



Peptide reactivity associated with skin sensitization: The QSAR Toolbox and TIMES compared to the DPRA



D. Urbisch^a, N. Honarvar^a, S.N. Kolle^a, A. Mehling^b, T. Ramirez^a, W. Teubner^c, R. Landsiedel^{a,*}

^a BASF SE, Experimental Toxicology and Ecology, Ludwigshafen, Germany

^b BASF Personal Care and Nutrition GmbH, Düsseldorf, Germany

^c BASF Schweiz AG, Basel, Switzerland

ARTICLE INFO

Article history:

Received 30 January 2016

Received in revised form 4 April 2016

Accepted 6 April 2016

Available online 14 April 2016

Keywords:

Skin sensitization

Protein binding

Direct peptide reactivity assay (DPRA)

Quantitative structure–activity relationship

(QSAR)

In silico

In chemico

ABSTRACT

The molecular initiating event (MIE) of skin sensitization is the binding of a hapten to dermal proteins. This can be assessed using the *in chemico* direct peptide reactivity assay (DPRA) or *in silico* tools such as the QSAR Toolbox and TIMES SS. In this study, the suitability of these methods was analyzed by comparing their results to *in vivo* sensitization data of LLNA and human studies.

Compared to human data, 84% of non-sensitizers and sensitizers yielded consistent results in the DPRA. *In silico* tools resulted in 'no alert' for 83%–100% of the non-sensitizers, but alerted only 55%–61% of the sensitizers. The inclusion of biotic and abiotic transformation simulations yielded more alerts for sensitizers, but simultaneously dropped the number of non-alerted non-sensitizers. In contrast to the DPRA, *in silico* tools were more consistent with results of the LLNA than human data. Interestingly, the new "DPRA profilers" (QSAR Toolbox) provided unsatisfactory results.

Additionally, the results were combined in the '2 out of 3' prediction model with *in vitro* data derived from LuSens and h-CLAT. Using DPRA results, the model identified 90% of human sensitizers and non-sensitizers; using *in silico* results (including abiotic and biotic activations) instead of DPRA results led to a comparable high predictivity.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The evaluation of skin sensitization potential is an essential step to define adequate safety measures for chemicals which may get into skin contact. Typically, animal tests are used to characterize the sensitization potential and internationally accepted procedures are described in the form of OECD test guidelines (OECD TG) 406 (*i.e.* the guinea pig maximization test (GPMT) & Buehler test) and 429, 442 A & B (*i.e.* the murine local lymph node assay (LLNA) and its non-radioactive variants). Research on the molecular and cellular mechanisms underlying skin sensitization has resulted in the OECD publishing a document in 2012 formally describing the key events leading to skin sensitization, the so-called Adverse Outcome Pathway (AOP) [OECD, 2012a,b]. Recently, *in chemico* and *in vitro* assays addressing the first three key events of this AOP were validated [OECD, 2015a,b; EURL-ECVAM, 2015]. These tests shall be used in so-called "Integrated Approaches to Testing and Assessment" (IATAs) to investigate skin sensitization potentials without the need for animal testing [Basketter et al., 2015].

The molecular initiation and therefore first key event of the AOP is the covalent binding of chemicals to skin proteins (haptens).

The relevance of a chemical's reactivity with skin proteins for its skin sensitization potential has been known for many years [Dupuis and Benezra, 1982; Landsteiner and Jacobs, 1936; Lepoittevin, 2006]. Haptens are electrophilic, low molecular weight chemicals or their reactive metabolites or spontaneously formed products that then covalently bind to nucleophilic centers of skin proteins. Amino acids containing nucleophilic heteroatoms, such as cysteine, histidine, lysine, methionine and tyrosine have been described to react with haptens [Dupuis and Benezra, 1982; Ahlfors et al., 2003; Lepoittevin et al., 1998]. The direct peptide reactivity assay (DPRA; OECD TG 442C adopted in February 2015) is an *in chemico* method to analyze a chemical's reactivity with synthetic peptides containing cysteine (Cys) or lysine (Lys) as amino acids with nucleophilic side chains which were shown to detect a majority of peptide reactive chemicals [Gerberick et al., 2007; Gerberick et al., 2004].

Long before the establishment of the harmonized testing protocol addressing protein reactivity, *in silico* tools have started making use of this mechanistic information. Several (Q)SAR tools for skin sensitization, for example Toxtree, DEREK Nexus (also compare [Macmillan et al., 2016]), TIMES SS and the QSAR Toolbox are based on structural alerts for protein binding and have been evaluated for skin sensitization in [Teubner et al., 2013]. The QSAR Toolbox containing profilers such as the protein binding profilers 'Protein binding by OASIS' version 1.3 and 'Protein binding by OECD' version 2.3 (termed "OASIS profiler" and

* Corresponding author.

E-mail address: robert.landsiedel@basf.com (R. Landsiedel).

“OECD profiler”, respectively, in this paper) can be used to address the MIE of skin sensitization. Both profilers generate alerts based on the molecular structure submitted to the tools. However, they can additionally be combined e.g., with the modules ‘autooxidation’ and ‘skin metabolism’ which simulate transformations of pre- or pro-haptens, respectively, in order to include the resulting products which may be peptide-reactive haptens [Lepoittevin, 2006]. In addition, version 3 of the QSAR Toolbox introduced two profilers for predicting reactivity with Lys- and Cys-containing model peptides in the DPRA (termed “DPRA profiler (Lys)” and “DPRA profiler (Cys)”, respectively, in this paper).

TIMES (Tissue Metabolism Simulator) — developed by the Laboratory of Mathematical Chemistry in Bourgas, Bulgaria — is a hybrid expert system that encodes structure–toxicity and structure–skin metabolism relationships [Patlewicz et al., 2014b; Dimitrov et al., 2005; Patlewicz et al., 2007]. The module for skin sensitization (TIMES SS) uses transformations defined on the basis of empiric and theoretical knowledge and expert peer-reviews. In a further step, covalent interactions with skin proteins of target chemicals and their putative metabolites are mimicked to estimate peptide reactivity.

Since the individual *in chemico* and *in silico* methods are not sufficient as ‘stand-alone’ methods for skin sensitization hazard assessment, combinations have been proposed in IATA case studies for regulatory purposes with non-animal test methods addressing other AOP key events [Basketter et al., 2015; Joint Cefic LRI/Cosmetics Europe/EPAA workshop, 2015]. One of the IATA case studies applies the ‘2 out of 3’ weight of evidence (WoE) approach (compare [Bauch et al., 2012]): Data on peptide reactivity (DPRA) are combined with results derived from cell-based test methods covering two further key events of the AOP. These are the two ARE-Nrf2 luciferase test methods ‘LuSens’ [Ramirez et al., 2014] or ‘KeratoSens™’ [Emter et al., 2010; OECD, 2015b] and the human cell line activation test (h-CLAT) [Ashikaga et al., 2006; EURL-ECVAM, 2015]. The ARE-Nrf2 luciferase test methods address keratinocyte activation (AOP key event 2) whereas the h-CLAT addresses dendritic cell activation (AOP key event 3). The combination of test methods used in this testing strategy covers the first three of the four AOP key events. For this ‘2 out of 3’ WoE approach, the prediction of skin sensitization is determined by the results of two concordant predictions of the three abovementioned tests [Bauch et al., 2012]. The classification as a sensitizer or non-sensitizer is therefore based on a weight of evidence pertaining to key events of the AOP.

Both, the *in chemico* and the *in silico* test methods, address the MIE of skin sensitization. It is, however, not known how they comparatively perform and whether they can be used interchangeably or complementarily. Such knowledge is crucial for their application in an IATA strategy and was therefore the subject of this investigation. For this purpose, 213 chemicals of a previously published dataset [Urbisch et al., 2015] were tested with the OASIS and OECD profilers of the QSAR Toolbox and with the expert system TIMES SS. Furthermore, experimental DPRA results for 199 of the 213 chemicals were compared to the *in silico* peptide reactivity estimates (with LLNA and human data as *in vivo* reference) (Section 3.1). In addition, the DPRA profilers of the QSAR Toolbox were analyzed for their suitability by comparing their *in silico* predictions to actual experimental DPRA results (Section 3.2). Finally, the results derived by the *in chemico* DPRA, by the *in silico* models QSAR Toolbox (i.e. OASIS and OECD profilers) and by TIMES SS were implemented into a ‘2 out of 3’ prediction model in combination with *in vitro* results derived from the LuSens assay and the h-CLAT (Section 3.3).

