Regulatory Toxicology and Pharmacology 108 (2019) 104473

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

In vitro RHE skin sensitisation assays: Applicability to challenging substances

A. Mehling[a], E. Adriaens[b], S. Casati[c], B. Hubesch[d], A. Irizar[e], M. Klaric[f], S. Letasiova[g], I. Manou[h], B.P. Müller[i], E. Roggen[j], E. van Vliet[f], D. Basketter[k,*]

[a] BASF Personal Care and Nutrition GmbH, Duesseldorf, Germany
[b] Adriaens Consulting BVBA, Aalter, Belgium
[c] European Commission, Joint Research Centre (JRC), Ispra, Italy
[d] Hubesch Consult BVBA, Sint-Pieters-Leeuw, Belgium, European Chemical Industry Council (Cefic), Brussels, Belgium
[e] The International Fragrance Association (IFRA), Geneva, Switzerland
[f] Cosmetics Europe, Brussels, Belgium
[g] MatTek In Vitro Life Science Laboratories, Bratislava, Slovakia
[h] European Partnership for Alternative Approaches to Animal Testing (EPAA), Brussels, Belgium
[i] Symrise AG, Holzminden, Germany
[j] 3Rs Management and Consulting ApS, Kongens Lydngby, Denmark
[k] DABMEB Consultancy Ltd, Sharnbrook, UK

ARTICLE INFO

Keywords:
Skin sensitisation
In vitro alternatives
Reconstructed human epidermis
Hazard identification

ABSTRACT

In the last 20 years, alternative approaches to the identification of skin sensitisation hazards have been at the forefront of the 3Rs and have helped refine the validation and acceptance processes. However, experience with the local lymph node assay showed that, post-validation, challenges still occurred, particularly when a wider diversity of chemical substances was addressed, a situation which will arise with validated in vitro alternatives. In the present work, a range of substances potentially challenging to assess in current nonanimal OECD test guidelines were evaluated in several of the emerging in vitro alternatives. Twelve such substances (of which just over half were known skin sensitisers) were assessed in 4 assays, all based on reconstructed human epidermis (RHE) models. For hazard identification, the overall predictive accuracy ranged around 70% for three assays, although for one (SensCeeTox), it fell below 50% when human data was used as the benchmark. In most cases, sensitivity was high, such that sensitisation was overpredicted. As the substances were challenging to assess in other nonanimal methods, the results indicate that the 3D RHE models may be a useful tool for assessing skin sensitisation potentials without needing to revert to animal use.

1. Introduction

Since the pace of the development and validation of alternatives was accelerated by the implementation of the European legislation which prohibited the use of animal tests for safety assessment of cosmetic products and their ingredients (European Union, 2009), there has been a steady adoption of nonanimal methods, most notably in the fields of irritation and sensitisation both in regulatory and risk assessment settings (OECD, 2015a, 2015b; 2016, 2018a; 2018b, 2018c; Ezendam et al., 2016; Strickland et al., 2017; Hoffmann et al., 2018; Kleinstreuer et al., 2018). The primary development of in vitro alternatives has depended upon research insights in the relevant field of toxicology. When this development has been successful and gained a degree of acceptance, the more formalised procedures of independent validation to demonstrate the relevance and reliability of a method have been initiated. Once validation has been achieved, the methods have then progressed to the Organisation for Economic Co-operation and Development (OECD), where they have been taken up and promulgated in globally accepted test guidelines (e.g. OECD, 2015a; 2015b, 2016). Subsequent to that, each method, generally in combination with others, is finally taken up into regulatory guidelines such as those in the European Union derived from the United Nations (European Union, 2008; United Nations, 2015). Once all of this has been achieved, relevant industries around the world are able to fully adopt and implement the nonanimal methods, confident that the results will be accepted by the authorities to whom they are required to submit the data.

* Correspondening author. DABMEB Consultancy Ltd, Sharnbrook, M44 1PR, UK.
E-mail address: dabmebconsultancyltd@me.com (D. Basketter).

https://doi.org/10.1016/j.yrtph.2019.104473
Received 15 May 2019; Received in revised form 30 July 2019; Accepted 4 September 2019
Available online 05 September 2019

Whilst some have been known to observe that the overall process outlined above can be cumbersome and consequently prolonged (e.g. Basketter et al., 2010), history will show that at least for the endpoints already addressed (skin and eye irritation, skin sensitisation, dermal penetration, phototoxicity, preliminary tests in genotoxicity and so on) there appears to have been a good deal of success. Nevertheless, there is one aspect of the entire development, validation and acceptance suite of processes that needs to be considered, regardless of whether it is an animal or a nonanimal method. This is the fact that those processes can never be broad enough to evaluate the vast spectrum of substances that the world of chemistry may present to them. In addition, the knowledge of the applicability domain of the test methods, including their technical limitations as defined in course of validation, will necessarily evolve as soon as new experience is acquired with the implementation of the methods by various laboratories.

