Replacement of animal testing: Evaluation of non-animal methods for assessing skin sensitization

Inaugural-Dissertation

Zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Daniel Urbisch

aus Bielefeld

Düsseldorf, Juli 2017

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Ellen Fritsche

2. Prof. Dr. Vlada B. Urlacher

Tag der mündlichen Prüfung: 09.10.2017

Danksagung

Da ist sie nun, die fertige Dissertation als Ergebnis einer ereignisreichen Zeit bei der BASF SE in Ludwigshafen in Kooperation mit der Heinrich-Heine-Universität in Düsseldorf. Diese Zeit hat nicht nur diese Arbeit als Ergebnis, sondern hat mich auch in meiner Persönlichkeit weitergebracht und mir einige Türen geöffnet. Daran waren viele Menschen beteiligt, denen ich meinen aufrichtigen Dank aussprechen möchte.

An erster Stelle möchte ich Robert Landsiedel danken, der mir die Bearbeitung dieses Projektes erst ermöglicht hat. In den vergangenen Jahren hat er das Projekt mit seinen Entscheidungen stets in die richtige Richtung gelenkt und nur seinem Verhandlungsgeschick ist es beispielsweise zu verdanken, dass das gemeinsame Projekt mit anderen Industrievertretern, welches die Grundlage dieser Dissertation darstellt, zustande gekommen ist. Vor allem möchte ich ihm aber dafür danken, dass er sich immer die Zeit für ein Gespräch genommen hat, auch wenn ich mal unangekündigt vor seiner Bürotür stand.

Weiterhin möchte ich Naveed Honarvar danken. Durch die wöchentlichen Projektbesprechungen mit Ihm habe ich eine hervorragende Betreuung genossen. Naveed hat mich in unseren lebhaften Diskussionen nicht nur fachlich enorm weitergebracht, sondern war auch immer eine große Hilfe, wenn es um persönliche Anliegen oder Ratschläge für diverse Lebenssituationen ging. Auch möchte ich ihm dafür danken, dass er mir durch seine Erfahrung und Ratschläge den Einstieg ins Berufsleben als regulatorischer Toxikologe enorm erleichtert hat.

Mein weiterer Dank gilt Frau Prof. Dr. Ellen Fritsche - meiner Doktormutter - für die Betreuung dieser Arbeit. Sie gab mir die Möglichkeit, mein Projekt in Ihrem Arbeitskreis zu diskutieren und hat mir zudem geholfen, mich bei der Vorstellung des Projektes auf das Wesentliche zu fokussieren. Die fachlichen Gespräche werden mir als bereichernder und konstruktiver Austausch in Erinnerung bleiben.

Ebenso danke ich Frau Prof. Dr. Vlada Urlacher für die Ihre Hilfsbereitschaft und die wissenschaftliche Betreuung als Zweitgutachterin. Auch Sie gab mir die Möglichkeit, mein Projekt mit Ihrem Arbeitskreis zu diskutieren und hat mich bei der strategischen Gestaltung der Dissertation sehr unterstützt.

Darüber hinaus gilt mein Dank allen Kollegen – allen voran Susanne, Tzutzuy, Wera, Britta, Juliane, Nadine (um nur einige zu nennen) - die mich in der täglichen Arbeit begleitet und unterstützt haben. Weiterhin danke ich meiner Mutter Monika, meinem Bruder Andrè sowie Freunden und Bekannten, die mir stets ein Rückhalt vor allem in schwierigen Phasen waren.

Nachdem ich die Abschlussprüfung des zweiten Semesters im Chemiegrundstudium (deutlich) nicht bestanden habe, hatte ich ein sehr offenes Gespräch mit Adelheit Godt, Professorin für organische Chemie an der Universität Bielefeld. In diesem Gespräch sagte Sie den folgenden Satz, der meine Einstellung und den weiteren Verlauf meines Lebens verändert hat:

”Sie haben zwei Optionen: Entweder, Sie bestehen ab jetzt jede Klausur mit mindestens „gut“, oder Sie suchen sich schnell etwas Anderes!“

Ich hatte mich für die erste Option entschieden.

Contents

1.1 Summary

To estimate the hazard potential of chemical substances to man, animal testing was routinely performed over decades. However, ethical concerns and legal requirements led to an increasing interest in non-animal test methods and progressed their development and validation. For example, non-animal methods for the assessment of local toxicity endpoints like genotoxicity or skin and eye irritation have recently been established. For more complex endpoints, however, animal testing is still required to date; here, skin sensitization is a vanguard for the toxicological assessment solely based on non-animal methods.

Allergic contact dermatitis (allergy of type IV) is the clinical manifestation of skin sensitization and provides the most prevalent form of immune toxicity in humans. Typical symptoms include itchiness, redness of skin and the appearance of blisters, which are triggered by considerable exposure to allergens like Nickel (e.g. in jewelry) and certain fragrances or dyes (e.g. in cosmetics). Such allergens were usually identified using animal testing, which is even nowadays still required by some legal bodies.

Meanwhile, non-animal methods were established, which may completely replace animal testing. The goal of this thesis was to assess the utility of selected alternative methods for identifying allergens and to uncover their strengths and limitations. For this purpose, a dataset compiling experimental non-animal data on more than 200 substances was created, which were compared to animal and human data from literature. It was shown that the single non-animal methods could identify human sensitizers and non-sensitizers with a comparable predictivity like animal tests.

One of the non-animal methods addresses the first step of the mechanism underlying skin sensitization (the "molecular initiating event"): the binding of an allergen to a dermal protein, which can be detected experimentally by the depletion of model peptides induced by the test substance. It was investigated, if peptide binding could also be modelled and predicted using computational tools (QSAR), which may be used if the experimental method is not applicable e.g. due to solubility issues.

Skin sensitization is based on a complex mechanism, and none of the single non-animal methods is intended to be used stand-alone. Instead, testing strategies have to be used, which combine several single test methods addressing different parts of the mechanism. Actually, the applied testing strategy compiling results of three non-animal test methods was even more predictive than the animal test itself. It was even possible to identify those substances, which require abiotic or metabolic activation to gain their allergic potential.

If the skin sensitization potential of a substance has been identified, further risk assessments require additional information on its potency. In this work, a peptide-based method to quantify peptide reactivity was established to reliably distinguish weak from strong sensitizers.

This work could prove the utility of non-animal methods to replace animal testing for skin sensitization. Moreover, the presented results could contribute to the regulatory acceptance of the applied non-animal testing strategy in the evaluation of chemicals.

1.2 Zusammenfassung

Um das Gefahrenpotenzial chemischer Substanzen für den Menschen abschätzen zu können, war die Durchführung von Tierversuchen lange Zeit unumgänglich. Ethische Bedenken und gesetzliche Vorgaben konnten jedoch ein immer stärker werdendes Interesse an Alternativmethoden wecken, das sich förderlich auf deren Entwicklung auswirkte. So wurden in den letzten Jahren Methoden etabliert, die beispielsweise das gentoxische, haut- oder augenreizende Potential zuverlässig und ohne den Einsatz von Versuchstieren vorhersagen. Endpunkte, denen komplexere biologische Prozesse zu Grunde liegen, können bis heute allerdings nicht in Gänze mit Alternativmethoden erfasst werden; eine Ausnahme bildet hier der toxikologische Endpunkt Hautsensibilisierung.

Die Hautsensibilisierung liegt dem Krankheitsbild der allergischen Kontaktdermatitis (Typ IV Allergie) zugrunde, welche die am weitesten verbreitete Form der Immuntoxizität beim Menschen darstellt. Die Symptomatik umfasst starken Juckreiz, Hautrötung und Bläschenbildung, welche durch signifikante Exposition mit Allergenen wie Nickel (z.B. in Schmuck) oder bestimmten Duft- oder Farbstoffen (z.B. in Kosmetika) ausgelöst wird. Um derartige Allergene zu identifizieren, wurden routinemäßig Tierversuche durchgeführt, die zum Teil auch heute noch von Behörden gefordert werden.

Mittlerweile wurden jedoch Alternativmethoden etabliert, die den Tierversuch vollständig ersetzen könnten. Im Rahmen dieser Arbeit wurde die Eignung ausgewählter Alternativmethoden zur Allergenidentifikation sowie deren Vor- und Nachteile umfassend analysiert. Dazu wurde ein Datensatz mit Ergebnissen experimenteller Alternativmethoden für über 200 Substanzen zusammengestellt und mit publizierten Tierversuchs- und Humandaten verglichen. Es konnte gezeigt werden, dass die einzelnen zell- oder peptidbasierten Tests die Humandaten ähnlich zuverlässig vorhersagen können wie der Tierversuch.

Eine der Testmethoden prüft den ersten Schritt der Hautsensibilisierung (das „molecular initiating event“): Die Bindung eines Allergens an ein Protein, was experimentell durch den Verbrauch zweier Modellpeptide durch die Testsubstanz gemessen wird. Es wurde geprüft, ob die Proteinbindung auch mit Hilfe zweier Computermodelle (QSAR) vorhergesagt werden kann, die dann den experimentellen Test ersetzen könnten, sollte dieser - beispielsweise mit unlöslichen Substanzen - nicht durchführbar sein.

Da jede Alternativmethode jeweils nur einen Teilschritt des komplexen Mechanismus` einer Hautsensibilisierung abbildet, erscheint es sinnvoll, mehrere Methoden zu kombinieren, um so die entscheidenden Schritte dieses Prozesses abbilden zu können. Tatsächlich war die Vorhersagegenauigkeit eines Prädiktionsmodells, dass auf dem Ergebnis dreier Testmethoden beruht, höher als die des Tierversuchs. Dabei konnten auch solche Substanzen identifiziert werden, die zunächst abiotisch oder metabolisch aktiviert werden müssen, um ihr allergenes Potential zu entfalten.

Eine Risikobewertung sensibilisierender Stoffe ist nur in Kenntnis derer Wirkstärke möglich. Mithilfe eines protein-basierten Tests konnte ein Zusammenhang zwischen der Wirkstärke eines Sensibilisierers und dessen Peptidreaktivität gezeigt und starke von schwachen Sensibilisierern unterschieden werden.

Diese Arbeit belegt die Eignung etablierter Alternativmethoden zur Allergenidentifikation und –bewertung. Die dargestellten Ergebnisse konnten wesentlich dazu beitragen, dass die hier analysierte Teststrategie als Ersatz für Tierversuche in der Bewertung von Chemikalien regulatorisch akzeptiert wurde.

2 Introduction

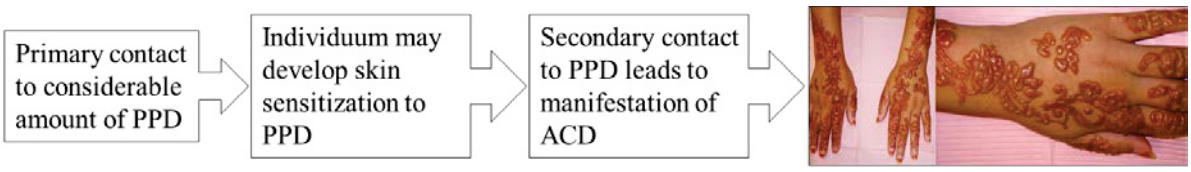
In 1959, the scientists Russel and Burch described the '3R concept' becoming a guiding principle for the reduction, refinement and replacement (3Rs) of animals testing [1]. Animal welfare considerations and progress of toxicological sciences have driven the 3R concept in the past decades. The rapidly increasing scientific knowledge of a variety of cellular processes and responses allowed the development and use of a number non-animal methods. These alternatives are mainly based on tissue – or cell cultures (*in vitro* methods), cell-free systems e.g. based on proteins (*in chemico* methods) or computational programs trained on a wealth of experimental data (*in silico* methods) [2]. However, before a non-animal alternative can be used for regulatory purposes, the predictivity and reliability has to be proven in time-consuming validation processes [3].

Recently, new European regulations are demanding the use of alternatives: animal testing to assess chemicals in the framework of REACH may only be performed as a "last resort", while for the hazard identification of cosmetic ingredients in Europe, animal testing was completely banned already in 2013 [4;5].

For genotoxicity and local toxicity endpoints (i.e. skin and eye irritation), single alternatives and combinations thereof are already accepted by legal bodies. However, more complex, systemic endpoints can generally not yet be fully assessed using exclusively non-animal methods [6]. Skin sensitization is the first complex toxicological endpoint, which can now be assessed solely based on non-animal methods.

2.1 Skin sensitization and allergic contact dermatitis (ACD)

Skin sensitization is the prerequisite for the development of the disease allergic contact dermatitis (ACD) – also known as type IV allergy. Once established, skin sensitization is a chronic and lifelong, but symptomless state of the acquired immunity. The sensitized individuals have to prevent the manifestation of ACD by avoiding considerable exposure to the specific allergen. Otherwise, parts of the acquired immune systems overreact and elicit symptoms of different severity like rash, inflammation and desquamation of the skin as well as the appearance of blisters (**Picture 2.1**). These symptoms affect the quality of work and social life [7].



Picture 2.1: Illustration of the development of ACD after considerable exposure to the skin sensitizers *para*-phenylene diamine (PPD). PPD is often used in illegally high concentrations as dye in Henna tattoos [8]. *The picture was reproduced with permission, Copyright Massachusetts Medical Society.*

Based on epidemiological data, ACD is the most prevalent form of immune toxicity with about 15 – 20% of the western population being sensitized at some point in the course of their lives to at least one allergen [9-11]. Risk factors for developing ACD may be inherent like individual predisposition, immune status, age and gender [12], or attributable like repeated exposure to allergens at work and frequent use of consumer products containing

excessive amounts of allergens. Most frequently, individuals develop ACDs to fragrances, preservatives, dyes and specific metal ions with nickel allergy being the most prominent form of ACD [9;13].

2.2 Mechanisms underlying ACD

The development of ACD can be subdivided into a sensitization phase (also called induction phase) and an elicitation phase.

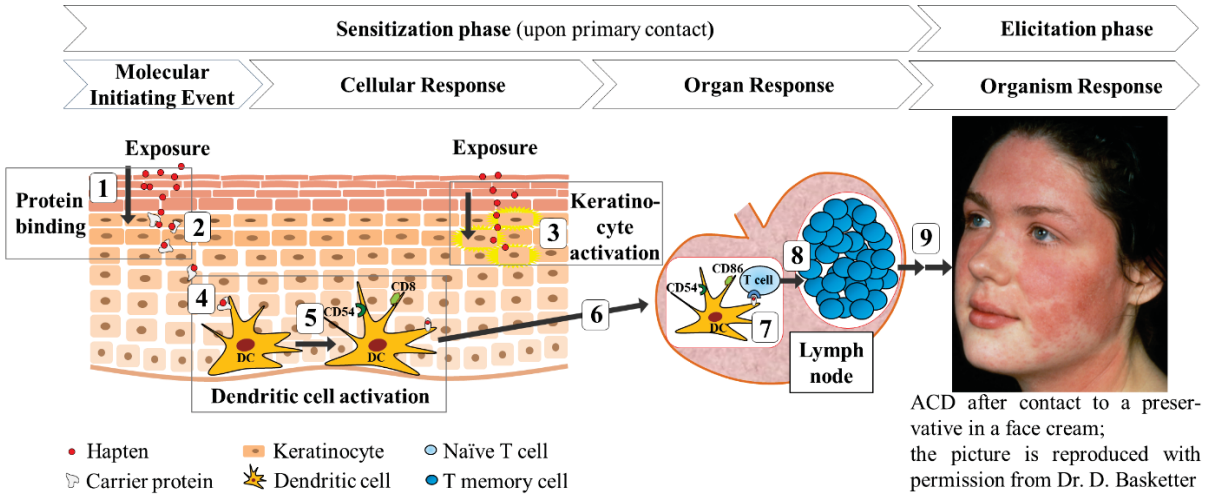
The induction phase is initiated upon the primary contact to specific allergens. Such allergens are typically small electrophilic molecules with low molecular weights - the so-called haptens [14-16]. Significant amounts of these haptens may penetrate the stratum corneum, which is the outermost layer of human skin representing a physical penetration barrier, and permeate into the living parts of the epidermis [17]. Components of the innate immune system in the skin (e.g. Langerhans cells, which are dendritic cells in the skin) may recognize small haptens, but fail to trigger defense mechanisms [18]. However, in a process known as haptenization [19-21], nucleophilic amino acid residues of dermal proteins (e.g. primary amine or sulfhydryl groups) may covalently bind electrophilic moieties of haptens to form hapten-protein complexes [22;23]. The 3D structure of the proteins in the resulting hapten-protein complexes is changed. This ultimate antigen triggers the inherent immune response as well as the following events inducing ACD [24].

Another event is the activation of keratinocytes (KC), which represent the main cell type of the epidermis. KCs interact with the immunogenic complex resulting in their release of cytokines like tumor necrosis factor α (TNF-α) and interleukin (IL)-1α, which act as a kind of ‘danger signals’. Amongst other functions, these signals may attract immature dendritic cells (DCs; e.g. Langerhans cells [18]). In addition to the release of danger signals, cyto-protective genes like (NADPH) quinone oxidoreductase (NQO1), Aldehyde dehydrogenase (ALDH3A) or thioredoxin (TXN) are expressed by activated KCs [25].

Attracted DCs scan the epidermis for foreign structures and may take up hapten-protein complexes by endocytosis, which initiates their activation and subsequent maturation [26-29]. Upon activation, DCs degrade the hapten-protein complex to a small peptide bearing the hapten, which is presented as antigen on the dendritic cell surface [30]. Mature DCs express various proteins such as CD40, CD54, CD80 and CD86 on their cell surfaces, which facilitate the following steps in the sensitization phase [21;29;31;32]. CD54 is an adhesion molecule (ICAM-1), which is associated with the migration of DCs to draining lymph nodes. In the lymph node, DCs present the antigen to naïve T cells (i.e. CD4+ T helper cells). If a T cell carries a cell-surface receptor matching and binding the antigen, the co-stimulatory cell surface molecule CD86 binds to its counterpart CD28 on the T cell [33-37]. This leads to the activation of the naïve T cell, which then proliferates by clonal expansion. The majority of the clones subsequently differentiates to a hapten-specific T memory cells [21;26].

The last step in the induction phase is the distribution of hapten-specific T memory cells in the blood stream, leading to sensitization of an individual to the specific hapten, usually for lifetime. The induction phase starting with the primary exposure to the haptens may require several days until hapten-specific T memory cells have developed and is usually free of any symptoms [18;27;31;38;39].

When it comes to secondary exposure to the same or structurally similar haptens (in terms of cross-allergy [40;41]), the elicitation phase may be initiated, in which a fast and intensive response occurs. In this phase, hapten-specific T memory cells are recruited into the exposed tissue and interaction of KCs, DCs and T memory cells trigger inflammatory reactions leading to the symptoms as illustrated in **Picture 2.1** and in **Figure 2.1** [26;42].



ACD after contact to a preservative in a face cream; the picture is reproduced with permission from Dr. D. Basketter

Figure 2.1: Mechanisms underlying ACD can be subdivided into several steps as considerably simplified in this figure. Sensitization includes dermal penetration (**1**), formation of an immunogenic hapten-protein complex (**2**), activation of keratinocytes (**3**) to release danger signals to attract dendritic cells (DCs), which take up and process the hapten-protein complex to an antigen (**4**) leading to their maturation and expression of cell surface markers (**5**). After migration to the next draining lymph node (**6**), DCs present the antigen to naïve T cells (**7**), which get activated, proliferate and differentiate to hapten-specific T memory cells (**8**). In the elicitation phase, distributed T memory cells are recruited into site of exposure to trigger inflammatory responses (**9**). These steps are explained in the context of an 'adverse outcome pathway' (AOP) in section 2.6.1.

2.3 Type IV allergens

2.3.1 Haptens

As already described in **Section 2.2**, haptens are represented by small molecules, which can covalently bind nucleophilic amino acid residues of dermal proteins leading to an immunogenic status [22;23].

Although haptens show a wide chemical diversity, they can be grouped into mechanistic groups according to a specific chemical mechanism by which they bind to proteins (i.e. Michael acceptors, acylating agents or substances reacting in nucleophilic substitutions (i.e. S_N2, S_NAr, S_N1)) or by specific structural features (i.e. reactive carbonyls (also often called Schiff' base formers) and quinone precursors) (see **Figure 2.2**).

