensitisers present in the LLNA-Basketter with respect to LLNA-HDSG were all FPs in the LLNA, resulting in a total of only 1 correctly predicted Basketter non-sensitiser out of 10. Therefore, the Basketter *et al.* data, although compiled according to less stringent criteria, confirms the trends observed in the LLNA-HDSG comparison.
Both datasets, LLNA-HDSG and LLNA-Basketter show an increased rate of FP vs human data in the LLNA-

positives for LogP≥3.5 substances. In the LLNA-HDSG case, only 1 out of 6 human non-sensitisers is correctly predicted by the LLNA (specificity=0.17), while for the LLNA-Basketter comparison, only 1 out of 10 human non-sensitisers is correctly predicted by the LLNA (specificity=0.1) in this physicochemical range.

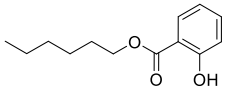
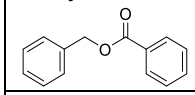
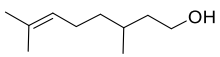
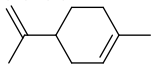
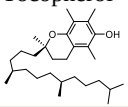
The use of the human data compiled by Basketter et al. (Basketter et al., 2014) has shed light on the predictivity of the LLNA vs. human. The LLNA-HDSG LogP<3.5 subset had a high FP rate (2 FPs out of 3 human non-sensitisers, *i.e.* specificity=0.33) possibly due to the small number of substances present in the subset. In fact, the data compiled by Basketter et al. shows that this was probably the case, as it contains 8 FP out of 18 human non-sensitisers, *i.e.* specificity=0.56 which is closer to the commonly reported values (Haneke et al., 2001; Hoffmann et al., 2018) for the specificity of LLNA vs. human. The prevalence of sensitisers/non-sensitisers is more balanced in the LLNA-Basketter as it is of 0.73 instead of 0.93. The LLNA-HDSG LogP≥3.5 subset showed an even higher FP rate with 5 FP out of 6 human non-sensitisers. In this case, however, the Basketter et al. data, has not changed the observation of a higher FP rate, else it has reinforced it, as the 4 extra human non-sensitisers found in the Basketter et al. data are all FP, and are reflected in a near to 0 predictivity of LLNA for human non-sensitisers with LogP≥3.5 of only 1 correctly predicted Basketter non-sensitiser out of 10, *i.e.* specificity=0.1. In addition, the prevalence of the LogP≥3.5 subset in the LLNA-Basketter comparison is more similar to that of LogP<3.5 subset, 0.6 vs 0.73, which is better for comparing two subgroups. This shows that the difference in prevalence was not the responsible for the higher rate of FPs of LLNA for LogP≥3.5 substances in the analysis of the subset with HDSG data.

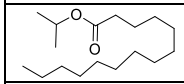
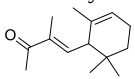
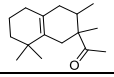
4. Analysis of the DASS reference database: Zooming in on single chemicals potentially false-positive in the LLNA

A human call was only made in the DASS reference database if the chemical had been tested in HRIPT or HMT test, and a negative call was only made if the maximal test concentration in this test was at least 25%. Next to HRIPT and HMT data, other evidence may be consulted on whether a chemical is a human non-sensitiser (use data, clinical patch test data). These data can be used in a weight-of-evidence to conclude whether a chemical is a skin sensitiser, although they do not meet the criteria set for inclusion in the DASS reference data which limited the evidence to high dose HRIPT and HMT testing. A more detailed documentation on a chemical-by-chemical level is provided here for a number of chemicals for which we consider sufficient evidence is available to rate them as human non-sensitisers. This documentation is more detailed as compared to the one provided by Basketter *et al.*(Basketter et al., 2014).