2. Material and methods

2.1. Test chemicals

The overall dataset consists of 213 chemicals for which experimental data on the DPRA and other *in vitro* studies on skin sensitization have been published recently [Urbisch et al., 2015]. The chemicals represent a wide diversity of mechanistic domains and uses and all of them have

been investigated in the LLNA. For a subset of 111 chemicals, human data allowed classification as a sensitizer/non-sensitizer and in some cases this classification deviates from the LLNA data: based on LLNA, this subset contains 83 sensitizers and 31 non-sensitizers whereas based on human data, set B contains 75 sensitizers and 36 non-sensitizers. Taking human data as gold standard, a total of 20 (18%) of the 111 chemicals were incorrectly identified as false positive (FP) or false negative (FN) by the LLNA as already presented in a previous paper [Urbisch et al., 2015]. Seven chemicals were assessed to be FN (benzaldehyde, benzyl alcohol, coumarin, streptomycin sulfate, kanamycin, benzocaine and nickel chloride) and 13 to be FP (pyridine, limonene, isopropyl myristate, citronellol, linalool, α -iso-methylionone, benzyl benzoate, sodium dodecyl sulfate, tocopherol, benzyl salicylate, hexyl salicylate, xylene and phthalic anhydride).

The performance of the LLNA to predict human skin sensitization revealed a sensitivity of 91%, a specificity of 64% and an overall accuracy of 82% [Urbisch et al., 2015].

Experimental DPRA data for 199 of the 213 chemicals were generated by P&G or BASF SE. Five chemicals were not considered for further analyses due to discordant results in the two independent laboratories.

For investigation of *in silico* tools in comparison to *in vivo* data, all 213 chemicals listed in Urbisch et al. (2015) were analyzed.

2.2. Direct Peptide Reactivity Assay (DPRA)

Peptide reactivity data have been published previously [Bauch et al., 2012; Natsch et al., 2013; Urbisch et al., 2015]. The procedure is in line with OECD TG 442C (Adopted 2015) and was originally described by Gerberick and coworkers in 2004 [Gerberick et al., 2004]. Briefly, chemicals or vehicle alone were incubated with model heptapeptides containing lysine (Ac-RFAAKAA-COOH) or cysteine (Ac-RFAACAACOOH) for 24 h. Peptide concentrations were determined by HPLC. Peptide reactivity was reported as percent depletion based on the decrease in non-reacted peptide concentration in the sample relative to the average concentration measured in the control. Criteria defining a negative response and reactivity classes are defined by the OECD TG 442C and are listed in Table 1.

2.3. QSAR Toolbox

In silico analysis was performed with the QSAR Toolbox version 3.3.2, which is freely available on the OECD website (<http://www.qsartoolbox.org/>). The QSAR Toolbox developed by OASIS in collaboration with the OECD and the European Chemicals Agency (ECHA) [Raunio, 2011] is a software application for filling gaps of (eco)toxicity data that are needed to assess potential hazards of chemicals. For this purpose, the QSAR Toolbox utilizes numerous databases with results from experimental studies for over 55,000 chemicals and profilers for specific screening properties. The primary function of the profilers is grouping of chemicals in the QSAR Toolbox databases for a read-across or chemical-tailored QSAR assessment of specific endpoints such as skin sensitization. Some of the profilers, however, address specific mechanistic steps and also have a high value on their own.

Table 1
Reactivity classes for peptide binding as defined for the *in chemico* DPRA (OECD TG 442C).

Reactivity class	Cys 1:10/Lys 1:50 prediction model according to OECD TG 442C [%]	Cys 1:10 prediction according to OECD TG 442C [%]
No or minimal	$D_{Cys/Lys} \leq 6.38$	$D_{Cys} \leq 13.89$
Low	$6.38 < D_{Cys/Lys} \leq 22.62$	$13.89 < D_{Cys} \leq 23.09$
Moderate	$22.62 < D_{Cys/Lys} \leq 42.47$	$23.09 < D_{Cys} \leq 98.24$
High	$42.47 < D_{Cys/Lys} \leq 100$	$98.24 < D_{Cys} \leq 100$

Cys = cysteine-containing heptapeptide; Lys = lysine-containing heptapeptide; D_{Cys} = Cys depletion [%]; $D_{Cys/Lys}$ = mean of Cys and Lys depletion [%].

2.3.1. Profilers for protein and peptide binding in the QSAR Toolbox

The two profilers 'Protein binding by OASIS' version 1.3 and 'Protein binding by OECD' version 2.3 were used in this study and are termed as "OASIS profiler" and "OECD profiler", respectively. Alerts being detected by such profilers are listed in Table 2. Both profilers were applied with and without the simulators for skin metabolism ('Skin metabolism simulator') and autooxidation ('Autooxidation simulator'). The primary function of these profilers in the workflow is grouping of chemicals for a read-across or chemical-tailored QSAR assessment. In this study, such grouping approach was however not performed. Instead, the predictivities for skin sensitization were assessed. Therefore, the following approximations were applied: *i.* alert present = skin sensitizer, *ii.* No alert present = non-sensitizer.

2.3.2. Profilers for reactivity estimation in the DPRA

The two profilers 'DPRA Cysteine peptide depletion' version 1.0 and 'DPRA Lysine peptide depletion' version 1.0 of the QSAR Toolbox predicting activities in the experimental DPRA were investigated. For ease of reading these profilers are termed "DPRA profilers". The profilers consist of 24 structural alerts extracted from 110 chemicals with experimentally measured Lys-depletion values and 32 structural alerts extracted from 112 chemicals with experimentally measured Cys-depletion values (compare information in QSAR Toolbox). Although not specifically mentioned in the QSAR Toolbox documentation, the training set data suggest that the 112 chemicals from which the DPRA profilers were derived are those tested by Gerberick and coworkers [Gerberick et al., 2007]. The DPRA profilers of the QSAR Toolbox also provide semi-quantitative potency information (3 classes: weak, moderate and strong), but the class "unreactive" is not available. For the DPRA profilers it should be noted that 13 of the 32 alerts for cysteine binding and 20 of 24 alerts for lysine binding are indicated to be 'under development'. The prediction model underpinning such profilers is listed in Table 3. The DPRA profilers have been built with an applicability domain so that the result of the profiling may be 'not possible to classify according to these rules'. These profilers are described separately from the OASIS and OECD profilers in Section 3.2.

2.4. TIMES SS

Analysis was performed with module 'Predicting skin sensitization with autooxidation' of TIMES SS v2.27.13 (<http://oasis-lmc.org/products/software/times.aspx>). TIMES SS reports structural alerts and assigns a test chemical to its mechanistic domain. The expert system also indicates whether the parent or its metabolite or autooxidation product is expected to be reactive. Beside the prediction of probable protein binding mechanisms, TIMES SS is able to predict the skin sensitization potential and semi-quantitative potency (3 classes: non-sensitizing, weak, strong) of a target chemical and its metabolites and estimates the reliability of the prediction and the domain adherence

Table 2
Proposed mechanistic domains assigned by the QSAR Toolbox.

OECD profiler	OASIS profiler
Acylation	Ac Acylation
Michael addition	MA Michael addition
Schiff base formation	SB Schiff base formation
Nucleophilic substitution type 2	S _N 2 Nucleophilic substitution type 2
Nucleophilic aromatic substitution	S _N Ar Nucleophilic aromatic substitution
The OECD profiler has 102 categories and was adopted in January 2011	Ionic interaction II Nucleophilic addition NA Radical reactions RR Nucleophilic substitution type 1 S _N 1 Nucleophilic substitution type 2 S _N 2i ionic S _N 2i Nucleophilic vinylic substitution S _N V
The OASIS profiler has 102 categories and was adopted in October 2010	

Table 3

Reactivity classes for peptide binding as defined for the two DPRA profilers within the QSAR Toolbox.

