Regulatory Toxicology and Pharmacology 141 (2023) 105402

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Specificity of the local lymph node assay (LLNA) for skin sensitisation

David W. Roberts [a,*], Ian Kimber [b], David A. Basketter [c]

[a] School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool, L3 3AF, UK
[b] Faculty of Biology, Medicine and Health, University of Manchester, Manchester, M13 9PT, UK
[c] DABMEB Consultancy Ltd, Abbey View, Abbey St, Kingswood, Wotton-under-Edge, GL12 8RN, UK

ARTICLE INFO

Handling Editor: Dr. Martin Van den berg

Keywords:
Skin sensitisation assays
Hydrophobicity
False positives
Potency categories

ABSTRACT

The local lymph node assay (LLNA) has provided a large dataset against which performance of non-animal approaches for prediction of skin sensitisation potential and potency can be assessed. However, a recent comparison of LLNA results with human data has argued that LLNA specificity is low, with many human non-sensitisers, particularly hydrophobic chemicals, being false positives. It has been suggested that such putative false positives result from hydrophobic chemicals causing cytotoxicity, which induces irritancy, in turn driving non-specific lymphocyte proliferation. This paper finds that the apparent reduced specificity of the LLNA largely reflects differences in definitions of the boundaries between weak skin sensitisers and non-sensitisers. A small number of LLNA false positives may be due to lymphocyte proliferation without skin sensitisation, but most alleged 'false' positives are in fact very weak sensitisers predictable from structure-activity considerations. The evidence does not support the hypothesis for hydrophobicity-induced false positives. Moreover, the mechanistic basis is untenable. Sound LLNA data, appropriately interpreted, remain a good measure of sensitisation potency, applicable across a wide hydrophilicity-hydrophobicity range. The standard data interpretation protocol enables detection of very low levels of sensitisation, irrespective of regulatory significance, but there is scope to interpret the data to give focus on regulatory significance.

1. Introduction

There is a major impetus to develop approaches for identifying skin sensitising chemicals and quantifying their potency without the use of animals, with potency prediction having proven a significant challenge (e.g. Li et al., 2019; Basketter and Gerberick, 2022; Natsch and Gerberick, 2022). To assess the predictive ability of new non-animal methods directed towards this goal, it is necessary to compare their predictions with existing, reliable, reference results. There already exists a large body of skin sensitisation test data produced using the murine local lymph node assay (LLNA). These data are potentially very useful, since they provide a quantitative index of potency, the EC3 value. These data have been subject to considerable retrospective analysis and curation (e.g. OECD, 2021a). On the other hand, a rather smaller number of chemicals have data, either from predictive sensitisation testing in humans (the human repeated insult patch test - HRIPT; the human maximization test - HMT) or from clinical diagnostic patch testing, which have been used to deliver potency predictions (Basketter et al., 2014; Api et al., 2017). At first sight such data might appear more relevant and useful than LLNA data, since the need for sensitisation evaluation is to protect humans, not mice or other experimental species. However, human potency determinations rely heavily on expert judgement, itself based often on data from testing in small group sizes, and thus the human tests have their own distinct limitations. An OECD expert group curated available human data to avoid reliance on expert judgement, although the rules they applied resulted in distinct loss of human information (OECD, 2021b). Most recently, substantial efforts have been made to collate all available potency information in a transparent and coherent manner, but the benefit from these works is yet to be seen (Irizar et al., 2022; Na et al., 2022).

There is an important further consideration: an assay intended to detect toxicity for any biological endpoint needs also to be able to detect the absence of toxicity. In other words, not only should it have high sensitivity (the ability to identify positives) but also high specificity (the ability to detect negatives). When the LLNA was developed and validated as an assay for skin sensitisation, with a dataset of 200 substances, the evidence indicated that it met both of these criteria satisfactorily (Gerberick et al., 2000; Dean et al., 2001). However, this has recently been brought into question based on comparisons between human data and LLNA data, and a specificity value as low as 22% has been reported

* Corresponding author.
E-mail address: d.w.roberts@ljmu.ac.uk (D.W. Roberts).

https://doi.org/10.1016/j.yrtph.2023.105402
Received 8 February 2023; Received in revised form 13 April 2023; Accepted 25 April 2023
Available online 26 April 2023
0273-2300/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

and then investigated further (OECD, 2021b; Natsch et al., 2023). Furthermore, it has been argued that hydrophobic chemicals are particularly prone to give false positives in the LLNA (Natsch et al., 2023). This is interpreted in terms of hydrophobic chemicals (i.e. chemicals with high logP values, P being the octanol-water partition coefficient) causing lymphocyte proliferation without sensitisation, as a result of logP-dependent cytotoxicity leading to irritant effects which in turn provoke a cascade of events leading to non-specific lymphocyte proliferation. Were this to be true, then the merits of the LLNA as a basis for the evaluation of non-animal alternatives to the identification and potency characterisation of skin sensitisers would be substantially undermined. Consequently, this paper addresses the issue of the specificity of the LLNA by considering three questions:

Is the specificity of the LLNA genuinely poor?
Are LLNA false positives due to proliferation without sensitisation?
How strong is the evidence for hydrophobicity-induced false positives?

