



Evaluating the performance of integrated approaches for hazard identification of skin sensitizing chemicals



Jochem W. van der Veen^{b,a}, Emiel Rorije^b, Roger Emter^c, Andreas Natsch^c, Henk van Loveren^{b,a}, Janine Ezendam^{b,*}

^a Department of Toxicogenomics, Maastricht University, PO Box 616, NL-6200 MD Maastricht, The Netherlands

^b National Institute for Public Health and the Environment (RIVM), PO Box 1, NL-3720BA Bilthoven, The Netherlands

^c Givaudan Schweiz AG, Ueberlandstrasse 138, CH-8600 Dübendorf, Switzerland

ARTICLE INFO

Article history:

Received 13 January 2014

Available online 9 May 2014

Keywords:

Skin sensitization

Integrated testing strategy

Keratinocytes

Allergic contact dermatitis

Animal-free testing

LLNA

QSAR

DPRA

Adverse outcome pathway

ABSTRACT

The currently available animal-free methods for the detection of skin sensitizing potential of chemicals seem promising. However, no single method is able to comprehensively represent the complexity of the processes involved in skin sensitization. To ensure a mechanistic basis and cover the complexity, multiple methods should be integrated into a testing strategy, in accordance with the adverse outcome pathway that describes all key events in skin sensitization. Although current majority voting testing strategies have proven effective, the performance of individual methods is not taken into account. To that end, we designed a tiered strategy based on complementary characteristics of the included methods, and compared it to a majority voting approach. This tiered testing strategy was able to correctly identify all 41 chemicals tested. In terms of total number of experiments required, the tiered testing strategy requires less experiments compared to the majority voting approach. On the other hand, this tiered strategy is more complex due the number of different alternative methods required, and predicted costs are similar for both strategies. Both the tiered and majority voting strategies provide a mechanistic basis for skin sensitization testing, but the strategy most suitable for regulatory decision-making remains to be determined.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Changes in EU legislations, such as the 7th amendment to the cosmetics directive and the REACH regulation (EC, 2006, 2008), have prompted the development of alternative methods that assess skin sensitizing potential of chemicals (Adler et al., 2011; Rovida et al., 2013; Vandebruel and van Loveren, 2010). Although several of these alternatives seem promising (Zuang et al., 2013), it is unlikely that a single method can capture the complexity of the processes involved in skin sensitization. It is therefore foreseen that multiple alternative methods need to be integrated into a testing strategy for reliable classification of chemicals. The Organization for Economic Co-operation and Development (OECD) has proposed an adverse outcome pathway (AOP) that describes the molecular initiating event and key events that lead to allergic contact dermatitis (ACD) (OECD, 2012). This AOP can guide the integration of methods that each represent a different key event in the development of ACD.

The binding of a chemical to proteins is considered the molecular initiating event of skin sensitization and is influenced by bio-availability of the chemical and the cellular metabolism that either activates or inactivates a chemical. The initiating event is the basis for methods that predict sensitizing potential of chemicals through the binding capability of chemicals, such as the ECVAM-validated direct peptide reactivity assay (DPRA) (Gerberick et al., 2004; EC-JRC, 2013a) or quantitative structure–activity relationships (QSARs). Subsequently, the epithelial cells respond to the haptens and hapten–protein complexes, mainly through the production and release of cytokines and stress markers by keratinocytes (KC). This includes interleukin IL-18, which can be used to predict the sensitizing potential of chemicals in epithelial cell lines and 3D-skin models (Gibbs et al., 2013; McKim et al., 2010; Corsini et al., 2009; Van Och et al., 2005). In addition, cellular stress caused by chemicals activates cytoprotective responses such as the Nrf2–Keap1 pathway, which serves as the basis of several reporter assays, such as the KeratinoSens assay for which the ECVAM-validation report is expected soon (EC-JRC, 2013b; Emter et al., 2010). Methods that assess multiple stress related cellular mechanisms are available, such as a gene signature that includes biomarker

* Corresponding author.

E-mail address: Janine.ezendam@rivm.nl (J. Ezendam).

genes with roles in oxidative stress, immune response and gene regulation (van der Veen et al., 2013).

The next key event outlines the activation of dendritic cells (DC), which is influenced by both the chemical and the mediators released by keratinocytes. Activation leads to up-regulation of co-stimulatory molecules, such as CD86. In the h-Clat and MUSST methods this is indicative of the skin sensitizing potential of chemicals (Nukada et al., 2011). The h-Clat has been validated by EURL ECVAM and results are expected in 2014. Upon activation, DCs migrate out of the skin towards the closest lymph node and present a peptide-hapten conjugate on the MHC-class II protein, here the T-cells can recognize the haptenated peptide and start their clonal expansion. This event is measured in the LLNA, but currently no routine *in vitro* assay is available to assess this step in the AOP (Martin et al., 2010; Adler et al., 2011; OECD, 2012).

These animal-free classification methods can be combined in a testing strategy in which molecular initiating and key events of the AOP are captured. Recently, a majority voting approach was proposed in which results from the DPRA, KeratinoSens and MUSST methods are used to classify a chemical according to the most prevalent prediction (Natsch et al., 2013; Bauch et al., 2012). In this study, we compared this majority voting strategy to a tiered testing strategy that is based on the predictive performance of the included methods.

2. Materials and methods

2.1. Chemicals

The sensitizing and non-sensitizing chemicals used in this study are shown in Table 1. The sensitizing compounds were selected based on human evidence and to reflect various potency classes. In addition, chemicals that have proven either false-positive or false-negative in the LLNA were included (Gerberick et al., 2005). All compounds were obtained from Sigma–Aldrich (Zwijndrecht, The Netherlands), except for 2-Mercaptobenzothiazole, which was obtained from Merck (Schiphol-Rijk, The Netherlands). Chemicals were dissolved in either absolute ethanol or dimethylsulfoxide (DMSO) and then added to the cells to a final solvent concentration of 1%.

2.2. Methods reflecting protein reactivity

The protein reactivity of chemicals was assessed using the individually effective methods of *in chemico* peptide reactivity methods and *in silico* QSAR methods.

2.3. Peptide reactivity

A literature search was performed to establish for which of the chemicals peptide reactivity has been assessed *in chemico* (Natsch et al., 2013; Bauch et al., 2012; Gerberick et al., 2007). The peptide reactivity of the 4 chemicals (tBHQ, TIBP, MA and HEG) that were not described in literature was evaluated using the DPRA (Gerberick et al., 2007). In short, the chemical was added to a synthetic peptide containing either a cysteine or a lysine. The mixture was then analyzed with HPLC using UV detection. After 24 h the mixture was analyzed again, a sensitizing chemical is designated by a significant decrease in the peak related to the unmodified peptide.

2.4. QSAR predictions for skin sensitization

In order to evaluate the predictive value of skin sensitization QSARs for our chemical set, the non-commercial QSAR models of

MultiCASE, CAESAR, DEREK and the OECD QSAR toolbox were applied to all chemicals to generate prediction of skin sensitization potential, these models are briefly described here.

MultiCASE (Klopman et al., 2005) generates QSAR models based on substructure fragments linked to biological activity. A MultiCASE implementation for skin sensitization from the Danish EPA (DTU, 2013) is used here. In the present study, only the positive and negative predictions within the applicability domain were taken into account. The CAESAR model (Chaudhry et al., 2010) uses atom centered fragments as descriptors in a multivariate statistical model. The model gives a prediction of active or inactive (as skin sensitizer), together with applicability domain information. Again, only predictions of active or inactive within the applicability domain are taken into account. The DEREK knowledgebase (Lhasa, 2013) is a collection of structural alerts linked to skin sensitization. The model only identifies skin sensitizers and is not meant to identify non-sensitizers. Despite this limitation, we have interpreted the absence of any structural alert as a prediction of non-sensitization. DEREK predictions of “certain”, “probable” and “plausible” were considered as positive predictions of skin sensitization. Predictions of “improbable”, “impossible” or “nothing to report” were interpreted as a prediction of non-sensitization. Equivocal predictions were not taken into account. In the OECD QSAR Toolbox software (OECD, 2013) an implementation of a set of protein binding reactivity alerts from Enoch et al. (2008) is present, as the “Protein binding” profile. These alerts are considered indicative of reactivity towards proteins, and subsequently any substance that has an alert in this profile is considered a skin sensitizer for this study. Absence of any of the alerts in this profile was taken as a prediction of non-sensitization.

2.5. Independent Bayesian approach to evaluate a battery of QSAR predictions

Instead of characterizing the individual predictivity of the QSAR models for our dataset, a classification of the four QSAR models combined (QSAR-battery) was generated, based on Bayesian statistics. A detailed description of the methodology, applied to skin sensitization, can be found in Rorije et al. (2013). In addition, the applicability domain information provided by the individual QSAR methods is taken into account. Furthermore, a newer version of the DEREK knowledge base and an entirely new model (CAESAR) are introduced here. The specificity and sensitivity of each model used in our Bayesian analysis, taking into account the applicability domain information in the case of MultiCASE and CAESAR, are based on the analysis of a large number of chemicals in comparison with the LLNA (Table 2).