In the present work, the toxicology endpoint that is addressed is that of skin sensitisation, where a suite of new nonanimal test methods based on key events in the adverse outcome pathway have become available in the last few years (OECD, 2015a, 2018b; 2018c). At least initially, the validated cell-based methods have required a substance to be placed in a hydrophilic medium compatible with the respective suspension cell culture (e.g. Ashikaga et al., 2006; Emter et al., 2010). This is already one well recognised limitation, as some skin sensitising chemicals may be very hydrophobic and thus not amenable to these test systems. Indeed, in consumer products they may well be applied to skin using vehicles or formulations which are not water miscible. Thus a further advantage of RHE assay is the scope to apply substances in “real-life” vehicles. With the existing validated in vitro alternatives, it is known already that a number of substances, for a variety of reasons, have proven somewhat difficult to evaluate (e.g. Sakaguchi et al., 2010). To the extent characterised so far, these issues are documented in the respective OECD Test Guidelines and may be updated in future as new knowledge on the applicability of the methods will be acquired. Accordingly, various efforts are underway to develop, validate and gain acceptance for approaches based on reconstructed human epidermis (RHE) models since these systems develop a form of stratum corneum at the air-liquid interface. Consequently, direct application of test materials in lipophilic solvents to such a surrogate skin surface obviates the key technical challenges associated with cells in suspension (e.g. Gibbs et al., 2013; Cottrez et al., 2016; Rodrigues Neves and Gibbs, 2018). What we have chosen to address in this study is how a range of upcoming assays based on RHE models perform with a selection of substances that may prove challenging to test in aqueous culture systems.

2. Materials and methods

2.1. Test substances

Table 1 records the 12 substances assessed in this study, their source and purity, the existing outcomes from in vivo and in vitro testing, together with a primary justification for their selection. The underlying data in the table is adapted from previous publications (Basketter et al., 2014; Urbisch et al., 2015; Api et al., 2017; Hoffman et al., 2018), together with consultation of the European Chemicals Agency (ECHA) dissemination website (ECHA, 2018). Although not stated explicitly in the table, a key requirement was a good body of existing data, including human information, this last mentioned also being presented in the same table, and which is based on the categories introduced previously (Basketter et al., 2014). Test substances were coded, with different coding being adopted for each laboratory. Also, note that vehicle and concentration selection was a matter for the individual test laboratories.

2.2. SensCeeTox assay

In this assay, changes in keratinocyte gene expression are measured alongside protein reactivity. The assay is fully described already (McKim et al., 2012), but in brief was carried out as follows: RHE from EpiDerm™ (MatTek Corporation, USA), equilibrated for a minimum of 1 h prior to dosing with test and reference substances, was exposed to a range of six concentrations of each test substance for 24 h 10% Triton X-100 was used as a viability control. RNA was isolated using QIAGEN's RNeasy 96 kit (QIAGEN, LN 151052720) according to Cyprotex standard operating procedure (SOP) 2062. Isolated RNA was quantitated with a Nanodrop ND-8000 spectrophotometer with 2 μL of eluted RNA. Gene induction by test substances and controls was determined by running quantitative reverse transcription-polymerase chain reaction (qRT-PCR) according to Cyprotex SOP-2050. The fold-over vehicle results for the four control genes were plotted in a box and whisker plot. The control gene that showed the lowest variation was selected. The target gene induction over vehicle was then divided by the selected control gene to determine a fold-induction over background for each target gene.

The second element of the assay required a measurement of protein reactivity and followed methods already published (Natsch et al., 2007; Gerberick et al., 2004). In brief, reaction mixes were prepared in 96-well micro plates with each test substance run. The reactions were incubated for 24 h at room temperature After 24 h, metaphosphoric acid (MPA) was added. Then 50 μL of reaction mix was added to a 96-well plate with 10 μL of 1.5 M triethanolamine. After 2 min of shaking, 150 μL glutathione (GSH) assay buffer was added to the wells. The plate was shaken for an additional 2 min and absorbance at 415 nm read on a plate reader. The mean absorbance for the sample triplicates was divided by the mean of the vehicle controls, and the percent of GSH depletion relative to vehicle control was calculated.

Outputs from the above assay components were used as inputs to the prediction model which employs an algorithm which is proprietary to SensCeeTox, grading the test substance into one of five categories, the lowest two of which do not trigger classification as a skin sensitiser (McKim et al., 2012).

2.3. SENS-IS assay

This assay is fully described already (Cottrez et al., 2016), but in brief was carried out as follows: test substance solution was applied to the surface of 2 week cultured reconstituted epidermis (SkinEthic, Episkin, Lyon) for 15 min at 37C. The 6 test concentrations employed were 0.1%, 1%, 10%, 50%, and 100% v/v, although 10% and 50% are typically tested first and, if necessary, subsequent concentrations chosen based on these initial results. After a washing with phosphate buffered saline, the epidermis was incubated without test substance exposure for a further 6 h. At the end of this incubation, the epidermis was snap frozen in liquid nitrogen and total RNA extracted for qRT-PCR analysis. Transcript measurements for 61 genes were normalized to the mean expression levels of 3 housekeeping genes. The genes were categorised into three groups: 23 genes as indicators of skin irritation, 21 “SENS-IS” genes, and 17 “redox” genes. The first group of genes was used for assay acceptance, and if $>$ 20 of these irritation genes were over-expressed ($>$ 1.25x control vehicle value) the test was rejected, and the substance was tested at a lower concentration. The relative expression of the “SENS-IS” and “redox” genes were used for the prediction of a test substance's skin sensitising potency. While the identity of irritation genes has been published (Cottrez et al., 2016), the identity of the genes of the “SENS-IS” and “redox” genes has only been disclosed in a patent. The cycle threshold value of the HSPAA1 gene was used for the evaluation of tissue destruction and had to be 110% to be acceptable. The prediction model, which was developed to predict the LLNA result, was based on the number of over-expressed genes in the SENS-IS and redox group and the test substance concentration. To be considered as a skin sensitiser, a test substance had to induce at least a 1.25-fold increase in the expression of at least seven genes in either the SENS-IS or redox group. The lowest concentration meeting these requirements determined the potency prediction: 100% or 50% = weak sensitiser,

Table 1
Substances selected with existing underlying data and brief rationale.