2. The evidence for 22% specificity

The 22% specificity figure is derived from a group of just 9 substances that were found to be negative in human predictive skin sensitisation studies, these being abstracted from a highly curated "gold list" of 56 substances with LLNA data and HMT or HRIPT data which are included in a larger "gold database" comprising Annex 2 of OECD 2021a (OECD 2021a; Natsch et al., 2023). These nine are shown, together with their LLNA results and logP values, in Table 1. The logP values shown in Table 1 are those given in the gold database – some of them are listed as experimental and others as calculated, but no references or information as to calculation methods are given. We have checked all these logP values by manual calculation using the Leo and Hansch method (Hansch and Leo 1979), and apart from two exceptions indicated, the agreement is within 0.2 log units.

Taken at face value, only two (green shading) of these nine chemicals are correctly identified by the LLNA as non-sensitisers and the other seven appear to be false positives, hence the specificity figure of 22%. Of these seven, five of them have high logP values, rendering them difficult to test in water-based in vitro methods - for example, logP >3.5 is the cut-off value in the applicability domain of the human cell line activation test (h-CLAT) (Takenouchi et al., 2013; OECD, 2018).

The picture changes when we consider human sensitisation potency (HSP) classifications assigned by Basketter et al. (2014) to the chemicals in Table 1. Fig. 1 displays the basic concepts underpinning both the human and regulatory classifications. Central to this is the reality that predictive tests focus on the identification and characterisations required by regulatory classification, whereas humans can and do react under some circumstances to much weaker sensitisers that often are only detected by careful clinical diagnostic work (Basketter et al., 2015; Basketter, 2023). Under that classification scheme, criteria for human data are applied to characterize chemicals into six categories of human sensitising potency, with HSP category 1 the most potent and category 5 the least potent, category 6 representing true non-sensitisers. Over 200 chemicals have been assigned HSP categories according to this scheme (Basketter et al., 2014; Api et al., 2017).

In the six-point classification scheme using human data, category 6 corresponds to non-sensitisers, with no evidence of sensitisation despite extensive human exposure, whereas category 5 corresponds to very

Table 1
Chemicals found to be negative HMT or HRIPT.

Name	CAS	LogP	LLNA EC3 & class[1]
Propylene glycol	57-55-6	-0.92	NC
n-Hexane	110-54-3	3.92	NC
Dimethyl sulfoxide (DMSO)	67-68-5	-1.35	72%, 1B
Sodium lauryl sulfate (SLS)	151-21-3	1.6	3.7%[2], 1B
Benzyl benzoate	120-51-4	3.97	17%, 1B
Citronellol	106-22-9	3.91[3]	43.5%, 1B
Hexyl salicylate	6259-76-3	5.5	0.18%, 1A
α-Iso methyl ionone	127-51-5	4.38[4]	21.8%, 1B
OTNE	54464-57-2	4.98	14.3%, 1B

Results in Table 1 are taken directly from Natsch et al. (2023).
[a] "Class" refers to regulatory classification according to UN GHS; NC = not classified as a skin sensitiser; 1B is a weaker skin sensitiser; 1A is a stronger skin sensitiser.
[b] This EC3 value is suspect as SLS has been tested on multiple occasions in the LLNA and mostly the EC3 values are about 10–15% (e.g. Gerberick et al., 2000).
[c] This logP figure is shown in the Gold list as an experimental value but without a reference. Based on its measured water solubility (Cal, 2006), on literature values calculated by ACD LogP software (Cal, 2006), by XlogP software (Pub.Chem, viewed 16 January 2023) and on manual calculation using the Leo and Hansch method (Hansch and Leo 1979), a logP value in the range 3.1–3.4 seems more plausible.
[d] This logP value is shown as a calculated value in the Gold list, but with no reference to the calculated method. We consider a value of 4.02, calculated by the Leo and Hansch method (Hansch and Leo 1979), to be more plausible.