Threshold values used to determine a reliable prediction from the battery of QSARs are applied as proposed in Rorije et al. (2013). These values are >80% or >90% probability for a positive or a negative conclusion respectively. This is based on the reliability with which the GPMT test predicts the LLNA outcome (or vice versa) in the official LLNA validation study (NICEATM-ICCVAM, 1999). If there are insufficient or conflicting results from the battery of QSAR models, these thresholds will not be reached and substances are considered equivocal.

2.6. Methods reflecting epithelial response

To assess which *in vitro* approach would be most effective in a tiered strategy, the expression of biomarker genes, activation of the Nrf2 transcription factor, and production of interleukin 18 were evaluated.

Table 1

Detailed results of the predictions based on the individual test methods.

Name	Abbr.	CAS-nr	CV80 (ug/mL)	EC3 (%)	Human	LLNA	Bayesian QSARs	Peptide Binding	Signature	Keratinosens	IL18	hCLAT
5-Chloro-2-methyl-4-isothiazolin-3-one	CMI	26172-55-4	149.6	0.009	S	S	S	S	S	S	NS	S
Diphenylcyclopropenone	DCP	886-38-4	2.1	0.003	S	S	O	S	S	S	NS	S
Oxazolone	OXA	15646-46-5	217.2	0.003	S	S	O	S	S	S	X	S
p-Benzoquinone	BQ	106-51-4	9.7	0.0099	S	S	O	S	S	S	NS	S
1-Chloro-2,4-dinitrobenzene	DNCB	97-00-7	4.1	0.04	S	S	S	S	S	S	NS	S
nitrobenzyl bromide	NBB	100-11-8	3.5	0.05	S	S	S	S	S	S	S	S
p-Phenylenediamine	PPD	106-50-3	162.2	0.11	S	S	O	S	S	S	S	S
Cobalt(II) chloride	CCl	7646-79-9	39.0	0.6	S	S	O	S	S	S	NS	S
Isoeugenol	IsoEug	97-54-1	88.7	1.5	S	S	S	S	S	S	X	O
2-Mercaptobenzothiazole	2-MBT	149-30-4	167.3	1.7	S	S	S	S	S	S	S	O
Cinnamaldehyde	Cal	104-55-2	300	3	S	S	S	S	S	S	X	S
Tetramethylthiuram disulfide	TMTD	137-26-8	48.1	5.2	S	S	O	S	S	S	NS	S
Resorcinol	RES	108-46-3	55.1	5.5	S	S	O	NS	S	NS	NS	S
Diethyl maleate	DM	141-05-9	43.0	5.8	S	S	S	S	S	S	S	S
Citral	Ci	5392-40-5	34.3	9.2	S	S	S	S	S	S	NS	S
α -Hexylcinnamaldehyde	HCA	101-86-0	30.3	9.7	S	S	S	NS	S	S	NS	O
Eugenol	EU	97-53-0	164.2	10.1	S	S	S	S	S	NS	X	S
Cinnamyl alcohol	CA	104-54-1	228.1	21	S	S	O	S	S	NS	X	S
Benzocaine	BENZ	94-09-7	470.8	22	S	S	O	NS	S	S	NS	S
Hydroxycitronellal	HC	107-75-5	689.0	23	S	S	S	S	S	S	S	S
Imidazolidinyl urea	IU	39236-46-9	38.8	24	S	S	O	S	S	S	NS	O
Ethylene glycol dimethacrylate	EGDM	97-90-5	79.3	28	S	S	S	S	S	S	NS	S
Ethyl acrylate	EA	140-88-5	40.0	28	S	S	S	S	S	S	NS	X
Methyl methacrylate	MMA	80-62-6	400.5	90	S	S	S	S	S	NS	NS	S
tert-Butylhydroquinone	tBHQ	1948-33-0	33.2	NA	S	S	O	S	S	S	NS	X
Nickel(II) chloride	NiCl	7718-54-9	155.5	NA	S	NS	O	S	S	S	NS	O
Triisobutylphosphate	TIBP	126-71-6	186.4	NA	S	NS	O	S	S	NS	NS	X
Chlorobenzene	DCB	108-90-7	225.1	NA	NS	NS	O	NS	NS	NS	X	S
2-Propanol	Iso	67-63-0	15.0	NA	NS	NS	NS	NS	NS	NS	NS	NS
Lactic acid	LA	50-21-5	3.6	NA	NS	NS	NS	NS	NS	NS	NS	NS
Methyl salicylate	MS	119-36-8	426.0	NA	NS	NS	NS	NS	NS	NS	S	NS
Salicylic acid	SA	69-72-7	345.3	NA	NS	NS	NS	NS	NS	NS	NS	O
Hexane	Hex	110-54-3	3.9	NA	NS	NS	NS	NS	NS	NS	NS	NS
Dextran	DEX	9004-54-0	2000.0	NA	NS	NS	O	NS	NS	NS	NS	NS
Propylene glycol	PG	57-55-6	304.4	NA	NS	NS	NS	NS	NS	NS	NS	NS
Tween80	T80	9005-65-6	1000.0	NA	NS	NS	O	NS	S	S	NS	NS
Xylenes	Xy	1330-20-7	21.2	95.8	NS	S	NS	NS	NS	NS	NS	NS
Benzalkonium chloride	BK	63449-41-2	6.0	0.07	NS	S	O	NS	S	S	X	NS
Sodium dodecyl sulfate	SDS	151-21-3	144.2	4.0	NS	S	NS	S	S	NS	NS	NS
Maleic acid	MA	110-16-7	464.3	7.0	NS	S	O	NS	NS	S	NS	X
hexaethylene glycol monodecyl ether	HEG	5168-89-8	169.0	NA	NS	S	O	S	S	NS	NS	X

The chemicals marked “S” are classified as sensitizer, while the “NS” marks a non-sensitizer. The predictions marked yellow are misclassified, the green marker “O” are equivocal calls and the orange “X” represents missing data. Classification is based on comparison with human data.

2.7. Gene signature

Gene expression was analyzed as described in our earlier work (van der Veen et al., 2013). In short, the HaCaT human keratinocyte cell line was exposed for 4 h to a concentration of the chemical that results in 80% cell viability (CV80). The total RNA was extracted and the expression levels of 10 genes (Supplementary Table I) were assessed using RT-PCR and normalized against HPRT1. The chemical class was then determined using the classification algorithms of Random forest, support vector machine and PAM-R (Breiman, 2001; Rifkin et al., 2003; Tibshirani et al., 2001). These algorithms compare changes in gene regulation of chemicals of unknown class to that induced by chemicals of known class. A chemical was assigned the most predominant class indicated by these algorithms.

2.8. KeratinoSens

To establish which chemicals had already been tested in the KeratinoSens assay a literature search was performed. The KeratinoSens assay (Emter et al., 2010; Natsch et al., 2011) was applied to evaluate the three chemicals (tBHQ, TIBP and HEG) that had not been previously described in literature. In short, KeratinoSens cells were exposed to a concentration range of the three chemicals. After 48 h of exposure the cells were lysed using passive lysis buffer (Promega, Leiden, The Netherlands) and the luminescence signal was evaluated. In a concurrent exposure the viability of the cells was determined using the MTT assay. Chemicals were considered sensitizers when they induced an luminescence intensity increase of 1.5-fold compared to vehicle control, at concentrations at which less than 30% cell death occurred.

Table 2
Performance of the individual QSARs.

Model	Sensitivity	Specificity
MultiCASE	0.660	0.667
CAESAR	0.943	0.321
DEREK	0.818	0.589
OECD Toolbox	0.778	0.523

Characterization of the individual predictive performance of the four QSAR models is based on comparison with LLNA data.

2.9. Interleukin 18

HaCaT keratinocytes were plated at 3.75×10^5 cells per well in a 12-well plate. The cells were allowed to adhere to the plate for 24 h. The medium was refreshed and the cells were then exposed for 24 h to the CV80 concentration of the chemical (van der Veen et al., 2013). After the exposure period, the cells were lysed using 0.5% Triton x-100 in PBS. The experiments were performed in triplicate for each chemical.