Reactivity class	Cys- or Lys-depletion according to QSAR Toolbox v3.3.2 [%]
No or minimal	n/a
Low	5 < D _{Cys} or D _{Lys} < 40
Moderate	40 < D _{Cys} or D _{Lys} < 80
High	80 < D _{Cys} or D _{Lys} < 100

Cys = cysteine-containing hepta peptide; Lys = lysine-containing hepta peptide; D_{Cys} = mean of Cys depletion; D_{Lys} = mean of Lys depletion; n/a = not available.

of the considered structure. The training set consists of 875 chemicals with experimental *in vivo* data from LLNA, GPMT or human studies (compare information within TIMES). In TIMES SS, each prediction is combined with a calculation of peptide reactivity, which was considered for the underlying study.

2.5. Application of protein binding modeling in the '2 out of 3' prediction model

Experimental data for peptide reactivity, keratinocyte activation and dendritic like cell line activation for the current test chemicals has previously been used in a weight of evidence approach to predict the skin sensitization potential [Urbisch et al., 2015]. For this '2 out of 3' WoE approach, any two congruent results of the non-animal tests addressing one of the key events determine the overall assessment: If at least two of the three tests considered were positive, the chemical was rated to be a skin sensitizer. If at least two of the three tests were negative, the chemical was rated to be a non-sensitizer. In this study, the results of the *in silico* tools were included in the '2 out of 3' prediction model with the existing *in vitro* studies on keratinocyte and dendritic like cell line activation. In contrast to the OASIS and OECD profilers of the QSAR Toolbox that were applied with and without the simulators for skin metabolism and autooxidation, the two DPRA profilers were not considered for the '2 out of 3' WoE approach, since these are not fully developed and established (for details refer to Section 3.2 and Discussion section) (Table 4).

2.6. Data analyses

2.6.1. Statistics

Sensitivities, specificities, positive predictive values, negative predictive values as well as accuracies were calculated for the individual *in chemico* and *in silico* methods as well as for the '2 out of 3' prediction model by using Cooper statistics [Cooper et al., 1979] with LLNA or human data as reference. However, the values determined here for the individual methods as a stand-alone test methods are only indicative since the test methods should be considered in combination with other sources of information in the context of an IATA. All parameters are based on a 2 × 2 contingency table counting the number of compounds that are "true positive" (TP), "false positive" (FP), "true negative" (TN) and "false negative" (FN):

$$\begin{aligned} \text{Sensitivity}[\%] &= \text{TP}/(\text{TP} + \text{FN}) * 100 \\ \text{Specificity}[\%] &= \text{TN}/(\text{TN} + \text{FP}) * 100 \\ \text{Positive predictive value}[\%] &= \text{TP}/(\text{TP} + \text{FP}) * 100 \\ \text{Negative predictive value}[\%] &= \text{TN}/(\text{TN} + \text{FN}) * 100 \\ \text{Accuracy} [\%] &= (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{TN} + \text{FN}) * 100. \end{aligned}$$

2.6.2. Interpretation of *in silico* profilers

The OASIS and OECD profilers of the QSAR Toolbox were applied both with and without the simulators for skin metabolism and autooxidation. To calculate the predictivity of the *in silico* tools for estimating skin sensitization potential, profiling results of all tests were assessed

Table 4

Overview of evaluations described in the current paper.

Study overview		Consideration of skin metabolism possible	Consideration of autoxidation possible	Evaluated in comparison to	Implementation in '2 out of 3' WoE approach
QSAR Toolbox v3.3 ("in silico")					
Profilers for peptide reactivity	OASIS profiler OECD profiler	Optional Optional	Optional Optional	LLNA and human data LLNA and human data	Yes Yes
Prediction model for DPRA reactivity	DPRA profiler Cys DPRA profiler Lys	No No	No No	DPRA data DPRA data	No No
TIMES SS v2.27.15 ("in silico")					
Prediction model for skin sensitization; also estimation of peptide reactivity		Optional	Optional	LLNA and human data	Yes
Direct peptide reactivity assay (DPRA) ("in chemico")					
Measurement of peptide reactivity using Lys or Cys containing peptides		No	Probably yes	LLNA and human data	Yes

as simple yes/no answers: If a structural alert for peptide reactivity was identified in the molecular structure of a target chemical, a skin sensitization potential of this chemical was assumed. If no structural alert was identified (*i.e.* 'no alert'), the chemical was treated as a non-sensitizer. To calculate the Cooper statistics, *in silico* results were compared to *in vivo* LLNA and human data used as reference data.

The two DPRA profilers of the QSAR Toolbox assign a chemical to be low, moderate or high reactive in the DPRA, whereas a non-reactivity is not predicted. The DPRA profilers have been built with an applicability domain, so that the result of the profiling may also be 'not possible to classify according to these rules'. Since a negative result is not predicted, only positive predictive values were calculated for the profilers and compared to the actual experimental result of the DPRA for the respective chemical.

TIMES SS simultaneously calculates peptide reactivity and predicts skin sensitization. Since the MIE of skin sensitization is investigated in this study, the peptide reactivity estimates were primarily considered in the following (*i.* alert present = skin sensitizer, *ii.* no alert present = non-sensitizer).

3. Results

3.1. Peptide or protein reactivity for predicting skin sensitization potential

213 chemicals of a previously published dataset [Urbisch et al., 2015] were tested by the OASIS and OECD profilers of the QSAR Toolbox and the expert system TIMES SS. Furthermore, experimental DPRA results

for 199 of the 213 chemicals were included in the study. The concordance of the *in chemico* and *in silico* estimates on peptide reactivity with *in vivo* data on skin sensitization is shown in Table 5, comparison is made to both, data of the murine LLNA and data derived from human diagnostic tests.

Compared to human data, the DPRA provided an accuracy of 84% and thereby exceeded the accuracies for the prediction of skin sensitization potential of the two QSAR Toolbox profilers and TIMES SS (for the external evaluation set). The DPRA predicted human data (Accuracy = 84%) with a higher accuracy than LLNA data (Accuracy = 75%).

In case of the *in silico* tools, different results on the presence or absence of alerts for peptide reactivity are provided for inclusion and exclusion of simulation of skin metabolism and autoxidation. A lower sensitivity was noted if such profilers were used without the modules for skin metabolism and autoxidation: For the parent chemicals alone, a protein binding potential for 67 or 61% (OASIS) and for 63 or 60% (OECD) of all sensitizing chemicals was detected; non-binding was detected for 82 or 88% (OASIS) and for 85 or 83% (OECD) of non-sensitizing chemicals, when compared to LLNA or human data, respectively. The accuracies were 71 or 70% (OASIS) and 69 or 67% (OECD), when compared to LLNA or human data, respectively (Table 5).

The combination of the two protein binding profilers with the skin metabolism and autoxidation simulators increased the number of chemicals predicted to be protein binding among the sensitizers to 92 or 89% (OASIS) and 88 or 88% (OECD), but decreased the fraction of non-binding chemicals among non-sensitizers to 65 or 65% (OASIS) and 69 or 61% (OECD), when compared to LLNA or human data,

Table 5Overview of the results on peptide reactivity for the assessment of skin sensitization derived by the *in silico* tools and the *in chemico* DPRA.

Concordance of <i>in silico</i> and <i>in chemico</i> estimates on peptide reactivity with LLNA and human data on skin sensitization [%]	<i>In silico</i> approach																<i>In chemico</i> approach	
	QSAR Toolbox v3.3.2								TIMES SS v2.27.13								DPRA (Direct Peptide Reactivity Assay)	
									TIMES – P		TIMES (P + M + AO)		TIMES (P)		TIMES (P + M + AO)			
	OASIS profiler (P)		OASIS profiler (P + M + AO)		OECD profiler (P)		OECD profiler (P + M + AO)		Part of training set & in domain (internal evaluation)		Part of training set & in domain (internal evaluation)		Not part of training set & in domain (external evaluation)		Not part of training set & in domain (external evaluation)			
<i>In vivo</i> reference	LLNA	Human	LLNA	Human	LLNA	Human	LLNA	Human	LLNA	Human	LLNA	Human	LLNA	Human	LLNA	Human	LLNA	Human
n	211	109	211	109	211	109	211	109	146	74	146	74	31	17	31	17	194	102
Sensitivity ^a	67	61	92	89	63	60	88	88	68	61	95	90	76	55	96	82	76	84
Specificity ^a	82	88	65	65	85	83	69	61	95	87	93	83	83	100	83	67	72	84
PPV ^a	90	92	87	85	91	88	88	83	97	91	97	92	95	100	96	82	87	90
NPV ^a	49	51	76	73	48	49	69	70	56	50	89	79	45	55	83	67	57	70
Accuracy ^a	71	70	84	82	69	67	82	79	76	69	95	88	77	71	94	76	75	84

^a To calculate cooper statistics for the single *in chemico* and *in silico* methods, the following approximations were applied: *i.* alert or positive result for peptide reactivity = skin sensitizer; *ii.* no alert or negative result for peptide reactivity = non-sensitizer; Cooper statistics were calculated with LLNA or human data as *in vivo* reference for both the parent (P) and the parent including skin metabolites and autoxidation products (P + M + AO). n = number of analyzed chemicals; NPV = negative predictive value; PPV = positive predictive value.

respectively. In summary, the inclusion of biotic and abiotic transformations led to a slight increase of the false positive predictions and to a clear decrease in the false negative rate at the same time (Table 5). The OASIS profiler as well as the OECD profiler predicted LLNA data with a higher accuracy than human data.