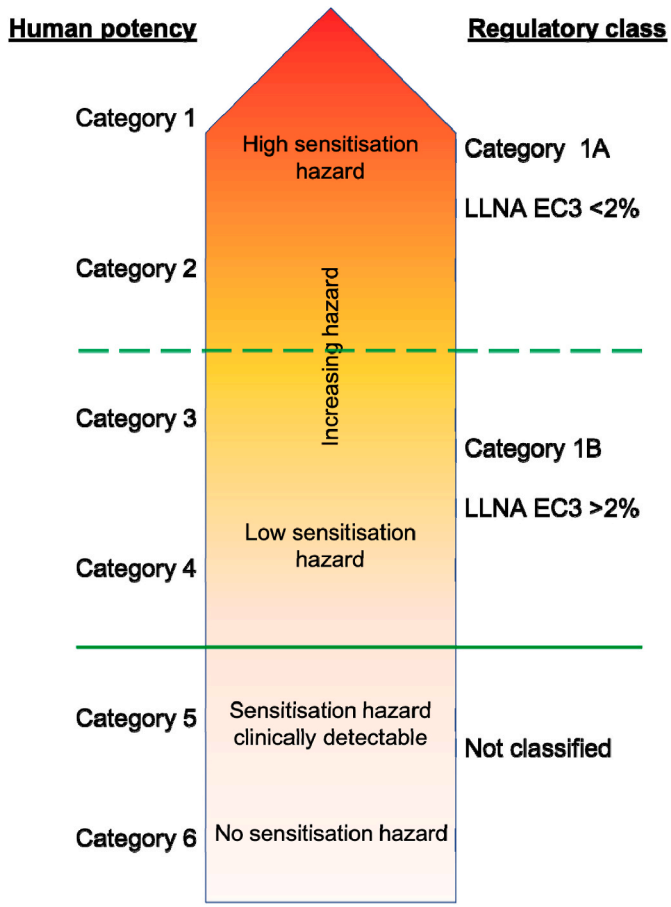


Fig. 1. Human potency versus regulatory classification for skin sensitisers
Six human potency categories, taken from Basketter et al. (2014) and Api et al. (2017) are displayed on the left hand side, with category 1 representing the most potent skin sensitisers. Across the vertical potency spectrum are two administrative regulatory thresholds, the continuous green line showing the boundary as used in the United Nations GHS scheme for distinguishing those substances that should be classified as skin sensitisers (i.e. above the green line). A second dashed green line divides weaker (1B) from stronger (1A) skin sensitisers. Note that human potency categories 5 and 6 fall below the classification threshold, but only category 6 represents true human non-sensitisers.

weak skin sensitisers which are not sufficiently potent to warrant regulatory classification (Fig. 1) (Basketter et al., 2014). As worded in that publication, the clinical evidence corresponding to categories 5 and 6 is as follows:

> **Category 5**: A rare cause of contact allergy except perhaps in special circumstances, eg, use in topical medicaments (examples: hexylcinnamal, isopropanol, parabens).
> **Category 6**: Essentially absent, with at least no systematic convincing evidence of contact allergy (examples: xylene, glycerol, sodium lauryl sulfate).

For category 5, Basketter et al. (2014) also state: "Human repeated insult patch test no observed effect level (NOEL) values are variable, or indeed absent, because of the inherent inaccuracy of determination of a threshold for such a weak sensitiser in a small panel size."

Table 2 shows the same 9 chemicals as in Table 1, this time with their HSP categories. Herein, LLNA positives are treated as true for Category 5 and false for Category 6.

On the basis of the expert judgment HSP classification of the human data (Basketter et al., 2014), only four of the nine chemicals negative in HMT/HRIPT are true non-sensitisers. Two of them are correctly predicted as such by the LLNA, and two are LLNA false positives. Notably, neither of the two LLNA false positives have high logP values.

The five chemicals that are sensitisers, which apart from hexyl salicylate are not classified as such for regulatory purposes, have structure-activity based alerts for skin sensitisation, and their LLNA potency corresponds well with what would be expected based on read-across or from quantitative mechanistic models (QMMs), as shown in Table 3. Hexyl salicylate represents an example where consideration of all the human evidence (as per Basketter et al., 2014) delivers a different classification decision to that of the OECD (OECD, 2021b). The key difference is that the OECD expert group did not take into account real life human experience, as evidenced by clinical diagnostic patch testing. In this situation, it is essential to keep in mind, as indicated in Fig. 1, that skin sensitising chemicals placed in human potency category 5 are not classified as sensitisers - in the regulatory sense, they are "not classified", a term which is not synonymous with "non-sensitiser".

In Table 3, the Schiff base QMM applied to α-isomethyl ionone and OTNE was derived from a series of aliphatic aldehydes and ketones, which can react by attack of a nucleophile at the carbonyl group as indicated by the red curved arrows (Roberts et al., 2006). If the nucleophile is a primary amino group, the reaction product has the substructure C=N and is known as a Schiff base, hence the name of the reaction mechanistic domain. It is not known if sensitisation by Schiff base domain chemicals involves Schiff base of formation or not – attack by ionised thiol to form a hemithioacetal or hemithioketal is also plausible. In any event, reactivity of the carbonyl group is modelled by the $\Sigma\sigma^*$ parameter. Simple monoketones are less reactive than aldehydes and only at high logP values does the combination of reactivity and hydrophobicity become large enough for them to be LLNA positive (Roberts et al., 2006), as is the case with α-isomethyl ionone and OTNE. It may be noted that α-isomethyl ionone also has a Michael acceptor alert (C=C double bond conjugated with C=O). However, the Michael acceptor reactivity will be low and Michael acceptor sensitisation potency is not hydrophobicity dependent (Roberts and Natsch, 2009), so we consider it more likely to sensitise as a Schiff base electrophile.

Regarding hexyl salicylate (Table 3), a manuscript on nucleophilic skin sensitisers is currently in preparation. In brief, a set of aromatic compounds with π-donating groups has been found to fit the QMM shown in the table, where $\Sigma\sigma^+$ models the π-donating activation at the reaction centre. This QMM has been applied to predict the LLNA potency of hexyl salicylate. The haptenation reaction is proposed to involve attack of the nucleophilic centre, as indicated for hexyl salicylate by the curved red arrow, on the S-S linkage of a cystine unit.