Reactive carbonyls are typically mono-aldehydes, 1,2- or 1,3-dicarbonyl compounds [43-45]. **Michael acceptors** contain α,β-unsaturated esters, ketones or aldehydes [46] which react by chemical additions. **Acylating agents** like organic acid anhydrides transfer an acyl moiety to nucleophiles in chemical additions [47]. Aliphatic or aromatic structures reacting in nucleophilic substitutions typically contain a (partially) positive charge with a leaving group being replaced in a concerted chemical addition-elimination (**S_N2, S_NAr**) or in stepwise reaction

with the intermediate formation of a carbocation (**S_N1**) [47;48]. The major mechanistic groups with their specific organic functional groups and examples are illustrated in **Figure 2.2**.

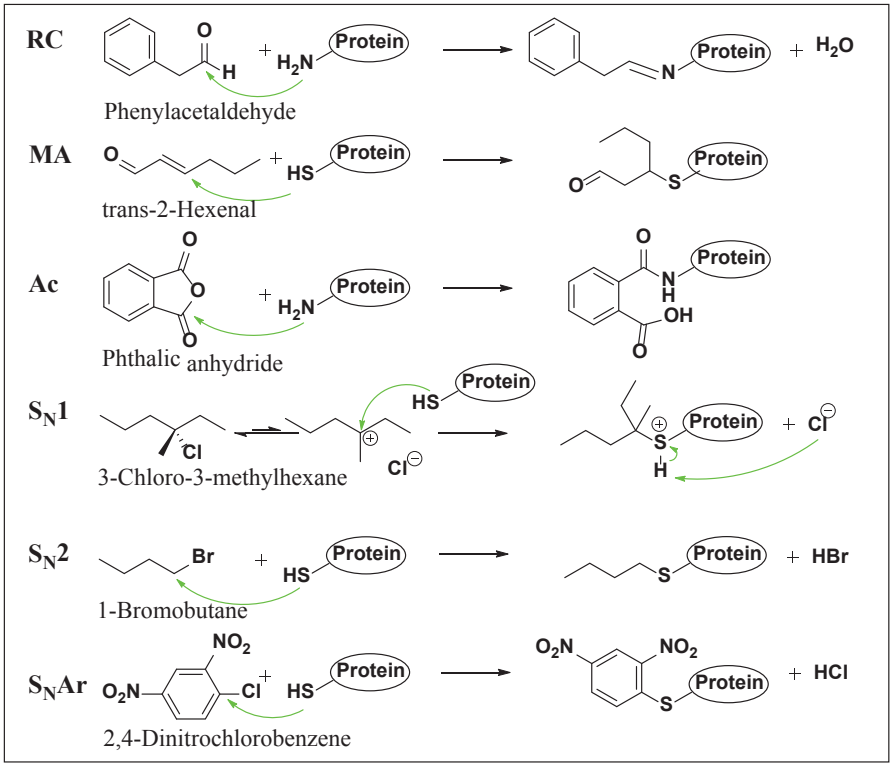


Figure 2.2: Examples for substances in the main mechanistic groups of reactive carbonyls (RC), Michael acceptors (MA), acylating agents (Ac), agents reacting in nucleophilic substitutions of type 1 (S_N1) and 2 (S_N2) or nucleophilic aromatic substitutions (S_NAr) and their resulting hapten-protein complexes. Further mechanistic groups may be defined e.g. based on radical reactions or the formation of coordination bonds [24;49].

2.3.2 Pre- and pro-haptens

A considerable number of skin sensitizers does not contain the above described specific functional groups in their parent structures, thus being intrinsically peptide non-reactive [50]. This implies that a certain form of activation is required to transform parts of the molecular structure to electrophilic moieties capable of protein binding. Depending on the nature of activation, two terms were defined for these specific types of sensitizing substances [24;51]:

i. Pro-haptens require non-metabolic activation such as autoxidation upon contact to air or dissolved oxygen to become electrophilic and thus reactive for nucleophilic protein residues.

ii. Pre-haptens require modification by metabolizing enzymes located in the skin to become reactive targets for nucleophilic protein residues.

Pre- and pro-hapten activation both results in the formation of haptens capable of binding cutaneous proteins (**Figure 2.3**). In many cases, substances can become peptide reactive as a result of both, metabolic as well as

non-metabolic activation, making it difficult to clearly assign a substance either to the class of pre-haptens or pro-haptens [52].

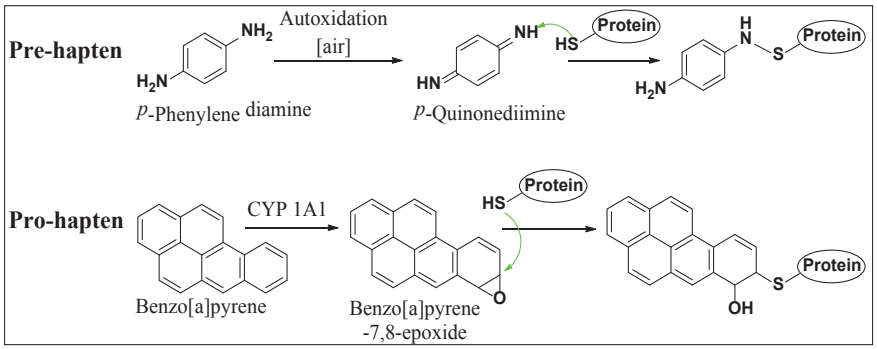


Figure 2.3: Illustration of non-metabolic activation of the pre-haptens *p*-phenylene diamine and metabolic activation of the pro-hapten benzo[a]pyrene with their resulting hapten-protein complexes.

2.4 Regulatory demands

As defined by the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN-GHS/CLP), a skin sensitizer refers to a substance that will lead to an allergic response following skin contact [53]. The sensitizing properties of almost all marketed or imported substances have to be assessed (hazard identification) and risk assessments and management measures have to be applied to protect the population from the above-described disease. In the regulatory context, however, huge differences in the hazard identification occur in the several industrial sectors of chemicals, cosmetics and agrochemicals.

In the industrial sector of **chemicals**, Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) is a European Union regulation entered into force in June 2007, with a phased implementation over the next decade [4]. The European Chemicals Agency (ECHA) has set three deadlines for registration of chemicals, which divides the period of assessing their toxicological properties (like skin sensitization) into three phases determined by tonnage manufactured or imported. For the REACH phases 1 and 2 (phase-in substances with an annual production volume of >1000 or >100 t/a, respectively), the skin sensitizing properties of approximately 3700 and 3000 chemicals, respectively, were assessed [54;55]. It is estimated that several thousand chemicals more will need to be evaluated during phase 3 (phase-in substances with an annual production volume of >1 t/a) with the deadline for registration submissions being in 2018. While animal testing has been the 'first-choice' for phases 1 and 2, the rapid development and validation of a number of non-animal methods led to a paradigm shift for phase 3. Since end of 2016, non-animal methods are defined to be the default methods for assessing skin sensitization, while animal testing can only be performed in exceptional cases with sound reasons [56;57].

In the industrial sector of **cosmetics**, the European Union imposed an animal testing ban on both cosmetic products and their ingredients with the 7th amendment to the Cosmetics Directive (now the Cosmetics Regulation

[58]). This animal testing ban was accompanied by a concurrent marketing ban, if animal tests were conducted after this date for the purpose of this Regulation. This regulation therefore led to a complete phasing out of animal testing in the cosmetics sector in Europe and the development and use of non-animal methods for new cosmetic substances became essential for new cosmetics.

While the increasing public interest in animal welfare and legal demand led to replace animal testing in parts or in general in the sectors of chemicals and cosmetics, respectively, the industrial sector of **agrochemicals** still foresees animal testing as 'first choice' [59].

These animal methods as well as non-animal methods will be described in the following sections.

2.5 Animal test methods

Traditionally, the skin sensitization hazard of a substance has been evaluated using animal tests either using guinea pigs or mice (according to OECD test guideline (TG) 406 or 429, respectively).

In both guinea pig-based tests described in the OECD TG, the animals are exposed repeatedly to mimic the sensitization and elicitation phase, finally resulting in symptoms of ACD in case the test substance is a sensitizer. In the Buehler test, animals are exposed in the sensitization phase to test substance concentrations causing mild to moderate skin irritation. After a resting phase to develop an immune response, the animals are challenged with another, non-irritating test substance concentration. Skin sensitizers typically elicit the above described symptoms, while non-sensitizers do not [60;61].

The main difference of the guinea pig maximization test (GPMT; the second test described in OECD TG 406) compared to the Buehler test is the addition of an adjuvant to the test substance solution to increase the inherent immunogenic response making the animals more susceptible for developing skin sensitization and thus more sensitive [60;62]. Further differences between the Buehler test and GPMT are summarized in **Table 2.1**. The sensitizing hazard of a test substance is evaluated in both tests in a subjective read-out by scoring the symptoms after the elicitation phase; potency can only be evaluated in a limited, semi-quantitative manner mainly considering the number of animals affected [2;60].

The murine LLNA represents another *in vivo* method, which is nowadays still 'first-choice' method for the regulation in certain industrial sectors (e.g. agrochemicals [59]). The test procedure of the LLNA is illustrated in **Figure 2.4**.

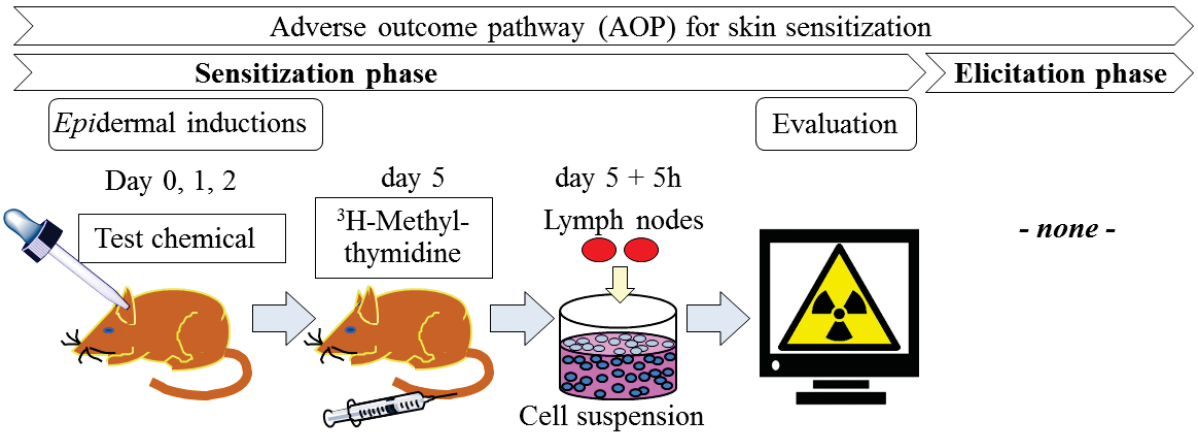


Figure 2.4: In the LLNA, mice are exposed to the dissolved substance by topical application on the dorsal side of each ear on three consecutive days. On day 5, ^{3}H-methyl thymidine is injected into the tail vein and after 5h, mice are killed and the lymph nodes of each ear are excised to prepare cell suspensions. By scintillation, incorporated 3H-methyl thymidine is detected and stimulation indices (SI) are calculated in comparison to control animals indicating lymphocyte proliferation in the lymph nodes. The LLNA protocol uses three different test substance concentrations. If any concentration induced a SI value of $\geqslant 3$, the substance is defined to be a skin sensitizer.

Compared to the guinea pig based tests, the number of animals in the LLNA is slightly reduced and the procedure is refined with respect to animal welfare, since less harm is caused to animals (i.e. no challenge/elicitation phase). In addition, the LLNA provides an objective read-out in a quantitative manner (**Table 2.1**), so that potency of a sensitizer can be calculated based on the concentration, which induced an SI value of 3 [63]. The lower this concentration, the more potent is the test substance.

Table 2.1: Comparison of the guinea pig maximization test (GPMT), Buehler test and local lymph node assay (LLNA).

***In vivo* test [OECD TG]**	**n animals (treated + controls)**	**Dura-tion [days]**	**Endpoint assessment**	**Evaluation of results**	**Thresholds for positive outcome**	**Estimation of potency**
GPMT [406A]	10-20 + 5-10	23	local inflammation	visual; subjective	30% of animals with symptoms	Semi quantitative
Buehler [406B]	20 + 10	34	local inflammation	visual; subjective	15% of animals with symptoms	Semi quantitative
LLNA [429][1]	15 + 5	7	incorporation of ^{3}H-thymidine	scintillation; objective	SI ≥ 3	quantitative

SI = stimulation index; [1] non-radioactive variants of LLNA described in OECD TG 442A & B; n = number. More recently, also a read-out simply measuring cell counts was proposed and termed as LNCC [64].

2.6 Adverse outcome pathway and non-animal methods

In accordance with the guiding principle for the reduction, refinement and replacement (3Rs) of animal testing [1], a number of non-animal methods addressing various biological pathways and toxicological endpoints have been developed in the past decades.

In the field of skin sensitization, the *in vivo* LLNA represents a '2R' method, since the number of required animals (reduction) and harm (refinement) is reduced compared to the guinea-pig based tests (see **Table 2.1**). The overall goal of the '3R concept', however, is a full replacement of animal testing.

The main issue for a rapid replacement of animal testing by non-animal methods is a fundamental difference of the two approaches: while animal tests directly address a given toxicological endpoint (i.e. the adverse outcome as the inflammation in case ACD), the non-animal methods can only address single events of the underlying mechanisms (i.e. (key) events of the adverse outcome pathway), which finally lead to the adverse outcome.

None of the non-animal methods can depict the whole complex and multifaceted mechanism underlying skin sensitization and ACD. Thus, a single non-animal alternative will not be sufficient to replace the LLNA or guinea pig-based tests. Instead, a combination of non-animal methods will be required, which addresses different (key) events of the adverse outcome pathway [65].

In order to make use of mechanistic toxicological information derived by non-animal methods for hazard and potency identification, the conceptual framework of the adverse outcome pathway (AOP) was created. Skin sensitization with ACD as manifested disease is the first adverse outcome for which the OECD has fully described an AOP [66;67]. This AOP is the structured representation of molecular and cellular mechanisms leading to adverse effects and linearly links existing knowledge along a series of causally connected key events [68]. Actually, this AOP can be described as a simplified and more structured version of the mechanisms underlying skin sensitization and ACD (as described in **Section 2.2**). **Figure 2.5** schematically illustrates the key events of the AOP for skin sensitization.

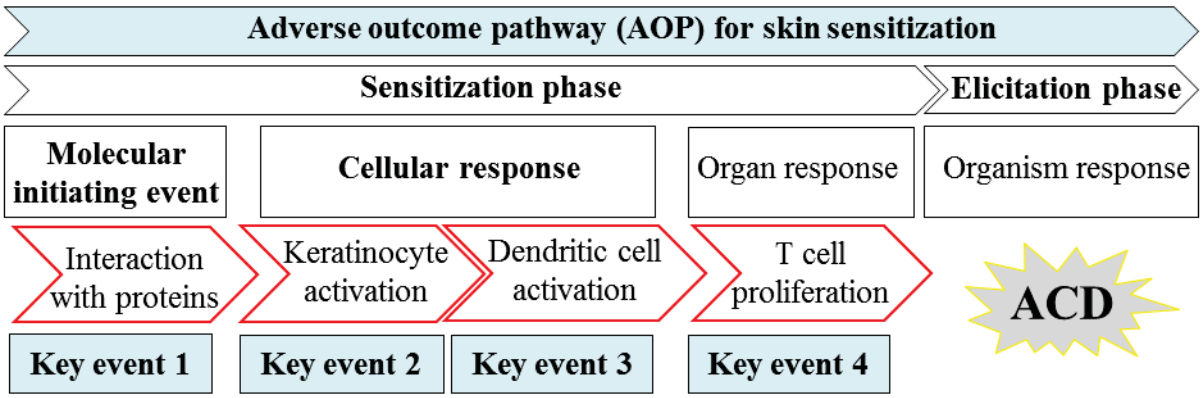


Figure 2.5: Flow diagram of the four key events (red boxes) essential for developing skin sensitization (modified from [67]). The molecular initiating event (MIE) in the AOP for skin sensitization reflects the process of haptenization and represents the first key event. The MIE is followed by the two cellular events 'keratinocyte activation' and 'dendritic cell activation', representing key events 2 and 3, respectively. The first three key events may subsequently lead to an organ response (i.e. T-cell proliferation in a draining lymph node - key event 4) and altogether eventually to an adverse outcome in the individual (i.e. ACD).

A number of promising non-animal methods covering single AOP key events for skin sensitization have been developed (e.g. reviewed in [2;69]). The following section introduces the non-animal methods, which can be used to address the four key events of skin sensitization.

2.6.1 Methods to address key event 1 – Interaction with proteins

In silico models

In silico refers to methods performed by a computer. (Q)SAR models are *in silico* methods which use physical and structural properties of a molecule to predict its biological/toxicological (in)activity ((quantitative) structure activity relationships ((Q)SAR). These models are generally trained on experimental data of a large number of substances [70].

As described in **Section 2.3**, molecular structures of haptens (direct haptens as well as activated pre- and pro-haptens) often contain electrophilic moieties which are capable of binding nucleophilic residues of cutaneous proteins. Several *in silico* models for skin sensitization include algorithms, which allow the identification of 'structural alerts' for a molecules ability to bind to proteins by a specified chemical reaction mechanism. For example, the QSAR Toolbox and TIMES SS (Tissue metabolism simulator for skin sensitization) are trained on a number of functional groups associated with peptide reactivity and can assign a target substance to its respective mechanistic group based on the respective structural alert (see **Section 2.3**) [71-73].

Although most allergens contain electrophilic moieties, the converse conclusion cannot be made that all electrophilic substances are automatically sensitizers, since further steps along the AOP are needed [70]. Furthermore, metabolic detoxification may occur or the substances cannot reach the target proteins due to high vapor pressure, insolubility or lacking dermal penetration [74].

In addition to the identification of structural alerts, further descriptors considering the molecular weight, lipophilicity/hydrophilicity (expressed as octanol/water partition coefficient; i.e. *log* K_{OW}) and the degree of the structural similarity between a target substance and the molecules of the QSAR model's training set as well as the number and quality of the experimental data underlying the training set can improve the overall prediction of skin sensitization hazard [74;75].

A further benefit is the possibility to simulate the activation of pre-haptens and pro-haptens by specific algorithms provided in the QSAR Toolbox or TIMES SS. TIMES SS for example was trained on more than 400 activation pathways [76]. This function is illustrated in **Figure 2.6** using *p*-phenylene diamine (PPD) as example again.