Taking a closer look at the results for the 213 chemicals, the profilers resulted in 12 discordant predictions (Table 6), indicating the OASIS profiler to have a slightly better accuracy than the OECD profiler for the underlying test set (Table 5). The broader definition of Schiff base formers in the OASIS profiler results both in more TP but also FP results compared to the OECD profiler. In case of acylating agents, in the OASIS profiler allows the correct identification of iodopropynyl butylcarbamate, whereas the OECD profiler gives an alert for 6-methylcoumarin, which was however not peptide reactive in the experimental DPRA and also a non-sensitizer according to *in vivo* tests. In case of Tween 80, a similar pattern appears to exist for the alert for nucleophilic substitution of type 2.

In TIMES SS, each prediction of a skin sensitization potential is connected to the most probable protein binding mechanism and coupled with a statement on the domain adherence. Of the 213 chemicals, 177 chemicals (83%) were noted to be 'in domain', while 30 chemicals were 'out of domain' (14%). For five inorganic metal salts and the mixture MCI/MI, 'no domain' was found. The 177 'in domain' chemicals could be subdivided into an internal evaluation set consisting of 146 chemicals being part of the training set (69% of the dataset) and an external evaluation set consisting of 31 chemicals (15% of the dataset) for which no experimental data were implemented in TIMES SS (Table 5). Three chemicals were part of the training set of TIMES SS but 'out of domain' and were not considered for further investigations.

As seen in Table 5, also with TIMES SS the accuracy benefits by the inclusion of biotic and abiotic transformations. The underlying algorithms result in a better performance than simply screening for absence or presence of a structural alert. Accuracies for the 177 'in domain' chemicals were 94 or 86%, when compared to LLNA or human data,

respectively (Table 5). For the external set (e.g. 'not part of the training set' and 'in domain'), the accuracy for human and LLNA data was 76% and 94%, respectively, and thus differed by 18% with a higher accuracy for LLNA data. The increased accuracies when taking LLNA data as reference indicate the potential bias by TIMES SS being trained on LLNA data.

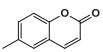
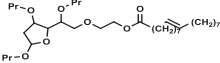
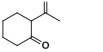
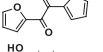
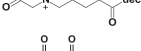
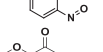
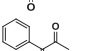
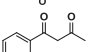
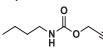
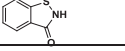
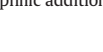
When comparing the results of the OASIS profiler of the QSAR Toolbox to TIMES SS, 13 chemicals with structural alerts identified by the OASIS profiler were correctly predicted as non-sensitizers by TIMES SS. In 12 cases, the structural alert was not fired by TIMES SS and in case of the highly volatile methyl methacrylate, a domain-specific model was activated and overruled the alert.

3.2. *In silico* profilers for DPRA reactivity

In Section 3.1, the OASIS and OECD profilers for generally identifying alert for peptide reactivity were described. Moreover, the QSAR Toolbox also offers two profilers – the DPRA profilers – which predict the putative DPRA result of a target chemical by differentiating a low, moderate or high reactivity with the Lys or Cys containing model peptide; while a negative result such as 'no reactivity' is not predicted by the DPRA profilers. It is has to be acknowledged that the underlying prediction model of the DPRA profilers (Table 3) clearly differs from that of the experimental DPRA as described in OECD TG 442C (Table 1).

To check the utility of the DPRA profilers, their predictions were compared to experimental DPRA data. For most of the analyzed chemicals, the outcome of the profiling was that it was 'not possible to classify according to these rules'. Therefore, predictions for Cys- and Lys-reactivities were only obtained for 79 and 41 chemicals of the dataset, respectively. For Cys-reactivity, 66 predictions were correct positive, whereas 13 were false positive (PPV = 84%). For Lys-reactivity, 27 predictions were correct positive and 14 false positive (PPV = 66%). Since the prediction of 'no reactivity' is not implemented for the DPRA profilers, other statistical parameters than PPV could not

Table 6
Discordant results of the two protein binding profilers of the QSAR Toolbox (OECD and OASIS algorithms) regarding the presence or absence of alerts for protein reactivity.

Chemical name	Chemical structure	OECD QSAR Toolbox v.3.3		<i>In chemico</i> ¹	<i>In vivo</i> reference ^a	
		OASIS profiler	OECD Profiler	DPRA	LLNA (EC3 [%])	Human
6-Methyl-coumarin		No alert	Ac/MA	Negative	Negative	Negative
Tween 80		No alert	S _N 2	Negative	Negative	Negative
2-Acetyl-cyclohexanone		SB	No alert	Positive	Negative	No data
Furil		SB	No alert	Positive	Negative	No data
Cocamido-propyl betaine		Ac/ionic interaction	No alert	Negative	Negative	No data
Ethyl benzoylacetate		NA	No alert	Negative	Negative	No data
N,N-dimethyl-4-nitrosoaniline		NA	No alert	Positive	Positive (0.48)	No data
Methyl pyruvate		NA/SB	No alert	Negative	Positive (2.4)	No data
1-Phenyl-1,2-propanedione		SB	No alert	Positive	Positive (1.3)	No data
1-Benzoyl-acetone		SB	No alert	No data	Positive (0.04)	No data
Iodopropynyl butylcarbamate		Ac	No alert	Positive	Positive (0.9)	Positive
1,2-Benziso-thiazolin-3-one		S _N 2	No alert	Positive	Positive (1.3)	Positive

SB = Schiff base formation, NA = nucleophilic addition, Ac = acylation, MA = Michael addition; S_N2 = nucleophilic substitution of type 2.

^a Data published [Urbisch et al., 2015].

be calculated. The results indicate a relatively high uncertainty of the DPRA profilers to predict the actual experimental DPRA result. However, most of the alerts addressing reactivity in the DPRA were noted to be under development in the supporting documentation of the QSAR Toolbox.

3.3. '2 out of 3' weight of evidence (WoE) approach

Results of the *in silico* tools QSAR Toolbox and TIMES SS and the *in chemico* DPRA were combined with results of the *in vitro* methods addressing the cellular key events of the AOP (LuSens and h-CLAT) and Cooper statistics were calculated for several combinations in an AOP-based '2 out of 3' WoE approach (Table 7).

The combination of the *in chemico* DPRA with the *in vitro* ARE-Nrf2 luciferase assays (KeratinoSens™ and LuSens alike) and h-CLAT provided an accuracy of 90%, when compared to human data [Urbisch et al., 2015]. Combining results of investigated *in silico* tools instead of the DPRA with results of the LuSens and h-CLAT also yielded high accuracies comparable to those previously described in the literature (*i.e.* accuracies ranging from 79% to 83% (compared to LLNA data) or from 90% to 94%, (compared to human data)) [Bauch et al., 2012; Urbisch et al., 2015; Natsch et al., 2013].

Compared to human data, the prediction model using the OASIS profiler in combination with simulators for skin metabolism and autoxidation led to an accuracy of 89%; the OECD profiler and TIMES SS performed similarly (88% and 86%, respectively). Without the consideration of autoxidation and metabolic transformation, the combinations were slightly less accurate in predicting human skin sensitizers (85%, 84% and 85% for OECD and OASIS profilers and TIMES SS, respectively).