The present re-analysis of the data for the nine chemicals on which the 22% specificity figure is based indicates that most of the seven proposed LLNA false positives are in reality true positives possessing very weak human sensitisation potency, including sensitising activity below the regulatory threshold, that were not detected by the HRIPT/HMT studies. This is in accordance with what was already written to clarify the characteristics of human sensitisation potency category 5: "Human repeated insult patch test NOEL values are variable, or indeed absent, because of the inherent inaccuracy of determination of a threshold for such a weak sensitiser in a small panel size.", (Basketter et al., 2014).

3. Other evidence on LLNA specificity

It is relevant to consider the predictive performance of the LLNA against the almost 100 chemicals in the "Gold database" that have human potency classifications and LLNA data. For this set of chemicals, the specificity of the LLNA has been reported to be 39% (Natsch et al., 2023). However, this figure is based on considering the chemicals in category 5 as non-sensitisers. When the 17 chemicals in category 5 are treated as sensitisers, the performance statistics shown in Table 4 are obtained, with specificity increasing markedly to 64%.

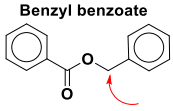
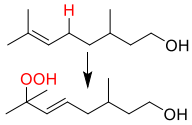
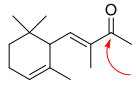
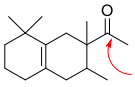
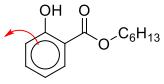
Of the four false positives, we consider DMSO, salicylic acid and SLS

Table 2
HMT/HRIPT negatives with human sensitisation potency categories [a].

Name	LogP	LLNA EC3 & class	HSP category
Propylene glycol	-0.92	NC	6
n-Hexane	3.92	NC	6
DMSO	-1.35	72%, 1B	6
SLS	1.6	3.7%, 1B	6
Benzyl benzoate	3.97	17%, 1B	5
Citronellol	3.91	43.5%, 1B	5
Hexyl salicylate	5.5	0.18%, 1A	4[2]
α-Iso methyl ionone	4.38	21.8%, 1B	5
OTNE	4.98	14.3%, 1B	5

[a] True negatives shaded green; false positives shaded orange; true positives shaded blue; data from Basketter et al. (2014).
[b] This was placed in the not classified category 5 by the OECD expert group (OECD, 2021b).

Table 3
Structural alerts for weak sensitisers negative in HMT/HPRIPT.

Compound/structure and logP	Reaction mechanistic domain	QMM or read-across reference chemicals	LLNA EC3	
			Estimated	Observed
Benzyl benzoate  logP = 3.97	S_N2, activated (benzylic) reaction centre with weak leaving group (carboxylate ion)	Allyl octanoate, similarly activated (allylic) reaction centre with similar weak leaving group (carboxylate ion) and similar logP (3.87)	EC3 of allyl octanoate is 6.4%	17%
Citronellol  logP = 3.91	Pre-hapten alert (allylic H able to give rise to a tertiary allylic hydroperoxide by autoxidation)	Linalool and limonene – these chemicals when pure are weakly positive in the LLNA, probably as a result of autoxidation during air exposure following open epicutaneous application	Linalool EC3 is 35.5% Limonene EC3 is 52.5%	43.5%
α-Iso methyl ionone  logP, 4.38; Σσ*, -0.2	Schiff-base electrophile domain	pEC3 = 1.12Σσ* + 0.42logP-0.62[a] (Roberts et al., 2006)	21% (30.6% based on logP = 4.02)	21.8%
OTNE  logP, 4.98; Σσ*, -0.3			17%	14.3%
Hexyl salicylate  logP = 5.5; Σσ+ = -0.60; RAI (= -Σσ+ + 0.4logP), 2.8	Nucleophilic sensitiser	pEC3 = 2.47RAI - 3.76 (see text)	0.16%	0.18%

[a] σ* is the Taft substituent constant, a parameter widely used in Physical Organic Chemistry as a quantitative index of electronegativity. Σσ* is the sum of the σ* values for the two substituents bonded to the carbonyl group (Roberts et al., 2006).

Table 4
Predictive performance of LLNA against 98 chemicals with human sensitisation potency category assignments [a].

Results:	TP	FP	TN	FN
	80	4	7	7
Sensitivity	92%			
Specificity	64%			
Accuracy	89%			
Balanced accuracy	86%			

False positives:	**EC3 (%)**	**logP**
DMSO	72	−1.35
Salicylic acid	12.2	2.26
SLS	3.7[b]	1.6
α-Tocopherol	7.4	9.4

[a] Chemicals in categories 1–5 are true human sensitisers, chemicals in category 6 are true human non-sensitisers.
[b] See first footnote to Table 1.

as likely to be genuinely false, since they have no obvious alerts for significant reactivity or potential to be activated to reactive species. None of these three chemicals has a high logP value (i.e. >3.5). Nevertheless, it is important to mention that of the four, at least salicylic acid has been reported very occasionally as a cause of allergic contact dermatitis(de Groot, 2021).

The fourth chemical, α-tocopherol, is an antioxidant, which means it is easily oxidised, which in turn means that it could be liable to be contaminated with its oxidation products or to become oxidised by air under the open application conditions of the LLNA. Without information on the purity of the LLNA sample that was tested, we cannot make a definite conclusion but we consider it most probable that the LLNA result reflects genuine sensitisation, not to α-tocopherol itself but to its oxidation products or oxidation products of its impurities such as the related tocotrienols that it is usually found together with. Arguably, this is consistent with the fact that α−tocopherol is occasionally identified as a cause of allergic contact dermatitis in humans (de Groot, 2018).