Substance	Typical use	CAS number	Supplier and reference	Purity	Sensitiser DPRA	Sensitiser KeratinoSens	Sensitiser hCLAT	Sensitiser LLNA	Sensitiser GPMT	Classifiable human skin sensitiser	Justification for testing
Hexaethylene glycol monododecyl ether	surfactant	3055-96-7	Sigma-Aldrich/ 52044	≥98.0%	no	no	no	yes	no	no indication in market data (2011)	high cytotoxicity; disruption of cell membranes
Resorcinol	dye coupler	108-46-3	Sigma-Aldrich/ 398047	≥99.0%	no	no	yes	yes	yes	yes (class 4)	pre/prohapten and false negative in DPRA and KeratinoSens
Isopropyl myristate	fatty ester	110-27-0	Sigma-Aldrich/ 172472	98%	no	no	yes	yes	no	no (class 5)	high logP; LLNA gives incorrect classification
Abietic acid	resin	514-10-3	Sigma-Aldrich/ 00010	~75%	yes	yes	no	yes	yes	yes (class 3)	prehapten (hydroperoxide former); high logP
Aniline	dye	62-53-3	Sigma-Aldrich/ 242284	≥99.5%	no	no	yes	yes	yes	yes (class 4)	pre/prohapten and false negative in DPRA and KeratinoSens
Propyl paraben	preservative	94-13-3	Sigma-Aldrich/ P53357	≥99.0%	no	yes	yes	no	no	no (class 5)	false positive in in vitro assays compared to human and animal data
2-Chloro-6-methyl-3-aminophenol	dye	84540-50-1	Extrachem GmbH, Bielefeld, Germany	98%	yes[b]	yes	yes	no	no data	no data	false positive in in vitro assays compared to LLNA; dye
2-Hydroxy-4-methoxy benzophenone	UV filter	131-57-7	Sigma-Aldrich/ H36206	98%	yes[b]	yes	yes	no	no	under review[c]	false positive in in vitro assays compared to LLNA
Farnesol	fragrance	4602-84-0	Sigma-Aldrich	≥95.0%	no	yes	yes	yes	equivocal	yes(class 3)	pre/prohapten; high logP
Benzoyl peroxide	bleaching/anti-acne agent	94-36-0	Sigma-Aldrich/ 517909	75%	yes	no	no	yes	yes	yes (class 3)	high logP; discordant results in vitro (2 out of 3) vs. in vivo (LLNA, human); oxidizing agent
Amylcinnamyl alcohol	fragrance	101-85-9	Sigma-Aldrich	≥94.0%	no	no	yes	no	yes	yes (class 4)	high logP; discordant results in vitro (2 x neg) vs. in vivo (LLNA, human “borderline")
Tween 80	surfactant	9005-65-6	Sigma-Aldrich/ P1754	100%[a]	yes	yes	no	no	no	no (class 6)	cytotoxicity by disruption of cell membranes

[a] Tween 80 is a mixture of long chain fatty acids with majority (65%–80%) oleic acid, the remainder being largely linoleic, palmitic and stearic acids.
[b] Unpublished data on blinded testing courtesy of Cosmetics Europe, Brussels.
[c] At study initiation this was regarded as class 5, whereas according to the Api et al., 2017 criteria this could now be categorised as class 4.

10% = moderate sensitiser, 1% = strong sensitiser, and 0.1% = extreme sensitiser; if none of the test concentrations met these criteria, the substance was considered to be a non-sensitiser (Hoffmann et al., 2018).

2.4. RHE IL-18 method (epiCS)

This method is fully described already (Corsini et al., 2013; Gibbs et al., 2013), but in brief was carried out as follows: 25 μl of a range of concentrations of test substance are applied to the surface of the RHE skin model (CellSystems® Biotechnologie, Troisdorf, Germany) for 24 h. After washing with phosphate buffered saline to terminate the exposure, cell viability is measured using an MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay. Levels of interleukin (IL) −18 in the culture medium are determined by enzyme linked immunosorbent assay (ELISA) in both starting and 24 h samples. In this method, a test substance is classified as contact allergen if it induces a 5-fold increase in stimulation index (SI) in IL-18 release compared to vehicle RHE at a cell viability between 5% and 40% in at least one of the concentrations tested. To optimise the prediction model, where there is no cytotoxicity, but the following criteria are met: (1) a 5-fold increase (SI) in IL-18 release compared to vehicle treated RHE, (2) in at least two independent runs and (3) at least at one of the concentrations tested in each run, then the test substance should be considered a skin sensitiser. In this situation, for relative potency ranking only the two-fold stimulation index (SI-2) can be used. In situation where there is no cytotoxicity at the maximum soluble concentration and no increase in IL-18 release, the test substance should be considered a non-sensitiser in the RHE IL-18 assay. In the case of two runs yielding discordant results, a third run is conducted and a "two out of three" approach for classification is used.