Applying the DPRA results in the '2 out of 3' prediction model generally predicted human data with higher accuracy than LLNA data. When applying OECD or OASIS profiler data instead, human and LLNA data were predicted with similar accuracies. When TIMES SS predictions were applied, LLNA data were predicted with higher accuracy – alas based on a considerably smaller set of chemicals.

4. Discussion

Previously, twelve IATA case studies were presented to the OECD of which at least seven included information on the molecular initiating

event (MIE) derived by statistical or mechanistic *in silico* tools in their prediction model (compare [Matheson, 2015; Hirota et al., 2015; Natsch et al., 2015; Jaworska et al., 2015; van der Veen et al., 2014; Patlewicz et al., 2014a; Takenouchi et al., 2015; Teissier and Alépée, 2015]). In this study we compared the performance of the OECD and OASIS profilers of the QSAR Toolbox and the expert system TIMES SS to the performance of the experimental DPRA. Both, the *in silico* and *in chemico* methods, were investigated for the prediction of peptide reactivity and their utility for hazard assessment of skin sensitization.

The *in silico* tools use mechanistic knowledge to provide information on the apparent protein binding mechanism for chemicals with a defined molecular structure. The freely available QSAR Toolbox utilizes numerous databases and profilers for specific screening properties. Of note, a number of features in the workflow of the QSAR Toolbox like profiling, grouping and data gap filling are not strictly predefined and can thus be applied in a substance-tailored manner. Hereby, the choice of inappropriate parameters can impact the validity of the results, thus, *in silico* analyses should be performed by experts.

Profiling for peptide reactivity is quickly done and thus an attractive method for an initial screening, if one is certain about the chemical identity and has a mono-constituent chemical with defined structure. Metal-containing salts are also considered, although interaction with nucleophilic peptide residues is different from covalent binding (coordination bonds). No information on amino acid selectivity, reaction rate and stability of protein conjugates is provided (Table 9). The OECD and OASIS profilers of the QSAR Toolbox showed high specificities and positive predictivities for sensitization potential (based on peptide reactivity), when considering the molecular structures of parent chemicals alone. Due to the low probability of obtaining false positive predictions, chemicals with alerts in their parent molecular structure are highly likely to give a positive response *in vivo* even when tested with chemicals not known to model developers (also compare Teubner et al., 2013). The poor sensitivities and negative predictive values could be increased by considering possible derivatives of the parent, which is relevant in cases where the chemical is either a pre- or pro-hapten. However, this consideration was accompanied by a slight drop in the specificities and positive predictive values as shown in Table 5. Since the QSAR Toolbox and its profilers are updated steadily, the performance of such profilers is likely to be improved by incorporation of new data of the constantly growing number of chemicals with experimental data on peptide reactivity and skin

Table 7

Overview of cooper statistics for the several combinations of the '2 out of 3' prediction model when compared to human and LLNA data.

Compared to human data		n	Cooper statistics [%]				
			Sens	Spec	PPV	NPV	Acc
LLNA		111	91	64	84	77	82
DPRA + <i>in vitro</i> results	DPRA + LuSens + h-CLAT	90	90	89	95	80	90
	DPRA + KeratinoSens™ + h-CLAT	101	90	90	96	79	90
	OASIS profiler (P) + LuSens + h-CLAT	85	81	93	96	69	85
	OASIS profiler (P + M + AO) + LuSens + h-CLAT	92	95	75	90	88	89
QSAR Toolbox + <i>in vitro</i> results	OECD profiler (P) + LuSens + h-CLAT	81	82	88	94	70	84
	OECD profiler (P + M + AO) + LuSens + h-CLAT	91	95	71	88	87	88
	TIMES SS (P) + LuSens + h-CLAT	13	78	100	100	67	85
TIMES SS (external evaluation set) + <i>in vitro</i> results	TIMES SS (P + M + AO) + LuSens + h-CLAT	14	100	60	82	100	86
Compared to LLNA data		n	Cooper statistics [%]				
			Sens	Spec	PPV	NPV	Acc
DPRA + <i>in vitro</i> results	DPRA + LuSens + h-CLAT	133	83	78	91	64	82
	DPRA + KeratinoSens™ + h-CLAT	180	82	72	89	59	79
	OASIS profiler (P) + LuSens + h-CLAT	128	82	86	94	67	84
	OASIS profiler (P + M + AO) + LuSens + h-CLAT	141	96	76	92	88	91
QSAR Toolbox + <i>in vitro</i> results	OECD profiler (P) + LuSens + h-CLAT	124	80	84	92	65	81
	OECD profiler (P + M + AO) + LuSens + h-CLAT	143	90	81	93	73	87
	TIMES SS (P) + LuSens + h-CLAT	19	93	100	100	83	95
TIMES SS (external evaluation set) + <i>in vitro</i> results	TIMES SS (P + M + AO) + LuSens + h-CLAT	20	100	100	100	100	100

n = number of analyzed chemicals; Sens = sensitivity; Spec = specificity; PPV = positive predictive value; NPV = negative predictive value; Acc = accuracy; P = only parent structure considered; P + M + AO = parent structure and transformation products simulated by autoxidation and/or skin metabolism.

sensitization as well as refining the algorithms of the profilers. Candidates for such improvements were identified as OECD and OASIS profilers classified them differently (Table 6). This reflects the fact that different profilers have different sensitivities and that the OECD profiler is more conservative than the OASIS profiler. For example, five chemicals were classified as Schiff base formers by the OASIS profiler, whereas no alert was identified by the OECD profiler at the same time (Table 8).

Direct acting Schiff base formers (SB) can be subdivided into α/β -saturated aldehydes, 1,2-dicarbonyls and 1,3-dicarbonyls; the five chemicals with discordant classifications are represented by dicarbonyls. The classification rules for dicarbonyls underlying the two profilers are listed in Fig. 1.

According to the classification rules of the OECD profiler (based on [Aptula and Roberts, 2006; Enoch et al., 2009; Roberts et al., 2006; Roberts et al., 2007]), **1** is not classified as SB due to the oxygen as one residue (residue **R** = C or H) and **3** is not classified as SB due to the two furan residues (both **R** = heteroaromatic). However, in the molecular structures of **2** and **4**, only one residue is aromatic and in **5** even none of the residues is aromatic. Obviously, the algorithm of the profiler is more stringent than described in the respective information. In case of the OASIS profiler, the rules also consider two aromatic residues for 1,2-dicarbonyls and one aromatic residue for 1,3-dicarbonyls. The formation of a covalent bond with an arginine residue is proposed for such dicarbonyls (compare profiler information in the QSAR Toolbox). Since all of the five discordant chemicals were positive in the LLNA, the OASIS profiler provides a higher sensitivity for SB formers.

While the profilers within the QSAR Toolbox only consider the molecular structure of a target chemical, TIMES SS additionally calculates physical properties of a target chemical such as vapor pressure and water solubility ($\log K_{ow}$) which were implemented into the prediction model (also compare Table 9). This is probably the reason, why the OASIS profiler of the QSAR Toolbox and TIMES SS provide different results on peptide reactivity (22 out of 213 chemicals classified differently; data not shown), although for both tools algorithms were developed by the Laboratory of Mathematical Chemistry in Bourgas (Bulgaria). For instance, methyl methacrylate and pyridine were classified by the OASIS profiler in combination with the skin metabolism and autoxidation profiler to be peptide reactive, whereas TIMES SS classified both chemicals not to bind peptides due to their high volatilities. Additionally, TIMES SS offers information on domain and training set affiliations of target chemicals (also compare Table 9). Many chemicals assessed in this study are part of the training set of TIMES SS ('internal data') and inclusion of these chemicals would have skewed the results to a better outcome. Only 15% of the 213 chemicals were 'in domain' and 'not part of the training set' at the same time. In order to avoid the above described bias, the evaluation was restricted to this smaller 'external' data subset. Interestingly, performances for the internal and external

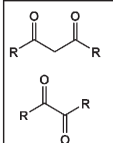
	OECD profiler R = hydrogen, any carbon (both R groups cannot be aromatic, heteroaromatic or heterocyclic, unless they are either mono- or di-ortho-substituted)
	OASIS profiler 1,2-dicarbonyl: both R = C (aromatic) 1,3-dicarbonyl: one R = C (aromatic) and the other R = C sp ³ (further rules are listed in the QSAR Toolbox)

Fig. 1. Different definitions of Schiff base formers considered in the algorithms of the OASIS and OECD profilers of the QSAR Toolbox.

sets were almost comparable (vs. LLNA data) indicating TIMES SS to be a robust and predictive *in silico* tool for assessing skin sensitization potential on the basis of peptide reactivity. However, TIMES SS as well as the QSAR Toolbox profilers clearly provide higher accuracies when taking LLNA data as reference instead of human data. These results indicated (OASIS & QSAR profilers) or confirmed (TIMES SS) that the investigated *in silico* tools were predominantly trained on animal data. The inclusion of (further) published human data and a higher weighing of those would probably further improve the accuracy of TIMES SS as well as of the profilers of the QSAR Toolbox. Patlewicz and coworkers have recently also identified a few cases where profilers would benefit from refinement [Patlewicz et al., 2014b].