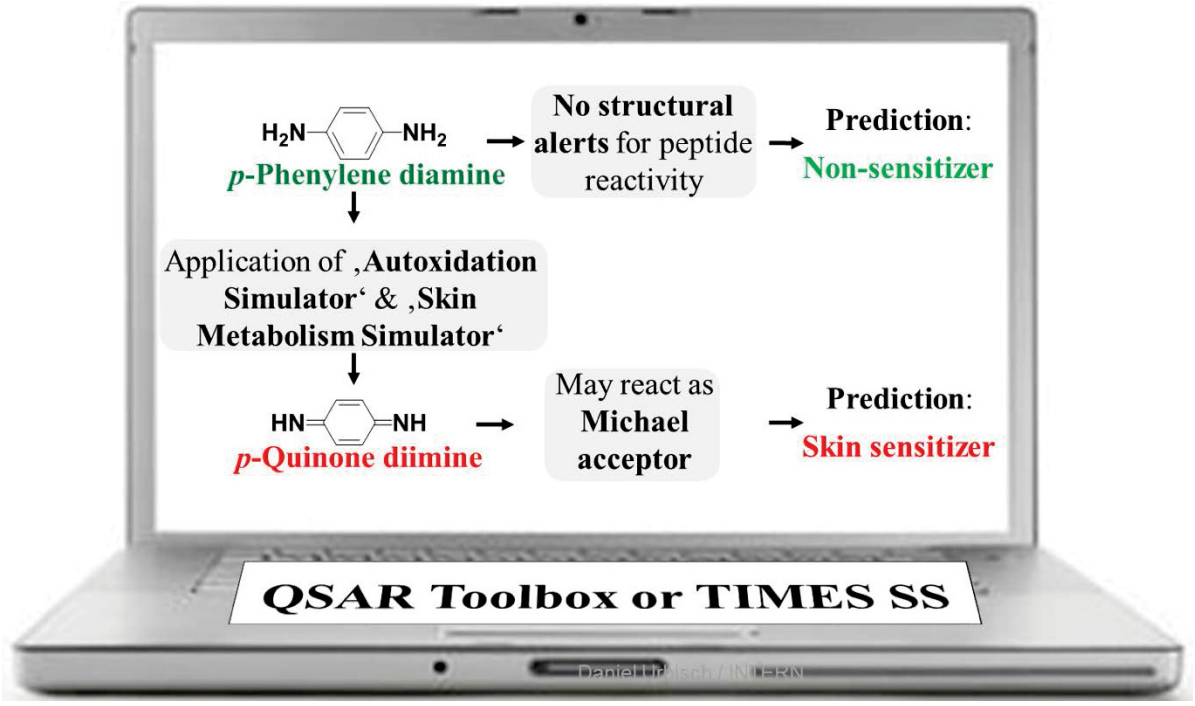


Figure 2.6: Illustration of the information output of the two *in silico* models QSAR Toolbox and TIMES SS. The parent molecular structure of the human sensitizer *p*-phenylene diamine (PPD) does not contain functional groups associated with peptide reactivity. However, PPD can be easily oxidized to *p*-quinone diimine, which can be simulated using the autoxidation simulator and the skin metabolism simulators. *p*-Quinone diimine contains structural alerts for peptide reactivity, is assigned to the mechanistic group of Michael acceptors and is subsequently predicted as skin sensitizer by the models.

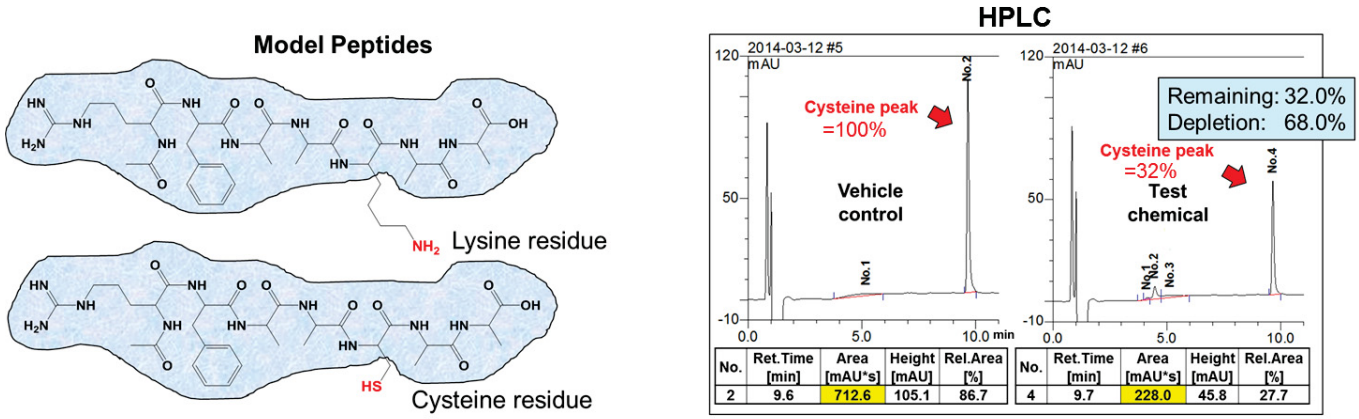
The utility of the two *in silico* models QSAR Toolbox and TIMES SS to predict the molecular initiating event of skin sensitization was evaluated in this work and is described in detail in **Section 4.1**.

***In chemico* methods: direct peptide reactivity assay (DPRA)**

Beside modelling with computational tools, the interaction between a test substance and proteins can also be experimentally detected in the cell-free (= *in chemico*) direct peptide reactivity assay (DPRA) [77;78], in which the protein is mimicked by peptides.

Two different model heptapeptides are used, each with an active amino acid residue containing either a primary amino group (Lys-peptide) or sulfhydryl group (Cys-peptide) as reactive nucleophilic center (**Figure 2.7**) [78]. According to the OECD TG 442C, a test substance is dissolved in a suitable solvent and then added to an aqueous solution containing one of the two peptides. The reaction of the test substance with either peptide is detected after 24 hours of incubation using high pressure liquid chromatography with ultra-violet light absorbance detection (HPLC-UV). Modification of the peptides by the test substance leads to a shift of their respective retention times reducing the area under the curve (AUC) of the unbound peptide. By comparing the AUCs of the treated and the control peptides, the percentage in depletion of the initial peptide is calculated. Based on the amount of peptide

depletion for a number of test substances with known sensitization potential, Gerberick and coworkers proposed a prediction model to assign substance to reactivity classes (see **Table 2.2**) and thus to determine, whether a substance is peptide reactive or not [79]. A mean peptide depletion > 6.38% of the used model peptides (compared to the solvent control) identifies a peptide reactive substance [78]. Gerberick and Coworkers also proposed a prediction model, to sub-classify peptide-reactive substances into weak, moderate and strong peptide binders (**Table 2.2**). This semi quantitative reactivity classification will be further analyzed in **Section 4.4**.



No.	Ret.Time [min]	Area [mAU*s]	Height [mAU]	Rel.Area [%]
2	9.6	712.6	105.1	86.7

No.	Ret.Time [min]	Area [mAU*s]	Height [mAU]	Rel.Area [%]
4	9.7	228.0	45.8	27.7

Figure 2.7: One the left-hand side, the molecular structures of the two heptapeptides is illustrated. The primary amine of the Lys-peptide and the sulfhydryl group of the Cys-peptide are highlighted in red. On the right-hand side, the chromatogram (HPLC-UV) of *para*-phenylene diamine (PPD, right) vs. the solvent control (left) is illustrated. PPD led to a considerable peptide depletion of the Cys-peptide depletion of 68% (right red arrow) compared to vehicle control (set equal to 100%; left red arrow). Since the peptide depletion clearly exceeds the cut-off value, PPD is classified to be peptide reactive accordingly.

Table 2.2: Reactivity classes for peptide binding as defined for the *in chemico* DPRA (OECD 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[1]	$D_{Cys/Lys} \leq 6.38$	$D_{Cys} \leq 13.89$
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 [%]; [1] values in this class are considered negative

The DPRA is also a method that does not require a cell culture laboratory possibly making it more accessible to laboratories. As not all substances can be readily characterized in terms of chemical structure, e.g. substances of unknown or variable composition (UVCBs) such as plant extracts, *in silico* tools are not applicable in all cases (strengths and limitations described in [80]). Since UVCBs were not the focus of this study as new methods are first evaluated using defined substances, testing methods will still be necessary.

2.6.2 Methods to address key event 2 - Keratinocyte activation

***In vitro* methods: KeratinoSens™ and LuSens**

The activation of keratinocytes represents the key event 2 of AOP for skin sensitization (see **Figure 2.5**). This key event can be addressed using the KeratinoSens™ [81] and LuSens [82] assays. The KeratinoSens™ is described in OECD TG 442D [83], while LuSens assay is proposed to be included as a "me-too" method in this OECD TG [84]. Recently, its scientific validity was proofed and confirmed by the EURL ECVAM Scientific Advisory Committee (ESAC) [85] and the inclusion into OECD TG 442D is expected for the end of 2017.

Both *in vitro* assays use keratinocyte lines which were stably transfected with reporter gene constructs containing antioxidant response elements (ARE) of the human aldoketo-reductase gene AKR1C2 (KeratinoSens™) or the rat NADPH:quinone oxidoreductase 1 gene (LuSens), which control the transcription of a downstream gene sequence encoding luciferase (originally obtained from firefly beetle), as illustrated in **Figure 2.8** [81;84]. In the KeratinoSens™ and LuSens assay, the reporter cells are exposed to test substances in various concentrations. After incubation and subsequent addition of oxyluciferin, the intensity of luminescence is detected and quantified. If the intensity of luminescence is significantly increased compared to control cells (> 150 %), the test is positive and the substance classified as KC activator [82;83].

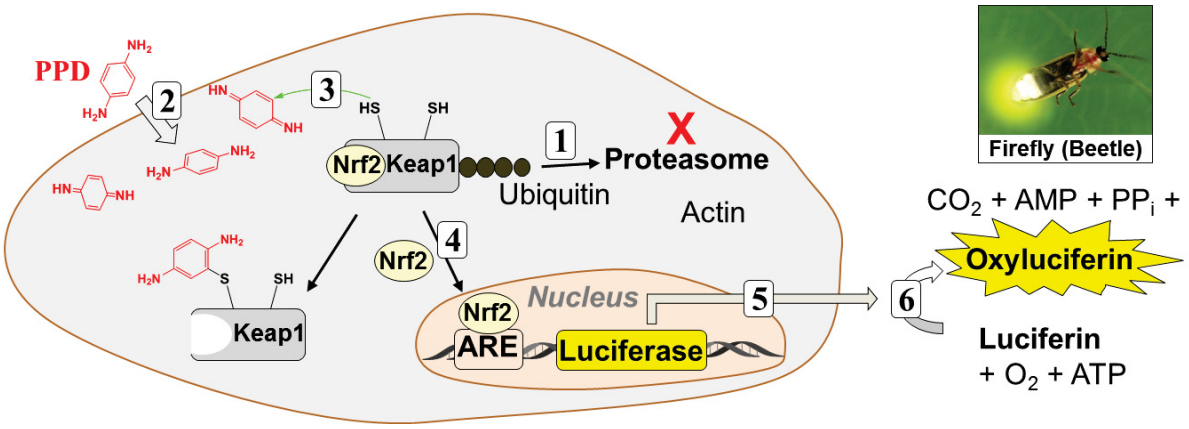


Figure 2.8: Illustration of a keratinocyte in the KeratinoSens™ or LuSens assay. Activation of the keratinocytes is based on the Nuclear-like 2 Kelch-like ECH associated protein-1 (Nrf2-Keap1) pathway, which is a cellular sensor for oxidative and electrophilic stress [86-88]. In unstressed cells, Nrf2 is bound to its inhibitor protein Keap1 and steadily degraded after ubiquitinoylation (**1**). If the cell comes into contact with an electrophilic stressor (**2**) like *p*-quinone diamine (being formed of *p*-phenylene diamine (PPD) under abiotic or biotic conditions), free cysteine residues of Keap1 can bind the electrophiles (**3**). Consequently, conformational changes of Keap1 trigger its dislocation and the release of Nrf2, which translocates into the nucleus (**4**). As a transcription factor, Nrf2 binds to antioxidant response element (ARE) and triggers the expression of luciferase (**5**) in these reporter cells [81;89]. In physiological keratinocytes, Nrf-2 regulates gene expression of Phase 2 enzymes like epoxide hydrolase, quinine reductase 1 and glutathione s transferase counteract electrophilic or oxidative stress [86-88]. Luciferin as a substrate of luciferase is then added to the cells and transformed to detectable oxyluciferin (**6**) [90;91].

The xenobiotic metabolizing enzyme activities in the two keratinocytic cell lines - KeratinoSens™ and LuSens - were recently investigated and compared to enzyme activities in primary keratinocytes. Overall, similarities in enzyme activities were observed for Flavin-containing monooxygenase (FMO), alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH) and N-acetyl transferase-1 (NAT1), while differences occurred in case of cytochrome P450 isoenzymes (CYP) and uridine diphosphate-glucuronosyltransferase (UGT), which were only detectable in primary keratinocytes [92]. In another study, the KeratinoSens™ assay was supplemented with rat liver S9 fractions [93], which contain much higher enzyme activities compared with those reported in skin [94;95]. Only for specific classes of putative pro-haptens being investigated (here: fragrances like creosol or estragole), the sensitivity of the "KeratinoSens-S9" was increased compared to the standard test, while also a decrease in the specificity was reported at the same time [93].

2.6.3 Methods to address key event 3 - Dendritic cell activation

***In vitro* methods: human cell-line activation test (h-CLAT)**

The human cell line activation test (h-CLAT) can be used to address key event 3 of the AOP and was recently adopted as OECD TG 442E [96]. The h-CLAT uses the immortalized cell line THP-1 (monocytes) to reflect DC activation [97-100]. The h-CLAT was intensively studied and found to be of high predictivity in identifying skin sensitizers [101-104]. In the h-CLAT, THP-1 cells are exposed to test substances in various concentrations. After incubation, the upregulation of the cell surface markers CD54 and CD86, which play an important role in the sensitization phase (see **Section 2.2**), is detected and quantified using antibody staining and flow cytometry. If the expression of at least one of the cell-surface markers is significantly increased compared to control cells (> 150 % for CD86 and > 200 % for CD54), the test is positive and the substance classified as DC activator [96].

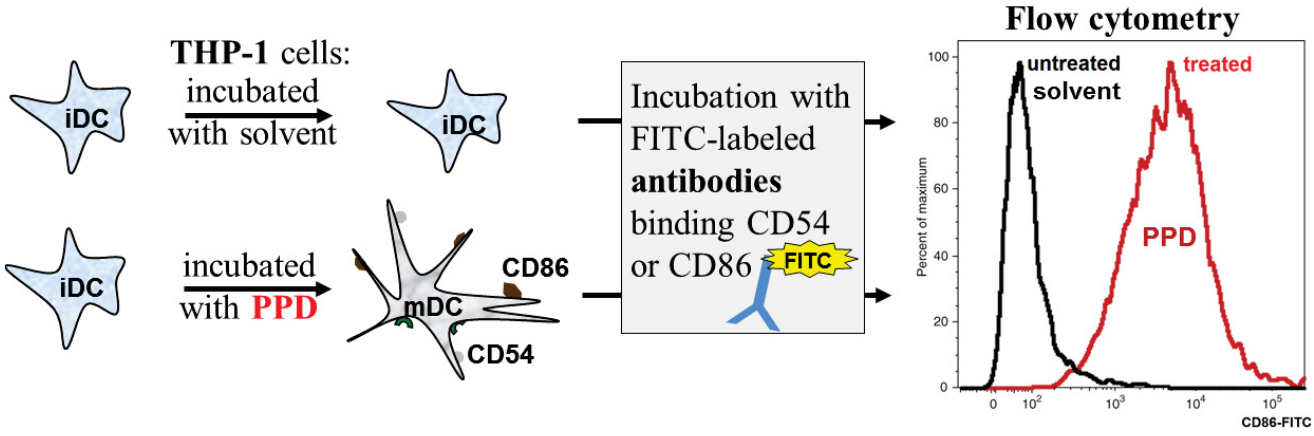


Figure 2.9: Illustration of the activation of THP-1 cells (immature dendritic cells) after incubation with *p*-phenylene diamine (PPD). While incubation with the solvent control does not increase expression of cell-surface markers in THP-1 cells, incubation leads to the maturation and increased expression of CD86. Subsequently added FITC-labeled antibodies can bind expressed CD54 and CD86, what is detected using flow cytometry.

Recently, THP-1 cells were investigated for their xenobiotic metabolizing enzyme activities and thus their capacity, to identify specific pro-haptens [92]. NAT-1, which is involved in the metabolic control (*i.e.* toxification and detoxification) of many sensitization reactions [94;105] was found to be highly active. The same was true for esterase, which is involved in the metabolic transformation of esters and amides being associated with the activation or detoxification of sensitizers [94]. In contrast, activities of several xenobiotic metabolizing enzymes were below the detection limit (i.e. CYP, FMO, ADH, ALDH and UGT activities).

Another DC-based assay, which addresses the third key events of the AOP, is the myeloid U937 skin sensitization test (MUSST). The MUSST uses the U937 cell-line [89;106] and labels upregulation of the cell-surface marker CD86. Since the pre-validation of the MUSST was stopped, the protocol of the MUSST had to be refined. The 'new' assay is called U-Sens and was recently submitted for pre-validation at ECVAM [107].

2.6.4 Methods to address key event 4 - T cell proliferation

Activation of naïve T cells to hapten-specific T memory cells upon contact with antigen-presenting cell (APC) is defined as key event 4 of the AOP and the last step in the sensitization phase [13;49;66]. This key event was recently addressed by a number of approaches, which tried to detect T cell proliferation after treatment with APCs measuring the proliferation of naïve T cells upon allergen treatment [108-112]. However, the available *in vitro* (*ex vivo*) systems are confronted with some drawbacks, as summarized by Maxwell and colleagues [112]:

- Inherent genetic variability in naïve T cells each carrying a specific T cell receptor (TCR) for one antigen
- Low sensitivity of *in vitro* systems for chemical allergens due to high variability
- Primary source of T cells required for each approach (stable cell lines entail lack of TCR variety [113;114])
- Difficulty of co-culturing primary DCs and T cells
- Poor reproducibility of results (e.g. seen in T cell priming assay)

Overall, the known protocols are not sufficiently advanced to analyze high number of substances for regulatory purposes [2;115] and a non-animal alternative capable of generating predictive information on sensitizer-induced T cell proliferation remains a key gap to date [112].

For this reason, non-animal methods addressing the first three of these four key events of the AOP for skin sensitization were analyzed in this work.

2.7 Testing strategies

As described above, single non-animal methods are usually covering only one key event of the adverse outcome pathway (AOP) leading to skin sensitization. To depict a more holistic picture of the AOP, results of several alternatives addressing different key events have to be combined within testing strategies to allow an accurate assessment of a substances' skin sensitization potential and potency [65]. Recently, a set of testing strategies was

proposed, which was termed by OECD as so-called “fixed data interpretation procedure” (DIP) within an “Integrated Approach to Testing and Assessment” (IATA) for safety evaluations of substances [116;117].

Simply expressed, the IATAs are a kind of operation procedures indicating which alternative had to be performed and how their results had to be evaluated to derive an overall outcome (e.g. sensitizer – yes/no). The evaluation is either done by a fixed data interpretation procedure (DIP) or a non-prescribed weight of evidence (WoE) approach or a combination of both [117].

The IATA may use *i.* a pre-defined procedure (a deterministic approach, similar to traditional and regulatory accepted endpoint methods [89;118-120]), *ii.* a case-by-case evaluation [120-122], or *iii.* pre-defined probabilistic models based on regression models [123], Bayesian networks [124;125], artificial neural networks [126-128] or support vectors machines [129;130]. According to the current status, twelve different IATA are available as case studies [131;132].

2.7.1 ‘2 Out of 3’ integrated testing strategy (ITS)

The simplest, yet effective IATA case study for hazard identification is represented by the ‘2 out of 3’ integrated testing strategy (ITS), which includes the results of single assays reflecting the first three key events of the AOP. The underlying prediction model uses *i.* DPRA, *ii.* KeratinoSens™ or LuSens and *iii.* h-CLAT data [118]. Any two congruent results of the three tests rule the overall assessment: If at least two of the three assays are positive, the substance is rated to be a skin sensitizer. If at least two of the three assays are negative, the substance is rated to be a non-sensitizer. The classification as a sensitizer or non-sensitizer is therefore based on a weight of evidence pertaining to the first three key events of the AOP (**Figure 2.10**; described in more detail in **Section 4.1**).

The predictivity of the classification is calculated in comparison to LLNA and human data, respectively [89;133;134]. Human data were mainly derived from *i.* historic prognostic tests like the human maximization test (HMT) [135], *ii.* confirmatory tests for safety evaluations like the human repeated insult patch test (HRIPT) [136], *iii.* diagnostic tests (i.e. contact delayed hypersensitivity allergy test) and *iv.* epidemiologic studies.