Overall, although the sensitivity/specificity figures are based on a small number of human potency category 6 chemicals, the LLNA performs well against the Basketter et al. (2014) HSP dataset, and the false positives show no bias towards high logP chemicals.

4. LLNA compared to Guinea pig data

There already exists literature debating potential false positives in the LLNA compared to older guinea pig outcomes, with a common, but obviously spurious, explanation given as skin irritation (Basketter and Kimber, 2011). As supporting evidence for the argument that high logP chemicals have a higher tendency to give false positives, studies have been cited in which lipophilic unsaturated fatty acids* and ethoxylated alcohol surfactants, as well as the hydrophobic unsaturated hydrocarbon squalene were found to be positive in the LLNA but negative in the GPMT (Kreiling et al., 2008, 2017; Ball et al., 2011; Natsch et al., 2023). These apparent LLNA false positives have been discussed in an earlier paper (Roberts et al., 2016) and there is therefore no need to repeat the discussion here but we simply restate the conclusion: "The LLNA positives all have alerts for autoxidation, and therefore should be considered as potential prohaptens. The LLNA offers a greater opportunity for these non-directly reactive compounds to express their potential to induce sensitisation. With the open epicutaneous application protocol of the LLNA the test chemical is exposed to a continuously replenished supply of oxygen, whereas in an occluded patch protocol the availability of oxygen is substantially restricted. Thus, the LLNA results are not false positives, but simply reflect the differences in protocols between the LLNA and GPMT."

*In fact the unsaturated acids, oleic, linoleic and linolenic would exist as their anions at skin pH, and only the oleic anion would have a logP >3.5.

5. The hypothetical mechanism for hydrophobicity related LLNA false positives

Here we consider the mechanism proposed by Natsch et al. (2023) whereby they suggest that lipophilic compounds have an increased tendency to produce false positives as a result of their alleged higher cytotoxicity, relative to less hydrophobic chemicals. This higher cytotoxicity, they argue, results in higher skin irritancy which in turn, they argue, induces non-specific lymph node cell proliferation.

The evidence that irritancy per se promotes lymph node cell proliferation is at best weak (Basketter et al., 2011). It has often been argued, partly based on old results with sodium lauryl sulfate (SLS), which is recognised both as an irritant and an LLNA false positive, to explain what were believed to be false positives with surfactants (e.g. Ball et al., 2011). However, many commercial surfactants, including nonionic surfactants of the type discussed by Ball et al. (2011) are substantially less irritating than SLS, such that 3 open applications on a mouse ear would be trivial with respect to irritancy. Some of the surfactants involved have, upon mature consideration, been demonstrated to be identifiable human sensitisers (e.g. Presley et al., 2021; Warshaw et al., 2022). At its core, the irritancy argument is not tenable and some of the materials are not actually false, but rather are true positives. Furthermore, cytotoxicity is at best a crude indicator for skin irritancy, as witnessed by the challenge associated with differentiating irritant potency (as opposed to skin irritant versus non-irritant) using in vitro alternatives (Kolle and Landseidel, 2021; OECD, 2021c).

The primary assumption, that lipophilic chemicals tend to be more cytotoxic than hydrophilic chemicals, seems to be based on a misunderstanding of the literature on quantitative structure-activity relationships (QSAR) in ecotoxicology and failure to distinguish between toxicity and cytotoxicity. In support of their claim that lipophilic chemicals (i.e. chemicals with high LogP values) in general have an increased cytotoxicity, Natsch et al. (2023) say: "Correlation between (cyto)toxicity and LogP is very well described in the literature on ecotoxicology, see e.g. (Tebby et al., 2011)." We note however that the words cytotoxicity or (cyto)toxicity do not appear anywhere in the Tebby et al., 2011 paper referred to, and this is precisely what would be expected based on experience in ecotoxicology QSAR.

Correlations between logP and aquatic toxicity, whereby toxicity increases with increasing logP are well known in the ecotoxicology literature, and date back to the early fish toxicity studies, wherein were developed QSARs for general (baseline) narcotic toxicity and polar narcotic toxicity respectively (Könemann, 1981; Saarikoski and Viluksela, 1982). These toxic effects are reversible (i.e. if the fish are removed from the test solutions and placed in clean water they revive immediately). They are believed to involve reversible partitioning of the toxicant into membranes (Roberts and Costello, 2003 and references therein). These are toxic effects, but they are not cytotoxic effects. We would not go so far as to deny that chemically unreactive (and non-sensitising) chemicals can have some degree of cytotoxicity and that this baseline cytotoxicity could be correlated with logP in the same way as narcotic toxicity is, but for reactive chemicals the trend, if anything is in the opposite direction, as illustrated by comparison of methyl methanesulfonate – hydrophilic, labelled as irritant (National Center for Biotechnology Information, 2023) – and methyl dodecansulfonate – hydrophobic, non-irritating to guinea pigs at 20% (Basketter and Roberts, 1990).