There is evidence that shows that the lipophilic chemicals of Table 4 are human non-sensitisers despite the fact that they are rated positive in the LLNA. Thus, there is strong evidence that Hexyl salicylate, Benzyl benzoate, Citronellol, (R)-(+)-Limonene, Tocopherol, Isopropyl myristate, iso-Methylionone and OTNE are not relevant human sensitisers based on HDSG analysis and /or weight of evidence (WoE) as discussed in detail for each chemical in Table 4; consequently, the LLNA for these chemicals is false-positive. Given that the DASS database contains 33 LLNA positives with a LogP ≥ 3.5, 8 *bona fide* false-positives vs. human data identified by this WoE analysis in Table 4 indicate a FDR rate of at least 24% vs. human data.
Among the 102 chemicals with a LogP < 3.5 and positive LLNA, we identified four potential false-positives vs. the HDSG and/or a WoE analysis (see Table 2). This indicates a FDR of at least 4% *bona fide* false-positives in the DASS-DB for chemicals with cLogP < 3.5. Since the FDR for higher LogP chemicals is roughly 20% higher than that of the low LogP, we would expect for any DA evaluated against the DASS reference database a reduced sensitivity vs. LLNA in the range of high LogP (if the approach is predictive for human data). Interestingly, the '2 out of 3' DA indeed rated Hexyl salicylate, (R)-(+)-Limonene, Tocopherol, Isopropyl myristate, iso-Methylionone and OTNE as negative, *i.e.* six of the eight chemicals for which the WoE / HDSG data would indicate no relevant human sensitisation potential, while Benzyl benzoate is borderline based on the Borderline Analysis (see Annex 7).

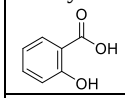
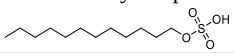
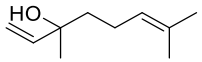
Table 4. Lipophilic chemicals in the DASS database which are rated as sensitisers by LLNA, but which are non-sensitisers in humans based on the DASS analysis and/or a WoE analysis

Chemical	CASRN	LLNA MLLP	HU GHS	LogP	Details and Reference
Hexyl salicylate 	6259-76-3	14.09	NC	3.18 (5.5 exp. REACH)	Considered human non-sensitiser by HDSG on available predictive tests, no structural alert
Benzyl benzoate 	120-51-4	17	NC	3.97	Considered human non-sensitiser by HDSG based on available predictive tests; very low SEQ in Schnuch *et al.*(Schnuch et al., 2015) (0; based on 1 positives in 2003 tests (Schnuch et al., 2007)) indicates very low frequency of clinical reaction compared to volume
Citronellol 	106-22-9	43.5	NC	3.91*	Considered human non-sensitiser by HDSG. Has been tested at 25% in 101 panellists with no positive reactions (see human subgroup report, RIFM08). SEQ 0.12/0.1 in Schnuch *et al.*(Schnuch et al., 2015) indicates very low frequency compared to use volume. (As for limonene and linalool, artificially oxidized samples are positive (Rudback et al., 2014))
(R)-(+)-Limonene 	5989-27-5	52.5	NC/1B	4.51	For Commercial Limonene very rare cases of contact allergy are reported, despite very widespread use. The SCCS opinion on fragrance materials (2012) lists patch test data on 5500 subjects with only ca. 15 reported cases globally over many years, despite very widespread use of limonene (present in majority of cosmetic products investigated, thus almost daily contact for the whole population). Schnuch et al. (Schnuch et al., 2015) indicated SEQ of 0.07/0.08, *i.e.* at the very low end of the scale. Very high exposure to almost pure limonene occurs regularly upon manually peeling of citrus fruits widely in the population and there are very rare / no reports on incompatibility of skin with citrus fruit handling by consumers (isolated case for citrus picker where high exposure is continuous over years mentioned in SCCS review). Next to the study mention in the database (8% on n=25 with no effects), there is a Kligman Maximisation test in RIFM database at 20% on n=25 with no effects. Based on both the clinical data (in relation to use) and the maximisation data, Limonene non-oxidized should be rated NC and not confused with oxidized limonene (oxidized by open stirring for weeks to months)(Karlberg et al., 1992). Also, the SCCS opinion states: "The allergenicity of limonene is closely related to oxidation".) Oxidized limonene also generates positive in vitro data.
Tocopherol 	59-02-9	7.4	NA	9.4	Vitamin E, endogenous antioxidant in human skin, very widely used in topical products with very rare cases of sensitisation compared to frequency of use. Sensitisation to endogenous compounds is generally rare (Kosari et al., 2010). Main reactions reported for esters of Tocopherol, but synthetic tocopheryl esters do have an additional alert, as phenols they are good acyl transfer agents.
Isopropyl myristate	110-27-0	44	NC/1B	6.9	Very widely used solvent in perfumery and cosmetics, very rare positive clinical reactions, no structural alert. Negative at 20% in human HRIPT (25 subjects only), Simple alkyl esters lack reactivity and acyl-transfer potential, considered 'extremely weak or non-sensitiser' based