The lysate was used to determine the total protein content. This was determined using BCA (Thermo Fisher scientific, Landsmeer, The Netherlands) and the IL-18 concentration using a custom ELISA. Nunc maxisorp 96-well plates (Thermo Fisher scientific) were coated overnight using anti-human IL-18 (Millipore, cat. D044-3, Amsterdam, The Netherlands), followed by 1% BSA in PBS for 1 h. After washing, the samples were incubated for 1 h at room temperature in a 10-fold dilution, and a dilution range of recombinant IL-18 (MBL, B001-5) standard was included. The detection antibody (Millipore, D045-6) was applied followed by streptavidin-HRP (Immunotools, cat. 31334248, Germany). Quantification of tetramethylbenzidine (Sigma–Aldrich) was assessed spectrophotometrically at 450 nm. A chemical inducing a 20% increase in pg IL-18/ μ g protein (Corsini et al., 2009) was considered a sensitizer.

2.10. Methods reflecting DC maturation

For DC maturation a literature search was performed to establish which chemicals had been tested in the h-Clat assay (Bauch et al., 2012; Ashikaga et al., 2010, 2006). In short, this assay uses THP-1 cells to measure the sensitizer induced expression of CD86 using FACS; up-regulated CD86 indicates sensitizing potential. Although no h-Clat data was available for several chemicals in our set, no additional chemicals were assessed using this method.

2.11. Calculation of the predictive performance and costs of the individual tests and testing strategies

The predictive performances of the individual test methods and the two testing strategies were characterized using Cooper statistics (Cooper et al., 1979); sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). The sensitivity and specificity are an indication of the diagnostic performance of a model, whereas the PPV and NPV are a measure of the reliability of a prediction. As opposed to sensitivity and specificity, the PPV and NPV are dependent on the ratio of skin sensitizers versus non-sensitizers in a dataset. The PPV and NPV were therefore balanced to an equal ratio of sensitizers and non-sensitizers. A detailed description of these statistics, specifically the use of balanced PPV and NPV, can be found elsewhere (Rorije et al., 2013). In the testing strategies equivocal calls can arise due to insufficient or contradictory data for classification, here no definite class can be assigned to the chemical. These results were excluded from statistical analysis.

Two different testing strategies were evaluated, a majority voting approach and a tiered approach (Fig. 1). Each strategy uses predictions of three individual methods. The majority voting strategy is comprised of the methods with high accuracy that reflects either protein reactivity, KC response or DC maturation. In practice, two equal predictions are sufficient for classification, therefore the protein reactivity and KC response are performed first, as they are cheaper and easier than DC maturation. In case of disagreement between protein reactivity and KC response, the DC activation method h-Clat is included for additional information (Natsch et al., 2013; Bauch et al., 2012).

In the tiered approach the methods are performed in sequence. The first tier reflects protein reactivity through a combination of the QSAR battery and peptide reactivity. Here peptide reactivity is used to provide additional information for the chemicals that the QSAR battery is unable to predict with sufficient probability. The result from this tier determines which epithelial response method is performed next. If a sensitizer is indicated, the chemical is tested in the epithelial method with the lowest number of false-positive results and thus the highest positive predictive value (PPV). When a non-sensitizer is indicated, the epithelial method with the lowest number of false-negative results and thus the highest negative predictive value (NPV) is used in the second tier. When the first and second tiers agree, a class can be assigned to the chemical. When contradictory results are obtained the h-Clat method determines the chemical class in the third tier.

The costs of both testing strategies were estimated, based on the number of times a method needs to be used in the strategy. The estimated cost for the individual methods to assess a single chemical was based on correspondence with the method developers and EURL ECVAM. The Bayesian QSAR battery is estimated at €25, the DPRA at €250, the KeratinoSens at €500, the gene signature at €530 and the h-Clat at €1.000. In addition, the costs for an LLNA were estimated at €2.500.

3. Results

3.1. Protein reactivity

The protein reactivity key event has been assessed using peptide binding and the Bayesian QSAR battery. This Bayesian QSAR battery correctly classifies all chemicals included in this study, although 19 chemicals cannot be classified with sufficient probability and are marked equivocal (Table 1, Supplementary Table III). For the set of chemicals selected in this study, the accuracy of peptide binding was 87.3%. The sensitizers Res, HCA, Benz and TIBP were false-negative, while SDS was false-positive (Table 1, Supplementary Tables II and III).

3.2. Epithelial response

The performance of the epithelial cell-based methods was evaluated for the set of 41 chemicals. The results are shown in Tables 1 and 3, detailed results are indicated in Supplementary Table II. IL-18 production by HaCaT cells resulted in a low accuracy due to a high number of false-negatives. Here 19 out of 24 sensitizing chemicals were incorrectly identified, and only NBB, pPD, 2-MBT, DM and HC induced a positive response. Both the gene signature and the KeratinoSens have accuracies similar to the LLNA (85.7%, 80.0% and 78.4%, respectively).

The gene signature resulted in four false-positive chemicals, of which BK, SDS and HEG were also misclassified by the LLNA. On the other hand, the KeratinoSens assay mainly had false-negative chemicals, including Res, EU, and TIBP. Both the KeratinoSens

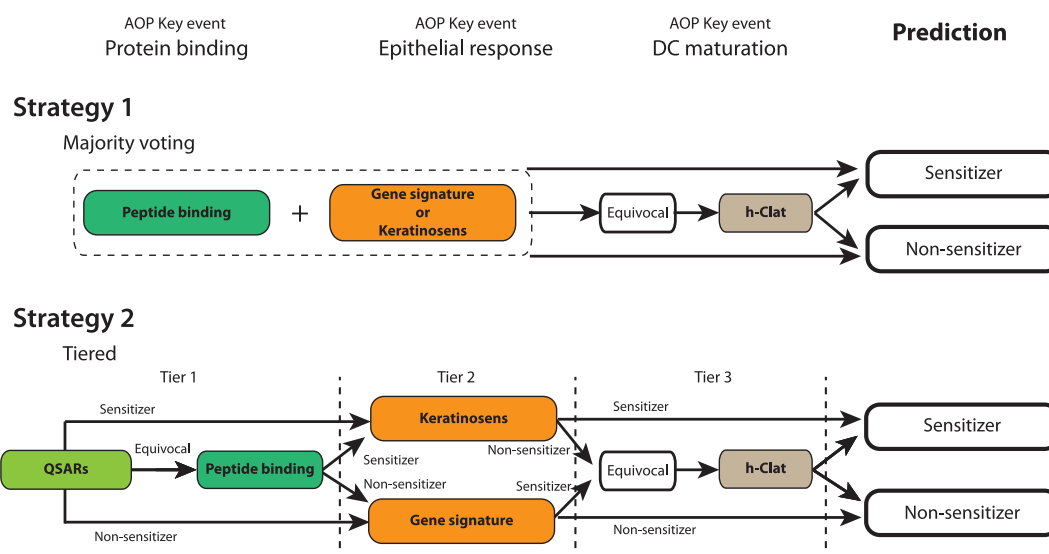


Fig. 1. The schematic setup of the majority voting and tiered testing strategies for skin sensitization. Representing the various testing strategies compared in this study, applying methods that each represent a different key events of the adverse outcome pathway.

Table 3

Performance of the individual test methods.

	LLNA (%)	Individual assays					
		QSARs (%)	Peptide binding (%)	Gene signature (%)	Keratinosens (%)	IL18 (%)	hCLAT (%)
Accuracy	78.4	100.0	87.3	85.7	80.0	57.2	95.5
Sensitivity	92.6	100.0	88.9	100.0	81.5	22.7	100.0
Specificity	64.3	100.0	85.7	71.4	78.6	91.7	90.9
npv	89.7	100.0	88.5	100.0	80.9	54.3	100.0
ppv	72.2	100.0	86.2	77.8	79.2	73.2	91.7
No prediction	0	19	0	0	0	0	11

The predictive capacity is based on comparison with human data.

and the gene signature misclassified T80. The gene signature had a higher NPV, while the KeratinoSens has a higher PPV (Table 3).

3.3. DC maturation

The key event of DC maturation was addressed by the h-Clat method. For this method the results were obtained from literature (Tables 1 and 3, Supplementary Table II) (Bauch et al., 2012; Ashikaga et al., 2010). However, results were unavailable for five chemicals: EA, tBHQ, TIBP, MA and HEG. In addition, contradictory results have been published for six chemicals: IEU, 2-MBT, HCA, IU, Ni and SA, these chemicals have been marked equivocal (Bauch et al., 2012; Ashikaga et al., 2010). For the remaining 30 chemicals, the h-Clat has a high accuracy (95.5%), with DCB as the only false-positive result.