Apart from the general profiling of target chemicals for their peptide reactivity with the OASIS and OECD profilers and also TIMES SS, the QSAR Toolbox also offers two specific DPRA profilers. Since these profilers directly confer to DPRA results, their predictions were compared directly to experimental DPRA data to evaluate the concordance of *in silico* data with *in chemico* data. Significant restrictions were observed during the analysis due to the facts that *i.* only positive (*i.e.* positive in DPRA) and no negative results were predicted and that *ii.* most of the target chemicals were not predictable. The comparison revealed an unsatisfactory performance of the DPRA profilers for predicting the outcome of the experimental assay. Both profilers are overpredictive what could be explained by the less stringent prediction model of the DPRA profilers (Table 3) when compared to the actual prediction model described in OECD TG 442C (Table 1). Since a profiler reflects the knowledge of data at the time it was developed, it is logically not as good as experimental data. Such profilers are based on about 112 chemicals, while today there are open literature test data for more than 200 chemicals. However, if a profiler is noted to be still under development (like many domains in the DPRA profilers), it should not be considered in a publicly available *in silico* tool such as the QSAR Toolbox. Especially when *i.* the performance is worse, *ii.* the underlying prediction model is implausible and *iii.* more robust and accurate profilers addressing a similar issue such as the OECD or OASIS profilers are already implemented in the same tool, the DPRA profilers will

Table 8

Discordant results, which were profiled as Schiff base formers (SB) or peptide unreactive (no alert).

No.	Chemical name	Chemical structure	Structural alert	OECD QSAR Toolbox v.3.3		<i>In chemico</i> DPRA
				OASIS profiler	OECD profiler	
1	Methyl pyruvate		1,2-Dicarbonyl	SB/NA	No alert	Negative
2	1-Phenyl-1,2-propanedione		1,2-Dicarbonyl	SB	No alert	Positive
3	Furil		1,3-Dicarbonyl	SB	No alert	Positive
4	1-Benzoyl-acetone		1,3-Dicarbonyl	SB	No alert	No data
5	2-Acetyl-cyclohexanone		1,3-Dicarbonyl	SB	No alert	Positive

SB = Schiff base formation, NA = nucleophilic addition,

Table 9

Overview of applicability and limitations of the protein-binding profilers within the QSAR Toolbox, TIMES SS and the DPRA.

	<i>In silico</i>		<i>In chemico</i>
	QSAR Toolbox ^a	TIMES SS	DPRA
Applicability to chemicals			
Defined chemical structure required	Yes	Yes	No ^b
Applicable to mixtures	Limited (qualitatively)	Limited (qualitatively)	Limited ^c (qualitatively & quantitatively)
Solubility considered	No	Yes (log Kow implemented in algorithm)	yes
Vapor pressure considered	No	Yes (calculated)	Not required (incubation in closed system)
Pre-hapten activation considered	Yes (simulation of abiotic activation)	Yes (simulation of abiotic activation)	Yes ^d
Pro-hapten activation considered	yes (simulation of biotic activation)	Yes (simulation of biotic activation)	No ^e
INFORMATION OUTPUT			
Information on reaction mechanism	Yes	Yes	Limited (by amino acid selectivity)
Information on amino acid selectivity	No	No	Yes (Cys & Lys)
Information on yield of conjugates	No	Indirectly by amount of non-depleted peptide (calculated)	Indirectly by amount of non-depleted peptide (measured)
Information on kinetics of conjugate formation	No	No	No ^e
Information on conjugate stability	No	No	No ^e
LIMITATIONS			
Information on applicability domain affiliation	No	Yes	Not required
Information on training set affiliation	No	Yes	Not required
Information on <i>in vivo</i> data used to define the profilers	No	Yes	Not required
Information on metal containing substances	Yes	No	No (compare OECD TG)
Potential artifacts (co-elution, non-relevant peptide depletion)	No	No	Yes

^a Protein binding profilers with OECD or OASIS algorithms combined with simulators of autoxidation and skin metabolism.^b A defined molecular structure is not required for the DPRA. However as the OECD TG 442C describes the protocol to test substances at a defined molar concentration, the molecular weight of the substances is required.^c According to the prediction model described in OECD TG complex mixtures cannot be assessed. However a gravimetric approach has been proposed but has not yet been validated yet.^d As described in OECD TG 442C metabolic capacity is not included in the test system. However we have shown that the DPRA conducted according to the OECD TG detects some pre-haptens correctly (Urbisch et al., submitted).^e Could be analyzed using further modifications.

probably cause uncertainty of users and authorities rather than providing valuable results.

Moving from computational to experimental data, intrinsic factors such as solubility, high molecular weight and HPLC-retention times of the test chemical overlapping with those of the model peptides (co-elution) can limit the application of the DPRA (also compare Table 9). Furthermore, the DPRA is also not applicable to putative pro-haptens and has been described not to be applicable to metal-containing chemicals (although providing a lot of true positive and true negative results) or oxidants (compare OECD TG 442C). Beside lysine and cysteine, also arginine, methionine, tyrosine, proline and histidine are potentially able to covalently bind specific electrophiles and chemicals predominantly reacting with these amino acid side chains have been suggested to be tested falsely negative in the DPRA [Gerberick et al., 2008]. However, no follow-up analysis on the actual (non-)detection of such chemicals in the DPRA has been done and at least Ni²⁺ ions [Romagnoli et al., 1991] and methyl methane sulfonate [Lepoittevin and Benezra, 1992] – both described to be predominantly histidine-reactive – were correctly detected as peptide reactive in the DPRA. Since the DPRA detects peptide depletions, no structural information on the formation of adducts is obtained during the standard procedure to possibly describe the type of reactivity. Hence, also artifacts like peptide oxidations, cleavages or transfer reactions might lead to peptide depletions and generate potentially false positive results [Natsch et al., 2007; Natsch and Gfeller, 2008; Natsch et al., 2012]. The biological role of these artifacts in the induction of a skin sensitization is scientifically unknown [Natsch and Gfeller, 2008; Roberts and Natsch, 2009]. However, despite these limitations, in this investigation the DPRA has proven to provide reliable results for the discrimination of human skin sensitizers from non-sensitizers.

In contrast to TIMES SS and the OECD and OASIS profilers of the QSAR Toolbox, the DPRA predicted human data more accurately than LLNA data (Table 5). In addition, the DPRA provided the highest accuracy in the prediction of human data in this study indicating its high suitability

for predicting skin sensitization potential. However, the DPRA provided negative predictive values of only 57 or 70%, when compared to LLNA or human data, respectively. This can be explained by a lack of metabolic activation of pro-haptens being detected as false negatives in the DPRA and an assay containing peroxidase has been developed to overcome this issue (peroxidase peptide reactivity assay, PPRA [Gerberick et al., 2009; Troutman et al., 2011]). In contrast, pre-haptens resulted in peptide depletions in the DPRA [Gerberick et al., 2009]. It should be noted that many chemicals requiring transformation to form a hapten can either be activated by abiotic or enzymatic processes and hence assigning the terms pre- and pro-hapten is often ambiguous. The overall accuracy of skin sensitization predictions of the DPRA based on a chemical set without pre- and pro-haptens was 79% when compared to LLNA data (n = 163, data not shown) or 89% when compared to human data (n = 83, data not shown); the negative predictive values were increased to values of 69% (LLNA) and 87% (human) in this specific set.