2.5. RHE IL-18 method (EpiDerm)

The EpiDerm IL-18 protocol has been developed on the basis of published research (Corsini et al. 2008; Deng et al., 2011; Gibbs et al., 2013) and further optimised by MatTek IVLSL (Letasiova et al., 2018) in collaboration with University Milano. Briefly, test substances stock solutions were prepared in the standard assay vehicles of Dulbecco's phosphate buffered saline (DPBS) or acetone:olive oil (4:1), and applied directly applied to the surface of EpiDerm RHE model (25 μl, without occlusion) for 24 h. At the conclusion of the exposure period, test substances were rinsed from the surface of the tissues with DPBS. Subsequently, the MTT assay was conducted in order to determine the EC-50 value, i.e. the effective chemical concentration required to reduce cell viability to 50% of RHE measured by metabolic activity compared to the maximum value of vehicle treated RHE. Culture medium from the test tissues was stored frozen at −80 °C until analysis for IL-18 content. IL-18 secretion was analysed using a method developed by University of Milano (self-made kit). A stimulation index (SI) was calculated for each of the 12 concentrations of a test substance. In this model, a test substance was classified as a skin sensitiser if it induced a 2.0-fold or greater increase (SI) in IL-18 release compared to vehicle-treated RHE at a cell viability < 50% in at least one of the concentrations tested.

Given its exploratory nature, all the study work reported in this paper was carried out to the spirit of, but not in full compliance with, formal good laboratory practice (GLP). The investigators in each individual laboratory were responsible for taking the steps necessary to assure the quality and integrity of the data.

3. Results

3.1. a. SensCeeTox assay

Key test details and the test results are presented in Table 2. All 12 of the test substances proved to give positive results in the SensCeeTox assay. As the classification "weak" would be considered to be a non-sensitiser in this test, five substances were classified as non-sensitizers and eight as sensitizers (positive predictions). This assay also delivers a prediction of potency; outcomes were almost evenly distributed between the weak and moderate categories, with the exception of the surfactant Tween 80, which was classified as extreme, but which is a non-sensitiser in humans.

3.2. b. SENS-IS assay

Key test details and the test results are presented in Table 3. Three of the test substances were shown to be non-sensitisers; the remainder proved positive in this assay. The proprietary nature of the prediction model limits a more detailed analysis, although potency predictions were evenly split between weak and moderate.

3.3. c. RHE IL-18 (EpiCS)

Key test details and the test results using an optimised prediction model are presented in Table 4. In this assay, three of the test substances were shown in all runs to be non-sensitisers; a fourth was negative once, but proved positive in another two runs; the remaining eight substances proved reproducibly positive in this assay.

3.4. d. RHE IL-18 (EpiDerm)

Key test details and the test results for this assay are presented in Table 5. Whereas three of the test substances reliably were shown to be non-sensitisers, a fourth produced discordant results. The remaining eight substances proved reproducibly positive in this assay.

However, concern over the integrity of two samples led to (unblinded) retesting. At the time of original testing, the first original sample of benzoyl peroxide (and which was negative in this assay) had passed its expiration date and also required heating to enable solubility (which may also have led to inactivation), whereas a fresh sample was clearly positive in the retest. Three further experiments were performed on the same tissue lot number, in parallel, with a fresh sample from the original supplier. When the 12 dilutions were prepared from the stock concentration 200 mg/ml (without heating so the compound was not well soluble) the overall prediction was negative. To address the limited solubility at 200 mg/ml, the other 2 experiments used stock concentrations of 100 mg/ml (solubility was much better, but not complete) and 50 mg/ml (completely soluble). In both cases, in each of 2 runs per experiment, these lower concentrations gave positive results. Probably, solubility and bio-availability had a significant impact on the prediction for this substance. In addition, a third experiment was conducted with 2-hydroxy-4-methoxybenzophenone using both the original sample and a replacement sample from the same lot as the original, run in parallel. Whereas the former remained negative, the replacement sample was clearly positive.

3.5. e. Comparison with OECD in vitro, existing in vivo and human data

Table 6 offers an overview of how results from the RHE-based in vitro assays compared with the known in vivo results from animal (LLNA) and human data (where the correlation is only approximately 2 out of 3). For each assay, sensitivity (rate of true positives), specificity (rate of true negatives) and overall accuracy (rate of correctly identified substances in the test dataset) were calculated.

The performance with respect either to the LLNA or to the human result is comparable for all of the assays except SensCeeTox, where the lack of any apparent specificity suggests that the prediction model of this assay may need to be modified and/or that the assay is not suited for use with certain challenging substances. So, with this exception, the RHE assays appear to offer an enhanced performance compared to the

Table 2
SensCeeTox results.

Substance	Vehicle/solvent	Depletion (%)		EC3 (%)	Predicted potency[a]	Prediction[b]
		Mean	SD			
Hexaethylene glycol monododecyl ether	PBS	1.5	0.7	2.05	Moderate	Sensitiser
Resorcinol	Water	−0.8	2.3	12.21	Weak	Non-sensitiser
Isopropyl myristate	Olive oil	18.7	3.3	6.46	Moderate	Sensitiser
Abietic acid	A/OO[c] 1:1	4.0	0.7	6.12	Moderate	Sensitiser
Aniline	A/OO 1:1	−5.1	7.0	10.28	Weak	Non-sensitiser
Propyl paraben	A/OO 1:1	1.9	5.5	7.54	Moderate	Sensitiser
2-Chloro-6-methyl-3-aminophenol	A/OO 1:1	−0.2	1.9	15.81	Weak	Non-sensitiser
2-Hydroxy-4-methoxybenzophenone	A/OO 1:1	1.1	2.9	9.37	Moderate	Sensitiser
Farnesol	A/OO 1:1	0.5	2.0	12.06	Weak	Non-sensitiser
Benzoyl peroxide	A/OO 1:1	2.5	1.4	6.27	Moderate	Sensitiser
Amylcinnamyl alcohol	A/OO 1:1	−2.4	4.2	10.46	Weak	Non-sensitiser
Tween 80	A/OO 1:1	4.5	1.5	0.02	Extreme	Sensitiser

[a] The proprietary nature of the prediction model means that the detail of the results is not available.
[b] In the SensCeeTox prediction model, a weak response is considered not to classify as a skin sensitiser.
[c] Acetone/olive oil.