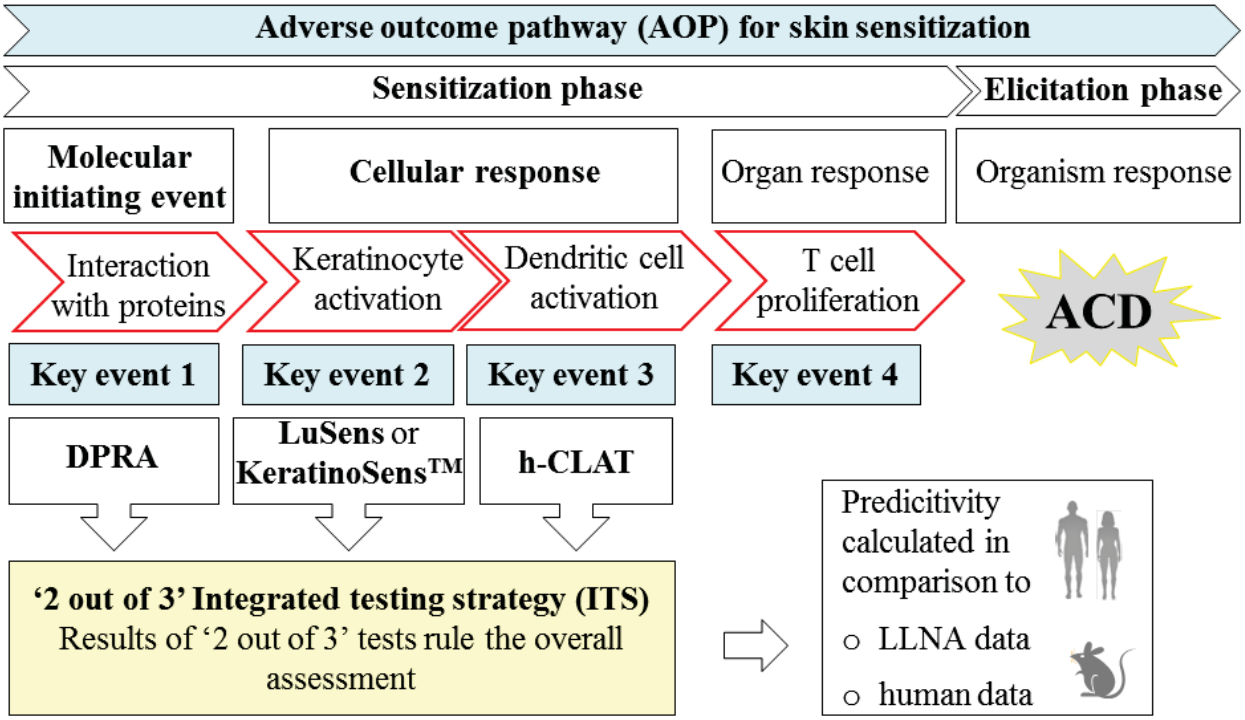


Figure 2.10: Schematic representation of the AOP for skin sensitization with validated non-animal methods addressing one of the first three key events. Results of the alternatives can be compiled in the '2 out of 3' integrated testing strategy (ITS) addressing the first three key events. The predictivity of this testing strategy can be compared to LLNA or human data as reference.

3 Aim of this thesis

Animal welfare considerations and progress of toxicological sciences have driven the replacement of animal testing by non-animal methods. For local toxicity endpoints, such as skin and eye irritation, strategies solely using non-animal methods are already regulatorily accepted. For more complex endpoints, skin sensitization is a vanguard of how to utilize non-animal methods in assessing the potential toxicity of substances.

Skin sensitization as the prerequisite for the development of allergic contact dermatitis is the most prevalent form of immune toxicity. To protect workers and consumers from skin sensitization, identification of skin sensitization potential of new substances is mandatory for hazard assessments; for risk assessments, also information on potency is required. Traditionally, the skin sensitization hazard of a substance has been identified using animal tests with guinea pigs or mice. However, new regulations are demanding the use of alternatives: animal testing to assess chemicals in the framework of REACH can only be performed as a "last resort", while for the hazard identification of cosmetic ingredients in Europe, animal testing was completely banned already in 2013.

This has encouraged the development of non-animal methods to test for skin sensitization, of which three were recently adopted as OECD test methods. In contrast to animal methods, the single non-animal methods are usually covering only one key event of the adverse outcome pathway (AOP) leading to skin sensitization. Therefore, a number of methods, which address different key events of this AOP, has to be combined to allow an accurate assessment of a substances' skin sensitization potential and potency.

The aim of this work was to analyze, how and in which cases existing non-animal methods for skin sensitization can be used to adequately replace animal testing and how such methods could be improved. The following five specific questions guided the research of the thesis:

1. Can the molecular initiating event be addressed using *in chemico* and *in silico* tools? (**Section 4.1**)
2. How predictive are the non-animal methods - either individually or in combination? (**Section 4.2**)
3. How can the applicability domain of these methods be defined? (**Sections 4.1** & **4.2**)
4. Can the individual methods or combinations thereof identify pre- and pro-haptens? (**Section 4.3**)
5. Can these methods evaluate the potency of sensitizers? (**Section 4.4**)

The outcome of this research to address these questions describes a straightforward use of non-animal methods to assess skin sensitization hazard and potency. Results have been published and were even partly considered and cited in the new ECHA guidance on REACH Information Requirements and Chemical Safety Assessment [57].

4 Scientific publications

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

Daniel Urbisch, Naveed Honarvar, Susanne N. Kolle, Annette Mehling, Tzutzuy Ramirez, Wera Teubner, Robert Landsiedel

Summary of the publication: This publication analyses the molecular initiating event (MIE) of skin sensitization using the experimental direct peptide reactivity assay (DPRA) as well as two computational programs. The results confirm the utility of the DPRA and most of the profilers within the QSAR models, but also uncover profilers which are not predictive.

Journal: Toxicology In Vitro (Elsevier)

Impact factor: 2.866 (Clarivate Analytics, 2017)

Type of authorship: first author

Status of publication: published in 2016

My contribution to the publication: approximately 80%

- Performance of the computational analyses for 213 substances using the models 'QSAR Toolbox' and 'TIMES SS'
- Discussion of the QSAR models and each of the protein-binding profilers
- Estimation of the utility of each profiler for the prediction of skin sensitization hazard
- Estimation of the utility of the metabolism and autoxidation simulators
- Calculation of Cooper statistics
- Identification of training set affiliations of the substances as well as estimation of the reliability of the single predictions
- Proposals on how to improve the respective QSAR models
- Preparation of the manuscript including all figures and tables

Toxicology in Vitro 34 (2016) 194–203

Contents lists available at ScienceDirect

Toxicology in Vitro

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

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.

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 (haptenization). 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).

http://dx.doi.org/10.1016/j.tiv.2016.04.005
0887-2333/

"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 'autoxidation' 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 'KeratinoSens™' [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 autoxidation ('Autoxidation 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 predictivites 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 autoxidation' 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 autoxidation 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 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.

Table 2
Proposed mechanistic domains assigned by the QSAR Toolbox.

OECD profiler		OASIS profiler	
Acylation	Ac	Acylation	Ac
Michael addition	MA	Michael addition	MA
Schiff base formation	SB	Schiff base formation	SB
Nucleophilic substitution type 2	S_N2	Nucleophilic substitution type 2	S_N2
Nucleophilic aromatic substitution	S_NAr	Nucleophilic aromatic substitution	S_NAr
The OECD profiler has 102 categories and was adopted in January 2011		Ionic interaction	II
		Nucleophilic addition	NA
		Radical reactions	RR
The OASIS profiler has 102 categories and was adopted in October 2010		Nucleophilic substitution type 1	S_N1
		Nucleophilic substitution type 2 ionic	S_N2i
		Nucleophilic vinylic substitution	S_NV

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.

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 autoxidation, 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):

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

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 autoxidation. 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	Optional	Optional	LLNA and human data	Yes
	OECD profiler	Optional	Optional	LLNA and human data	Yes
Prediction model for DPRA reactivity	DPRA profiler Cys	No	No	DPRA data	No
	DPRA profiler Lys	No	No	DPRA data	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 5
Overview 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 *in silico* and *in chemico* estimates on peptide reactivity with LLNA and human data on skin sensitization [%]	*In silico* approach																*In chemico* approach	
	QSAR Toolbox v3.3.2								TIMES SS v2.27.13									
	OASIS profiler (P)		OASIS profiler (P + M + AO)		OECD profiler (P)		OECD profiler (P + M + AO)		TIMES − P		TIMES (P + M + AO)		TIMES (P)		TIMES (P + M + AO)		DPRA (Direct Peptide Reactivity Assay)	
									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)			
In vivo 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.

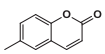
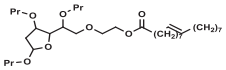
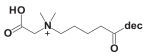
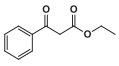
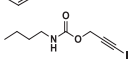
When comparing the results of the OASIS profiler of the QSAR Toolbox to TIMES SS, 13 chemicals with structural alerts identified by the OASIS profiler were correctly predicted as non-sensitizers by TIMES SS. In 12 cases, the structural alert was not fired by TIMES SS and in case of the highly volatile 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		*In chemico*[1]	*In vivo* 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_N2	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_N2	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 + *in vitro* results	DPRA + LuSens + h-CLAT	90	90	89	95	80	90
	DPRA + KeratinoSens™ + h-CLAT	101	90	90	96	79	90
QSAR Toolbox + *in vitro* results	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
	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 (external evaluation set) + *in vitro* results	TIMES SS (P) + LuSens + h-CLAT	13	78	100	100	67	85
	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 + *in vitro* results	DPRA + LuSens + h-CLAT	133	83	78	91	64	82
	DPRA + KeratinoSens™ + h-CLAT	180	82	72	89	59	79
QSAR Toolbox + *in vitro* results	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
	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 (external evaluation set) + *in vitro* results	TIMES SS (P) + LuSens + h-CLAT	19	93	100	100	83	95
	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 $\alpha\beta$-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.

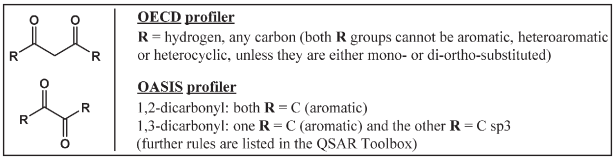


Fig. 1. Different definitions of Schiff base formers considered in the algorithms of the OASIS and OECD profilers of the QSAR Toolbox.

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 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 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		*In chemico*
				OASIS profiler	OECD profiler	DPRA
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.

	In silico		*In chemico*
	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 *in vivo* 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 Ni^{2+} 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
$S_N1/2$	chemicals reacting by nucleophilic substitutions of type 1 or 2
S_NAr	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., Hubesch, B., Jaworska, J., de, K.J., Landsiedel, R., Manou, I., Mehling, A., Petersohn, D., Rorije, E., Rossi L, H., Steiling, W., Teissier, S., Worth, A., 2015. Alternatives for skin sensitisation: hazard identification and potency categorisation: report from an EPAA/CEFIC 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 dermatits: The molecular basis. Springer, Berlin.

Lepoittevin, J.P., 2006. Metabolism versus chemical transformation or pro- versus prehaptens? 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.

4.2 Assessing skin sensitization hazard in mice and men using non-animal test methods

Daniel Urbisch, Naveed Honarvar, Susanne N. Kolle, Annette Mehling, Katharina Guth, Tzutzuy Ramirez, Wera Teubner, Robert Landsiedel, Joanna Jaworska, Petra Kern, Frank Gerberick, Andreas Natsch, Roger Emter, Takao Ashikaga, Masaaki Miyazawa, Hitoshi Sakaguchi

Summary of the publication: In this publication, a comprehensive dataset with 213 substances is compiled, for which non-animal, human and/or animal data are available. The utility of single experimental non-animal methods as well as combinations thereof to predict skin sensitization is analyzed. Moreover, this dataset is also used to define different mechanistic domains of the non-animal methods. Results of this publication is even partly considered and cited in the new ECHA guidance on REACH Information Requirements and Chemical Safety Assessment.

Journal: Regulatory Toxicology and Pharmacology (Elsevier)

Impact factor: 2.221 (Clarivate Analytics, 2017)

Type of authorship: first author

Status of publication: published in 2015

My contribution to the publication: approximately 40%

- Compilation of experimental data from internal and external databases as well as from the literature
- Data gap analysis and performance of experimental testing to close these gaps
- Calculation of all Cooper statistics
- Performance of QSAR analyses to assign each substance to its respective protein-binding mechanism
- Detailed discussion of false positive and false negative results of the non-animal methods
- Preparation of the manuscript including all figures and tables

Regulatory Toxicology and Pharmacology 71 (2015) 337–351

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

Assessing skin sensitization hazard in mice and men using non-animal test methods

Daniel Urbisch [a], Annette Mehling [b], Katharina Guth [a], Tzutzuy Ramirez [a], Naveed Honarvar [a], Susanne Kolle [a], Robert Landsiedel [a,*], Joanna Jaworska [c], Petra S. Kern [d], Frank Gerberick [e], Andreas Natsch [f], Roger Emter [f], Takao Ashikaga [g], Masaaki Miyazawa [h], Hitoshi Sakaguchi [h]

[a] BASF SE, Experimental Toxicology and Ecology, Ludwigshafen, Germany
[b] BASF Personal Care and Nutrition GmbH, Düsseldorf, Germany
[c] Procter & Gamble NV, 100 Temselaan, 1853 Strombeek-Bever, Belgium
[d] Procter & Gamble Technology (Beijing) Co., Ltd., China
[e] Procter & Gamble Company, Cincinnati, OH, USA
[f] Givaudan Schweiz AG, Ueberlandstraße 138, CH-8600 Dübendorf, Switzerland
[g] Shiseido Research Center, Shiseido Co., Ltd., 2-2-1 Hayabuchi, Tsuzuki-ku, Yokohama-shi, Kanagawa 224-8558, Japan
[h] Safety Science Research Laboratories, Kao Corporation, 2606 Akabane, Ichikai, Haga, Tochigi 321-349, Japan

ARTICLE INFO

Article history:
Received 22 September 2014
Available online 23 December 2014

Keywords:
Skin sensitization
Allergic contact dermatitis (ACD)
Adverse outcome pathway (AOP)
DPRA
KeratinoSens™
LuSens
h-CLAT
(modified) MUSST
Database
Integrated testing strategy (ITS)

ABSTRACT

Sensitization, the prerequisite event in the development of allergic contact dermatitis, is a key parameter in both hazard and risk assessments. The pathways involved have recently been formally described in the OECD adverse outcome pathway (AOP) for skin sensitization. One single non-animal test method will not be sufficient to fully address this AOP and in many cases the use of a battery of tests will be necessary. A number of methods are now fully developed and validated. In order to facilitate acceptance of these methods by both the regulatory and scientific communities, results of the single test methods (DPRA, KeratinoSens™, LuSens, h-CLAT, (m)MUSST) as well for a the simple '2 out of 3' ITS for 213 substances have been compiled and qualitatively compared to both animal and human data. The dataset was also used to define different mechanistic domains by probable protein-binding mechanisms. In general, the non-animal test methods exhibited good predictivities when compared to local lymph node assay (LLNA) data and even better predictivities when compared to human data. The '2 out of 3' prediction model achieved accuracies of 90% or 79% when compared to human or LLNA data, respectively and thereby even slightly exceeded that of the LLNA.

1. Introduction

Allergic contact dermatitis (ACD[1]) is the clinically relevant outcome of skin sensitization and it is estimated that 15–20% of the general population will be sensitized at some point in the course of their lives (Thyssen et al., 2007; Bruckner et al., 2000). ACD can be associated with morbidity, affect the quality of life, and sensitization – the prerequisite for the development of ACD – is usually a lifelong effect. Sensitization is considered to be one of the key human health endpoints of toxicological assessments of a substance both in the

* Corresponding author. Fax: +49 621 60 58134.
E-mail address: robert.landsiedel@basf.com (R. Landsiedel).

[1] *Abbreviations:* ACD, allergic contact dermatitis; Ac, acylating agent; AOP, adverse outcome pathway; ARE, antioxidant response element; CD, cluster of differentiation; CV, cell viability; DC, dendritic cell; DPRA, direct peptide reactivity assay; EC, effective concentration; ECHA, European Chemical Agency; ECVAM, European Centre for the Validation of Alternative Methods; ESAC, ECVAM Scientific Advisory Committee; GPTs, guinea-pig based tests; FN, false negative; FP, false positive; h-CLAT, human cell-line activation test; IATA, integrated testing approaches and assessments; IC, inhibitory concentration; ITS, integrated testing strategy; KC, keratinocyte; Keap-1, Kelch-like ECH-associated protein 1; LC–MS, liquid chromatography coupled with mass spectrometry; LLNA, local lymph node assay; LMW, low molecular weight; MA, Michael acceptor; MHC, major histocompatibility complex; MIT, minimal induction threshold; (m)MUSST, (modified) myeloid U937 skin sensitization test; MTT, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazoliumbromid; *n* = number of substances; NC, not classified; NQO1 = NADPH:quinone oxidoreductase 1; Nrf-2, nuclear factor (erythroid-derived 2)-like 2; OECD, Organization for Economic Co-operation and Development; QP, quinone precursor; QSAR, quantitative structure activity relationship; REACh, Registration, Evaluation, Authorization and Restriction of Substances; RFI, relative fluorescence intensity; RIFM, Research Institute for Fragrance Materials; SB, Schiff 'base former; $S_N1/2$, substances reacting by nucleophilic substitutions of type 1 or 2; S_NAr, Aromatics reacting by nucleophilic substitutions; TN, true negative; TP, true positive; WoE, weight of evidence.

http://dx.doi.org/10.1016/j.yrtph.2014.12.008
0273-2300/

occupational and consumer setting. Sensitization and ACD are the result of a complex multifactorial sequence of events. The chemical and biological pathways involved are relatively well characterized and can be simplified and structured into an adverse outcome pathway (AOP). The key events involved have now been formally described by the OECD in a document titled "The Adverse Outcome Pathway for Skin Sensitization Initiated by Covalent Binding to Proteins" with the goal of facilitating the development of methods and approaches addressing the relevant events (ENV/JM/MONO(2012)10/PART 1 and 2).

Contact allergies develop in two stages: (1) the sensitization phase in which antigen/allergen specific T-cells are generated, and (2) the elicitation phase in which renewed contact with the allergen leads to the allergic response (Goebel et al., 2012). Following entry into the skin, the AOP described by the OECD identifies eleven events involved whereby four are considered to be key events. The initiating event of the sensitization process is the covalent binding of a hapten, a low molecular weight (LMW) and typically electrophilic substance, to the skin proteins. This is the essential event to transform an otherwise non-immunogenic molecule into a potential allergen. The protein reactivity of the substance may be inherent or can develop following metabolic or abiotic transformations of the pro- or pre-haptens, respectively (Gerberick et al., 2008; Jäckh et al., 2012). Key event 2 is the activation of keratinocytes. Keratinocytes are the main cell population of the epidermis and are the first cells to come into contact with the potential allergen. Keratinocytes may respond to the contact with the hapten and/or the complete allergen. Among these responses, the oxidative and electrophilic stress-driven expression of genes under the control of the antioxidant response element (ARE) as part of the Keap1/Nrf2 pathway is well described (Natsch and Emter, 2008). The third key event is the activation of dendritic cells. Dendritic cells take up and process antigens and present fragments in form of major histocompatibility complexes (MHC) on their surfaces. Dendritic cells mature during this process and migrate to the lymph nodes. Mature dendritic cells are characterized by the up-regulation of cell surface markers such as CD54 and CD86 in order to activate naïve T-cells. The final event of the sensitization phase is the proliferation of the antigen-specific T-cells and the generation of antigen-specific memory T-cells (Goebel et al., 2012).

In the regulatory context, the skin sensitization potential of a chemical has traditionally been evaluated using animal tests, in particular the guinea-pig based tests (GPTs) described in OECD 406 or the mouse-based local lymph node assay (LLNA) described in OECD 429 (or OECD 442 A + B). The 7th amendment of the Cosmetic Directive (Council Directive 76/768/EEC of 1976-07-27; now Cosmetics Regulation: REGULATION (EC) No. 1223/2009), implemented a phasing out of animal testing for the purposes of this legislation. This in turn has made the development and use of non-animal tests for new cosmetic substances indispensable. Other legislations have followed or are in the process of following suit, and explicitly permit animal testing only as a last resort, e.g. the current European substances legislation Nr. 1907/2006 [Registration, Evaluation, Authorization and Restriction of Chemicals (REACh)]. Of the human health endpoints to be assessed under REACh, skin sensitization must be evaluated for all substances to be registered. In REACh phase 1 more than 3700 substances were assessed for skin sensitization (Angers-Loustau et al., 2011) and several thousand more are expected for REACh phase 3 until 2018.