Chemical	CASRN	LLNA MLLP	HU GHS	LogP	Details and Reference
					on extended patch-testing (16 reaction in 12'600 patients; less frequent than *e.g.* propylene glycol)(Uter et al., 2004).
iso-Methylionone 	127-51-5	21.8	NC	4.38	Negative by HDSG analysis. Has been tested at 60% in 106 and 23 panellists with no positive reactions (see human subgroup report, RIFM08). Very low frequency of positive reactions (n=1 of 2004; (Schnuch et al., 2007)in clinic despite low use leading to high SEQ 0.23/0.08 Schnuch *et al.*(Schnuch et al., 2015)). Has a Michael acceptor alert, but was shown to be completely shielded leading to not even traces of peptide adducts (Natsch and Emter, 2017)
OTNE 	54464-57-2	14.2	NC	5	Considered human non-sensitiser by the assessment of the HDSG. Has been tested at 40% in 101 and 22.5% in 53 panellists with no positive reactions (see human subgroup report, RIFM08). no structural alert

* LogP from DASS database; ECHA reports cLogP of 3.55 and experimental LogP of 3.41, thus this chemical is borderline for the 3.5 threshold.

Table 5. Chemicals with a LogP < 3.5 which are false-positive in the LLNA based on HDSG criteria and/or WoE

Chemical	CASRN	LLNA MLLP	HU GHS	LogP	Details and Reference
Salicylic acid 	69-72-7	12.2	NC/1B	2.26	Non-sensitiser in humans despite widespread topical use. Negative in human predictive test at 20%, see HDSG report. SCCS made a WoE assessment and concluded that salicylic acid is a non-sensitiser (Safety), 2019). Salicylic acid is very widely used up to 2% in leave-on cosmetics as active ingredient, but allergic reactions are not reported. Negative in Buehler test with 25% induction concentration.
Sodium lauryl sulphate 	151-21-3	3.7	NC	1.6	SDS is widely documented as false-positive in the LLNA(Loveless et al., 1996). Negative by HDSG.
Linalool 	78-70-6	35.5	NC/1B	2.97	SEQ 0.08/0.1 in Schnuch (Schnuch et al., 2015) indicates very low frequency compared to volume despite very high patch test concentration (10%). Linalool has an HRIPT NOEL of 13793 µg/cm^2 (Basketter et al, 2014; Gerberick et al, 2001) and HMT NOEL of 55176 µg /cm^2 (Greif, 1967) with no cases of sensitisation reported in both tests conducted at high concentration, thus it is close from being labelled as a human NC by HDSG criteria (20% top test concentration as compared to 25% set as threshold). 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 (Brared Christensson et al., 2012), which is not relevant for the regulatory assessment of the parent compound and also appears not to represent commercial use(Kern et al., 2014). For Linalool there is often a confusion made between artificially oxidized Linalool

					(open stirring for months (Sköld et al., 2004)) and the peroxide formed and normal, commercial Linalool which we assess here. Oxidized Linalool is also positive in *in vitro* assays.
DMSO 	67-68-5	72	NC	-1.35	Considered human non-sensitiser by the HDSG on available predictive tests; (Marren, 2011)

This analysis is not comprehensive: There may be other chemicals in the database, in both physicochemical ranges, for which the LLNA may be false-positives, and the chemicals shown here are just those for which we consider the evidence sufficient. Therefore, the exact difference in LLNA specificity and LLNA FDR between the two physicochemical ranges cannot be known for the full database, but this additional analysis indicates that such a difference does exist.

5. Analysis of LLNA data compared to other references (Guinea pig and *in silico*)

To understand whether the observed somewhat lower specificity of the LLNA for human data in the discussed physicochemical range is a new finding, it is instructive to also look at data beyond the curated DASS DB and to consult the peer-reviewed literature. The primary reference data when validating the LLNA, next to human data, were guinea pig data (GPMT and Buehler test) (Dean et al., 2001; Haneke et al., 2001; Sailstad et al., 2001). The advantage of the guinea pig data is that it truly measures the elicitation phase of skin sensitisation as compared to the LLNA, which measures cell-proliferation in the lymph node during induction, a proxy for the sensitisation phase. False-positive LLNA results vs. the guinea pig data have been repeatedly discussed in the literature both during validation but also after the LLNA became an OECD guideline. Here we indicate the results from two key studies:

- Ball et al. (Ball et al., 2011) reported, next to sodium dodecyl sulphate (SDS), five surfactants that are FP in the LLNA when assessed vs. guinea pig maximization test. Of these, 4/5 have a LogP ≥ 3.5

- Kreiling et al. (Kreiling et al., 2017; Kreiling et al., 2008) reported 6 ‘unsaturated compounds’ FP in the LLNA when assessed vs. guinea pig maximization test. Of these, 4/6 have a LogP ≥ 3.5 and include simple fatty acids and the most prominent hydrocarbon endogenous to the human skin, namely squalene.