Table 3
SENS-IS results.

Substance	Vehicle	Predicted potency[a]	Prediction[a]
Hexaethylene glycol monododecyl ether	PBS or DMSO (both negative)	Non-sensitiser	Non-sensitiser
Resorcinol	Water	Moderate	Sensitiser
Isopropyl myristate	Olive Oil	Weak	Sensitiser
Abietic acid	DMSO	Weak	Sensitiser
Aniline	DMSO	Weak	Sensitiser
Propyl paraben	DMSO	Weak	Sensitiser
2-Chloro-6-methyl-3-aminophenol	DMSO	Weak	Sensitiser
2-Hydroxy-4-methoxybenzophenone	DMSO	Non-sensitiser	Non-sensitiser
Farnesol	DMSO	Moderate	Sensitiser
Benzoyl peroxide	DMSO	Moderate	Sensitiser
Amylcinnamyl alcohol	DMSO	Moderate	Sensitiser
Tween 80	DMSO	Non-sensitiser	Non-sensitiser

[a] The proprietary nature of the prediction model means that further detail of the results is not available.

existing OECD in vitro assays where predictions were based on the "2 out of 3" model (Bauch et al., 2012). The overall performance in terms of sensitivity, specificity and accuracy for each of the 4 RHE assays, as well as the "2 out of 3" approach for this limited set of chemicals, compared to LLNA (n = 12) and human (N = 11) data is contained in Table 7.

4. Discussion

Perhaps the most important caveat to be kept in mind when considering the results from this study is that the substances were chosen because, for a variety of reasons (see Table 1) they were deemed to be difficult to evaluate using the currently adopted in vitro OECD test methods (OECD 442C, D and E). Furthermore, the intention of this study was a comparative feasibility trial with several methods, so only a small number of chemicals were tested. Accordingly, it is reasonable to expect when testing "challenging" substances that measures such as sensitivity, specificity and overall accuracy are unlikely to match those found with the broad spectrum of chemistry normally deployed in skin sensitisation validation studies, i.e. typically in the range of 75%–90% (e.g. Gerberick et al., 2000; Sakaguchi et al., 2010; Emter et al., 2010; Cottrez et al., 2016). Also, it should be borne in mind that as the choice of test concentrations and vehicles was the responsibility of the test institution (and the requirements of the relevant SOP), this means that there are additional variables to consider when relative test performance is assessed.

Despite the above mentioned shortcomings, it is interesting to note a few trends. The SensCeeTox method proved to be too sensitive, identifying all tested substances as skin sensitising (albeit with varying potency predictions, but classifiable as non-sensitising for the weakly positive results in their prediction model). This was regardless of whether a substance proved initially positive in the LLNA or compared to human evidence. Whilst the other three assays (SENS-IS and IL-18 in two RHE test systems) provided more varied results, these did not always match the assessment made with the LLNA or the human situation. Nevertheless, it is worth remembering that the LLNA also did not correctly predict the human sensitisation potential of all substances (e.g. 74% or 82% accuracy, depending on the study and materials tested (Hoffmann et al., 2018 or Urbisch et al., 2015 respectively)). In this regard it is interesting to note that overall, there is a tendency for higher sensitivity, specificity, and accuracy of the RHE assays (excluding SensCeeTox) when human data are used as a basis for comparison versus the comparison based on the LLNA. Furthermore, in comparison to the predictions stemming from the validated cell-culture based assays ('2 out of 3' approach), the overall sensitivity, specificity, and accuracy of the RHE-based assays (except for SensCeeTox) were clearly superior, at least for this test set.

The results obtained in this small study highlight also the wider challenges of bioavailability in in vitro assays, reinforce the need to ensure sample integrity and in general demonstrate that it is unlikely any in vitro (or in vivo) predictive assay in toxicology will ever be perfect (Basketter et al., 2009). Indeed, if anything, they strengthen the argument that toxicology results should always be subject to weight of evidence expert judgment.

Overall, it is considered that the tests with reconstructed human epidermis can assist in a weight of evidence approach to identify the skin sensitisation potential of substances for which testing in aqueous medium proves difficult. The results with the tested substances in these 3D RHE skin models suggest that the positive experience with in vitro methods in culture media may be extended by such 3D skin models. Enhancing the dataset and experience is necessary to move such 3D models to an OECD guideline stage to complement the existing validated cell-culture based models.

In conclusion, the hereby presented results from 3 RHE assays (SENS-IS, and IL-18 with two different skin models) suggest a superior performance for a small set of substances that prove difficult to test in aqueous in vitro assays. This is particularly true when the results are compared to skin sensitisation data in humans. The results of this study

Table 4
RHE-IL18 epiCS.