Thus, the argument for lipophilic chemicals producing LLNA false positives via irritancy resulting from cytotoxicity is untenable. Firstly, irritancy has little correlation with allergy (Auton et al., 1995; Basketter et al., 2007). Secondly, the suggestion that lipophilic chemicals have higher cytotoxicity and higher irritancy is a false premise. If anything, the trends are in the opposite direction.

6. Re-evaluation of LLNA performance against the Basketter et al. (2014) HSPC classifications

The arguments presented so far indicate that the apparent low specificity of the LLNA does not reflect a deficiency in the LLNA but is rather the result of conflicting decisions made by different working groups at different times on where to draw the line between what are considered weak sensitisers and what are considered non-sensitisers. In essence, the LLNA data interpretation protocol may, at least for some substances, identify levels of true sensitisation for which there may not be regulatory significance, whereas the criteria for interpreting the human data are aimed at detecting only those chemicals that are potent enough to be recognised as sensitisers for regulatory purposes. This is true also for the older guinea pig models (Basketter, 2023). Table 4, discussed in section 3, shows the results of comparing LLNA results against HSP categories, with category 5 being treated as positive. This corresponds to both the LLNA data and the human data being considered from the standpoint of detecting any level of potency, irrespective of regulatory significance. We carried out a further evaluation, this time with category 5 being treated as negative and LLNA EC3 values above an arbitrary value of 15% being treated as negative. This corresponds to the LLNA and the human data both being considered from the standpoint of only detecting potency of regulatory significance. Table 5 shows the results of this re-evaluation.

Although our choice of 15% as the EC3 cut-off was completely arbitrary, Table 5 shows clearly that the LLNA can perform well against human data from the perspective of identifying not only chemicals that should be regarded as sensitisers for regulatory purposes but also chemicals that can be treated as non-sensitizers for regulatory purposes. It should be kept in mind that the question of potency and EC3 values formed no part of the formal validation of the LLNA undertaken in 1999 (Gerberick et al., 2000; Dean et al., 2001). These findings complement those in Table 4, which show that the LLNA can perform well against human data from the perspective of identifying sensitisers, irrespective of whether or not they need to be treated as such for regulatory purposes.

7. Discussion

We are now in a position to answer the three questions posed in the Introduction:

Is the specificity of the LLNA genuinely poor? The answer is that the apparent low specificity of the LLNA is not genuine. Most of the so-called LLNA false positives are not true non-sensitisers but are weak sensitisers, correctly identified as such by the LLNA, that are deemed for regulatory and risk assessment purposes to be treatable in the same way as true non-sensitisers. In an absolute sense, these so-called LLNA false positives could very reasonably be regarded as HMT/HRIPT false negatives. The apparent low specificity of the LLNA is the result of different decisions made by different working groups at different times on where to draw the line between what are considered weak sensitisers and what are considered non-sensitisers. As shown in Tables 4 and 5, when the data interpretation criteria are adjusted so as to be more mutually consistent, the specificity becomes much higher.

Are the LLNA false positives due to proliferation without sensitisation? A small proportion of apparent LLNA positives may well be due to proliferation without sensitisation, but not necessarily associated with either cytotoxicity or irritancy. These include DMSO and SLS, and there is no evidence of correlation with hydrophobicity. Most of the apparent LLNA false positives are chemicals that genuinely cause sensitisation but are not sufficiently potent to be considered as sensitisers for regulatory purposes. These include pre-haptens such as limonene that can cause sensitisation due to their autoxidation products, which are more readily formed under the open epicutaneous application protocol of the LLNA than under the occlusion conditions of human patch testing. A similar conclusion was reached in comparison of the LLNA against GPMT data (Roberts et al., 2016).

How strong is the evidence for hydrophobicity-induced false positives? The argument that LLNA false positives (i.e. lymphocyte proliferation without sensitisation) are the result of irritancy, which in turn is the result of cytotoxicity, which in turn is dependent on hydrophobicity, is not tenable. Firstly, irritancy has little correlation with allergy. Secondly, cytotoxicity is a very poor marker for irritation. Thirdly, the suggestion that lipophilic chemicals have higher cytotoxicity and higher irritancy is a false premise. If anything, the trend is in the opposite direction.

Table 5
Predictive performance of LLNA, with EC3 > 15% treated as negative, against 98 chemicals with HSP categories for human potency
Chemicals in categories 1–4 are sensitisers, chemicals in category 5 and 6 are non-sensitisers.