Details of the false-positives with a LogP≥ 3.5 as compared to guinea-pig data in these two detailed studies are shown in Table 6. These two studies indicate that at a high LogP, a number of chemicals return false-positive results in the LLNA when compared to guinea pig data.

In a study on the predictivity of an *in silico* model to predict reactive skin sensitisers vs. LLNA data (Patlewicz et al., 2007), a low sensitivity (56%) for an external test set of chemicals tested in the LLNA was found. While the true human sensitisation status of those chemicals is not known, it is interesting to note that among the seven chemicals (7 of 16 tested) considered false-negative vs. LLNA in the *in silico* model, there are four chemicals with no apparent structural alert for skin sensitisation, including a simple alkane. All these four chemicals have a LogP ≥ 3.5 and, based on our understanding of skin sensitisation mechanism, these are putative false-positives in the LLNA and not false-negatives in the *in silico* model.

The studies summarized in Table 6 and Table 7 may all be considered reports on a low number of cases per study, but taken together they show that repeated, independent studies found false-positives vs. guinea pig data or reactivity alerts, and these false-positives cluster in the range of LogP≥3.5, thus giving *supporting evidence* to the analysis conducted above. These chemicals include a number of surfactants, but also other highly lipophilic chemicals which are clearly not surfactants such as squalene.

In summary, different lines of evidence summarized above indicate that indeed the specificity of the LLNA at high LogP tends to be lower and the FDR higher. This fact explains why the sensitivity of *e.g.* h-CLAT *vs.* human data is equally good at both LogP ranges, while when using LLNA as reference the sensitivity at high LogP is reduced. Namely, what we *empirically* observe is decreased specificity of the LLNA to correctly

identify sensitisers at high LogP (higher FDR of LLNA in the high LogP range) and not only limited sensitivity of h-CLAT (or other *in vitro* methods*)*. What could be the *mechanistic* reason for this?

Table 6. Lipophilic chemicals which were rated as false-positive vs. guinea pig in the studies by Kreiling *et al.* and Ball *et al.*

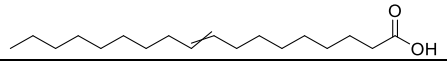
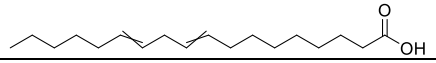
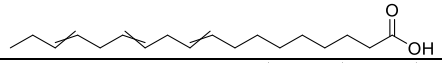
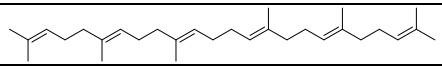
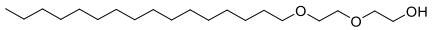
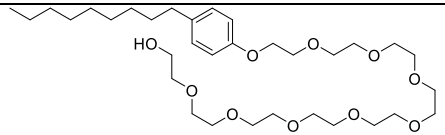
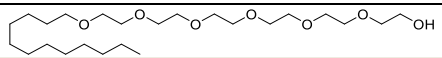
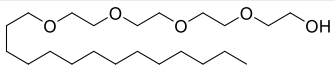
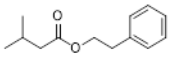
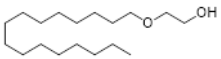
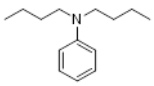
CAS	Structure	Name	cLogP Chem -Draw	LLNA Result (EC3 %)	Guinea pig result
Study of Kreiling et al.					
112-80-1		Oleic acid	7.8	10.4	negative
60-33-3		Linoleic acid	7.3	14	negative/ambiguous
463-40-1		Linolenic acid	6.8	10	negative
111-02-4		Squalene	12.9	<10%	negative
Study of Ball et al					
n.a.		Hexadecan-1-ol, Ethoxylated (2 EO)	7	1.2	negative (0/20)
n.a.		Decylphenolpolyethylene glycol ether	4.8	2.0	negative (0/20)
n.a.		Hexaethylene glycol monododecyl ether	4.2	7.4	negative (0/20)
n.a.		Tetraethylene glycol monotetradecyl ether	5.6	2.8	negative (0/20)

Table 7. Lipophilic chemicals positive in the LLNA and mispredicted by TIMES SS as non-sensitisers in (Patlewicz et al., 2007)