Substance	Vehicle	Experiment 1				Experiment 2				Experiment 3[b]				Final prediction
		Conc. (mg/mL)	Viability (%)	IL-18 SI	Pred.[a]	Conc. (mg/mL)	Viability (%)	IL-18 SI	Pred.[a]	Conc. (mg/mL)	Viability (%)	IL-18 SI	Pred.[a]	
Hexaethylene glycol monododecyl ether	AOO	0.78	54.7	5.6	**S**[c]	1.56	47.1	5.0	**S**					**Sensitiser**
Resorcinol	AOO	200	4.0	−0.1	**NS**[c]	0.39	120.2	5.21	**S**	0.1	72.2	5.7	**S**	**Sensitiser**
Isopropyl myristate	AOO	200	104.1	3.1	**NS**	200	107.1	0.8	**NS**					**Non-sensitiser**
Abietic acid	Ethanol	200	5.6	6.9	**S**	200	49.9	5.2	**S**					**Sensitiser**
Aniline	AOO	100	30.8	7.4	**S**	100	15.8	14.3	**S**					**Sensitiser**
Propyl paraben	AOO	12.5	56.1	6.3	**S**	12.5	40.8	5.1	**S**					**Sensitiser**
2-Chloro-6-methyl-3-aminophenol	AOO	50	74	7.6	**S**	50	50.1	14.5	**S**					**Sensitiser**
2-Hydroxy-4-methoxybenzophenone	AOO	200	98.9	0.8	**NS**	200	115.9	−0.2	**NS**					**Non-sensitiser**
Farnesol	AOO	25	13.7	5.5	**S**	25	29.3	7.0	**S**					**Sensitiser**
Benzoyl peroxide	AOO	200	102.2	0.1	**NS**	200	104.5	−1.2	**NS**					**Non-sensitiser**
Amylcinnamyl alcohol	AOO	1.56	98.1	9.5	**S**	25	15.1	6.6	**S**					**Sensitiser**
Tween 80	AOO	200	95.2	0.5	**NS**	200	80.9	−0.2	**NS**					**Non-sensitiser**

[a] RHE-IL18 epiCS optimised prediction model: test substances that induce stimulation index ≥ 5 are predicted sensitiser (S) and test substances that induce stimulation index < 5 are predicted non-sensitiser (NS).
[b] In the case of two runs yielding discordant results, a third run is conducted and a "two out of three" approach for classification is used.
[c] S = Sensitiser; NS = Non-sensitiser.

Table 5
RHE-IL18 EpiDerm.

Substance	Vehicle	Experiment 1				Experiment 2				Experiment 3[b]				Final prediction
		Conc. (mg/mL)	Viability (%)	IL-18 SI	Pred. [a]	Conc. (mg/mL)	Viability (%)	IL-18 SI	Pred. [a]	Conc. (mg/mL)	Viability (%)	IL-18 SI	Pred. [a]	
Hexaethylene glycol monododecyl ether	AOO (4:1)	6.25	14.8	3.2	**S**	3.125	37.5	2.1	**S**					**Sensitiser**
Resorcinol	AOO (4:1)	50	36.1	5.9	**S**	25	45.7	6.5	**S**					**Sensitiser**
Isopropyl myristate	AOO (4:1)	200	96.9	0.8	**NS**	200	102.7	0.4	**NS**					**Non-sensitiser**
Abietic acid	AOO (4:1)	12.5	15.6	6.25	**S**	12.5	23.4	4.7	**S**					**Sensitiser**
Aniline	AOO (4:1)	50	46.2	3.6	**S**	25	32.5	5.7	**S**					**Sensitiser**
Propyl paraben	AOO (4:1)	6.25	29.7	4.9	**S**	3.125	45.4	3.7	**S**					**Sensitiser**
2-Chloro-6-methyl-3-aminophenol	AOO (4:1)	25	35.9	5.3	**S**	25	16.1	4.1	**S**					**Sensitiser**
2-Hydroxy-4-methoxybenzophenone	AOO (4:1)	50	42.0	2.9	**S**	200	62.9	0.4	**NS**	200	72.5	0.8	**NS**[c]	**Non-sensitiser** [c]
Farnesol	AOO (4:1)	25	33.2	4.3	**S**	25	33.5	3.3	**S**					**Sensitiser**
Benzoyl peroxide	AOO (4:1)			< 2.0	**NS**			< 2.0	**NS**					**Non-sensitiser** [d]
Amylcinnamyl alcohol	AOO (4:1)	25	39.4	4.2	**S**	25	41.7	4.5	**S**					**Sensitiser**
Tween 80	AOO (4:1)	200	74.2	0.7	**NS**	200	79.4	0.8	**NS**					**Non-sensitiser**

[3] S = Sensitiser; NS = Non-sensitiser.
[a] RHE-IL18 EpiDerm prediction model: test substances that induce stimulation index ≥ 2 are predicted sensitiser (S) in IL-18 release compared to vehicle-treated RHE at a cell viability < 50% in at least one of the concentrations tested; test substances that induce a stimulation index < 2 are predicted non-sensitiser (NS).
[b] In the case of two runs yielding discordant results, a third run is conducted and a "two out of three" approach for classification is used.
[c] Unblinded retesting.
[d] Unblinded retesting, due to solubility issues in the first experiments this test substance was retested and it was identified as S in two independent experiments at a concentration of 50 mg/ml – see Results section.

Table 6
Results summary and comparison with OECD-validated[1] in vitro, existing in vivo and human outcomes.