3.4. Test strategy

Two testing strategies were analyzed, a majority voting and a tiered strategy (Fig. 2), of which the results are shown in Tables 4 and 5. In the majority voting strategy, protein reactivity is assessed using peptide reactivity. In the majority voting approach it is of importance that the included methods are able to predict each chemical tested, which is not the case for the QSAR battery. The epithelial response is addressed by either the KeratinoSens or the gene signature, and, in case of disagreement between the previous methods, the DC maturation is assessed using the h-Clat assay. In this majority voting approach, the h-Clat method was required for HCA, EU, Benz, MMA, T80 and SDS using the KeratinoSens

(Fig. 2A) and for Res, HCA, Benz, TIBP, T80, BK and HEG when using the gene signature (Fig. 2B). The majority voting approach using the gene signature has a balanced accuracy of 96.2%, as it misclassifies SDS. In addition, this approach classifies HCA, TIBP and HEG as equivocal, due to missing or inconsistent data from the h-CLAT. The majority voting using the KeratinoSens misclassifies TIBP and RES, leading to a balanced accuracy of 96.2% (Table 5). In this strategy, only HCA is an equivocal result.

In the tiered approach (Fig. 2C) all chemicals were first evaluated in the Bayesian QSAR battery, which was able to classify 22 chemicals. For the remaining 19 chemicals results from the peptide binding methods were used. In this tier, combining data from the QSAR battery and peptide reactivity, indicated 24 sensitizing chemicals and 17 non-sensitizers. This combination of QSAR battery and peptide reactivity misclassified two sensitizers (Benz and TIBP) and has an overall accuracy of 95.1%. In the second tier, these results were verified by testing the 24 predicted sensitizers in the KeratinoSens, while the gene signature was applied for the predicted non-sensitizers. In this tier, 31 chemicals were already correctly identified as either sensitizers or non-sensitizers, but 10 equivocal calls remained. For these equivocal ones the h-Clat assay was used as a third tier. After the third tier, the accuracy was 100%, but TIBP and HEG remained equivocal because these chemicals have not been previously assessed using the h-Clat assay (Table 5).

The different testing strategies have a comparable accuracy. Nevertheless, the tiered strategy requires less experiments compared to the majority voting approach. A total of 69 experiments was required to classify the chemicals; 19 DPRA, 24 KeratinoSens, 17 gene signature and 10 h-Clat assays. The majority voting

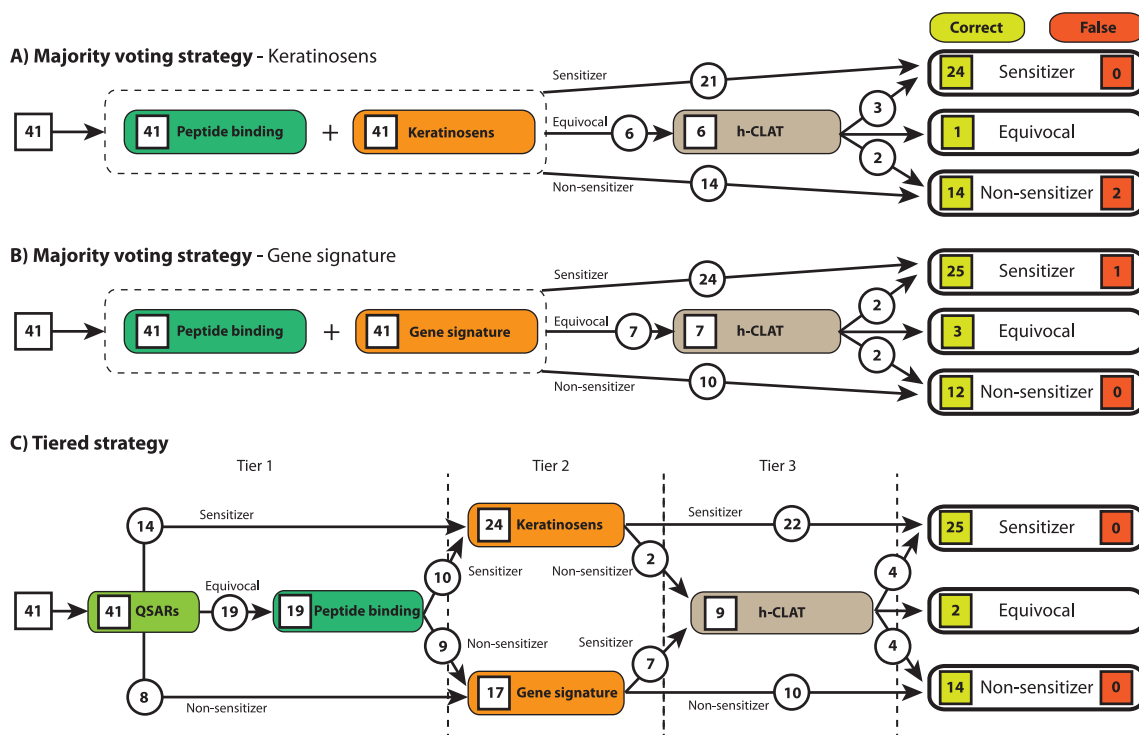


Fig. 2. Performance of the majority voting and tiered testing strategies. The number of chemicals tested in an assay is indicated in the squares, while the number predicted chemicals is indicated in circles. The number of correct final predictions is indicated in green, while false predictions are indicated in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

required 88 experiments: 41 for both the DPRA and KeratinoSens and 6 h-Clat assays. The difference of 19 experiments is mainly due to the inclusion of the QSAR battery in the first tier, which limits the need for *in chemico* peptide reactivity experiments. In addition to accuracy and number of tests, the costs associated with a testing strategy can also be of importance for implementation of a strategy. Therefore the cost for evaluation of a single chemical was estimated for each approach. The total costs for evaluating the 41 chemicals in the tiered approach is €36.785, while the majority voting costs €36.750. In comparison, the LLNA has an estimated total cost of €102.500.

4. Discussion

This proof of principle study showed that both integrated testing strategies were able to predict skin sensitizing potential of the set of chemicals evaluated here with a higher accuracy than the LLNA. In addition, the AOP-based testing strategies generate more knowledge concerning mechanistic events triggered by the chemicals than is obtained in individual animal-free methods. The tiered testing strategy includes non-testing methods and takes the statistical characteristics, i.e. the NPV and PPV of individual methods into account. Ultimately, this results in an adequate classification and a lower number of required experiments, compared to a majority voting strategy.

The Bayesian QSAR battery that was incorporated to reduce the number of experiments has a good predictive capacity, but 19 chemicals are indicated equivocal. This is due to thresholds of probability in the Bayesian method. The probability is influenced by the information provided by the individual QSARs (Rorije et al., 2013), contradictory results between the individual QSARs as well as missing information due to applicability domain constraints. Predictions for the chemicals indicated equivocal by the QSAR battery were generated by also performing the DPRA in

the first tier. It should be noted that the accuracy of the QSAR battery might be biased as many of the included chemicals have also been used in the development of the QSARs. Therefore, their predictive capacity for unknown chemicals may not be as good as for the set of chemicals used here, which is in line with previous observations (Teubner et al., 2013).

In the second tier, methods that measure activation of keratinocytes were included. Measurement of IL-18 was not included, since in the current study this proved unreliable in determining the sensitizing potential of chemicals. This is remarkable, since the HaCaT cell line has been shown to produce IL-18 upon exposure to sensitizers (Van Och et al., 2005) and a preliminary cell line assessment has indicated that HaCaT would also be suitable to identify sensitizers based on IL-18 (Galbiati et al., 2011), as was previously described for the human KC cell line NCTC-2544 (Corsini et al., 2009). The other keratinocyte-based methods did provide accurate predictions. In the tiered approach, the selection of the methods in the second tier was based on the PPV and NPV, which ensured that the classification of chemicals was confirmed using the method that provides the most certain outcome. Combined, the first and second tier were able to classify the majority of the chemicals, and the h-Clat was only required for a minority of the chemicals. Unfortunately, equivocal calls remain due to lacking h-Clat data. These equivocal predictions will no longer arise when a reliable and consistent h-Clat prediction is available for all chemicals.

Overall, both approaches were able to correctly predict sensitizers, including the selected prohaptens. Nevertheless, it is possible that other prohaptens will not be recognized, as this is a well-known limitation of the peptide reactivity assays and QSARs in the first tier. In addition, the THP-1 cells from tier three might not have sufficient metabolic activity (Gerberick et al., 2004, 2009; Ashikaga et al., 2010). In tier two, the HaCaT cells have some intrinsic metabolic capacity, with the presence of Phase I and Phase

Table 4

Detailed results of the predictions based on the tiered testing strategy and the majority voting approach.