CAS		cLogP	LLNA	TIMES[1]	Note
140-26-1		3.84 LogP (ECHA)	1	0	Clear hydrophobic LLNA FP, no indication that aliphatic esters are sensitiser
2136-71-2		6.4	1	0	Clear hydrophobic LLNA FP, surfactant
629-50-5		6.3	1	0	Clear hydrophobic LLNA FP, simple alkane
613-29-6		4.7	1	0	Hydrophobic potential FP, no structural alert, eye and skin irritant according ECHA

[1)] TIMES result at the time of the publication

6. A potential explanation: Mechanistic rationale

There is an established sequence of events that might lead to a false-positive in the LLNA for a high LogP molecule, although this has not been investigated in detail at the mechanistic level:

1. High LogP molecules have increased cytotoxicity. Correlation between (cyto)toxicity and LogP is very well described in the literature on ecotoxicology, see *e.g.* (Tebby et al., 2011).
2. Cytotoxic molecules tend to have increased skin irritation potential – note that cytotoxicity is the read out in the OECD skin- and eye irritation methods.
3. Skin irritation by cytotoxic molecules involves release of interleukin-1α from damaged skin cells (note that IL-1α release is an add-on endpoint in skin irritation tests (Cotovio et al., 2005; Spielmann et al., 2007)), but also other cytokines such as TNF-alpha (Kock et al., 1990) or IL-6 (Kirnbauer et al., 1991) are released by damaged keratinocytes.
4. IL-1α (and maybe other cytokines) can trigger irritant induced dendritic cell emigration from the skin to the lymph nodes similarly as IL-1β can do for allergens (Cumberbatch et al., 2002) and it can induce lymph node cell proliferation.

Just to name an example, ethoxylated surfactants tested diluted in the EpiSkin® model triggered dramatic levels of IL-1α release already at concentrations below strong cytotoxicity and they were strongly false-positive in the LLNA (Ball et al., 2011), see Table 6.

This sequence of events, which are individually well established but have not been fully proven *in vivo* in the LLNA situation, might explain why high LogP molecules might trigger by their cytotoxic / irritant nature IL-1α / cytokine-induced lymph node cell proliferation. As the LLNA is based on general cell proliferation in the lymph nodes, the LLNA may, therefore, not perfectly discriminate some lipophilic irritants / cytotoxic molecules from sensitisers when tested at a high dose. In this respect, it is important to note that many of the potential LLNA FP reported here were positive only at high test concentrations.

7. Analysis of the LLNA validation set

Finally, a key question remains: if such a somewhat limited specificity of the LLNA at LogP ≥ 3.5 exists, why had it not been recognized earlier? To answer this question, we investigated the list of chemicals analysed when validating the LLNA (Haneke et al., 2001). We could retrieve the chemical structure and hence a LogP for 194 chemicals. 129 of those have a LogP < 3.5. Out of these, 96 have guinea pig data with 27 being negative in guinea pig tests. 24 are also negative in the LLNA, showing a high specificity of the LLNA in the range of LogP <3.5 (89%). The dataset contains many chemicals with a LogP ≥ 3.5 (n=63), however only for 20 of those, guinea pig data are available, and these include only four negative chemicals. Two of those 4 are positive in the LLNA, which would give a specificity of 50%, but given this low number it is not surprising, that this issue was never of concern during validation and upon implementation of OECD TG429. Also, it has to be re-iterated that during validation most negatives in the guinea pig test were only tested in the LLNA up to 25% (Kolle et al., 2020), while the DASS curation criterium required a test concentration of at least 50% to call a negative LLNA study. The majority of the LLNA FP in the Basketter *et al.* compilation are positive in the LLNA at > 25% too, hence, beyond the validated range.