The results of the abovementioned individual methods were also implemented into the '2 out of 3' prediction model to address the first three key events of the AOP instead of only the first key event (compare Table 7). The accuracy was generally increased, when comparing results from the single methods to results of the '2 out of 3' prediction model, indicating a further decrease in the uncertainty of the predictions. All combinations of *in silico* or *in chemico* results taken together with results of the LuSens and h-CLAT were more predictive (i.e. 84–90%) than the LLNA (i.e. 82%), when taking human data as *in vivo* reference. The increased accuracies of the combinations '*in silico* + LuSens + h-CLAT result' again showed the above outlined bias, since LLNA data (used as main *in vivo* reference in the training sets of the *in silico* tools) are predicted with a higher accuracy (i.e. 84%–100%) than human data (84%–89%). As likewise seen in the individual *in silico* models, the accuracies of the '2 out of 3' prediction model were increased when the simulation of the metabolic capacity of the skin and the potential of a chemical to be oxidized was integrated. The combination of TIMES

SS including skin metabolism and autoxidation simulation with LuSens and h-CLAT results led to an accuracy of 100%. However, since the majority of the investigated chemicals is also part of the training set of TIMES SS, only those few chemicals being ‘in domain’ but ‘not part of the training set’ ($n = 20$) could have been considered for this approach. A decrease in the accuracy when moving to larger datasets could be explained by limitations in the applicability domain or experimental inaccuracies [Natsch et al., 2013].

5. Conclusions

In this study, computationally (*in silico* tools ‘QSAR Toolbox’ and ‘TIMES SS’) and experimentally (*in chemico* DPRA) derived data on peptide reactivity were evaluated for their usefulness in predicting the skin sensitization potential of chemicals in humans. In general, alerts from *in silico* tools were more consistent with the results of LLNA studies than actual human skin sensitization potentials – reflecting their training sets. These findings emphasize the importance of considering human data while developing *in silico* tools. In contrast to the *in silico* results, DPRA results were more consistent with human data than with LLNA data.

Interestingly, the general *in silico* profilers of protein reactivity performed more accurately than specific profilers to predict the outcome of the DPRA.

Selected *in silico* tools have the potential to address peptide reactivity as the molecular initiating event of skin sensitization as accurately as *in chemico* methods suggesting that they can be used complementarily, as each method has its own limitations: The DPRA is the only regulatory accepted method addressing the MIE of skin sensitization. However, chemicals which cannot be tested by DPRA (insoluble chemicals, pro-haptens, chemicals co-eluting with the model peptide) can still be tested by *in silico* tools. In addition *in silico* tools offer information on the reaction mechanism while the DPRA offers only limited information on reactivity with Lys or Cys. While autoxidation of pre-haptens is considered in the DPRA and can be simulated by *in silico* methods, simulating metabolic activation of pro-haptens is only available *in silico*. Interestingly, *in silico* tools with metabolic simulation did not yield more positive results for human sensitizers than the DPRA.

While protein reactivity is the MIE and first key event of the AOP, adding information on other key events leading to skin sensitization can be used to increase the weight of evidence and to decrease uncertainty within a hazard or risk assessment. If the strengths and limitations of *in silico* tools are taken into account, their proper use in a fixed data interpretation procedure within an ITS – such as the ‘2 out of 3’ prediction model – leads to an assessment of skin sensitization potential in humans with a comparable or even higher accuracy than provided by the LLNA.

Conflict of interest statement

The authors declare no competing financial interest.

Abbreviations

Acc	accuracy
Ac	acylating agent
AOP	adverse outcome pathway
ARE	antioxidant response element
Cys	cysteine
DPRA	direct peptide reactivity assay
ECHA	European Chemical Agency
FN	false negative
FP	false positive
GPMT	guinea pig maximization test
h-CLAT	human cell-line activation test
HPLC	high performance liquid chromatography
IATA	integrated testing approaches and assessments

JRC EURL ECVAM	European Union Reference Laboratory for Alternatives to Animal Testing hosted by Joint Research Center
LLNA	local lymph node assay
Lys	lysine
MA	Michael acceptor
MCI/MI	mixture of methylchloroisothiazolinone/methylisothiazolinone
MIE	molecular initiating event
NA	nucleophilic addition
NC	not classified
NPV	negative predictive value
Nrf-2	Nuclear factor (erythroid-derived 2)-like 2
OECD	Organization for Economic Co-operation and Development
OECD TG	OECD test guideline
PPV	positive predictive value
QP	quinone precursor
(Q)SAR	(quantitative) structure activity relationship
REACH	Registration, Evaluation, Authorization and Restriction of Chemicals
SB	Schiff base former
Sens	sensitivity
S _N 1/2	chemicals reacting by nucleophilic substitutions of type 1 or 2
S _N Ar	Aromatics reacting by nucleophilic substitutions
Spec	specificity
TIMES SS	tissue metabolism simulator for the endpoint skin sensitization
TN	true negative
TP	true positive
WoE	weight of evidence

Transparency document

The [Transparency document](#) associated with this article can be found, in the online version.

References

- Ahlfors, S.R., Sterner, O., Hansson, C., 2003. Reactivity of contact allergenic haptens to amino acid residues in a model carrier peptide, and characterization of formed peptide-hapten adducts. *Skin Pharmacol. Physiol.* 16, 59–68.
- Aptula, A.O., Roberts, D.W., 2006. Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity. *Chem. Res. Toxicol.* 19, 1097–1105.
- 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.
- Basketter, D., Ashikaga, T., Casati, S., Hubsch, B., Jaworska, J., de, K.J., Landsiedel, R., Manou, I., Mehling, A., Petersohn, D., Rorije, E., Rossi, L., Steiling, W., Teissier, S., Worth, A., 2015. Alternatives for skin sensitisation: hazard identification and potency categorisation: report from an EPAA/CEFC LRI/Cosmetics Europe cross sector workshop, ECHA Helsinki, April 23rd and 24th 2015. *Regul. Toxicol. Pharmacol.* 73, 660–666.
- 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.
- Cooper, J.A., Saracci, R., Cole, P., 1979. Describing the validity of carcinogen screening-tests. *Br. J. Cancer* 39, 87–89.
- Dimitrov, S.D., Low, L.K., Patlewicz, G.Y., Kern, P.S., Dimitrova, G.D., Comber, M.H.I., Phillips, R.D., Niemela, J., Bailey, P.T., Mekenyan, O.G., 2005. Skin sensitization: modeling based on skin metabolism simulation and formation of protein conjugates. *Int. J. Toxicol.* 24, 189–204.
- Dupuis, G., Benezra, C., 1982. *Allergic Contact Dermatitis to Simple Chemicals: A Molecular Approach*, 200pp. Marcel Dekker, New York, NY, USA & Basel, Switzerland.
- 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., Cronin, M.T., 2009. Electrophilic reaction chemistry of low molecular weight respiratory sensitizers. *Chem. Res. Toxicol.* 22, 1447–1453.
- EURL-ECVAM, 2015. EURL ECVAM Recommendation on the Human Cell Line Activation Test (H-CLAT) for Skin Sensitisation Testing.
- Gerberick, F.M., Basketter, D., Casati, S., Karlberg A. T., Kern, P., Kimber, I., Lepoittevin J. P., Natsch, A., Ovigne J. M., Rovida, C., Sakaguchi, H., Schultz, T., 2008. Chemical Reactivity Measurement and the Predictive Identification of Skin Sensitisers. *Atla-Alternatives to Laboratory Animals Vol. 36 pp.* 215–242.
- Gerberick, G.F., Troutman, J.A., Foertsch, L.M., Vassallo, J.D., Quijano, M., Dobson, R.L.M., 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.