Name	Abbr.	CAS-nr	EC3 (%)	Human	LLNA	Majority voting		Tiered		
						Gene Signature	KeratinoSens	TIER 1	TIER 2	TIER 3
5-Chloro-2-methyl-4-isothiazolin-3-one	CMI	26172-55-4	0.009	S	S	S	S	S	S	S
Diphenylcyclopropenone	DCP	886-38-4	0.003	S	S	S	S	S	S	S
Oxazolone	OXA	15646-46-5	0.003	S	S	S	S	S	S	S
p-Benzoquinone	BQ	106-51-4	0.0099	S	S	S	S	S	S	S
1-Chloro-2,4-dinitrobenzene	DNCB	97-00-7	0.04	S	S	S	S	S	S	S
nitrobenzyl bromide	NBB	100-11-8	0.05	S	S	S	S	S	S	S
p-Phenylenediamine	PPD	106-50-3	0.11	S	S	S	S	S	S	S
Cobalt(II) chloride	CCI	7646-79-9	0.6	S	S	S	S	S	S	S
Isoeugenol	IsoEug	97-54-1	1.5	S	S	S	S	S	S	S
2-Mercaptobenzothiazole	2-MBT	149-30-4	1.7	S	S	S	S	S	S	S
Cinnamaldehyde	Cal	104-55-2	3	S	S	S	S	S	S	S
Tetramethylthiuram disulfide	TMTD	137-26-8	5.2	S	S	S	S	S	S	S
Resorcinol	RES	108-46-3	5.5	S	S	S	NS	NS	O	S
Diethyl maleate	DM	141-05-9	5.8	S	S	S	S	S	S	S
Citral	Ci	5392-40-5	9.2	S	S	S	S	S	S	S
α-Hexylcinnamaldehyde	HCA	101-86-0	9.7	S	S	O	O	S	S	S
Eugenol	EU	97-53-0	10.1	S	S	S	S	S	O	S
Cinnamyl alcohol	CA	104-54-1	21	S	S	S	S	S	O	S
Benzocaine	BENZ	94-09-7	22	S	S	S	S	NS	O	S
Hydroxycitronellal	HC	107-75-5	23	S	S	S	S	S	S	S
Imidazolidinyl urea	IU	39236-46-9	24	S	S	S	S	S	S	S
Ethylene glycol dimethacrylate	EGDM	97-90-5	28	S	S	S	S	S	S	S
Ethyl acrylate	EA	140-88-5	28	S	S	S	S	S	S	S
Methyl methacrylate	MMA	80-62-6	90	S	S	S	S	S	O	S
tert-Butylhydroquinone	tBHQ	1948-33-0	NA	S	S	S	S	S	S	S
Nickel(II) chloride	NiCl	7718-54-9	FN	S	NS	S	S	S	S	S
Triisobutylphosphate	TIBP	126-71-6	FN	S	NS	O	NS	S	O	O
Chlorobenzene	DCB	108-90-7	NA	NS	NS	NS	NS	NS	NS	NS
2-Propanol	Iso	67-63-0	NA	NS	NS	NS	NS	NS	NS	NS
Lactic acid	LA	50-21-5	NA	NS	NS	NS	NS	NS	NS	NS
Methyl salicylate	MS	119-36-8	NA	NS	NS	NS	NS	NS	NS	NS
Salicylic acid	SA	69-72-7	NA	NS	NS	NS	NS	NS	NS	NS
Hexane	Hex	110-54-3	NA	NS	NS	NS	NS	NS	NS	NS
Dextran	DEX	9004-54-0	NA	NS	NS	NS	NS	NS	NS	NS
Propylene glycol	PG	57-55-6	NA	NS	NS	NS	NS	NS	NS	NS
Tween80	T80	9005-65-6	NA	NS	NS	NS	NS	NS	O	NS
Xylenes	Xy	1330-20-7	FP	NS	S	NS	NS	NS	NS	NS
Benzalkonium chloride	BK	63449-41-2	FP	NS	S	NS	NS	NS	O	NS
Sodium dodecyl sulfate	SDS	151-21-3	FP	NS	S	S	NS	NS	O	NS
Maleic acid	MA	110-16-7	FP	NS	S	NS	NS	NS	NS	NS
hexaethylene glycol monodecyl ether	HEG	5168-89-8	FP	NS	S	O	NS	S	O	O

The chemicals marked “S” are classified as sensitizer, while the “NS” marks a non-sensitizer. The predictions marked yellow are misclassified, while the green marker “O” are equivocal calls. Classification is based on comparison with human data.

Table 5

Performance of the tiered testing strategy and majority voting approach.

	LLNA (%)	Majority voting		Tiered		
		Gene signature (%)	KeratinoSens (%)	TIER 1 (%)	TIER 2 (%)	TIER 3 (%)
Accuracy	78.4	96.2	96.2	92.7	100.0	100.0
Sensitivity	92.6	96.2	96.2	92.7	100.0	100.0
Specificity	64.3	100.0	92.3	92.6	100.0	100.0
npv	81.8	100.0	92.9	92.6	100.0	100.0
ppv	83.3	92.9	100.0	91.4	100.0	100.0
No prediction	0	3	1	0	10	2

The predictive capacity is based on comparison with human data.

II metabolizing enzymes. The expression of CYP450 has been shown to be comparable to human skin, combined with abundant CYP2S1 expression (Hewitt et al., 2013; Fabian et al., 2013). Prohaptens are therefore only likely to be detected in the second tier, and correct classification of prohaptens might fail in both testing strategies. To address this issue, both testing strategies should be further evaluated by testing additional chemicals that require metabolic capacity. It should also be explored whether inclusion of other test methods, for example the peroxidase peptide reactivity assay, which includes limited metabolism, can be used to replace the DPRA (Lalko et al., 2012) and to use a KeratinoSens with a S9-metabolic component which has been recently described to further enhance prohaptens detection in HaCaT based models (Natsch and Haupt, 2013). Another important limitation of the individual methods and the testing strategies is that they are unable to assign sensitizing potency to a chemical, which is one of the main advantages of the LLNA (Kimber et al., 1994; Kimber and Dearman, 2010). Currently, methods based on full thickness skin models are being explored to resolve this issue (Gibbs et al., 2013; Saito et al., 2013). In addition, the potency might also be estimated through advanced statistical methods, such as Quantitative Mechanistic Modeling (Enoch and Roberts, 2013) or Bayesian statistics (Jaworska et al., 2013).

Although both testing strategies are based on the AOP, neither of them address all key events. Based on the data presented here and in previously published ITS (Natsch et al., 2013; Bauch et al., 2012), it might be unnecessary to address additional key events, such as T-cell proliferation, as the current ITS are already proficient in the prediction of sensitizing potential of chemicals. As the costs and prediction accuracy of the testing strategies are similar, other characteristics become more important. The tiered approach is better able to correct for the limitations of the individual methods as only the most favorable characteristics of the KC methods are applied. On the other hand, the majority voting strategy is more straightforward, as it requires less different methods and is easier to evaluate. In conclusion, the combination of non-animal testing methods in a testing strategy provides a mechanistic basis for classification and high prediction accuracy, although the strategy most suitable for implementation remains to be determined.

Conflict of interest

None declare.

Acknowledgments

We would like to acknowledge H. Hodemaekers and E. Gremmer of the RIVM for their excellent technical support. We also thank Dr. S. Casati of the European Union Reference Laboratory for alternatives to animal testing, and Dr. T. Ashikaga of the Shiseido Company for providing additional information regarding the costs of the methods used in this study. This study was funded by a grant of the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research (NWO) No.: 050-060-510. In addition, the Dutch Ministry of Infrastructure and the Environment is acknowledged for supporting this work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.yrtph.2014.04.018>.