Substance	LLNA	Human2	DPRA	Keratinosens	h-CLAT	2 out of 3 in vitro prediction	SensCeeTox	SENS-IS	IL-18 epiCS®3	IL-18 EpiDerm
Hexaethylene glycol monododecyl ether	S[4]	NS[4]	NS	NS	NS	NS	S	NS	S	S
Resorcinol	S	S	NS	NS	S	NS	NS	S	S	S
Isopropyl myristate	S	NS	NS	NS	S	NS	S	S	NS	NS
Abietic acid	S	S	S	S	NS	S	S	S	S	S
Aniline	S	S	NS	NS	S	NS	NS	S	S	S
Propyl paraben	NS	NS	NS	S	S	S	S	S	S	S
2-Chloro-6-methyl-3-aminophenol	NS	Not known	S	S	S	S	NS	S	S	S
Farnesol	S	S	NS	S	S	S	NS	S	S	S
2-Hydroxy-4-methoxybenzophenone	NS	S	No data	S	S	S	S	NS	NS	NS[4]
Benzoyl peroxide	S	S	S	NS	NS	NS	S	S	NS	NS
Amylcinnamyl alcohol	NS	S	NS	NS	S	NS	NS	S	S	S
Tween 80	NS	NS	S	S	NS	S	S	NS	NS	NS

[1] The local lymph node assay (LLNA), direct peptide reactivity assay (DPRA) and human cell line activation test (h-CLAT) were the validated alternatives and the time of the study.
[2] It is essential to remember that this judgment is made with respect to regulatory classification; substances described as NS may still sensitise, albeit rarely.
[3] Outcome includes the results with the optimised prediction model.
[4] S = Sensitiser; NS = Non-sensitiser
[5] Correctly identified as a skin sensitiser when a fresh, but non-blinded sample was evaluated.

Table 7
Overall conclusisson on sensitivity, specificity and accuracy.

Assay	Compared to the LLNA			Compared to human data[a]		
	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy
SensCeeTox	**57**	**40**	**50**	**43**	**0**	**27**
SENS-IS	**86**	**40**	**67**	**86**	**50**	**73**
RHE-IL18 epiCS	**71**	**40**	**58**	**71**	**50**	**64**
RHE-IL18 Epiderm	**71**	**25**	**55**	**83**	**50**	**70**
Prediction (2 out of 3) from aqueous in vitro testing	**29**	**20**	**25**	**43**	**50**	**45**

[a] See footnote 3 on Table 1. For this analysis, 2-Hydroxy-4-methoxybenzophenones has been assumed to be class 4.

also indicate that use of 3D RhE models may well prove to be a viable alternative to the aqueous in vitro tests when testing challenging substances, expanding the applicability (domain) of nonanimal test methods for skin sensitisation.

Declaration of interests

DAB was compensated by EPAA for the time spent in the preparation, review and submission of this manuscript; EA was compensated by EPAA for the statistical analysis of the results; BH is a retained consultant for CEFIC; AI is a retained consultant for IFRA; all other authors are fully paid employees of their respective organisations.

Acknowledgement

The authors would like to acknowledge the valuable contribution of Dr Oliver Engelking, previously of CellSystems Biotechnologie, Troisdorf, Germany, regarding the EpiCS model.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yrtph.2019.104473.

References

Api, A.M., Parakhia, R., O'Brien, D., Basketter, D.A., 2017. Fragrances categorised according to their relative human skin sensitisation potency. Dermatitis 28, 299–307.
Ashikaga, T., Yoshida, Y., Hirota, M., Yoneyama, K., Itagaki, H., Sakaguchi, H., Miyazawa, M., Ito, Y., Suzuki, H., Toyoda, H., 2006. Development of an in vitro skin sensitisation test using human cell lines: the human cell line activation test (h-CLAT). I. Optimization of the h-CLAT protocol. Toxicol. in Vitro 20, 767–773.
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 sensitising potency. Dermatitis 25, 11–21.
Basketter, D.A., Kimber, I., Hartung, T., 2010. The evolution of validation. Cutan. Ocul. Toxicol. 29, 1–3.
Basketter, D.A., McFadden, J.F., Gerberick, G.F., Cockshott, A., Kimber, I., 2009. Nothing is perfect, not even the local lymph node assay. A commentary and the implications for REACH. Contact Dermatitis 60, 65–69.
Bauch, C., Kolle, S.N., Ramirez, T., Eltze, T., Fabian, E., Mehling, A., Teubner, W., van Ravenzwaay, B., Landsiedel, R., 2012. Putting the parts together: combining in vitro methods to test for skin sensitizing potentials. Regul. Toxicol. Pharmacol. 63, 489–504.
Corsini, E., Galbiati, V., Mitjans, M., Galli, C.L., Marinovich, M., 2013. NCTC2544 and IL-18 production: a tool for the identification of contact allergens. Toxicol. in Vitro 27, 1127–1134.
Corsini, E., Mitjans, M., Galbiati, V., Lucchi, L., Galli, C.L., Marinovich, M., 2008. Use of IL-18 production in a human keratinocyte cell line to discriminate contact sensitisers from irritants and low molecular weight respiratory allergens. Toxicol. in Vitro 23, 789–796. https://doi.org/10.1016/j.tiv.2009.04.005.
Cottrez, F., Boitel, E., Ourlin, J.C., Peiffer, J.L., Fabre, I., Henaoui, I.S., Mari, B., Vallauri, A., Paquet, A., Barbry, P., Auriault, C., Aeby, P., Groux, H., 2016. SENS-IS, a 3D reconstituted epidermis based model for quantifying chemical sensitisation potency: reproducibility and predictivity results from an inter-laboratory study. Toxicol. in Vitro 32, 248–260. https://doi.org/10.1016/j.tiv.2016.01.007.
Deng, W., Oldach, J., Armento, A., Ayehunie, S., Kandarova, H., Letasiova, S., Klausner, M., Hayden, P.J., 2011. Il-18 secretion as a marker for identification of contact sensitisers in the Epiderm in vitro human skin model. SOT 2011 poster abstract. Toxicologist 120 (Suppl. 12), 550.
ECHA, 2018. https://echa.europa.eu/information-on-chemicals/registered-substances, Accessed date: 4 December 2018.
Emter, R., Ellis, G., Natsch, A., 2010. Performance of a novel keratinocyte-based reporter cell line to screen skin sensitisers in vitro. Toxicol. Appl. Pharmacol. 245, 281–290.
European Union, 2008. Regulation (EC) No 1272/2008 on the classification, labelling and packaging of substances and mixtures. 4th December 2018. https://eur-lex.europa.eu/legal-content/en/TXT/PDF/?uri=CELEX:02008R1272-20180301&from=EN.
European Union, 2009. Regulation (EC) No 1223/2009 on cosmetic products. 4th December 2018. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32009R1223.
Ezendam, J., Braakhuis, H.M., Vandebriel, R.J., 2016. State of the art in non-animal