The current consensus among the scientific community is that one single non-animal test might not be sufficient as a stand-alone method to cover the endpoint skin sensitization and that the use of an integrated testing strategy (ITS; in this context ITS includes sequential testing strategies (STS), etc.) will be necessary (Mehling et al. 2012, Rovida et al. 2014; Basketter et al., 2013). Currently, two tests, namely the direct peptide reactivity assay (DPRA) (Gerberick et al., 2004) and the ARE–Nrf2 luciferase test method KeratinoSens™ (Emter et al., 2010) have passed the validation process at ECVAM and a statement of the ECVAM Scientific Advisory Committee (ESAC) has been published regarding their use within integrated testing approaches and assessments (IATA) (EURL ECVAM, 2013; EURL ECVAM, 2014); the OECD draft guidelines have been drafted and are in the final stages of the commentation process. The ECVAM recommendation and ESAC statement as well as an OECD draft guideline are available for a third test, namely the human cell-line activation test (h-CLAT) (Ashikaga et al., 2006). In addition, validation studies are ongoing for the LuSens assay (Bauch et al., 2012; Ramirez et al., 2014), an ARE–Nrf2 luciferase test method similar to the KeratinoSens™. Data on over a hundred substances is available for the myeloid U937-based skin sensitization tests MUSST (myeloid U937 skin sensitization test) (Ade et al., 2006) and the modified MUSST (mMUSST) (Bauch et al., 2011, 2012). The MUSST is also in the validation at ECVAM (ECVAM Test Method Submissions 2008–2014).

Integrated approaches to testing and assessment (IATA) and in particular the integrated testing strategies (ITS) used within the IATA should consist of methods with a mechanistic relevance for the endpoint being assessed. Placed in the context of the AOP, the DPRA evaluates key event 1 – the protein/peptide reactivity of a substance, the KeratinoSens™ and LuSens assays represent key event 2 and give a measure keratinocyte activation, and the h-CLAT, MUSST and mMUSST describe key event 3 – dendritic cell activation. Thus, when used together, they cover the first three of the four key events of the sensitization process, thus being of mechanistic relevance and supporting the scientific rationale for using a combination of these methods in an AOP-based ITS. Indeed, these tests are already being successfully used in combination in a number of ITS which include, but are not limited to, the '2 out of 3' approach (also sometimes termed majority vote or '2 out of 3' weight of evidence (WoE) approach (Bauch et al., 2012; Natsch et al., 2013; van der Veen et al., 2014)), an ITS based on Bayesian Networks (Jaworska et al., 2013) and an ITS which includes an assessment of potency (Nukada et al., 2013; Tsujita-Inoue et al., 2014). Currently, the OECD in conjunction with EURL ECVAM is developing a document describing the key elements of an (AOP-based) IATA and designing templates for reporting an ITS. According to the roadmap proposed by EURL ECVAM submission of the document to the OECD should take place in 2016 (Kinsner-Ovaskainen et al., 2012). Whether this document will be accepted in time for the testing phase preceding the REACh phase 3 (substances with an annual production tonnage of 1–100 tons) deadline is currently unknown.

As risk assessments are based on hazard assessments, the first goal is to adequately address the skin sensitization hazard potential of a substance. A study conducted by ECVAM on all registered substances in the new substances data base, revealed that 2745 of the evaluated 3792 substances were not classified (NC; approx. 72%) for sensitization and that 1047 (about 28%) were classified as sensitizers (Angers-Loustau et al., 2011). This indicates that if a sensitization potential (hazard) can be excluded, further information on potency for risk assessments or classification according to the Globally Harmonized System (GHS) would only be needed for approximately 30% of the substances. Regulatory acceptance, e.g. for GHS classification and/or REACh registration, and the use of non-animal tests for safety assessments is critically dependent on the correct predictivities of a method or an ITS. This in turn is supported by the availability of data allowing the evaluation of the types of substances tested and the limitations/strengths of the methods or ITS. Only such data driven analysis will help to build confidence to use the new approaches. This was also a key message from a workshop at which regulators and industry dis-

cussed how to best facilitate acceptance of non-animal methods and ITS (Basketter et al., 2013).

This manuscript compiles and evaluates the available data – both previously published and unpublished – obtained with the above mentioned *in vitro* methods to help facilitate the acceptance of non-animal testing approaches for the toxicological endpoint 'skin sensitization'. The predictivities of the individual methods are evaluated as is their use in one ITS for skin sensitization hazard assessment, namely the '2 out of 3' approach. The evaluation was limited to this example of an ITS as the design of an ITS needs to be adapted depending on the assessments being made, e.g. an ITS for risk assessments will usually differ from an ITS for hazard assessments. Additionally, flexibility is needed, e.g. if the ITS is part of a read across approach, if some methods may not be available at contract research institutes, or if an ITS has proprietary elements. Therefore, only the weight of evidence based '2 out of 3' approach for hazard assessments which is based on the first three key events of the adverse outcome pathway was assessed. In this study, the applicability of the test methods and ITS for different substance properties and mechanisms of reactivity was also assessed. The compilation of non-animal test method results for over 200 substances now provides a very comprehensive data base which exceeds that used for the formal validation of the "benchmark" animal method – the LLNA.

2. Materials and methods

2.1. Data set

The data set consists of both previously published (Natsch et al., 2013; Nukada et al., 2012; Bauch et al., 2012) and additional unpublished data for 213 test substances. All substances, physicochemical properties, types of use, test data, proposed reaction mechanisms as well as related references and data bases are listed in Supplementary Table (supporting information). Fig. 1 shows the use categories of these substances.

For 208 substances, high quality LLNA data were described in the literature or data bases; for 5 substances hitherto unpublished LLNA data were used. Of the 213 substances evaluated in this study, 151 (71%) are considered to be sensitizers and 62 (29%) to be non-sensitizers according to available LLNA data. In addition, human data were available for 114 of these substances. For all substances within this data compilation, results were available for at least two of the investigated non-animal test methods. The set of non-animal test data comprises data originating from the DPRA (results for 199 substances), KeratinoSens™ (results for 195 substances), LuSens (results for 77 substances), h-CLAT (results for 166 substances), MUSST (results for 145 substances) and mMUSST data (results for 65 substances). In order to create mechanistic domains from a chemical perspective, the putative reaction mechanisms for peptide reactivity of the 213 substances was evaluated using the freely available OECD QSAR Toolbox v3.2 and searching the scientific literature.

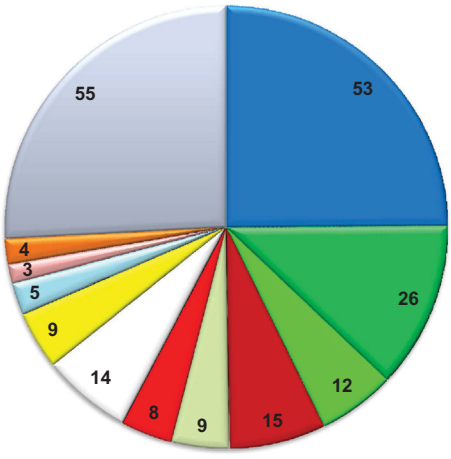


Fig. 1. Schematic representation of the uses of the investigated substances. The numbers of substances for each use are indicated in parentheses as well as in the pie chart itself. Blue – fragrances ($n = 53$), dark green – preservatives/disinfectants ($n = 26$), light green – dyes ($n = 12$), dark red – monomers ($n = 15$), mint green – pesticides ($n = 9$), red – solvents ($n = 8$), white – cosmetics ($n = 14$), yellow – pharmaceuticals ($n = 9$), light blue – surfactants ($n = 5$), pink – plasticizers ($n = 3$), orange – food/feed ($n = 4$), gray – other uses ($n = 55$).

2.2. Non-animal test methods

2.2.1. Direct peptide reactivity assay (DPRA)

Peptide reactivity data were generated using a method to measure reactivity of a test chemical with model hepta-peptides containing lysine (Ac-RFAAKAA-COOH) or cysteine (Ac-RFAACAACOOH) (Gerberick et al., 2004). 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.

Peptides were prepared and purified by the SynPep Corporation (Dublin, CA, USA) to >90% purity as measured by HPLC, and molecular weight confirmation was determined by flow injection positive-ion electrospray mass spectrometry. Briefly, 400 μL of a 1.25 mM peptide stock solution prepared in buffer and a 100 mM test chemical stock solution prepared in either acetonitrile or DMSO/acetonitrile were added to 100 mM ammonium acetate buffer (pH 10.2) for the lysine peptide or 100 mM sodium phosphate buffer (pH 7.5) for the cysteine peptide. The final reaction, containing 0.5 mM of the peptide and 5 or 25 mM of the test chemical, representing 1:10 and 1:50 M ratios, was mixed and incubated in the dark for 24 h at 25 °C. Control samples and standards used for defining the calibration curve for each analysis were prepared without test chemical for each peptide and ranged from 0.0156 to 1.0 mM. All samples were prepared in triplicate. Following incubation, the peptide was quantified by reverse-phase HPLC (Waters 2695 Alliance) on a Zorbax SB-C18 column (3.5 lm, 100 × 2.1 mm) with UV detection at 220 nm (Waters 996 PDA detector) using an external standard linear calibration curve. The UV spectrum was collected from 210 to 400 nm to permit verification of the peptide peak identity. Results for 199 substances were generated by P&G (referred to in Supplementary Table as DPRA I) or BASF (DPRA II) and thus available for this study. Five substances were not considered for further analyses due to discordant results in the two independent labs.

2.2.2. KeratinoSens™ assay

The standard operating procedure described (Natsch et al., 2011) and published online (ECVAM, 2014) was used to test additional substances in the KeratinoSens™ assay. Briefly, cells were grown for 24 h in 96-well plates. The medium was then replaced with medium containing the test substance and a final level of 1% of the solvent DMSO. Each test substance was subsequently tested at 12 twofold dilutions (0.98–2000 μM). In each repetition, three parallel replicate plates were run for luciferase determination and a fourth parallel plate was prepared for cytotoxicity determination. Cells were incubated for 48 h with the test substances, and then luciferase activity and cytotoxicity (with the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazoliumbromid (MTT) assay (Mosmann, 1983)) were determined. For each chemical the EC1.5, EC2 and EC3 values (concentration in μM for 1.5, 2 and 3-fold induction of the luciferase activity) were calculated along with

IC50 values for the concentration yielding 50% reduction in cellular viability. Substances were tested in at least two independent experiments. A substance is considered to have a sensitizing potential if an induction equal to or exceeding 1.5-fold compared to the vehicle control is observed at a concentration below 1000 μM and at which cells remain >70% viable. If the results of the two experiments were concordant, a prediction according to the prediction model was derived. Substances with discordant results or results close to the 1.5-fold threshold (borderline) were tested in additional independent experiments. The number of experiments and the number of positive results for each chemical is given in the database. Results for 195 substances were generated by Givaudan (referred to in Supplementary Table as KeratinoSens assay I) or BASF (KeratinoSens assay II) and thus available for this study. Eight substances were not considered for further analyses due to discordant results in the two independent labs.

2.2.3. LuSens assay

The LuSens assay is a keratinocyte-based assay which utilizes the luciferase gene under the control of the antioxidant response element (ARE) originating from the rat NQO1 gene as the reporter construct. The LuSens assay used in this paper is similar to that described in Bauch et al., 2012 with some modifications (Ramirez et al., 2014). In brief, a cytotoxicity range finding experiment (consisting of twelve concentrations) was performed, to calculate the concentration in which cell viability corresponds to no less than 75% (CV75). Following the range finder experiment, a main experiment was set up using six concentrations of test substance (in triplicates), the highest tested concentration was 1.2× CV75 (or 2000 μM if no cytotoxicity was observed). After 48 h treatment, luminescence and cytotoxicity were measured. A test substance is considered to have skin sensitization potential when the luciferase induction is above or equal to 1.5-fold compared to the vehicle control in two (or more than) consecutive non-cytotoxic tested concentrations whereby at least three tested concentrations must be non-cytotoxic (viability $\geqslant$70%). A test compound is considered not to have sensitizing potential if the above effects are not observed. The skin sensitization potential of a test substance is determined by the result of the majority of the repetitions of an experiment. If two of two or two of three repetitions are negative/positive, the substance is considered as negative/positive. In order to come to a conclusion on the skin sensitization hazard of a substance, one complete experiment needs to be conducted. A complete experiment consists of two valid independent repetitions (Ramirez et al., 2014). Results for 77 substances were obtained by BASF and considered for this study.

2.2.4. Human cell line activation test (h-CLAT)

In the h-CLAT assay, THP-1 cells (American Type Culture Collection, Manassas, VA, USA) were used as surrogate for dermal dendritic cells. For dose finding, cytotoxicity tests were conducted and the concentration resulting in 75% cell viability, termed CV75, was calculated based on the analysis of viable cells. THP-1 cells were treated with eight different concentrations, decided based on dose finding cytotoxicity test, for 24 h. After removing the test substance, the expression of CD86 and CD54 on the cell surface was measured by flow cytometry. The relative fluorescence intensity (RFI) was used as an indicator of CD86 and CD54 expression. If the RFI of CD86 or CD54 was greater than 150% or 200% at any dose in at least two out of three experiments, the substance was judged as a sensitizer. Otherwise, it was considered a non-sensitizer (Ashikaga et al., 2006). From the dose-dependency curves of three experiments, the median concentration inducing 150% of CD86 RFI and/or 200% of CD54 RFI (EC150 or EC200) was calculated like EC3 value determination in the LLNA. The lower EC value was defined as minimal induction threshold (MIT) (Nukada et al., 2013). Results for 166 substances were available for this study and for the determination of Cooper statistics; data generated by the respective method developer was used.

2.2.5. Myeloid U937 skin sensitization test (MUSST) and modified MUSST

The MUSST uses the U937 cell-line purchased from the American Type Culture Collection (Rockville, MD, USA). Four to six concentrations are chosen based on preliminary propidium iodide cytotoxicity experiments and are applied in duplicate for 48 h. The highest tested concentration in the main experiment is twice the concentration causing a cytotoxicity of 25% (CV75) determined in a pretest. A test substance is predicted to have a dendritic cell activating potential indicative of being a sensitizer when CD86 induction (measured by flow cytometry) exceeds the threshold of 1.5-fold with respect to vehicle treated cells at any tested concentration showing sufficient cell viability ($\geqslant$70%) in at least two independent experiments (Natsch et al., 2013). Results for 145 substances were made available for this study by P&G (referred to in Supplementary Table as U-937 Test).

A modified version of the MUSST (mMUSST) uses the U937 cell line from German Resource Center for Biological Material DSMZ, Braunschweig, Germany. In the mMUSST, a test substance is predicted to have a dendritic cell activating potential when CD86 induction exceeds a threshold of 1.2-fold (Bauch et al., 2012). Data for 65 substances were generated in the mMUSST by BASF (referred to in Supplementary Table as mMUSST).

For the analyses within this study, the results from the MUSST and mMUSST were taken together to create a dataset of 161 substances. Data for 12 substances were not further considered for analyses due to discordant results being obtained in both tests.

2.2.6. '2 out of 3' prediction model

The least complicated way to assess the skin sensitization hazard potential of a substance is to use the results of single assays which reflect key steps of the AOP within a '2 out of 3' prediction model. For the assays addressing the three key events described in the OECD AOP on skin sensitization mentioned above, a '2 out of 3' assessment was introduced for the first time by Bauch et al. (2012). In the current study, this prediction model was applied using DPRA, KeratinoSens™ and h-CLAT data. Any two congruent results of the three tests rule the overall assessment: If at least two of the three assays were positive, the substance was rated to be a skin sensitizer. If at least two of the three assays were negative, the substance was rated to be a non-sensitizer. The classification as a sensitizer or non-sensitizer is therefore based on a weight of evidence pertaining to key events of the AOP. Cooper statistics for this classification were determined in comparison to LLNA or human data. Results for 180 or 101 substances, respectively, were obtained using this prediction model.

2.2.7. OECD QSAR Toolbox

The QSAR Toolbox developed by OECD in collaboration with the European Chemical Agency (ECHA) (Raunio, 2011) is a standalone software application for filling gaps of (eco)toxicity data that are needed to assess the potential hazards of substances. *In silico* analysis was performed with the OECD QSAR Toolbox in the version 3.2., which is freely available on the OECD website (http://toolbox.oasis-lmc.org/?section=download&version=latest). In order to support the identification of a chemical's toxicity, the OECD QSAR Toolbox contains numerous databases with results from experimental studies for over 55.000 substances and profilers for calculating specific properties (references are added at each profiler within this tool). Two such profilers are the protein-binding profilers based on OECD and OASIS algorithms (''Protein binding by OECD'', ''Protein binding by OASIS v1.2''). In order to also identify

substances, which require abiotic or metabolic activation, the "auto-oxidation profiler" and the "skin metabolism profiler" were used Teubner et al., 2013.

2.2.8. Cooper statistics

The Cooper statistics of predictivity were calculated for the single assays as well as for the '2 out of 3' prediction model (Cooper et al., 1979). 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):

$$\text{Sensitivity } [\%] = \frac{TP}{TP + FN} * 100$$

$$\text{Specificity } [\%] = \frac{TN}{TN + FP} * 100$$

$$\text{Positive predictive value } [\%] = \frac{TP}{TP + FP} * 100$$

$$\text{Negative predictive value } [\%] = \frac{TN}{TN + FN} * 100$$

$$\text{Accuracy } [\%] = \frac{TP + TN}{TP + FP + TN + FN} * 100$$

3. Results

3.1. Availability of in vivo data

Due to the fact that human data were available for only a limited number of substances, two different sets of substances were defined (Table 1). Murine LLNA data were available for all 213 substances (set A) whereas additional human data were available for a subset of 114 substances (subset B). No reliable human data were found in the literature or data bases for 99 substances (data were not further analyzed for this group). In subset B, direct comparisons of non-animal test results to both human and LLNA data were possible.

3.2. Predictivities compared to LLNA and human data

Detailed results (including concentration–response data, cytotoxicity information, peptide depletions) of the investigated assays are listed in Supplementary Table (supporting information). When calculating the Cooper statistics (Table 2), the assessments were reduced to yes/no ratings according to the prediction models of the individual assays. It should be noted that data for all 213 substances was not available for every assay. The number of tested substances being considered for further analyses hence ranges from $n = 77$ (LuSens) to $n = 194$ (DPRA). Accuracies of the single assays compared to LLNA data range from 73% to 76% (Table 2, set A). The '2 out of 3' approach (DPRA, KeratinoSens™ and h-CLAT data) provides an accuracy of 79% when compared to LLNA data.

Using the data from subset B, a direct comparison between the predictivities of the non-animal test methods and human or LLNA data is possible (Table 3). Interestingly, when comparing the Cooper statistics of almost all single non-animal test methods with human data, the percentages were higher than comparisons with LLNA data (Table 3). The '2 out of 3' approach (DPRA, KeratinoSens™ and h-CLAT data) provides somewhat higher accuracies compared to the predictivities of the single assays, whereby the accuracies compared to human or LLNA data were 90% or 82%, respectively. These results indicate that the investigated non-animal methods generally predict human data more accurately than LLNA data.

To determine the performance of the LLNA to predict skin sensitization hazard of this data set, LLNA data were directly compared to the available human data for the substances within subset B. A total of 20 (18%) of the substances were incorrectly identified as FN or FP by the LLNA. Seven substances 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, SDS, tocopherol, benzyl salicylate, hexyl salicylate, xylene and phthalic anhydride). In total, comparison to human data revealed a sensitivity of 91%, a specificity of 64% and an overall accuracy of 82% of the LLNA (Table 3). For comparison, the '2 out of 3' prediction model revealed a sensitivity of 90%, a specificity of 90% and an overall accuracy of 90% for the same set of substances suggesting predictions to be more relevant for humans.