References

- Adler, S., Basketter, D., Creton, S., Pelkonen, O., van Benthem, J., Zuang, V., Andersen, K.E., Angers-Loustau, A., Aptula, A., Bal-Price, A., Benfenati, E., Bernauer, U., Bessems, J., Bois, F.Y., Boobis, A., Brandon, E., Bremer, S., Broschard, T., Casati, S., Coecke, S., Corvi, R., Cronin, M., Daston, G., Dekant, W., Felter, S., Grignard, E., Gundert-Remy, U., Heinonen, T., Kimber, I., Kleinjans, J., Komulainen, H., Kreiling, R., Kreysa, J., Leite, S.B., Loizou, G., Maxwell, G., Mazzatorta, P., Munn, S., Pfuhler, S., Phrakonkham, P., Piersma, A., Poth, A., Prieto, P., Repetto, G., Rogiers, V., Schoeters, G., Schwarz, M., Serafimova, R., Tahti, H., Testai, E., van Delft, J., van Loveren, H., Vinken, M., Worth, A., Zaldivar, J.M., 2011. Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010. *Arch. Toxicol.* 85, 367–485.
- Ashikaga, T., Yoshida, Y., Hirota, M., Yoneyama, K., Itagaki, H., Sakaguchi, H., Miyazawa, M., Ito, Y., Suzuki, H., Toyoda, H., 2006. Development of an in vitro skin sensitization test using human cell lines: the human Cell Line Activation Test (h-CLAT). I. Optimization of the h-CLAT protocol. *Toxicol. In Vitro* 20, 767–773.
- Ashikaga, T., Sakaguchi, H., Sono, S., Kosaka, N., Ishikawa, M., Nukada, Y., Miyazawa, M., Ito, Y., Nishiyama, N., Itagaki, H., 2010. A comparative evaluation of in vitro skin sensitisation tests: the human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA). *Altern. Lab. Anim.* 38, 275–284.
- 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.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Chaudhry, Q., Piclin, N., Cotterill, J., Pintore, M., Price, N.R., Chretien, J.R., Roncaglioni, A., 2010. Global QSAR models of skin sensitizers for regulatory purposes. *Chem. Cent. J.* 4 (Suppl. 1), S5.
- Cooper 2nd, J.A., Saracci, R., Cole, P., 1979. Describing the validity of carcinogen screening tests. *Br. J. Cancer* 39, 87–89.
- Corsini, E., Mitjans, M., Galbiati, V., Lucchi, L., Galli, C.L., Marinovich, M., 2009. Use of IL-18 production in a human keratinocyte cell line to discriminate contact sensitizers from irritants and low molecular weight respiratory allergens. *Toxicol. In Vitro* 23, 789–796.
- DTU, (2013). Danish QSAR database [Online]. Danish QSAR group, National Food Institute, Technical University of Denmark, (accessed June 2013).
- EC, 2006. Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directive 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC.
- EC, 2008. EC Consolidated version of Cosmetics Directive 76/768/EEC. European Commission (EC) (Internet publication at <<http://ecb.jrc.it>>).
- EC-JRC, 2013a. EURL ECVAM Recommendation on the Direct Peptide Reactivity Assay (DPRA) for Skin Sensitisation Testing. In: Protection, I. F.H.A.C. (Ed.), July 2013 Draft.
- EC-JRC, 2013b. EURL ECVAM Recommendation on the KeratinoSens assay for skin sensitisation testing. November 2013 Draft.
- Emter, R., Ellis, G., Natsch, A., 2010. Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers in vitro. *Toxicol. Appl. Pharmacol.* 245, 281–290.
- Enoch, S.J., Roberts, D.W., 2013. Predicting skin sensitization potency for michael acceptors in the LLNA using quantum mechanics calculations. *Chem. Res. Toxicol.* 26, 767–774.
- Enoch, S.J., Madden, J.C., Cronin, M.T., 2008. Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. *SAR QSAR Environ. Res.* 19, 555–578.
- Fabian, E., Vogel, D., Blatz, V., Ramirez, T., Kolle, S., Eltze, T., van Ravenzwaay, B., Oesch, F., Landsiedel, R., 2013. Xenobiotic metabolizing enzyme activities in cells used for testing skin sensitization in vitro. *Arch. Toxicol.* 87, 1683–1696.
- Galbiati, V., Mitjans, M., Lucchi, L., Viviani, B., Galli, C.L., Marinovich, M., Corsini, E., 2011. Further development of the NCTC 2544 IL-18 assay to identify in vitro contact allergens. *Toxicol. In Vitro* 25, 724–732.
- Gerberick, G.F., Vassallo, J.D., Bailey, R.E., Chaney, J.G., Morrall, S.W., Lepoittevin, J.P., 2004. Development of a peptide reactivity assay for screening contact allergens. *Toxicol. Sci.* 81, 332–343.
- Gerberick, G.F., Ryan, C.A., Kern, P.S., Schlatter, H., Dearman, R.J., Kimber, I., Patlewicz, G.Y., Basketter, D.A., 2005. Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. *Dermatitis* 16, 157–202.
- 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. *Toxicol. Sci.* 97, 417–427.
- Gerberick, G.F., Troutman, J.A., Foertsch, L.M., Vassallo, J.D., Quijano, M., Dobson, R.L., Goebel, C., Lepoittevin, J.P., 2009. Investigation of peptide reactivity of pro-hapten skin sensitizers using a peroxidase-peroxide oxidation system. *Toxicol. Sci.* 112, 164–174.
- 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 sensitizers. *Toxicol. Appl. Pharmacol.* 272, 529–541.

- Hewitt, N.J., Edwards, R.J., Fritsche, E., Goebel, C., Aeby, P., Scheel, J., Reisinger, K., Ouedraogo, G., Duche, D., Eilstein, J., Latil, A., Kenny, J., Moore, C., Kuehl, J., Barroso, J., Fautz, R., Pfuhler, S., 2013. Use of human in vitro skin models for accurate and ethical risk assessment: metabolic considerations. *Toxicol. Sci.* 133, 209–217.
- Jaworska, J., Dancik, Y., Kern, P., Gerberick, F., Natsch, A., 2013. Bayesian integrated testing strategy to assess skin sensitization potency: from theory to practice. *J. Appl. Toxicol.*
- Kimber, I., Dearman, R.J., 2010. The local lymph node assay and skin sensitization testing. *Methods Mol. Biol.* 598, 221–231.
- Kimber, I., Dearman, R.J., Scholes, E.W., Basketter, D.A., 1994. The local lymph node assay: developments and applications. *Toxicology* 93, 13–31.
- Klopman, G., Gilles, I., Ivanov, J., Saiakhov, R., Chakravarti, S., 2005. MC4PC – an artificial intelligence approach to the discovery of structure toxic activity relationships (STAR). In: Helma, C., (Ed.), *Predictive Toxicology*, pp. 423–457.
- Lalko, J.F., Dearman, R.J., Gerberick, G.F., Troutman, J.A., Api, A.M., Kimber, I., 2012. Reactivity of chemical respiratory allergens in the peroxidase peptide reactivity assay. *Toxicol. In Vitro*.
- Lhasa, NEXUS 1.5 Knowledge Suite: Derek Nexus, an expert, knowledge-based software for toxicity predictions [Online]. Available: <https://www.lhasalimited.org/derek_nexus/>, 2013 (accessed June 2013).
- Martin, S.F., Esser, P.R., Schmucker, S., Dietz, L., Naisbitt, D.J., Park, B.K., Vocanson, M., Nicolas, J.F., Keller, M., Pichler, W.J., Peiser, M., Luch, A., Wanner, R., Maggi, E., Cavani, A., Rustemeyer, T., Richter, A., Thierse, H.J., Sallusto, F., 2010. T-cell recognition of chemicals, protein allergens and drugs: towards the development of in vitro assays. *Cell. Mol. Life Sci.* 67, 4171–4184.
- McKim Jr., J.M., Keller 3rd, D.J., Gorski, J.R., 2010. A new in vitro method for identifying chemical sensitizers combining peptide binding with ARE/EpRE-mediated gene expression in human skin cells. *Cutan. Ocul. Toxicol.* 29, 171–192.
- Natsch, A., Haupt, T., 2013. Utility of rat liver S9 fractions to study skin-sensitizing prohaptenes in a modified KeratinoSens assay. *Toxicol. Sci.* 135, 356–368.
- Natsch, A., Bauch, C., Foertsch, L.M., Gerberick, G.F., Norman, K., Hilberer, A., Inglis, H., Landsiedel, R., Onken, S., Reuter, H., Schepky, A., Emter, R., 2011. The intra- and inter-laboratory reproducibility and predictivity of the KeratinoSens assay to predict skin sensitizers in vitro: results of a ring-study in five laboratories. *Toxicol. In Vitro* 25, 733–744.
- Natsch, A., Ryan, C.A., Foertsch, L., Emter, R., Jaworska, J., Gerberick, F., Kern, P., 2013. A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. *J. Appl. Toxicol.*
- NICEATM-ICCVAM, 1999. The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds. The Results of an Independent Peer Review Evaluation Coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program Center for the Evaluation of Alternative Toxicological Methods (NICEATM). NIH Publication No. 99-4494.
- Nukada, Y., Ashikaga, T., Sakaguchi, H., Sono, S., Mugita, N., Hirota, M., Miyazawa, M., Ito, Y., Sasa, H., Nishiyama, N., 2011. Predictive performance for human skin sensitizing potential of the human cell line activation test (h-CLAT). *Contact Dermatitis* 65, 343–353.
- OECD, 2012. Organisation for economic co-operation and development (OECD): the adverse outcome pathway for skin sensitisation initiated by covalent binding to proteins. Part 1: scientific evidence series on testing and assessment no. 168. ENV/JM/MONO(2012)10/PART1, <[http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono\(2012\)10/part1&doclanguage=en](http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono(2012)10/part1&doclanguage=en)>.
- OECD, Toolbox for grouping chemical in categories v3.1 [Online]. Available: <<http://www.qsartoolbox.org/index.html>>, 2013 (accessed June 2013).
- Rifkin, R., Mukherjee, S., Tamayo, P., Ramaswamy, S., Yeang, C.H., Angelo, M., Reich, M., Poggio, T., Lander, E.S., Golub, T.R., Mesirov, J.P., 2003. An analytical method for multiclass molecular cancer classification. *SIAM Rev.* 45, 706–723.
- Rorije, E., Aldenberg, T., Buist, H., Kroese, D., Schuurmann, G., 2013. The OSIRIS weight of evidence approach: its for skin sensitisation. *Regul. Toxicol. Pharmacol.*
- Rovida, C., Martin, S.F., Vivier, M., Weltzien, H.U., Roggen, E., 2013. Advanced tests for skin and respiratory sensitization assessment. *ALTEX* 30, 231–252.
- Saito, K., Nukada, Y., Takenouchi, O., Miyazawa, M., Sakaguchi, H., Nishiyama, N., 2013. Development of a new in vitro skin sensitization assay (Epidermal Sensitization Assay; EpiSensA) using reconstructed human epidermis. *Toxicol. In Vitro*.
- Teubner, W., Mehling, A., Schuster, P.X., Guth, K., Worth, A., Burton, J., van Ravenzwaay, B., Landsiedel, R., 2013. Computer models versus reality: how well do in silico models currently predict the sensitization potential of a substance. *Regul. Toxicol. Pharmacol.* 67, 468–485.
- Tibshirani, R., Hastie, T., Narasimhan, N., Chu, G., 2001. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *PNAS* 99, 6567–6572.
- van der Veen, J.W., Pronk, T.E., van Loveren, H., Ezendam, J., 2013. Applicability of a keratinocyte gene signature to predict skin sensitizing potential. *Toxicol. In Vitro* 27, 314–322.
- Van Och, F.M., Van Loveren, H., Van Wolfswinkel, J.C., Machielsen, A.J., Vandebriel, R.J., 2005. Assessment of potency of allergenic activity of low molecular weight compounds based on IL-1alpha and IL-18 production by a murine and human keratinocyte cell line. *Toxicology* 210, 95–109.
- Vandebriel, R.J., van Loveren, H., 2010. Non-animal sensitization testing: state-of-the-art. *Crit. Rev. Toxicol.* 40, 389–404.
- Zuang, V., Schäffer, M., Tuomainen, A.M., Amcoff, P., Bernasconi, C., Bremer, S., Casati, S., Castello, P., Coecke, S., Corvi, R., Griesinger, C., Roi, A.J., Kirmizidis, G., Prieto, P., Worth, A., Munn, S., Berggren, E., Whelan, M., 2013. EURL ECVAM progress report on the development, validation and regulatory acceptance of alternative methods (2010–2013). In: JRC (Ed.).