8. Summary of lines of evidence that support a higher FP rate of LLNA for high LogP substances

- LLNA vs HDSG data (n=56) show that the FDR (FP/FP+TP) of LLNA in the LogP<3.5 range is 0.05 (2/40) while for LogP≥3.5 range is 0.45 (5/11), the latter, however, is based on a limited number of substances.
- If we consider the 95% CI it shows that: FDR LogP<3.5 = 2/40= 0.05 with the 95%CI [0.01,0.17] and FDR LogP>3.5 = 5/11= 0.45 with the 95%CI [0.17,0.77]
- LLNA vs Basketter human data (n=96) show the FDR in the range LogP<3.5 is 0.13 (8/60) while for the LogP≥3.5 range is 0.38 (9/24)
- There is strong evidence that at least 8/33 LLNA positive results (*i.e.* Hexyl salicylate, Benzyl benzoate, Citronellol, (R)-(+)-Limonene, Tocopherol, Isopropyl myristate, iso-Methylionone and OTNE) in the DASS database with LogP ≥3.5 are not relevant human sensitisers based on HDSG analysis and /or WoE (see Table 4). This translates into an FDR of the LLNA of ≥ 0.24 in the full database at LogP ≥3.5
- Ball et al. and Kreiling et al. reported 11 FP of the LLNA with respect to GPMT data, which, unlike the LLNA, measures the elicitation phase of skin sensitisation. Of these, 8/11=0.73 have LogP≥3.5
- The following mechanistic rational can explain a LLNA high FP rate LogP≥3.5 substances: High LogP substances might trigger by their cytotoxic / irritant nature IL-1α / cytokine-induced lymph node cell proliferation.

9. Conclusions

The analysis presented here does not indicate that the LLNA is wrong at higher LogP – but different lines of evidence indicate that the false-positive rate of the LLNA is higher for lipophilic chemicals. This could explain the observed apparent lower sensitivity of ca. 10-15% calculated for the defined approaches in this physicochemical range. Thus, we consider it proven that there is an uncertainty for the LLNA positive *in vivo* reference data at high LogP.

Rejecting *in vitro* data for high LogP substances does therefore not appear to be an appropriate solution because

it may probably lead to conducting LLNA studies, exactly in the physicochemical range at which its predictivity is lower as shown above.

Currently, a limitation of the h-CLAT for accepting negative results for chemicals at LogP ≥ 3.5 forms part of OECD TG 442E. This then affects negative calls made with a DA based on h-CLAT. The analysis presented here indicates, that such cases should be evaluated based on a weight-of-evidence approach, especially also taking structural alerts into account. For chemicals, for which adequate exposure of the cells is proven by cytotoxicity and which do lack a clear structural alert for reactivity, negative calls should be accepted based on a WoE rather than reject it and proceed to LLNA testing.

References

Ball, N., Cagen, S., Carrillo, J. C., Certa, H., Eigler, D., Emter, R., Faulhammer, F., Garcia, C., Graham, C., Haux, C., Kolle, S. N., Kreiling, R., Natsch, A., Mehling, A., 2011. Evaluating the sensitisation potential of surfactants: Integrating data from the local lymph node assay, guinea pig maximization test, and in vitro methods in a weight-of-evidence approach. Regul. Toxicol. Pharmacol. 60**,** 389-400.10.1016/j.yrtph.2011.05.007

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 sensitising potency. Dermatitis. 25**,** 11-21.10.1097/DER.0000000000000003

Brared Christensson, J., Andersen, K. E., Bruze, M., Johansen, J. D., Garcia-Bravo, B., Giménez Arnau, A., Goh, C. L., Nixon, R., White, I. R., 2012. Air-oxidized linalool-a frequent cause of fragrance contact allergy. Contact Dermatitis. 67**,** 247-259.10.1111/j.1600-0536.2012.02134.x

Cotovio, J., Grandidier, M. H., Portes, P., Roguet, R., Rubinstenn, G., 2005. The in vitro skin irritation of chemicals: optimisation of the EPISKIN prediction model within the framework of the ECVAM validation process. Altern Lab Anim. 33**,** 329-49

Cumberbatch, M., Dearman, R. J., Groves, R. W., Antonopoulos, C., Kimber, I., 2002. Differential regulation of epidermal Langerhans cell migration by interleukins (IL)-1? and IL-1? during irritant- and allergen-induced cutaneous immune responses. Toxicology and Applied Pharmacology. 182**,** 126-135

Dean, J. H., Twerdok, L. E., Tice, R. R., Sailstad, D. M., Hattan, D. G., Stokes, W. S., 2001. ICCVAM evaluation of the murine local lymph node assay. Conclusions and recommendations of an independent scientific peer review panel. Regul Toxicol Pharmacol. 34**,** 258-73.10.1006/rtph.2001.1497S0273230001914977 [pii]

Gerberick, G. F., Vassallo, J. D., Foertsch, L. M., Price, B. B., Chaney, J. G., Lepoittevin, J. P., 2007. Quantification of chemical peptide reactivity for screening contact allergens: A classification tree model approach. Toxicological Sciences. 97**,** 417-427