Results	TP	FP	TN	FN
	53	7	20	18
Sensitivity	76%			
Specificity	74%			
Accuracy	75%			
Balanced accuracy	75%			

8. Conclusions

Recently, LLNA results argued to be false positives have been rationalised by assuming a chemico-biological interaction confounder - alleged irritancy masquerading as sensitisation (Natsch et al., 2023). Here we refute the theoretical basis hypothesised for this interpretation. We show that differences between LLNA predictions and human data arise not from chemico-biological interaction effects, but rather from differences in data interpretation protocols. In the currently adopted LLNA data interpretation protocol, applied by Natsch et al. (2023), any chemical giving SI $\geq$ 3 at any concentration is positive; for a chemical to be judged negative then an SI < 3 at a test concentration of 50% or higher has to be recorded (unless there are compelling reasons why testing at such concentrations was not possible). This protocol performs well in terms of predicting maximum potential to sensitise, as shown by our analysis against HSP categories with category 5 counted as positive. We have now demonstrated that the LLNA is also capable of performing well in terms of discrimination between chemicals having potency of regulatory significance and chemicals that are too weakly sensitising to be significant in the regulatory context. This is illustrated by the analysis with HSP category 5 substances counted as negative and an EC3>15% cut-off for LLNA negative. This 15% cut-off was arbitrarily chosen for illustrative purposes and of course it is not the only way in which an LLNA data interpretation protocol for potency of regulatory significance could be defined (eg a different EC3 cut-off, an EC4 < 100% …). A re-analysis of the dose-response data for the LLNA results on chemicals with human data would be needed to define an optimum protocol for this purpose.

The low water solubility of hydrophobic chemicals remains a limitation of the applicability domains of water-based in vitro assays (e.g. DPRA, KeratinoSens™, h-CLAT) and there is no valid justification to relax the logP = 3.5 cut-off for applicability of these assays.

Overall, sound LLNA data, appropriately interpreted, continue to constitute the best quantitative measure of skin sensitisation potency, applicable across a wide hydrophilicity-lipophilicity range, against which the performance of NAMs can be tested.

CRediT authorship contribution statement

David W. Roberts: Conceptualization, Formal analysis, Writing – review & editing. **Ian Kimber:** Conceptualization, Writing – review & editing. **David A. Basketter:** Conceptualization, Writing – review &

editing.

Declaration of competing interest

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.

Data availability

No new data were used for the research described in the article.

References

Api, A.M., Parakhia, R., O'Brien, D., Basketter, D.A., 2017. Fragrances categorized according to relative human skin sensitization potency. Dermatitis 28, 299–307.

Auton, T.R., Botham, P.A., Kimber, I., 1995. Retrospective appraisal of the relationship between skin irritancy and contact sensitization potential. J. Toxicol. Environ. Health 46, 149–154.

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 sensitization potential of surfactants: integrating data from the local lymph node assay, Guinea pig maximization test, and in vitro methods in a weight-of-evidence approach. Regul. Toxicol. Pharmacol. 60, 389–400.

Basketter, D.A., 2023. Risk Management of Skin Sensitizers. Regul. Toxicol. Pharmacol. (submitted).

Basketter, D.A., Alepee, N., Ashikaga, T., Barroso, J., Gilmour, N., Goebel, C., Hibatallah, J., Hoffmann, S., Kern, P., Martinozzi-Teissier, S., Maxwell, G., Millet, M., Reisinger, K., Sakaguchi, H., Schepky, A., Tialhardat, M., Templier, M., 2014. Categorisation of chemicals according to their relative human skin sensitizing potency. Dermatitis 25, 11–21.

Basketter, D.A., Gerberick, G.F., 2022. Skin sensitization testing: the ascendancy of non-animal methods. Cosmetics 9, 38. https://doi.org/10.3390/cosmetics9020038.

Basketter, D.A., Kimber, I., 2011. Skin irritation, false positives and the local lymph node assay: a guideline issue. Regul. Toxicol. Pharmacol. 61, 137–140.

Basketter, D.A., Kan-King-Yu, D., Dierkes, P., Jowsey, I.R., 2007. Does irritation potency contribute to the skin sensitization potency of contact allergens? Cutan. Ocul. Toxicol. 26, 279–286.

Basketter, D.A., Roberts, D.W., 1990. Structure/activity relationships in contact allergy. Int. J. Cosmet. Sci. 12, 81–90.

Basketter, D.A., White, I.R., McFadden, J.P., Kimber, I., 2015. Skin sensitization: implications for integration of clinical data into hazard identification and risk assessment. Hum. Exp. Toxicol. 34, 1222–1230.

Cal, K., 2006. Aqueous solubility of liquid monoterpenes at 293 K and relationship with calculated Log P value. Yakugaku Zasshi 126, 307–309. https://doi.org/10.1248/yakushi.126.307.

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, 258–273.

de Groot, A.C., 2018. Tocopherol. In: Monographs in Contact Allergy Volume 1 : Non-fragrance Allergens in Cosmetics," Part 2. CRC Press, Boca Raton, pp. 1253–1257.

de Groot, A.C., 2021. Salicylic acid. In: Monographs in Contact Allergy Volume 3: Topical Drugs. CRC Press, Boca Raton, pp. pp714–715.

Gerberick, G.F., Ryan, C.A., Kimber, I., Dearman, R.J., Lea, L.J., Basketter, D.A., 2000. Local lymph node assay validation assessment for regulatory purposes. Am. J. Contact Dermatitis 11, 3–18.

Hansch, C., Leo, A.J., 1979. Substituent Constants for Correlation Analysis in Chemistry and Biology. Wiley and Sons, New York.

Irizar, A., Bender, H., Griem, P., Natsch, A., Vey, M., Kimber, I., 2022. Reference Chemical Potency List (RCPL): a new tool for evaluating the accuracy of skin sensitization potency measurements by New Approach Methodologies (NAMs). Regul. Toxicol. Pharmacol. 134, 105–244. https://doi.org/10.1016/j.yrtph.2022.105244. Epub 2022 Aug. 3.