Supplementary table I: The gene signature. The gene signature was defined through analysis of gene profiling performed in HaCaT cells (van der Veen et al. 2013).

Entrez ID	Gene name		Function	
1789	DNA (cytosine-5-)-methyltransferase 3 beta	DNMT3b	DNA methylation	(Hervouet et al. 2009)
3162	Hemeoxygenase 1	HMOX1	Oxidative stress	(Morse and Choi 2002)
8614	Stanniocalcin 2	STC2		(Law and Wong 2010)
133	Adrenomedullin	ADM		(Katsuki et al. 2003)
140809	Sulfiredoxin 1 homolog	SRXN1		(Singh et al. 2009)
2353	FBJ murine osteosarcoma viral oncogene homolog	FOS		(Zenz et al. 2008)
8061	FOS-like antigen 1	FosL1	Development and inflammation	(Zenz et al. 2008)
10181	RNA binding motif protein 5	RBM5	Alternative splicing	(Kotlajich and Hertel 2008)
51755	Cyclin-dependent kinase 12	CDK12		(Bartkowiak et al. 2010)
353322	Ankyrin repeat domain 37	ANKRD37	Unknown	(Benita et al. 2009)

Supplementary table II; Detailed results of individual assays. The results of each individual assay have been detailed, indicating literature references and experimental results. Results indicated in yellow represent misclassified chemicals.

Name	CAS-nr	Peptide Reactivity				Gene Signature	KeratiNoSens			IL18		hCLAT	
		Result	Ref*	Cys (% rem.)	Lys (% rem.)		Results	Ref*	EC1.5	Result	Fold change	Result	Ref*
5-Chloro-2-methyl-4-isothiazolin-3-one	26172-55-4	S	B , N			S	S	B		NS	0.9	S	B
Diphenylcyclopropanone	886-38-4	S	Ge , N			S	S	N		NS	0.5	S	A
Oxazolone	15646-46-5	S	B , Ge , N			S	S	B , N		X		S	B , A
p-Benzoquinone	106-51-4	S	B , Ge , N			S	S	B , N		NS	0.7	S	B , A
1-Chloro-2,4-dinitrobenzene	97-00-7	S	B , Ge			S	S	B , N		NS	1.0	S	B , A
nitrobenzyl bromide	100-11-8	S	N			S	S	N		S	1.3	S	A
p-Phenylenediamine	106-50-3	S	B , N			S	S	B , N		S	1.3	S	B , A
Cobalt(II) chloride	7646-79-9	S	B			S	S	B		NS	0.8	S	B , A
Isoeugenol	97-54-1	S	B			S	S	B , N		X	0.8	O	B , A
2-Mercaptobenzothiazole	149-30-4	S	B, Ge			S	S	B , N		S	1.4	O	B , A
Cinnamaldehyde	104-55-2	S	Ge			S	S	N		X		S	A
Tetramethylthiuram disulfide	137-26-8	S	N			S	S	N		NS	0.7	S	A
Resorcinol	108-46-3	NS	Ge			S	NS	N		NS	0.9	S	A
Diethyl maleate	141-05-9	S	B , Ge			S	S	B , N		S	3.7	S	B , A
Citral	5392-40-5	S	B			S	S	B , N		NS	0.8	S	B , A
α-Hexylcinnamaldehyde	101-86-0	NS	B , Ge			S	S	B , N		NS	0.9	O	B , A
Eugenol	97-53-0	S	B			S	NS	N		X		S	B , A
Cinnamyl alcohol	104-54-1	S	B , N			S	NS	N		X		S	B , A
Benzocaine	94-09-7	NS	N			S	S	N		NS	0.7	S	A
Hydroxycitronellal	107-75-5	S	Ge			S	S	N		S	1.5	S	A
Imidazolidinyl urea	39236-46-9	S	B , Ge			S	S	B , N		NS	0.8	O	B , A
Ethylene glycol dimethacrylate	97-90-5	S	B , Ge			S	S	B		NS	1.1	S	B , A
Ethyl acrylate	140-88-5	S	B			S	S	N		NS	1.0	X	
Methyl methacrylate	80-62-6	S	B			S	NS	B		NS	0.8	S	B
tert-Butylhydroquinone	1948-33-0	S	-	70.0	1.5	S	S	-	8.7	NS	1.0	X	
Nickel(II) chloride	7718-54-9	S	B			S	S	N		NS	0.8	O	B , A
Triisobutylphosphate	126-71-6	S	-	15.4	100.0	S	NS	-	>2000	NS	0.6	X	
Chlorobenzene	108-90-7	NS	Ge			NS	NS	N		X		S	A
2-Propanol	67-63-0	NS	B , Ge			NS	NS	B , N		NS	0.8	NS	B , A
Lactic acid	50-21-5	NS	B , Ge			NS	NS	B , N		NS	0.9	NS	B , A
Methyl salicylate	119-36-8	NS	B , Ge			NS	NS	B , N		S	1.3	NS	B , A

Salicyclic acid	69-72-7	NS	B , Ge			NS	NS	B		NS	0.9	O	B , A
Hexane	110-54-3	NS	B , Ge			NS	NS	N		NS	0.8	NS	B , A
Dextran	9004-54-0	NS	N			NS	NS	-	>2000	NS	0.9	NS	A
Propylene glycol	57-55-6	NS	Ge			NS	NS	N		NS	0.6	NS	A
Tween80	9005-65-6	NS	N			S	S	-	129.2	NS	0.7	NS	A
Xylenes	1330-20-7	NS	B			NS	NS	B		NS	0.8	NS	B
Benzalkonium chloride	63449-41-2	NS	N			S	S	N		X		NS	A
Sodium dodecyl sulfate	151-21-3	S	B			S	NS	B		NS	0.4	NS	B , A
Maleic acid	110-16-7	NS	-	100.0	100.0	NS	S	-	1034	NS	0.8	X	
hexaethylene glycol monodecyl ether	5168-89-8	S	-	17.6	51.9	S	NS	-	>2000	NS	1.0	X	

*: References: A - (Ashikaga et al. 2010) ; B - (Bauch et al. 2012) ; Ge - (Gerberick et al. 2007) ; N - (Natsch et al. 2013)

Supplementary table III; Predictions of the individual QSARs and the bayesian statistics. The chemical set was evaluated using qualitative structure-activity relationships of DEREK, CEASAR, MultiCase and the OECD toolbox. The results of the individual QSARs were combined using bayesian statistics (rorije 2013) to obtain the probability of prediction. The chemicals marked “S” are classified as sensitizer, while the “NS” marks a non-sensitizer. The predictions marked yellow are misclassified, the green marker “O” are equivocal calls and the orange “X” represents missing data.