3.3. Comparing results of single assays in different labs

3.3.1. Interlaboratory reproducibilities

For some of the 213 substances the DPRA, KeratinoSens™ and (m)MUSST were conducted in two labs enabling tentative analyses of interlaboratory reproducibilities. Within this study, 45 substances were tested in the DPRA in two labs (P&G and BASF). Among these, 40 substances gave congruent (89%) and five substances gave incongruent results. 52 Substances were tested in the KeratinoSens™ assay in two labs (Givaudan and BASF). Here, 44 substances gave congruent (85%) and eight substances gave incongruent qualitative outcomes. Comparing MUSST and mMUSST results obtained at P&G and BASF, the analyses of 45 substances resulted in 32 congruent (71%) and 13 incongruent qualitative outcomes (this comparison needs to be viewed in the light of slightly different protocols and use of the 1.5-fold and 1.2-fold CD86 induction cut-off, respectively). Substances with incongruent results were excluded from subsequent analyses.

3.3.2. Interchangeabilities

Since both the KeratinoSens™ and the LuSens address keratinocyte activation and the h-CLAT and (m)MUSST address dendritic cell activation, interchangeabilities of the assays representing the same AOP key event were analyzed based on predicitivities of the assays and in the '2 out of 3' approach.

A set of 69 substances was tested in the KeratinoSens™ (Givaudan) and the LuSens assay (BASF). The same overall outcome for 61 of the 69 substances resulted in an interchangeability of 88%. The discordant substances among the two ARE activation assays are listed in Table 11. The KeratinoSens™ offered an accuracy of 75% or 83% and the LuSens an accuracy of 71% or 79%, when comparing the results to LLNA or human data, respectively. The implementation of the KeratinoSens™ into a '2 out of 3' approach with DPRA and h-CLAT data for the set of 69 substances resulted in an accuracy of 85% or 91%, when comparing the predictions to LLNA or

Table 1
Sensitizers and non-sensitizers among substances with LLNA data and with human data on skin sensitization.

Chemical set and reference data	Set A	Subset B	
	LLNA data	Human data	LLNA data
Sensitizer	151	75	83
Non-sensitizer	62	36	31
Borderline	0	3	0
Total	213	114	114

(A) All substances within the data set ($n = 213$) with available LLNA data. (B) Substances within the data set ($n = 114$) with available human data.

Table 2
Cooper statistics for the single non-animal test methods and the '2 out of 3' approach for the whole data set (set A).

Chemical set and reference data	Set A LLNA data			
Cooper statistics	Sensitivity [%]	Specificity [%]	Accuracy [%]	*n*
'2 out of 3' approach	82	72	79	180
DPRA	76	72	75	194
KeratinoSens™	77	63	73	188
h-CLAT	81	64	76	166
LuSens	76	67	73	77
(m)MUSST	73	74	73	150

n = number of substances analyzed.

Table 3
Cooper statistics for the single non-animal test methods and the '2 out of 3' approach when compared to human and LLNA data (subset B).

Chemical set and reference data	Subset B Human data[a]				LLNA data			
Cooper statistics	Se [%]	Sp [%]	Acc [%]	*n*	Se [%]	Sp [%]	Acc [%]	*n*
'2 out of 3' approach	90	90	90	101	81	83	82	103
DPRA	84	84	84	102	77	85	79	105
KeratinoSens™	82	84	82	102	74	73	74	103
h-CLAT	89	64	82	98	86	68	81	101
LuSens	78	79	79	60	73	70	71	62
(m)MUSST	74	88	78	85	71	83	75	87
LLNA	91	64	82	111	–	–	–	–

Se = sensitivity; Sp = specificity; Acc = accuracy; *n* = number of substances analyzed.
[a] α-Amyl cinnamic aldehyde, α-hexyl cinnamic aldehyde and 6-methyl coumarin could not finally be evaluated as human sensitizers or non-sensitizers and were thus rated as "borderline".

Table 4
Substances with discordant test results among KeratinoSens™ and LuSens.

Name	Cas #	KeratinoSens™			LuSens		
		Final	vs. LLNA	vs. Human	Final	vs. LLNA	Vs. Human
Farnesal	502-67-0	1	TP	–	0	FN	–
Ethylenediamine	107-15-3	1	TP	TP	0	FN	FN
Propyl gallate	121-79-9	1	TP	TP	0	FN	FN
Nickel chloride	7718-54-9	1	FP	TP	0	TN	FN
Methyl salicylate	119-36-8	0	TN	TN	1	FP	FP
Benzyl alcohol	100-51-6	0	TN	FN	1	FP	TP
Eugenol	97-53-0	0	FN	FN	1	TP	TP
α-Hexyl cinnamic aldehyde	101-86-0	Bord.	–	–	1	TP	–

1, positive overall outcome; 0, negative overall outcome; bord., borderline outcome; TN/TP, true negative/positive; FN/FP, false negative/positive.

human data, respectively. If the LuSens assay was used for this approach instead of the KeratinoSens™, Cooper statistics estimated an accuracy of 83% or 93%, when compared to LLNA or human data, respectively.

Regarding the h-CLAT (Kao and Shiseido) and the (m)MUSST (BASF and P&G), a common set of 105 substances was tested to investigate dendritic cell activation. The same overall outcome for 76 of 105 substances resulted in an interchangeability of 72%. The 29 substances with discordant results can be found in Supplementary Table (supporting information). The h-CLAT offered an accuracy of 80% or 86% and the (m)MUSST an accuracy of 75% or 78%, when comparing the predictions to LLNA or human data, respectively. The implementation of the h-CLAT into a '2 out of 3' approach with DPRA and KeratinoSens™ data for the set of 105 substances resulted in an accuracy of 85% or 91%, when comparing the predictions to LLNA or human data, respectively. If the (m)MUSST was used for this approach instead of the h-CLAT, Cooper statistics estimated an accuracy of 81% or 90%, when compared to LLNA or human data, respectively.

3.4. Predictivities for groups of substances

3.4.1. Mechanistic domains by protein-binding mechanisms

Several means for grouping of substances into specific chemical classes are possible, and the chosen approach may depend on the specific purpose (Aptula et al., 2005). The molecular initiating event (MIE) starting the skin sensitization process is the binding of haptens to proteins. In order to analyze if the non-animal test methods may detect classes of substances with different reaction mechanisms with similar performances, nine different mechanistic domains were defined (Table 4). The probable protein-binding mechanisms were assigned by the OECD QSAR Toolbox v3.2. Protein-binding mechanisms described in the scientific literature were added to build a mechanistic chemistry framework for the 213 substances (Roberts and Aptula, 2014; Karlberg et al., 2013; Aptula et al., 2009, 2007, 2005; Patlewicz et al., 2008; Roberts et al., 2007a) (Supplementary Table). For 11 substances, two possible protein-binding mechanisms were proposed to be probable. For example, maleic anhydride is supposed to react both as an acylat-

Table 5
Proposed mechanistic domains assigned by the OECD QSAR Toolbox, literature search and expert judgment as well as number and percentage of positive LLNA results within each mechanistic domain.

Protein-binding mechanism	*n*	*n* of positives in LLNA
No protein-binding mechanisms assigned (no alert)	65	17 (26%)
Acylating agents (Ac)	21	19 (91%)
Michael acceptors (MA)	45	41 (91%)
Quinone precursors (QP)	22	19 (86%)
Schiff 'base formers (SB)	30	25 (83%)
Substances reacting by nucleophilic substitutions of type 1 or 2 ($S_N1/2$)	30	26 (87%)
Aromatics reacting by nucleophilic substitutions (S_NAr)	6	6 (100%)
Substances reacting by a nucleophilic addition	1	1 (100%)
Metal ions forming coordination bonds	4	2 (50%)
Total (including two possible protein-binding mechanism for 11 substances)	224	–

n = number of substances analyzed.

ing agent and as a Michael acceptor. Furthermore, in some cases it was not possible to assess whether a chemical may react either in a S_N1 or S_N2 reaction. Thus, both organic reaction types were placed into one mechanistic domain. With exception of the one substance reacting in a nucleophilic addition and the four metal containing substances forming coordination bonds, all mechanistic domains are described in more detail in the following sections.

3.4.1.1. No reaction mechanism assigned (no alert). This substance group contains 65 substances whose molecular structures do not contain any obvious alerts for protein reactivity. Many substances in this group represent aliphatic alcohols, organic acids, nitriles, amides, α,β-saturated esters or simple dialkylketones. Accuracies for the non-animal tests calculated within this substance group were compared to LLNA and human data. One aspect which should be taken into account is that the predictivities are influenced by the number and type of substances being assessed and data sets may therefore vary in this study. DPRA data, which directly indicate peptide reactivity, were available for 61 substances in this group. The accuracy of the DPRA was 67% when compared to LLNA data or 76% when compared to human data. Regarding keratinocyte activation, the KeratinoSens™ offered an accuracy of 65% or 82% and the LuSens an accuracy of 60% or 80%, when comparing the predictions to LLNA or human data, respectively. Regarding dendritic cell activation, the accuracy of the h-CLAT was 67% or 59% and that of the (m)MUSST was 68% or 82% when compared to LLNA or human data, respectively. Using the '2 out of 3' approach, 73% or 80% of the substances were correctly identified in this domain when compared to LLNA or human data, respectively. The number of tested substances and predictions is given in Table 5.

Table 6
Total number of analyzed substances, number of TP, FN, TN, FP and accuracies for each non-animal test method and the '2 out of 3' approach for substances whose molecular structures do not contain any obvious alerts for protein reactivity.

No reaction mechanism assigned (*n* = 65)		Set A	Subset B	
		LLNA data	Human data	LLNA data
DPRA	*n* total	61	34	34
	TP/FN	5/11	2/4	3/8
	TN/FP	36/9	24/4	20/3
	Accuracy [%]	67	76	68
KeratinoSens™	*n* total	55	33	33
	TP/FN	5/9	3/3	2/8
	TN/FP	31/10	24/3	19/4
	Accuracy [%]	65	82	64
LuSens	*n* total	30	25	25
	TP/FN	1/5	2/1	1/4
	TN/FP	17/7	18/4	15/5
	Accuracy [%]	60	80	64
h-CLAT	*n* total	48	29	29
	TP/FN	10/4	3/3	6/3
	TN/FP	22/12	14/9	14/6
	Accuracy [%]	67	59	69
(m)MUSST	*n* total	47	28	28
	TP/FN	4/6	3/2	3/5
	TN/FP	28/9	20/3	17/3
	Accuracy [%]	68	82	71
'2 out of 3' approach	*n* total	51	30	30
	TP/FN	7/8	2/3	3/7
	TN/FP	30/6	22/3	18/2
	Accuracy [%]	73	80	70

n, number of substances analyzed; TN/TP, true negative/positive; FN/FP, false negative/positive.

3.4.1.2. Acylating agents (Ac). In an acylation, acyl moieties can be transferred to the nucleophilic centers of proteins. This mechanistic domain contains 21 acylating agents like esters of acidic alcohols (e.g. phenol esters) or carboxylic anhydrides. Accuracies for the non-animal tests calculated for the substances within this domain were compared to LLNA and human data. The accuracy of the DPRA was 100% or 82% when compared to LLNA or human, respectively. The ARE-based assays did not allow a reliable prediction of the sensitization potential of substances in this domain. The KeratinoSens™ offered an accuracy of 56% or 58% and the LuSens an accuracy of 44% or 50%, when comparing the predictions to LLNA or human data, respectively. Regarding dendritic cell activation test methods, Cooper statistics estimated an accuracy of 88% or 83% for the h-CLAT and an accuracy of 69% or 56% for the (m)MUSST, when compared to LLNA or human data, respectively. Using the '2 out of 3' approach, 88% or 83% of the substances were correctly identified in this domain when compared to LLNA or human data, respectively. The number of substances, accuracies and number of correct or incorrect predictions can be found in Table 6.

3.4.1.3. Michael acceptors (MA). MAs contain α,β-unsaturated ester, ketone or aldehyde functions. α,β-unsaturated alcohols can also react as MAs after the alcohol group is oxidized to an aldehyde (Karlberg et al., 2013). Due to their high probability to react as MAs, quinone precursors could also be placed into this domain (Aptula et al., 2009), but are handled as a separate substance domain on account of their specific structural characteristics. This mechanistic domain contains 45 Michael acceptors. Accuracies for the non-animal tests calculated within this mechanistic domain were compared to LLNA and human data. In this domain, the DPRA provided an accuracy of 81% when compared to LLNA data or 86% when compared to human data. The KeratinoSens™ offered an accuracy of 85% or 100% and the LuSens an accuracy of 88% or

100%, when comparing the predictions to LLNA or human data, respectively. Regarding dendritic cell activation, Cooper statistics estimated an accuracy of 92% or 90% for the h-CLAT and an accuracy of 86% or 80% for the (m)MUSST, when compared to LLNA or human data, respectively. Using the '2 out of 3' approach, 88% or 95% of the substances were correctly identified in this domain when compared to LLNA or human data, respectively. The number of substances, accuracies and number of correct or incorrect predictions can be found in Table 7.

3.4.1.4. Quinone precursors. This mechanistic domain contains di- or poly-substituted aromatic compounds with alkoxy, hydroxyl or amine residues. These compounds are pre- or pro-haptens and thus require abiotic or metabolic oxidation to form quinones, quinone imines or quinone methides. These oxidation products are most likely to react as Michael acceptors (free radical binding via Wuerster-type radical may also be possible) (Aptula et al., 2009). 22 substances were identified as possible quinone precursors. Accuracies for the non-animal test methods calculated within this mechanistic domain were compared to LLNA and human data. In this domain, the DPRA provided an accuracy of 71% when compared to LLNA data or 91% when compared to human data. Regarding keratinocyte activation, the KeratinoSens™ offered an accuracy of 79% or 90% and the LuSens an accuracy of 71% or 71%, when comparing the predictions to LLNA or human data, respectively. However, only 7 substances (set A and B) in this domain have LuSens data. Regarding dendritic cell activation, Cooper statistics estimated an accuracy of 94% or 91% for the h-CLAT and an accuracy of 83% or 80% for the (m)MUSST, when compared to LLNA or human data, respectively. Using the '2 out of 3' approach, 83% or 91% of the substances were correctly identified in this domain when compared to LLNA or human data, respectively. The number of substances, accuracies and number of correct or incorrect predictions can be found in Table 8.

3.4.1.5. Schiff 'base formers (SB). Aldehydes and activated ketones predominantly react with hard nucleophiles to form imines (Schiff 'bases). Pre/pro-SB formers like primary amines were also included in this domain (Foussereau et al., 1983) as well as 1,3 dicarbonyl compounds due to their tendency to enolize (Roberts et al., 2007b). Aromatic aldehydes of the general formula ArCHO were assigned to this domain, although a non-sensitizing effect of this alert is discussed in the literature (Patlewicz et al., 2001). In total, 30 of the 213 substances were supposed to react as Schiff 'base formers. Regarding peptide reactivity, the DPRA provided an accuracy of 65% when compared to LLNA data or 77% when compared to human data. The KeratinoSens™ offered an accuracy of 66% or 79% and the LuSens an accuracy of 75% or 86%, when comparing the predictions to LLNA or human data, respectively. However, only 8 (set A) or 7 substances (subset B) in this domain have LuSens data. Regarding dendritic cell activation, Cooper statistics estimated an accuracy of 75% or 93% for the h-CLAT and an accuracy of 64% or 75% for the (m)MUSST, when compared to LLNA or human data, respectively. Using the '2 out of 3' approach, 69% or 92% of the substances were correctly identified in this domain when compared to LLNA or human data, respectively. The number of substances, accuracies and number of correct or incorrect predictions can be found in Table 9.

3.4.1.6. Nucleophilic substitutions ($S_N1/2$). Nucleophilic substitutions exchange a moiety of a molecule (leaving group) for an attacking nucleophilic group. This can be a concerted reaction (SN2) or stepwise with the intermediate formation of a carbocation (SN1). Since a clear distinction of the reaction order is not always possible and the reaction products are identical (except for stereochemistry), both reaction types were placed into one mechanistic domain. The accuracy of the DPRA was 73% when compared to LLNA data or 100% when compared to human data. The KeratinoSens™ offered an accuracy of 88% or 83% and the LuSens an accu-

Table 7
Total number of analyzed substances, number of TP, FN, TN, FP and accuracies for each non-animal test method and the '2 out of 3' approach for possible acylating agents.

Acylating agents (n = 21)		Set A	Subset B	
		LLNA data	Human data	LLNA data
DPRA	n total	19	11	11
	TP/FN	17/0	9/1	10/0
	TN/FP	2/0	0/1	1/0
	Accuracy [%]	100	82	100
KeratinoSens™	n total	18	12	12
	TP/FN	10/6	6/5	5/6
	TN/FP	0/2	1/0	0/1
	Accuracy [%]	56	58	42
LuSens	n total	9	6	6
	TP/FN	4/4	2/3	2/4
	TN/FP	0/1	1/0	0/0
	Accuracy [%]	44	50	33
h-CLAT	n total	16	12	12
	TP/FN	12/2	9/2	9/2
	TN/FP	2/0	1/0	1/0
	Accuracy [%]	88	83	83
(m)MUSST	n total	16	9	9
	TP/FN	9/5	5/4	5/3
	TN/FP	2/0	0/0	1/0
	Accuracy [%]	69	56	67
'2 out of 3' approach	n total	16	12	12
	TP/FN	12/2	9/2	9/2
	TN/FP	2/0	1/0	1/0
	Accuracy [%]	88	83	83

n, number of substances analyzed; TN/TP, true negative/positive; FN/FP, false negative/positive.

Table 8
Total number of analyzed substances, number of TP, FN, TN, FP and accuracies for each non-animal test method and the '2 out of 3' approach for possible Michael acceptors.

Michael acceptors (n = 45)		Set A	Subset B	
		LLNA data	Human data	LLNA data
DPRA	n total	42	22	22
	TP/FN	32/6	18/3	18/3
	TN/FP	2/2	1/0	1/0
	Accuracy [%]	81	86	86
KeratinoSens™	n total	41	21	21
	TP/FN	35/2	20/0	19/1
	TN/FP	0/4	1/0	0/1
	Accuracy [%]	85	100	90
LuSens	n total	17	10	10
	TP/FN	15/1	10/0	10/0
	TN/FP	0/1	0/0	0/0
	Accuracy [%]	88	100	100
h-CLAT	n total	37	20	20
	TP/FN	30/3	18/1	19/0
	TN/FP	4/0	0/1	1/0
	Accuracy [%]	92	90	100
(m)MUSST	n total	29	15	15
	TP/FN	22/3	12/3	12/2
	TN/FP	2/1	0/0	1/0
	Accuracy [%]	83	80	87
'2 out of 3' approach	n total	43	22	22
	TP/FN	36/3	20/1	20/1
	TN/FP	2/2	1/0	1/0
	Accuracy [%]	88	95	95

n, number of substances analyzed; TN/TP, true negative/positive; FN/FP, false negative/positive.

Table 9
Total number of analyzed substances, number of TP, FN, TN, FP and accuracies for each non-animal test method and the '2 out of 3' approach for possible quinone precursors.

Quinone precursors (n = 22)		Set A	Subset B	
		LLNA data	Human data	LLNA data
DPRA	n total	21	11	11
	TP/FN	14/5	10/1	10/1
	TN/FP	1/1	0/0	0/0
	Accuracy [%]	71	91	91
KeratinoSens™	n total	19	10	10
	TP/FN	14/3	9/1	9/1
	TN/FP	1/1	0/0	0/0
	Accuracy [%]	79	90	90
LuSens	n total	7	7	7
	TP/FN	4/2	4/2	4/2
	TN/FP	1/0	1/0	1/0
	Accuracy [%]	71	71	71
h-CLAT	n total	16	11	11
	TP/FN	13/1	9/1	9/1
	TN/FP	2/0	1/0	1/0
	Accuracy [%]	94	91	91
(m)MUSST	n total	18	10	10
	TP/FN	14/2	8/2	8/2
	TN/FP	1/1	0/0	0/0
	Accuracy [%]	83	80	80
'2 out of 3' approach	n total	18	11	11
	TP/FN	14/3	10/1	10/1
	TN/FP	1/0	0/0	0/0
	Accuracy [%]	83	91	91

n, number of substances analyzed; TN/TP, true negative/positive; FN/FP, false negative/positive.