Kolle, S.N., Landsiedel, R., 2021. Human-derived in vitro models used for skin toxicity testing under REACh. Handb. Exp. Pharmacol. 65, 3–27.

Könemann, H., 1981. Quantitative structure-activity relationships in fish toxicity studies Part 1: relationship for 50 industrial pollutants. Toxicology 19, 209–221.

Li, H., Bai, J., Zhong, G., Lin, H., He, C., Dai, R., Du, H., Huang, L., 2019. Improved defined approaches for predicting skin sensitization hazard and potency in humans. ALTEX 36, 363–372. https://doi.org/10.14573/altex.1809191. Epub 2019 Jan 23.

Na, M., O'Brien, D., Lavelle, M., Lee, I., Gerberick, G.F., Api, A.M., 2022. Weight of evidence approach for skin sensitization potency categorization of fragrance ingredients. Dermatitis 33, 161–175.

National Center for Biotechnology Information, 2023. PubChem Compound Summary for CID 4156, Methyl Methanesulfonate. Retrieved March 21, 2023 from. https://pubchem.ncbi.nlm.nih.gov/compound/Methyl-methanesulfonate.

Natsch, A., Gerberick, G.F., 2022. Integrated skin sensitization assessment based on OECD methods (II): hazard and potency by combining kinetic peptide reactivity and the "2 out of 3" Defined Approach. ALTEX 39 (4), 647–655. https://doi.org/10.14573/altex.2201142. Epub 2022 Apr 11.

Natsch, A., Kleinstreuer, N., Asturiol, D., 2023. Reduced specificity for the local lymph node assay for lipophilic chemicals: implications for the validation of new approach methods for skin sensitization. Regul. Toxicol. Pharmacol. https://doi.org/10.1016/j.yrtph.2023.105333.

OECD, 2018. In vitro skin sensitization assays addressing the key event on activation of dendritic cells on the adverse outcome pathway for skin sensitization. OECD Guidelines for the Testing of Chemicals, Section 4, 442e. https://doi.org/10.1787/9789264264359-en.

OECD, 2021a. Supporting Document to the OECD Guideline 497 on Defined Approaches for Skin Sensitisation. Series on Testing and Assessment, No. 336. https://one.oecd.org/document/ENV/CBC/MONO(2021)11/En/pdf. (Accessed 15 March 2023).

OECD, 2021b. Analysis of LLNA Reference Data to Conclude on Predictivity of Alternative Methods for Skin Sensitization for Lipophilic Chemicals. Annex 6, Series on Testing and Assessment, No. 336, Paris. https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV-CBC-MONO(2021)11/ann6%20&doclanguage=en. (Accessed 27 January 2023).

OECD, 2021c. Test No. 439. In: Vitro Skin Irritation: Reconstructed Human Epidermis Test Method. Paris.

Presley, C.L., Militello, M., Barber, C., Ladd, R., Laughter, M., Ferguson, H., Dewey, J., Pulsipher, K.J., Rundle, C.W., Dunnick, C.A., 2021. The history of surfactants and review of their allergic and irritant properties. Dermatitis 32, 289–297.

PubChem PubChem [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004-. PubChem Compound Summary for CID 7793, (-)-Citronellol; [cited 2023 Jan. 16]. Available from:https://pubchem.ncbi.nlm.nih.gov/compound/Citronellol.

Roberts, D.W., Aptula, A.O., Patlewicz, G., 2006. Mechanistic applicability domains for non-animal based toxicological endpoints. QSAR analysis of the Schiff Base applicability domain for skin sensitization. Chem. Res. Toxicol. 19, 1228–1233.

Roberts, D.W., Costello, J.F., 2003. Mechanisms of action for general and polar narcosis: a difference in dimension. QSAR Comb. Sci. 22, 226–233.

Roberts, D.W., 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, 592–603.

Roberts, D.W., Schultz, T.W., Api, A.M., 2016. Chemical applicability domain of the Local Lymph Node Assay (LLNA) for skin sensitization potency. Part 3. Apparent discrepancies between LLNA and GPMT sensitization potential: false positives or differences in sensitivity? Regul. Toxicol. Pharmacol. 80, 260–267.

Saarikoski, J., Viluksela, M., 1982. Relation between physicochemical properties of phenols and their toxicity and accumulation in fish. Ecotoxicol. Environ. Saf. 6, 501–512.

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.

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. Sci. Total Environ. 409, 3334–3343. https://doi.org/10.1016/j.scitotenv.2011.05.029.

Warshaw, E.M., Xiong, M., Atwater, A.R., DeKoven, J.G., Pratt, M.D., Maibach, H.I., Taylor, J.S., Belsito, D.V., Silverberg, J.I., Reeder, M.J., Zug, K.A., Sasseville, D., Fowler Jr., J.F., DeLeo, V.A., Houle, M.C., Dunnick, C.A., 2022. Patch testing with glucosides: the north American contact dermatitis group experience, 2009-2018. J. Am. Acad. Dermatol. 87, 1033–1041.