- 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., 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.
- Hirota, M., Fukui, S., Okamoto, K., Kurotani, S., Imai, N., Fujishiro, M., Kyotani, D., Kato, Y., Kasahara, T., Fujita, M., Toyoda, A., Sekiya, D., Watanabe, S., Seto, H., Takenouchi, O., Ashikaga, T., Miyazawa, M., 2015. Evaluation of combinations of in vitro sensitization test descriptors for the artificial neural network-based risk assessment model of skin sensitization. *J. Appl. Toxicol.* 35, 1333–1347.
- Matheson, J., 2015. ICCVAM integrated decision strategy for skin sensitization. <http://ntp.niehs.nih.gov/iccvam/meetings/sot15/matheson-poster-text-508.pdf>.
- Jaworska, J.S., Natsch, A., Ryan, C., Strickland, J., Ashikaga, T., Miyazawa, M., 2015. Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: a decision support system for quantitative weight of evidence and adaptive testing strategy. *Arch. Toxicol.* 89, 2355–2383.
- Joint Cefic LRI/Cosmetics Europe/EPAA workshop HHA2, 2015. <http://cefic-lri.org/events/joint-cross-sector-workshop-on-alternatives-for-skin-sensitization-testing-and-assessment/>.
- Landsteiner, K., Jacobs, J., 1936. Studies on the sensitization of animals with simple chemical compounds. *J. Exp. Med.* 64, 625–639.
- Lepoittevin, J., Basketter, D.A., Goossens, A., Karlberg, A.T., 1998. Allergic contact dermatitis: The molecular basis. Springer, Berlin.
- Lepoittevin, J.P., 2006. Metabolism versus chemical transformation or pro- versus pre-haptens? *Contact Dermatitis* 54, 73–74.
- Lepoittevin, J., Benezra, C., 1992. C-13-enriched methyl alkanesulfonates – new lipophilic methylating agents for the identification of nucleophilic amino acids of proteins by NMR. *Tetrahedron Lett.* 33, 3875–3878.
- Macmillan, D.S., Canipa, S.J., Chilton, M.L., Williams, R.V., Barber, C.G., 2016. Predicting skin sensitisation using a decision tree integrated testing strategy with an in silico model and in chemico/in vitro assays. *Regul. Toxicol. Pharmacol.* 76, 30–38.
- Natsch, A., Emter, R., Gfeller, H., Haupt, T., Ellis, G., 2015. Predicting skin sensitizer potency based on in vitro data from KeratinoSens and kinetic peptide binding: global versus domain-based assessment. *Toxicol. Sci.* 143, 319–332.
- Natsch, A., Gfeller, H., 2008. LC-MS-based characterization of the peptide reactivity of chemicals to improve the in vitro prediction of the skin sensitization potential. *Toxicol. Sci.* 106, 464–478.
- Natsch, A., Gfeller, H., Haupt, T., Brunner, G., 2012. Chemical reactivity and skin sensitization potential for benzaldehydes: can Schiff base formation explain everything? *Regul. Toxicol. Pharmacol.* 25, 2203–2215.
- Natsch, A., Gfeller, H., Rothaupt, M., Ellis, G., 2007. Utility and limitations of a peptide reactivity assay to predict fragrance allergens in vitro. *Toxicol. in Vitro* 21, 1220–1226.
- 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.* 33, 1337–1352.
- OECD, 2012a. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Evidence.
- OECD, 2012b. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 2: Use of the AOP to Develop Chemical Categories and Integrated Assessment and Testing Approaches.
- OECD, 2015a. OECD TG 442C: *In chemico* Skin Sensitization: Direct Peptide Reactivity Assay (DPRA).
- OECD, 2015b. OECD TG 442D: *In Vitro* Skin Sensitization: ARE-Nrf2 Luciferase Test Methods.
- Patlewicz, G., Dimitrov, S.D., Low, L.K., Kern, P.S., Dimitrova, G.D., Comber, M.I., Aptula, A.O., Phillips, R.D., Niemela, J., Madsen, C., Wedebye, E.B., Roberts, D.W., Bailey, P.T., Mekenyan, O.G., 2007. TIMES-SS – a promising tool for the assessment of skin sensitization hazard. A characterization with respect to the OECD validation principles for (Q)SARs and an external evaluation for predictivity. *Regul. Toxicol. Pharmacol.* 48, 225–239.
- Patlewicz, G., Kuseva, C., Kesova, A., Popova, I., Zhechev, T., Pavlov, T., Roberts, D.W., Mekenyan, O., 2014a. Towards AOP application—implementation of an integrated approach to testing and assessment (IATA) into a pipeline tool for skin sensitization. *Regul. Toxicol. Pharmacol.* 69, 529–545.
- Patlewicz, G., Kuseva, C., Mehmed, A., Popova, Y., Dimitrova, G., Ellis, G., Hunziker, R., Kern, P., Low, L., Ringeissen, S., Roberts, D.W., Mekenyan, O., 2014b. TIMES-SS—recent refinements resulting from an industrial skin sensitisation consortium. *SAR QSAR Environ. Res.* 25, 367–391.
- Ramirez, T., Mehling, A., Kolle, S.N., Wruck, C.J., Teubner, W., Eltze, T., Aumann, A., Urbisch, D., van, R.B., Landsiedel, R., 2014. LuSens: a keratinocyte based ARE reporter gene assay for use in integrated testing strategies for skin sensitization hazard identification. *Toxicol. in Vitro* 28, 1482–1497.
- Raunio, H., 2011. In silico toxicology – non-testing methods. *Front. Pharmacol.* 2, 33.
- Roberts, D.W., Aptula, A.O., Patlewicz, G., 2006. Mechanistic applicability domains for non-animal based prediction of toxicological endpoints. QSAR analysis of the Schiff base applicability domain for skin sensitization. *Chem. Res. Toxicol.* 19, 1228–1233.
- Roberts, D.W., Natsch, A., 2009. High throughput kinetic profiling approach for covalent binding to peptides: application to skin sensitization potency of Michael acceptor electrophiles. *Chem. Res. Toxicol.* 22, 592–603.
- Roberts, D.W., Patlewicz, G., Kern, P.S., Gerberick, F., Kimber, I., Dearman, R.J., Ryan, C.A., Basketter, D.A., Aptula, A.O., 2007. Mechanistic applicability domain classification of a local lymph node assay dataset for skin sensitization. *Chem. Res. Toxicol.* 20, 1019–1030.
- Romagnoli, P., Labhardt, A.M., Sinigaglia, F., 1991. Selective interaction of Ni with an MHC-bound peptide. *EMBO J.* 10, 1303–1306.
- Teissier, Silvia, Alépée, Nathalie, 2015. Case study 5: L'OREAL approach & decision strategy. <http://cefic-lri.org/wp-content/uploads/2014/03/G-Case-Study-3-LOreal-April-2015.pdf>.
- Takenouchi, O., Fukui, S., Okamoto, K., Kurotani, S., Imai, N., Fujishiro, M., Kyotani, D., Kato, Y., Kasahara, T., Fujita, M., Toyoda, A., Sekiya, D., Watanabe, S., Seto, H., Hirota, M., Ashikaga, T., Miyazawa, M., 2015. Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin sensitizing potential and potency of chemicals. *J. Appl. Toxicol.* 35, 1318–1332.
- Teubner, W., Mehling, A., Schuster, P.X., Guth, K., Worth, A., Burton, J., van, R.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.
- Troutman, J.A., Foertsch, L., Kern, P.S., Dai, H.J., Quijano, M., Dobson, R.L.M., Lalko, J.F., Lepoittevin, J.P., Gerberick, G.F., 2011. The incorporation of lysine into the peroxide-peptide reactivity assay for skin sensitization assessment. *Toxicol. Sci.* 122, 422–436.
- Urbisch, D., Mehling, A., Guth, K., Ramirez, T., Honarvar, N., Kolle, S., Landsiedel, R., Jaworska, J., Kern, P.S., Gerberick, F., Natsch, A., Emter, R., Ashikaga, T., Miyazawa, M., Sakaguchi, H., 2015. Assessing skin sensitization hazard in mice and men using non-animal test methods. *Regul. Toxicol. Pharmacol.* 71, 337–351.
- Urbisch, D., Becker, M., Honarvar, N., Kolle, S.N., Mehling, A., Teubner, W., Wareing, B., Landsiedel, R., 2016. Assessment of pre- and pro-haptens using non-animal test methods for skin sensitization. *Chem. Res. Toxicol.* (submitted).
- van der Veen, J.W., Rorije, E., Emter, R., Natsch, A., van, L. H., Ezendam, J., 2014. Evaluating the performance of integrated approaches for hazard identification of skin sensitizing chemicals. *Regul. Toxicol. Pharmacol.* 69, 371–379.