Haneke, K. E., Tice, R. R., Carson, B. L., Margolin, B. H., Stokes, W. S., 2001. ICCVAM evaluation of the murine local lymph node assay. Data analyses completed by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods. Regul Toxicol Pharmacol. 34**,** 274-86.10.1006/rtph.2001.1498 [doi]S0273230001914989 [pii]

Hoffmann, S., Kleinstreuer, N., Alepee, N., Allen, D., Api, A. M., Ashikaga, T., Clouet, E., Cluzel, M., Desprez, B., Gellatly, N., Goebel, C., Kern, P. S., Klaric, M., Kuhnl, J., Lalko, J. F., Martinozzi-Teissier, S., Mewes, K., Miyazawa, M., Parakhia, R., van Vliet, E., Zang, Q., Petersohn, D., 2018. Non-animal methods to predict skin sensitisation (I): the Cosmetics Europe database(). Crit. Rev. Toxicol. in press**,** 1-15.10.1080/10408444.2018.1429385

Karlberg, A. T., Magnusson, K., Nilsson, U., 1992. Air oxidation of d-limonene (the citrus solvent) creates potent allergens. Contact Dermatitis. 26**,** 332-40

Kern, S., Dkhil, H., Hendarsa, P., Ellis, G., Natsch, A., 2014. Detection of potentially skin sensitising hydroperoxides of linalool in fragranced products. Anal Bioanal Chem. 406**,** 6165-78.10.1007/s00216-014-8066-3

Kirnbauer, R., Kock, A., Neuner, P., Forster, E., Krutmann, J., Urbanski, A., Schauer, E., Ansel, J. C., Schwarz, T., Luger, T. A., 1991. Regulation of epidermal cell interleukin-6 production by UV light and corticosteroids. J Invest Dermatol. 96**,** 484-9

Kock, A., Schwarz, T., Kirnbauer, R., Urbanski, A., Perry, P., Ansel, J. C., Luger, T. A., 1990. Human keratinocytes are a source for tumor necrosis factor alpha: evidence for synthesis and release upon stimulation with endotoxin or ultraviolet light. J Exp Med. 172**,** 1609-14

Kolle, S. N., Landsiedel, R., Natsch, A., 2020. Replacing the refinement for skin sensitisation testing: Considerations to the implementation of adverse outcome pathway (AOP)-based defined approaches (DA) in OECD guidelines. Regul Toxicol Pharmacol. 115**,** 104713.10.1016/j.yrtph.2020.104713

Kosari, P., Alikhan, A., Sockolov, M., Feldman, S. R., 2010. Vitamin E and allergic contact dermatitis. Dermatitis. 21**,** 148-53

Kreiling, R., Gehrke, H., Broschard, T. H., Dreeßen, B., Eigler, D., Hart, D., Höpflinger, V., Kleber, M., Kupny, J., Li, Q., Ungeheuer, P., Sauer, U. G., 2017. In chemico, in vitro and in vivo comparison of the skin sensitising potential of eight unsaturated and one saturated lipid compounds. Regulatory Toxicology and Pharmacology. 90**,** 262-276.10.1016/j.yrtph.2017.09.023

Kreiling, R., Hollnagel, H. M., Hareng, L., Eigler, D., Lee, M. S., Griem, P., Dreessen, B., Kleber, M., Albrecht, A., Garcia, C., Wendel, A., 2008. Comparison of the skin sensitising potential of unsaturated compounds as assessed by the murine local lymph node assay (LLNA) and the guinea pig maximization test (GPMT). Food Chem Toxicol. 46**,** 1896-904.S0278-6915(08)00041-0 [pii]
10.1016/j.fct.2008.01.019 [doi]

Loveless, S. E., Ladics, G. S., Gerberick, G. F., Ryan, C. A., Basketter, D. A., Scholes, E. W., House, R. V., Hilton, J., Dearman, R. J., Kimber, I., 1996. Further evaluation of the local lymph node assay in the final phase of an international collaborative trial. Toxicology. 108**,** 141-52

Marren, K., 2011. Dimethyl sulfoxide: an effective penetration enhancer for topical administration of NSAIDs. Phys Sportsmed. 39**,** 75-82.10.3810/psm.2011.09.1923

Natsch, A., Emter, R., 2017. Reaction Chemistry to Characterize the Molecular Initiating Event in Skin Sensitisation: A Journey to Be Continued. Chem. Res. Toxicol. 30**,** 315-331.10.1021/acs.chemrestox.6b00365