approaches for skin sensitisation testing: from individual test methods towards testing strategies. Arch. Toxicol. 90, 2861–2883.
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.
Gerberick, G.F., Vassallo, J.D., Bailey, R.E., Chaney, J.G., Morrall, S.W., 2004 Oct. Lepoittevin JP.Development of a peptide reactivity assay for screening contact allergens.Toxicol. Sci 81 (2), 332–343.
Gibbs, S., Corsini, E., Spiekstra, S.W., Galbiati, V., Fuchs, H.W., Degeorge, G., Troese, M., Hayden, P., Deng, W., Roggen, E., 2013. An epidermal equivalent assay for identification and ranking potency of contact sensitisers. Toxicol. Appl. Pharmacol. 272, 529–541. https://doi.org/10.1016/j.taap.2013.07.003.
Hoffmann, S., Kleinstreuer, N., Alépée, N., Allen, D., Api, A.M., Ashikaga, T., Clouet, E., Cluzel, M., Desprez, B., Gellatly, N., Goebel, C., Kern, P.S., Klaric, M., Kühnl, 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. 48, 344–358. https://doi.org/10.1080/10408444.2018.1429385.
Kleinstreuer, N.C., Hoffmann, S., Alépée, N., Allen, D., Ashikaga, T., Casey, W., Clouet, E., Cluzel, M., Desprez, B., Gellatly, N., Göbel, C., Kern, P.S., Klaric, M., Kühnl, J., Martinozzi-Teissier, S., Mewes, K., Miyazawa, M., Strickland, J., van Vliet, E., Zang, Q., Petersohn, D., 2018. Non-animal methods to predict skin sensitisation (II): an assessment of defined approaches ". Crit. Rev. Toxicol. 48, 359–374. https://doi.org/10.1080/10408444.2018.1429386.
Letasiova, S., Galbiati, V., Corsini, E., Kandarova, H., 2018. Assessment of the skin sensitisation using reconstructed human skin model EpiDerm and Interleukin-18. In: MS in Preparation for Special Issue of TIV.
McKim Jr., J.M., Keller 3rd, D.J., Gorski, J.R., 2012. An in vitro method for detecting chemical sensitisation using human reconstructed skin models and its applicability to cosmetic, pharmaceutical, and medical device safety testing. Cutan. Ocul. Toxicol. 31, 292–305. https://doi.org/10.3109/15569527.2012.667031.
Natsch, A., Gfeller, H., Rothaupt, M., 2007 Oct. Ellis G.Utility and limitations of a peptide reactivity assay to predict fragrance allergens in vitro.Toxicol In Vitro 21 (7), 1220–1226.
OECD, 2015a. OECD guideline for testing chemicals No. 442C. In: Chemico Skin Sensitisation: Direct Peptide Reactivity Assay (DPRA). Paris, France.
OECD, 2015b. OECD guideline for testing of chemicals No. 442D. In: Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method, Paris, France.
OECD, 2016. OECD Guideline for Testing of Chemicals No. 442E. Human Cell Line Activation Test (h-CLAT) Paris, France.
OECD, 2018a. OECD Guideline for Testing of Chemicals No 442E: in Vitro Skin Sensitisation Assays. (Paris, France).
OECD, 2018b. OECD Guideline for Testing of Chemicals No 442d: in Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method. (Paris, France).
OECD, 2018c. OECD Guideline for Testing of Chemicals No 442E: in Vitro Skin Sensitisation Assays Addressing the Key Event on Activation of Dendritic Cells on the Adverse Outcome Pathway for Skin Sensitisation. (Paris, France).
Rodrigues Neves, C., Gibbs, S., 2018. Progress on reconstructed human skin models for allergy research and identifying contact sensitisers. Curr. Top. Microbiol. Immunol. https://doi.org/10.1007/82_2018_88. Jun 23.
Sakaguchi, H., Ryan, C., Ovigne, J.M., Schroeder, K.R., Ashikaga, T., 2010. Predicting skin sensitisation potential and inter-laboratory reproducibility of a human Cell Line Activation Test (h-CLAT) in the European Cosmetics Association (COLIPA) ring trials. Toxicol. In Vitro 24, 1810–1820. https://doi.org/10.1016/j.tiv.2010.05.012.
Strickland, J., Zang, Q., Paris, M., Lehmann, D.M., Allen, D., Choksi, N., Matheson, J., Jacobs, A., Casey, W., Kleinstreuer, N., 2017. Multivariate models for prediction of human skin sensitisation hazard. J. Appl. Toxicol. 37, 347–360. https://doi.org/10.1002/jat.3366.
United Nations, 2015. In: Globally Harmonized System of Classification and Labelling of Chemicals, 6th revised edition. . www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev06/English/03e_part3.pdf, Accessed date: 11 October 2018.
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, 337–351. https://doi.org/10.1016/j.yrtph.2014.12.008.