Table 10
Total number of analyzed substances, number of TP, FN, TN, FP and accuracies for each non-animal test method and the '2 out of 3' approach for Schiff 'base formers.

Schiff 'base formers (n = 30)		Set A	Subset B	
		LLNA data	Human data	LLNA data
DPRA	n total	26	13	13
	TP/FN	15/7	12/1	11/1
	TN/FP	2/2	0/0	0/0
	Accuracy [%]	65	77	85
KeratinoSens™	n total	29	14	14
	TP/FN	18/7	11/3	10/3
	TN/FP	1/3	0/0	0/1
	Accuracy [%]	66	79	71
LuSens	n total	8	7	7
	TP/FN	6/1	5/1	5/1
	TN/FP	0/1	1/0	1/0
	Accuracy [%]	75	86	86
h-CLAT	n total	24	14	14
	TP/FN	15/5	12/1	11/1
	TN/FP	3/1	1/0	1/1
	Accuracy [%]	75	93	86
(m)MUSST	n total	22	12	12
	TP/FN	13/6	9/3	8/3
	TN/FP	1/2	0/0	0/1
	Accuracy [%]	64	75	67
'2 out of 3' approach	n total	26	13	13
	TP/FN	17/6	12/1	11/1
	TN/FP	1/2	0/0	0/1
	Accuracy [%]	69	92	85

n, number of substances analyzed; TN/TP, true negative/positive; FN/FP, false negative/positive.

racy of 100% or 100%, when comparing the predictions to LLNA or human data, respectively. However, only 6 (set A) or 5 substances (subset B) in this domain have LuSens data. Regarding dendritic cell activation, Cooper statistics estimated an accuracy of 59% or 92% for the h-CLAT and an accuracy of 64% or 75% for the (m)MUSST, when compared to LLNA or human data, respectively. Using the '2 out of 3' approach, 76% or 100% of the substances were correctly identified in this domain when compared to LLNA or human data, respectively. The number of substances, accuracies and number of correct or incorrect predictions can be found in Table 10.

3.4.1.7. Nucleophilic substitutions in aromatic compounds (S_NAr). A specific case of nucleophilic substitution takes place in aromatic compounds with electron-withdrawing groups. Within this study, only six substances were assigned to this domain whereby only two had human data. Therefore, Cooper statistics were only calculated against LLNA data (set A). The accuracies of the single assays ranged from 75% to 100% (DPRA: 100%, KeratinoSens™: 83%, LuSens: 100%, h-CLAT: 100%, (m)MUSST: 75%) and the accuracy of the '2 out of 3' approach achieved 100% within this mechanistic domain.

3.4.1.8. Pre- and pro-haptens. The class of pre- and pro-haptens represents substances which require abiotic (e.g. via autoxidation) or biotic (e.g. via metabolic pathways) activation, respectively, in order to become electrophilic. Opposite to pro-haptens, the activation of pre-haptens could be prevented by precautionary measures in the handling and storage of the substances (Gerberick et al., 2008). Nevertheless, pre- and pro-haptens are not always distinct since autoxidation and metabolic oxidation can result in the same product, although the underlying mechanisms may differ (Karlberg et al., 2013). Since pre-haptens may also be pro-haptens and vice versa no distinction between substances requiring biotic and abiotic activation was made. The data set contains at least 30 pre/pro-haptens (compare Supplementary Table). The '2 out of 3' approach could be performed for 25 substances within this domain and resulted in 21 correct predictions, when compared to LLNA data (accuracy = 84%).

4. Discussion

With the compilation of currently available non-animal test method results for the distinction of skin sensitizers and non-sensitizers the database could be enlarged to a set of 213 substances. Included in the present dataset are 54 and 145 substances published previously by Bauch et al. (2012) and Natsch et al. (2013), respectively, whereby 41 substances were reported in both studies.

4.1. Predictivities

Compared to the predictivity described by Bauch et al. (2012) and Natsch et al. (2013), the performance of the investigated non-animal test methods is generally slightly lower when analyzing the data set of 213 substances (Table 12). For example, the DPRA provided an accuracy of 79% for 52 substances in Bauch et al. (2012) and 75% for 194 substances in the current study (both compared to LLNA data). This difference may simply reflect the variations occurring with various datasets on the one hand, but on the other hand newly developed methods are often first compared against 'gold-standard' lists of chemicals with the most consistent *in vivo* evidence from multiple tests and test methods, while upon expansion of the database chemicals are included for which, e.g. only a single LLNA study is available. In this regard, the degree of variability of the *in vivo* method needs to be kept in mind. In the light of hazard prediction 15.7% of the investigated non-sensitizers were falsely predicted as sensitizers in repeated experiments and 3.1% of sensitizers were falsely predicted as

Table 11
Total number of analyzed substances, number of TP, FN, TN, FP and accuracies for each non-animal test method and the '2 out of 3' approach for substance reacting in nucleophilic substitutions of type 1 or 2.

Nucleophilic substitutions ($S_N1/2$) ($n = 30$)		Set A	Subset B	
		LLNA data	Human data	LLNA data
DPRA	*n* total	26	11	11
	TP/FN	18/4	9/0	9/0
	TN/FP	1/3	2/0	0/2
	Accuracy [%]	73	100	82
KeratinoSens™	*n* total	26	12	12
	TP/FN	22/0	10/0	10/0
	TN/FP	1/3	0/2	2/0
	Accuracy [%]	88	83	100
LuSens	*n* total	6	5	5
	TP/FN	6/0	5/0	5/0
	TN/FP	0/0	0/0	0/0
	Accuracy [%]	100	100	100
h-CLAT	*n* total	27	12	12
	TP/FN	14/9	9/1	9/2
	TN/FP	2/2	2/0	0/1
	Accuracy [%]	59	92	75
(m)MUSST	*n* total	17	9	9
	TP/FN	10/5	7/1	7/1
	TN/FP	2/0	1/0	0/1
	Accuracy [%]	64	75	78
'2 out of 3' approach	*n* total	25	12	12
	TP/FN	18/3	10/0	10/0
	TN/FP	1/3	2/0	0/2
	Accuracy [%]	76	100	83

n, number of substances analyzed; TN/TP, true negative/positive; FN/FP, false negative/positive.

non-sensitizers (Hoffmann, 2014). Similarly, by retesting 22 LLNA performance standards in the standard LLNA protocol, a reproducibility of only 77% was found for the investigated substances (Kolle et al., 2013).

In addition to evaluating the results of the single test methods, the overall yes/no-ratings of the '2 out of 3' prediction models from the previously published studies and the current study was compared (Table 12). The '2 out of 3' prediction model in Bauch et al. (2012) consisted of the DPRA, LuSens and mMUSST. An overall accuracy of 83% and 94% compared to LLNA or human data, respectively, was achieved. Natsch et al. (2013) used the DPRA, KeratinoSens™ and MUSST in their '2 out of 3' prediction model and reported a similar accuracy for 145 substances (81% compared to LLNA). In the current study, the results of the DPRA, KeratinoSens™ and h-CLAT were used for applying this prediction model. Table 12 illustrates, that all three '2 out of 3' prediction models provide very similar and high overall accuracies. Even though the dataset was expanded from 54 to 180 substances the accuracy is still similar when compared to the LLNA; 79% compared to 83% in the Bauch et al. study and 81% in the Natsch et al. study. Compared to human data, the accuracies reported in Bauch et al. (2012) and in the current study are very similar (94% and 90%, respectively), although the current analysis is based on an almost twofold larger dataset. The higher accuracy when comparing to human instead of LLNA data indicates that the investigated non-animal test methods predict human data more accurately than LLNA data.

In addition, the direct comparison of LLNA data to human data revealed an overall accuracy of 82% of the LLNA. The '2 out of 3' prediction model however revealed an accuracy of 90% when compared to human data for the same set of substances. This indicates that the '2 out of 3' predictions might be more relevant for humans than the LLNA predictions. However, some of the substances with different *in vivo* outcomes are pre-haptens (e.g. limonene, citronellol or linalool). This could be one reason for discordant LLNA and human results, since pre-hapten activation could be prevented by precautionary measures in handling or storage (Gerberick et al., 2008). But this does not imply that pre-haptens are not possibly a risk for humans. Substances leading to different predictions in the LLNA, human tests and the '2 out of 3' are discussed in more detail in the following part.

4.2. False negative and false positives

The discussion of false predictions provided by the single non-animal test methods would be beyond the scope of this study. Hence, the analysis was restricted to the detailed investigation of false positive and false negative predictions by the '2 out of 3' approach, when compared to LLNA and human data. The 28 substances leading to false negative results are divided into 4 groups and possible explanations for the discrepancy are discussed individually (Table 13). In brief, in group FN-1 negative human data confirm the negative overall results of the non-animal test methods for 11 substances (the LLNA is over-predictive for this set of substances). For the three substances in group FN-2 (diethylenetriamine, resorcinol and benzoyl peroxide) concordant human and

Table 12
Performances of the investigated non-animal test methods and the '2 out of 3' approach in different datasets.

		Bauch et al. (2012)		Natsch et al. (2013)		This paper	
		Acc [%]	*n*	Acc [%]	*n*	Acc [%]	*n*
Compared to LLNA data							
Peptide reactivity	DPRA	79	54	80	145	75	194
KC activation	KeratinoSens™	81	54	77	145	73	188
	LuSens	77	54	–	–	73	77
DC activation	(m)MUSST	74	54	71	141	73	149
	h-CLAT	–	–	–	–	76	166
Compared to human data							
Peptide reactivity	DPRA	86	51	–	–	84	102
KC activation	KeratinoSens™	80	51	–	–	82	102
	LuSens	84	51	–	–	79	61
DC activation	(m)MUSST	86	51	–	–	78	85
	h-CLAT	–	–	–	–	82	98
Prediction model							
'2 out of 3' approach[a] (vs. LLNA data)		83	54	81	145	79	180
'2 out of 3' approach[a] (vs. human data)		94	51	–	–	90	101

Acc, accuracy; *n*, number of analyzed substances; KC, keratinocyte; DC, dendritic cell; "–", no data available or data not considered in this study.
[a] 2 out of 3' prediction model in Bauch et al. (2012) (DPRA, LuSens, mMUSST) and Natsch et al. (2013) (DPRA, KeratinoSens™, MUSST) were slightly different compared to the prediction model underlying this paper (DPRA, KeratinoSens™, h-CLAT).

Table 13
False negative predictions within the '2 out of 3' approach.

Chemical name	CAS #	LLNA EC3	Discussion
Group FN-1	'2 out of 3' approach (−), LLNA (+), human (−)		
Phthalic anhydride	85-44-9	0.16	Respiratory sensitizer, but no skin sensitizer (Dearman et al., 2013); LLNA probably FP
α-*iso*-Methylionone	127-51-5	21.8	In HRIPT[a] no induction at 70,866 μg/cm^2; LLNA probably FP
Xylene	1330-20-7	95.8	No evidence for skin sensitization in humans (Basketter et al., 2014); LLNA probably FP
Pyridine	110-86-1	71.2	No evidence for skin sensitization in humans (Schneider and Akkan, 2004); LLNA probably FP
Isopropyl myristate	110-27-0	44	No evidence for skin sensitization in humans (Basketter et al., 2014); LLNA probably FP
Linalool	78-70-6	30	No or weak evidence in humans with no induction in HRIPT[a] at 15,000 μg/cm^2; LLNA possibly FP
Sodium dodecyl sulfate (SDS)	151-21-3	14	Well characterized false positive response in LLNA, no evidence for skin sensitization in humans (Basketter et al., 2014); LLNA probably FP
Hexyl salicylate	6259-76-3	0.18	In HRIPT[a] no induction at 35,433 μg/cm^2 and in HMT[a] no induction at 20,654 μg/cm^2; LLNA probably FP
DL-α-Tocopherol	10191-41-0	7.4	No evidence for skin sensitization in humans (Basketter et al., 2014); LLNA probably FP
Benzyl benzoate	120-51-4	17	No or only weak evidence in humans (Basketter et al., 2014) with no induction in HRIPT[a] at 59000 μg/cm^2; LLNA probably FP
Benzyl salicylate	118-58-1	2.9	No or only weak evidence in humans (Basketter et al., 2014); in HRIPT[a] no induction at 17,717 μg/cm^2 and in HMT[a] no induction at 20,690 μg/cm^2; LLNA probably FP
Group FN-2	'2 out of 3' approach (−), LLNA (+), human (+)		
Benzoyl peroxide	94-36-0	0.004	Well-known contact allergen in humans (Basketter et al., 2014), clearly positive in DPRA
Resorcinol	108-46-3	5.5	Known contact allergen in humans after considerable exposure (Basketter et al., 2014); positive in h-CLAT and MUSST[b]
Diethylenetriamine	111-40-0	5.8	Positive in HMT (Kligman, 1966b), all non-animal tests negative, putative pro-hapten
Group FN-3	'2 out of 3' approach (−), LLNA (+), human (no or conflicting data)		
α-Hexyl cinnamic aldehyde	101-86-0	11.97	Inconclusive human data; borderline in KeratinoSens™, positive in LuSens[2] and MUSST[b]
N,N-Dibutylaniline	613-29-6	19.6	Putative pro-hapten, all non-animal tests negative
2-Methoxy-4-methylphenol	93-51-6	5.8	Putative pro-hapten, positive in h-CLAT and MUSST[b]
3-Aminophenol	591-27-5	3.2	Putative pro-hapten, positive in h-CLAT and MUSST
2,2,6,6-Tetramethyl-3,5-heptanedione	1118-71-4	27	All non-animal tests negative
3-Methyl-1-phenylpyrazolone	89-25-8	8.5	All non-animal tests negative
Undec-10-enal	112-45-8	6.8	Positive in KeratinoSens™
Squaric acid diethyl ester	5231-87-8	0.9	All non-animal tests negative
Methyl pyruvate	600-22-6	2.4	All non-animal tests negative
Benzyl cinnamate	103-41-3	18.4	Positive in KeratinoSens™
Group FN-4	'2 out of 3' approach (−), LLNA (−), human (+)		
Coumarin	91-64-5	NC	Well-known contact allergen in humans (Basketter 2014-62), positive in KeratinoSens™
Streptomycin sulfate	3810-74-0	NC	Positive in HMT (Kligman, 1966a)
Kanamycin	8063-07-8	NC	Known contact allergen in humans after considerable exposure (Schneider and Akkan, 2004)
Benzyl alcohol	100-51-6	NC	Positive in HMT[a] at 8858 μg/cm^2

[a] HRIPT and HMT data were found in the data base of the Research Institute for Fragrance Materials (RIFM).
[b] DPRA, KeratinoSens™ and h-CLAT results were considered for '2-out-of-3' approach.

LLNA data are available. In this case, *in vivo* evidence gives a strong hint that the negative outcome in the '2 out of 3' approach is false. Diethylenetriamine and resorcinol are putative pro-haptens. They are not detectable in the cell-free DPRA and may not be activated in the cellular assays due to a limited metabolic capacity (Fabian et al., 2013; Oesch et al., 2014). Benzoyl peroxide was strongly positive in the DPRA but negative in the KeratinoSens™ and h-CLAT. The ten substances in group FN-3 were tested positively by LLNA, but no human data are available. Group FN-4 contains four substances which are true negatives when compared to LLNA data but false negatives, when compared to human data. Among these are the two water soluble antibiotics streptomycin and kanamycin. Both are negative in all investigated non-animal tests and also in the LLNA, but decades of human use indicate an incidence for an ACD after considerable exposure (Kligman, 1966b; Schneider and Akkan, 2004). These oligoaminoglucosides are structurally significantly different from all other known haptens and pre-/pro-haptens and it is still not clear whether they sensitize by the classical hapten-based mechanism.

Similar to the discussion of the false negative results, also the overall 15 false positive substances could be divided into 4 groups (Table 14). In group FP-1 positive human data confirm the positive overall results of the non-animal test methods and identify the LLNA to be under-predictive for the three substances benzocaine, benzaldehyde and nickel chloride in this group. Benzocaine and benzaldehyde are known contact allergens in humans but only after considerable exposure (Griem et al., 2003) (RIFM database). Hence, the possibility of a false negative response in the LLNA has to be considered. Nickel chloride directly activates the human Toll-like receptor 4 (TLR4). As the mouse cannot mount this TLR4 based response (Schmidt et al., 2010), this indicates a true and well recognized false negative LLNA result. Group FP-2 with concordant

negative human and negative LLNA data just contains propyl paraben, which was tested positively in the KeratinoSens™ and h-CLAT. Although some positive human patch test results are described in the literature, the frequent use of parabens in general corroborates this class to be non-sensitizers from a regulatory point of view (Schnuch et al., 2011; Basketter et al., 2006). Group FP-3 contains nine substances which were predicted to be sensitizers within the '2 out of 3' approach but are non-sensitizers according to LLNA data. For these substances no human data were found in the literature. Peptide adduct formation was detected in a modified peptide reactivity test using LC–MS detection (Natsch et al., submitted for publication) for six substances within this group (i.e. 2-hydroxypropyl methacrylate, 2-acetyl-cyclohexanone, furil, 1-bromobutane, 1-iodohexane and methyl-3-bromopropionate) which indicates that the positive outcome of the '2 out of 3' prediction model might possibly be correct as these are clearly protein-modifying haptens. Group FP-4 contains R(+)-limonene and D,L-citronellol. For both, negative human data are described in the literature (Basketter et al., 2014). Thus, human data are supposed to overrule the positive overall outcome of the non-animal test methods as well as the positive LLNA results. For citronellol, also a recent LLNA on highly pure material was negative (Rudback et al., 2014). However, under the EU Cosmetics Regulation both limonene and citronellol are considered to be allergens although only oxidative metabolites may be reactive suggesting probable pre-/pro-haptens.

4.3. Integrated and sequential testing strategies

From the 144 chemicals with clear-cut results in all three tests (DPRA, KeratinoSens™ and h-CLAT), a congruent result in all three tests was obtained in 76 cases ('3 out of 3'), while for 68 cases a '2 out of 3' assessment was made. For an additional set of 36 chemicals the prediction model is based on a '2 out of 2' assessment from either the combination of DPRA and KeratinoSens™ ($n = 24$), KeratinoSens™ and h-CLAT ($n = 11$) or DPRA and h-CLAT ($n = 1$).

The high overall accuracy of the '2 out of 3' approach indicates that in many cases positives or negatives in single assays are actually FP or FN, respectively, what underlines the importance of making an majority voting. False-positives in the different assays might be due to different mechanisms; thus in KeratinoSens™ unspecific activation of the antioxidant response due to other mechanisms than covalent modification of Keap1 is possible, while false-positives in DPRA may be generated by unspecific peptide oxidation. In the h-CLAT, non-sensitizing irritants such as octanoic acid may also lead to surface marker expression. False negatives results might occur due to solubility issue or limited metabolic or oxidative activation. Among the 68 substances relying on two concordant of three results ('2 out of 3'), the DPRA and KeratinoSens™ rule the '2 out of 3' overall prediction in 24 cases, the DPRA and h-CLAT in 29 cases and the KeratinoSens™ and h-CLAT in 15 cases.

At the same time this analysis indicates that to arrive at the final conclusions often not all three tests are needed, as with two congruent tests the third assay can be waived as it would not change the assessment. This was actually applied in 36 cases with a congruent result in 2 out of 2 tests.

4.4. Interchangeabilities

Within the current dataset both the KeratinoSens™ and the LuSens assay cover the AOP key event 'activation of keratinocytes'. For a common subset of 69 substances an interchangeability of 88% could be calculated. Only 8 test substances did not provide concordant data among both assays (Table 11). Reasons for the different results might be due to differences in the test procedures (e.g. cytotoxicity range finders), the nature of the used ARE sequence in the keratinocyte cell line (e.g. in KeratinoSens™ and LuSens derived from human or rat, respectively), putative differential metabolic capacities of the cell lines (four substances were pro-haptens) or incidental (borderline read-out in one case). In addition, some of these differences could be also related to the different prediction models used, for instance, the KeratinoSens™ only requires one single concentration of the test substance yielding an induction higher than 1.5-fold whereas the LuSens requires at least two consecutive concentrations. A detailed discussion of the similarity of both assays can be found in Ramirez et al. (2014). Compared to human and LLNA data, both assays provide comparable Cooper statistics. And also the use of the LuSens assay instead of the KeratinoSens™ within a '2 out of 3' approach leads to similar accuracies, although the available dataset of the LuSens is smaller compared to the KeratinoSens™.