OECD, 2018. In vitro skin sensitisation assays addressing the key event on activation of dendritic cells on the adverse outcome pathway for skin sensitisation. OECD testing guidelines. 442e.10.1787/9789264264359-en

Patlewicz, G., Dimitrov, S. D., Low, L. K., Kern, P. S., Dimitrova, G. D., Comber, M. I., Aptula, A. O., Phillips, R. D., Niemela, J., Madsen, C., Wedebye, E. B., Roberts, D. W., Bailey, P. T., Mekenyan, O. G., 2007. TIMES-SS--a promising tool for the assessment of skin sensitisation hazard. A characterization with respect to the OECD validation principles for (Q)SARs and an external evaluation for predictivity. Regul Toxicol Pharmacol. 48**,** 225-39.S0273-2300(07)00039-6 [pii]10.1016/j.yrtph.2007.03.003

Rudback, J., Hagvall, L., Borje, A., Nilsson, U., Karlberg, A. T., 2014. Characterization of skin sensitisers from autoxidized citronellol - impact of the terpene structure on the autoxidation process. Contact Dermatitis. 70**,** 329-339.10.1111/cod.12234

Safety), S. S. C. o. C., 2019. OPINION ON salicylic acid (CAS 69-72-7). https://op.europa.eu/en/publication-detail/-/publication/75be19bf-5390-11ea-aece-01aa75ed71a1, visited last 16.6.2020.

Sailstad, D. M., Hattan, D., Hill, R. N., Stokes, W. S., 2001. ICCVAM evaluation of the murine local lymph node assay. The ICCVAM review process. Regul Toxicol Pharmacol. 34**,** 249-57.10.1006/rtph.2001.1496
S0273230001914965 [pii]

Schnuch, A., Uter, W., Geier, J., Lessmann, H., Frosch, P. J., 2007. Sensitisation to 26 fragrances to be labelled according to current European regulation. Results of the IVDK and review of the literature. Contact Dermatitis. 57**,** 1-10.COD1088 [pii]
10.1111/j.1600-0536.2007.01088.x

Schnuch, A., Uter, W., Lessmann, H., Geier, J., 2015. Risk of sensitisation 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**,** 208-217.10.1002/ffj.3241

Sköld, M., Börje, A., Harambasic, E., Karlberg, A. T., 2004. Contact allergens formed on air exposure of linalool. Identification and quantification of primary and secondary oxidation products and the effect on skin sensitisation. Chem Res Toxicol. 17**,** 1697-1705

Spielmann, H., Hoffmann, S., Liebsch, M., Botham, P., Fentem, J. H., Eskes, C., Roguet, R., Cotovio, J., Cole, T., Worth, A., Heylings, J., Jones, P., Robles, C., Kandarova, H., Gamer, A., Remmele, M., Curren, R., Raabe, H., Cockshott, A., Gerner, I., Zuang, V., 2007. The ECVAM international validation study on

in vitro tests for acute skin irritation: report on the validity of the EPISKIN and EpiDerm assays and on the Skin Integrity Function Test. Altern Lab Anim. 35**,** 559-601

Takenouchi, O., Miyazawa, M., Saito, K., Ashikaga, T., Sakaguchi, H., 2013. Predictive performance of the human Cell Line Activation Test (h-CLAT) for lipophilic chemicals with high octanol-water partition coefficients. J Toxicol Sci. 38**,** 599-609.10.2131/jts.38.599

Tebby, C., Mombelli, E., Pandard, P., Péry, A. R. R., 2011. Exploring an ecotoxicity database with the OECD (Q)SAR Toolbox and DRAGON descriptors in order to prioritise testing on algae, daphnids, and fish. Science of the Total Environment. 409**,** 3334-3343.http://dx.doi.org/10.1016/j.scitotenv.2011.05.029

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 sensitisation hazard in mice and men using non-animal test methods. Regul. Toxicol. Pharmacol. 71**,** 337-51.10.1016/j.yrtph.2014.12.008

Uter, W., Schnuch, A., Geier, J., Lessmann, H., 2004. Isopropyl myristate recommended for aimed rather than routine patch testing. Contact Dermatitis. 50**,** 242-4.10.1111/j.0105-1873.2004.00352.x

Yamamoto, Y., Wanibuchi, S., Sato, A., Kasahara, T., Fujita, M., 2019. Precipitation of test chemicals in reaction solutions used in the amino acid derivative reactivity assay and the direct peptide reactivity assay. Journal of Pharmacological and Toxicological Methods. 100.10.1016/j.vascn.2019.106624