Name	Abrr.	Human	CAS-nr	CEASAR	DEREK	MultiCase	SMARTs	Bayesian	Probability	
5-Chloro-2-methyl-4-isothiazolin-3-one	CMI	S	26172-55-4	S	S	O	S	S	81,8%	18,2%
Diphenylcyclopropanone	DCP	S	886-38-4	O	S	O	S	O	76,4%	23,6%
Oxazolone	OXA	S	15646-46-5	NS	S	S	S	O	53,5%	46,5%
p-Benzoquinone	BQ	S	106-51-4	S	S	NS	S	O	69,7%	30,3%
1-Chloro-2,4-dinitrobenzene	DNCB	S	97-00-7	S	S	S	S	S	89,9%	10,1%
4-Nitrobenzylbromide	NBB	S	100-11-8	S	S	S	S	S	89,9%	10,1%
p-Phenylenediamine	PPD	S	106-50-3	S	S	S	NS	O	69,9%	30,1%
Cobalt(II) chloride	CCl	S	7646-79-9	O	S	X	NS	O	45,8%	54,2%
Isoeugenol	IsoEug	S	97-54-1	S	S	S	S	S	89,9%	10,1%
2-Mercaptobenzothiazole	2-MBT	S	149-30-4	O	S	S	S	S	86,5%	13,5%
Cinnamaldehyde	Cal	S	104-55-2	S	S	S	S	S	89,9%	10,1%
Tetramethylthiuram disulfide	TMTD	S	137-26-8	S	S	S	NS	O	69,9%	30,1%
Resorcinol	RES	S	108-46-3	S	S	S	NS	O	69,9%	30,1%
Diethyl maleate	DM	S	141-05-9	S	S	S	S	S	89,9%	10,1%
Citral	Ci	S	5392-40-5	S	S	O	S	S	81,8%	18,2%
α-Hexylcinnamaldehyde	HCA	S	101-86-0	S	S	O	S	S	81,8%	18,2%
Eugenol	EU	S	97-53-0	S	S	S	S	S	89,9%	10,1%
Cinnamyl alcohol	CA	S	104-54-1	S	NS	S	S	O	58,0%	42,0%
Benzocaine	BENZ	S	94-09-7	NS	S	S	NS	O	23,1%	76,9%
Hydroxycitronellal	HC	S	107-75-5	S	S	S	S	S	89,9%	10,1%
Imidazolidinyl urea	IU	S	39236-46-9	O	O	O	S	O	62,0%	38,0%
Ethylene glycol dimethacrylate	EGDM	S	97-90-5	S	S	S	S	S	89,9%	10,1%
Ethyl acrylate	EA	S	140-88-5	S	S	S	S	S	89,9%	10,1%
Methyl methacrylate	MMA	S	80-62-6	S	S	S	S	S	89,9%	10,1%
tert-Butylhydroquinone	tBHQ	S	1948-33-0	S	S	S	NS	O	69,9%	30,1%
Nickel(II) chloride	NiCl	S	7718-54-9	O	S	O	NS	O	45,8%	54,2%
Triisobutylphosphate	TIBP	S	126-71-6	O	S	O	S	O	76,4%	23,6%
Chlorobenzene	DCB	NS	108-90-7	S	NS	S	NS	O	26,5%	73,5%

2-Propanol	Iso	NS	67-63-0	NS	NS	NS	NS	NS	1,2%	98,8%
Lactic acid	LA	NS	50-21-5	NS	NS	NS	NS	NS	1,2%	98,8%
Methyl salicylate	MS	NS	119-36-8	NS	NS	NS	NS	NS	1,2%	98,8%
Salicyclic acid	SA	NS	69-72-7	NS	NS	NS	NS	NS	1,2%	98,8%
Hexane	Hex	NS	110-54-3	S	NS	NS	NS	NS	8,5%	91,5%
Dextran	DEX	NS	9004-54-0	NS	O	X	S	O	22,6%	77,4%
Propylene glycol	PG	NS	57-55-6	NS	NS	S	NS	NS	4,4%	95,6%
Tween80	T80	NS	9005-65-6	O	NS	X	S	O	33,5%	66,5%
Xylenes	Xy	NS	1330-20-7	S	NS	NS	NS	NS	8,5%	91,5%
Benzalkonium chloride	BK	NS	63449-41-2	O	NS	X	NS	O	11,6%	88,4%
Sodium dodecyl sulfate	SDS	NS	151-21-3	S	NS	NS	NS	NS	8,5%	91,5%
Maleic acid	MA	NS	110-16-7	S	NS	S	NS	O	26,5%	73,5%
hexaethylene glycol monodecyl ether	HEG	NS	5168-89-8	S	NS	X	NS	O	15,4%	84,6%

References

- Ashikaga, T., Sakaguchi, H., Sono, S., Kosaka, N., Ishikawa, M., Nukada, Y., Miyazawa, M., Ito, Y., Nishiyama, N. and Itagaki, H. (2010). A comparative evaluation of in vitro skin sensitisation tests: the human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA). *Altern Lab Anim* **38**(4): 275-284.
- Bartkowiak, B., Liu, P., Phatnani, H.P., Fuda, N.J., Cooper, J.J., Price, D.H., Adelman, K., Lis, J.T. and Greenleaf, A.L. (2010). CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev* **24**(20): 2303-2316.
- Bauch, C., Kolle, S.N., Ramirez, T., Eltze, T., Fabian, E., Mehling, A., Teubner, W., van Ravenzwaay, B. and Landsiedel, R. (2012). Putting the parts together: combining in vitro methods to test for skin sensitizing potentials. *Regul Toxicol Pharmacol* **63**(3): 489-504.
- Benita, Y., Kikuchi, H., Smith, A.D., Zhang, M.Q., Chung, D.C. and Xavier, R.J. (2009). An integrative genomics approach identifies Hypoxia Inducible Factor-1 (HIF-1)-target genes that form the core response to hypoxia. *Nucleic Acids Res* **37**(14): 4587-4602.
- Gerberick, G.F., Vassallo, J.D., Foertsch, L.M., Price, B.B., Chaney, J.G. and Lepoittevin, J.P. (2007). Quantification of chemical peptide reactivity for screening contact allergens: a classification tree model approach. *Toxicol Sci* **97**(2): 417-427.
- Hervouet, E., Vallette, F.M. and Cartron, P.F. (2009). Dnmt3/transcription factor interactions as crucial players in targeted DNA methylation. *Epigenetics* **4**(7): 487-499.
- Katsuki, A., Sumida, Y., Urakawa, H., Gabazza, E.C., Maruyama, N., Morioka, K., Kitagawa, N., Hori, Y., Nakatani, K., Yano, Y. and Adachi, Y. (2003). Increased oxidative stress is associated with elevated plasma levels of adrenomedullin in hypertensive patients with type 2 diabetes. *Diabetes Care* **26**(5): 1642-1643.
- Kotlajich, M.V. and Hertel, K.J. (2008). Death by splicing: tumor suppressor RBM5 freezes splice-site pairing. *Mol Cell* **32**(2): 162-164.
- Law, A.Y. and Wong, C.K. (2010). Stanniocalcin-2 is a HIF-1 target gene that promotes cell proliferation in hypoxia. *Exp Cell Res* **316**(3): 466-476.
- Morse, D. and Choi, A.M. (2002). Heme oxygenase-1: the "emerging molecule" has arrived. *Am J Respir Cell Mol Biol* **27**(1): 8-16.
- Natsch, A., Ryan, C.A., Foertsch, L., Emter, R., Jaworska, J., Gerberick, F. and Kern, P. (2013). A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. *J Appl Toxicol*.
- Singh, A., Ling, G., Suhasini, A.N., Zhang, P., Yamamoto, M., Navas-Acien, A., Cosgrove, G., Tuder, R.M., Kensler, T.W., Watson, W.H. and Biswal, S. (2009). Nrf2-dependent sulfiredoxin-1 expression protects against cigarette smoke-induced oxidative stress in lungs. *Free Radic Biol Med* **46**(3): 376-386.
- van der Veen, J.W., Pronk, T.E., van Loveren, H. and Ezendam, J. (2013). Applicability of a keratinocyte gene signature to predict skin sensitizing potential. *Toxicol In Vitro* **27**(1): 314-322.
- Zenz, R., Eferl, R., Scheinecker, C., Redlich, K., Smolen, J., Schonhaler, H.B., Kenner, L., Tschachler, E. and Wagner, E.F. (2008). Activator protein 1 (Fos/Jun) functions in inflammatory bone and skin disease. *Arthritis Res Ther* **10**(1): 201.