An interchangeability of only 72% was calculated for the h-CLAT and the (m)MUSST that both cover the AOP key event 'activation of dendritic cells'. The MUSST and the mMUSST use slightly different protocols, prediction models and cell lines. For the analyses within this study, the results from the MUSST and mMUSST were taken together and discordant results were excluded. When comparing the results of a common subset to *in vivo* data, the (m)MUSST provides a slightly increased specificity whereas the h-CLAT provides an increased sensitivity and overall accuracy. Reasons for the higher sensitivity of the h-CLAT might be the additional marker CD54 and the suitability for this specific common subset of substances. A detailed analysis of the CD54 and CD86 induction might provide further valuable information.

4.5. Mechanistic domains

In order to analyze if the non-animal test methods may detect classes of substances with different reaction mechanisms with similar performances, nine different mechanistic domains were defined by probable protein-binding mechanisms of the 213 substances underlying this study (Table 4). One aspect which should be taken into account is that the predictivities are influenced by the number and type of substances being assessed and data sets may therefore vary in this study. No reaction mechanism was assigned for a group of substances ($n = 65$) with a lack of obvious structural characteristics associated with skin sensitization. However, some of the substances in this group are sensitizing in the LLNA or in humans (e.g. hexyl salicylate). *In vivo* evidence for hexyl salicylate resulting from a single LLNA test provides the lowest EC3 value within this substance group (EC3 = 0.18, compare RIFM data base). Like the structurally similar methyl salicylate, which was tested negatively in the LLNA, the chemical structure of hexyl salicylate reveals no obvious alert for peptide reactivity. Irritation is a confounding factor in the LLNA since it leads to overestimations of sensitization potentials (Ball et al., 2011). In addition, hexyl salicylate was negative even at high concentrations in HRIPT (NOEL $\approx$ 35,400 $\mu g/cm^2$) and HMT (NOEL $\approx$ 20,600 $\mu g/cm^2$) (compare RIFM database). Therefore, the very low EC3 may well be due to its irritating properties or possibly also due to sensitizing impuities. Another substance which was sensitizing in both human clinical trials and in the murine LLNA is abietic acid. Nevertheless, abietic acid itself is considered to be a non-sensitizer, but depending on storage, sensitization via hydro peroxides derived from autoxidation is probable (Roberts et al., 2007a). A further chemical with a positive LLNA outcome in this group is the well-characterized irritant SDS (EC3 = 14%) (Gerberick et al., 2005; Ball et al., 2011). SDS is considered to be the classic example of a substance yielding a false positive response in the LLNA, what is also confirmed by negative human patch tests (Basketter et al., 2014). Likewise, anhydrous oxalic acid was tested positively in the LLNA (EC3 = 15). This substance was further analyzed for its capacity to

Table 14
False positive predictions within the '2 out of 3' approach.

Chemical name	CAS #	LLNA EC3	Discussion
Group FP-1	'2 out of 3' approach (+), LLNA (−), human (+)		
Nickel chloride	7718-54-9	NC	Well characterized false negative response in murine LLNA, most common contact allergen in humans (Griem et al., 2003); LLNA probably FN
Benzocaine	94-09-7	NC	Known contact allergen in humans after considerable exposure (Griem et al., 2003), all non-animal tests positive; LLNA probably FN
Benzaldehyde	100-52-7	NC	Positive HRIPT1[a] at 2760 μg/cm^2, positive in KeratinoSens™, h-CLAT and MUSST; LLNA probably FN
Group FP-2	'2 out of 3' approach (+), LLNA (−), human (−)		
Propyl paraben	94-13-3	NC	No or only weak evidence in humans (Basketter et al., 2014)
Group FP-3	'2 out of 3' approach (+), LLNA (−), human (no or conflicting data)		
2-Hydroxypropyl methacrylate	923-26-2	NC	Only h-CLAT negative; peptide adduct formation in LC–MS test (Natsch et al., submitted for publication)
3-Phenoxypropiononitrile	3055-86-5	NC	Only KeratinoSens™ negative
2-Acetyl-cyclohexanone	874-23-7	NC	All non-animal tests positive; peptide adduct formation in LC–MS test (Natsch et al., submitted for publication)
4-Methyl-2-nitroanisole	119-10-8	NC	Only DPRA negative
2-Fluoro-5-nitroaniline	369-36-8	NC	Only h-CLAT negative
Furil	492-94-4	NC	Only h-CLAT negative; peptide adduct formation in LC–MS test (Natsch et al., submitted for publication)
1-Bromobutane	109-65-9	NC	Only KeratinoSens™ and MUSST[b] negative; highly volatile; peptide adduct formation in LC–MS test (Natsch et al., submitted for publication)
1-Iodohexane	638-45-9	NC	Only MUSST negative; peptide adduct formation in LC–MS test (Natsch et al., submitted for publication)
Methyl-3-bromopropionate	3395-91-3	NC	Only h-CLAT negative; peptide adduct formation in LC–MS test (Natsch et al., submitted for publication)
Group FP-4	'2 out of 3' approach (+), LLNA (+), human (−)		
R(+)-Limonene	5989-27-5	69	No or only weak evidence in humans (Basketter et al., 2014); putative pro-hapten
D,L-Citronellol	106-22-9	43.5	No or only weak evidence in humans (Basketter et al., 2014); putative pro-hapten, recent negative LLNA (Rudback et al., 2014)

[a] HRIPT and HMT data were found in the data base of the Research Institute for Fragrance Materials (RIFM).
[b] DPRA, KeratinoSens™ and h-CLAT results were considered for '2-out-of-3' approach.

covalently bind to reactive side chains of model peptides but did not form any adducts in an LC–MS test (Natsch et al., submitted for publication) and also provided negative results in the DPRA.

For this mechanistic group, accuracies of the single non-animal test methods are in a range of 65–68% when compared to LLNA data and 56–84% when compared to human data. The accuracies provided by the '2 out of 3' approach are 73% and 80% when compared to LLNA or human data, respectively. If such substances as exemplified above with known *in vivo* sensitization potential and a theoretical non-binding capacity were excluded from this group, cooper statistics would have been higher. However, these data cannot be disregarded. Thus, the aim should be extension of the parameters used for assessment of the binding capacities and include factors such as abiotic or enzymatic activation processes.

In contrast to the above discussed group of substances with a lack of obvious structural characteristics being associated with peptide reactivity, the following mechanistic domains containing Michael acceptors, acylating agents, Schiff 'base formers, quinone precursors and substances reacting in nucleophilic substitutions are discussed. The substances within the domain of Michael acceptors were predicted with a generally high accuracy of at least 80% by all of the non-animal test methods. In this domain, human and LLNA data are also concordant for most of the substances.

Somewhat lower accuracies were calculated for acylating agents. The low accuracies of the KeratinoSens™ (56%, compared to LLNA data) and LuSens (44%, compared to LLNA data) are related to a lack of activation of the Keap1/Nrf-2 pathway. Molecular pathway activation triggered in KeratinoSens™ and LuSens is linked to cysteine reactivity with the Keap 1 sensor protein. However, acylating agents like anhydrides transfer their acyl moiety predominantly to lysine residues (Emter et al., 2013; Aptula et al., 2005). Seven substances of the false negatives in the keratinocyte-based assays show considerably increased lysine reactivity in the DPRA. In this mechanistic domain more weight should be given to DPRA results, since the accuracy in this domain was 100% or 82%, when compared to LLNA or human data, respectively. Similar to acylating agents, short chain aldehydes and longer chain saturated alkanals in the domain of Schiff 'base formers represent hard electrophiles preferring to react with hard nucleophiles like lysine residues instead of cysteine residues (LoPachin and Gavin, 2014). In general, the battery of non-animal test methods appears to be more sensitive to cysteine-reactive substances, as the DPRA has a readout depending on cysteine reactivity, the KeratinoSens™ and LuSens are also dependent on cysteine binding to a significant degree, but also the CD86 expression may be associated with cysteine reactivity (Natsch et al., 2013). This might explain the slightly lower accuracies of most of the non-animal test methods within the domain of Schiff 'base formers. Quinone precursors act as pro-Michael acceptors and must first be activated in order to become electrophilic, but other protein-binding mechanisms cannot be ruled out for some of the substances (Roberts et al., 2007a). Compared to the other non-animal test methods in this substance domain, the DPRA provided the lowest accuracy of 71%, when compared to LLNA data. This can be explained by the fact, that some members of this group might require enzymatic activation, which is absent in the *in chemico* assay. The cell-based test methods also have only limited metabolic capacities (Oesch et al., 2014; Fabian et al., 2013) and thus have limitations in detecting putative pro-haptens.

Regarding the domain of nucleophilic substitutions (S_N1 and 2), the DPRA yields false negative results for one pro-hapten (i.e. dimethylbenz[α]anthracene) and four benzylic esters with common structural alerts (i.e. benzyl benzoate, benzyl salicylate, butylbenzyl phthalate and benzyl cinnamate). Concerning the latter, the benzylic sp3 carbon atom is supposed to react as electrophile. The resulting benzylic cation is resonance stabilized what indicates a

S_N1 mechanism to be probable. Resulting highly reactive S_N1 electrophiles might spontaneously react with water or other rival nucleophilic reaction partners of the model peptides such as hydroxyl ions or solvents what would explain the negative outcome and the slightly decreased accuracy of the DPRA in this domain. This would also be an explanation for the slightly decreased accuracies found in the dendritic cell-based assays when compared to LLNA data, although DC activation is not necessarily be associated with protein binding potential of a compound. The domain of aromatics that react by nucleophilic substitutions (S_NAr, $n = 6$) is quite small in this data set, although further data are described in the literature (Roberts and Aptula, 2014). Also the number of metal ions forming coordination bonds is small ($n = 4$). In contrast to the DC and KC based assays, the applicability domain of DPRA is not defined for compounds containing metal ions.

5. Conclusions

The compilation of non-animal test results provides a comprehensive reference dataset with additional information such as physicochemical properties, types of use, proposed organic reaction mechanisms as well as related *in vivo* reference data for 213 substances.

This study confirms the utility of the five investigated non-animal test methods, i.e. DPRA, KeratinoSens™, LuSens, h-CLAT and (m)MUSST, to predict the respective AOP key event with a high accuracy. When implemented into a '2 out of 3' test strategy, skin sensitizers can be discriminated from non-sensitizers with a high reliability be the use of these alternative methods. For several substances the '2 out of 3' approach does not predict the outcome of the LLNA. For those, human and LLNA data only show a limited concordance. The direct comparison of both *in vivo* references demonstrates that the non-animal test methods predict human data more accurately (Accuracy = 90%) than LLNA data (Accuracy = 82%).

The expanded dataset was further used to define different mechanistic domains by probable protein-binding mechanisms. This approach shows that Michael acceptors, substances reacting in nucleophilic substitutions and quinone precursors were predicted with the highest accuracies. In the domain of Schiff 'base formers as well as in the group of substances with a lack of obvious alerts for peptide reactivity, accuracies were slightly decreased. In the domain of acylating agents, the keratinocyte based assays show mechanistically justifiable decreased predictivities. If a chemical is supposed to react by acylation, more weight should be given to the DC-based assays and especially the Lys reactivity in the DPRA. The number of substances tested in some specific groups is still low (only six aromatics reacting by nucleophilic substitutions (S_NAr) and four metal-containing complexes within this dataset). Overall, assigning a test substance to a domain according to its protein binding mechanisms offers a way to obtain a more accurate estimate of the predictive performance of the individual non-animal test methods as well as the overall '2 out of 3' prediction.

6. Outlook

In consideration of the obtained data, the presented strategy can be integrated in the regulatory assessment of the skin sensitization hazard potential. For this purpose the "2 out of 3" results should be interpreted under the consideration of the impact the mechanistic domain of the pertinent compound has on the outcome of the individual test method. Thus, the obtained experimental results of the non-animal test methods together with the reliability of the data based on the mechanistic domain provide a weight of evidence for predicting the hazard potential for the induction of skin sensitization.

Beside information for hazard assessment the reference standard dataset also contains concentration–response data on the non-animal test methods and potency information based on LLNA and in part on human data. Therefore, this reference database may further be used to develop prediction models for skin sensitization potential and in particular it may be used to analyze how to arrive at skin sensitization potency predictions based on current data.

Conflict of interest statement

The authors are not aware of any conflicts of interest.

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.2014.12.008.

References

Ade, N., Martinozzi-Teissier, S., Pallardy, M., Rousset, F., 2006. Activation of U937 cells by contact sensitizers: CD86 expression is independent of apoptosis. J. Immunotoxicol. 3, 189–197.
Angers-Loustau, A., Tosti, L., Casati, S., 2011. The regulatory use of the local lymph node assay for the notification of new chemicals in Europe. Regul. Toxicol. Pharmacol. 60, 300–307.
Aptula, A.O., Enoch, S.J., Roberts, D.W., 2009. Chemical mechanisms for skin sensitization by aromatic compounds with hydroxy and amino groups. Chem. Res. Toxicol. 22, 1541–1547.
Aptula, A.O., Patlewicz, G., Roberts, D.W., 2005. Skin sensitization: reaction mechanistic applicability domains for structure–activity relationships. Chem. Res. Toxicol. 18, 1420–1426.
Aptula, A.O., Roberts, D.W., Pease, C.K., 2007. Haptens, prohaptens and prehaptens, or electrophiles and proelectrophiles. Contact Dermatitis 56, 54–56.
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.
Ball, N., Cagen, S., Carrillo, J.C., Certa, H., Eigler, D., Emter, R., Faulhammer, F., Garcia, C., Graham, C., Haux, C., Kolle, S.N., Kreiling, R., Natsch, A., Mehling, A., 2011. Evaluating the sensitization potential of surfactants: integrating data from the local lymph node assay, guinea pig maximization test, and in vitro methods in a weight-of-evidence approach. Regul. Toxicol. Pharm. 60 (3), 389–400.
Basketter, D., Alepee, N., Casati, S., Crozier, J., Eigler, D., Griem, P., Hubesch, B., de, K.J., Landsiedel, R., Louekari, K., Manou, I., Maxwell, G., Mehling, A., Netzeva, T., Petry, T., Rossi, L.H., 2013. Skin sensitisation – moving forward with non-animal testing strategies for regulatory purposes in the EU. Regul. Toxicol. Pharmacol. 67, 531–535.
Basketter, D.A., Alepee, N., Ashikaga, T., Barroso, J., Gilmour, N., Goebel, C., Hibatallah, J., Hoffmann, S., Kern, P., Martinozzi-Teissier, S., Maxwell, G., Reisinger, K., Sakaguchi, H., Schepky, A., Tailhardat, M., Templier, M., 2014. Categorization of chemicals according to their relative human skin sensitizing potency. Dermatitis 25, 11–21.
Basketter, D.A., McFadden, J., Evans, P., Andersen, K.E., Jowsey, I., 2006. Identification and classification of skin sensitizers: identifying false positives and false negatives. Contact Dermatitis 55, 268–273.
Bauch, C., Kolle, S.N., Fabian, E., Pachel, C., Ramirez, T., Wiench, B., et al., 2011. Intralaboratory validation of four *in vitro* assays for the prediction of the skin sensitizing potential of chemicals. Toxicol. In Vitro 25 (6), 1162–1168.
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.
Bruckner, A.L., Weston, W.L., Morelli, J.G., 2000. Does sensitization to contact allergens begin in infancy? Pediatrics 105, e3.
Cooper, J.A., Saracci, R., Cole, P., 1979. Describing the validity of carcinogen screening-tests. Br. J. Cancer 39, 87–89.
Dearman, R., Basketter, D., Kimber, I., 2013. Inter-relationships between different classes of chemical allergens. J. Appl. Toxicol. 33, 558–565.
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.
Emter, R., van der Veen, J.W., Adamson, G., Ezendam, J., van Loveren, H., Natsch, A., 2013. Gene expression changes induced by skin sensitizers in the KeratinoSens cell line: discriminating Nrf2-dependent and Nrf2-independent events. Toxicol. In Vitro 27, 2225–2232.
Fabian, E., Vogel, D., Blatz, V., Ramirez, T., Kolle, S., Eltze, T., van, R.B., Oesch, F., Landsiedel, R., 2013. Xenobiotic metabolizing enzyme activities in cells used for testing skin sensitization in vitro. Arch. Toxicol. 87, 1683–1696.

Foussereau, J., Samsoen, M., Hecht, M.T., 1983. Occupational dermatitis to Ampholyt G in hospital personnel. Contact Dermatitis 9, 233–234.
Gerberick, F., Aleksic, 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. Altern. Lab. Anim. 36, 215–242.
Gerberick, G.F., Ryan, C.A., Kern, P., Schlatter, H., Dearman, R., Kimber, I., Patlewicz, G., Basketter, D., 2005. Compilation of historical local lymph node data for evalutaion of skin sensitization alternative methods. Dermatitis 16, 157–202.
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.
Goebel, C., Aeby, P., Ade, N., Alepee, N., Aptula, A., Araki, D., Dufour, E., Gilmour, N., Hibatallah, J., Keller, D., Kern, P., Kirst, A., Marrec-Fairley, M., Maxwell, G., Rowland, J., Safford, B., Schellauf, F., Schepky, A., Seaman, C., Teichert, T., Tessier, N., Teissier, S., Weltzien, H.U., Winkler, P., Scheel, J., 2012. Guiding principles for the implementation of non-animal safety assessment approaches for cosmetics: skin sensitisation. Regul. Toxicol. Pharmacol. 63, 40–52.
Griem, P., Goebel, C., Scheffler, H., 2003. Proposal for a risk assessment methodology for skin sensitization based on sensitization potency data. Regul. Toxicol. Pharmacol. 38, 269–290.
Hoffmann, S., 2014. LLNA variability: an essential ingredient for a comprehensive assessment of non-animal skin sensitisation test methods and strategies. Altex Proceedings 3, 1/14, Prague 2014, II-12-110.
Jäckh, C., Fabian, E., van Ravenzwaay, B., Landsiedel, R., 2012. Relevance of xenobiotic enzymes in human skin in vitro models to activate pro-sensitizers. J. Immunotoxicol. 9 (4), 426–438.
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. 33 (11), 1353–1364.
Karlberg, A.T., Borje, A., Duus, J.J., Liden, C., Rastogi, S., Roberts, D., Uter, W., White, I.R., 2013. Activation of non-sensitizing or low-sensitizing fragrance substances into potent sensitizers – prehaptens and prohaptens. Contact Dermatitis 69, 323–334.
Kinsner-Ovaskainen, A., Maxwell, G., Kreysa, J., Barroso, J., Adriaens, E., Alepee, N., Berg, N., Bremer, S., Coecke, S., Comenges, J.Z., Corvi, R., Casati, S., Dal, N.G., Marrec-Fairley, M., Griesinger, C., Halder, M., Heisler, E., Hirmann, D., Kleensang, A., Kopp-Schneider, A., Lapenna, S., Munn, S., Prieto, P., Schechtman, L., Schultz, T., Vidal, J.M., Worth, A., Zuang, V., 2012. Report of the EPAA–ECVAM workshop on the validation of integrated testing strategies (ITS). Altern. Lab. Anim. 40, 175–181.
Kligman, A.M., 1966a. Identification of contact allergens by human assay. 3. Maximization test – a procedure for screening and rating contact sensitizers. J. Invest. Dermatol. 47, 393.
Kligman, A.M., 1966b. The identification of contact allergens by human assay. II. Factors influencing the induction and measurement of allergic contact dermatitis. J. Invest Dermatol. 47, 375–392.
Kolle, S.N., Basketter, D.A., Casati, S., Stokes, W.S., Strickland, J., van, R.B., Vohr, H.W., Landsiedel, R., 2013. Performance standards and alternative assays: practical insights from skin sensitization. Regul. Toxicol. Pharmacol. 65, 278–285.
LoPachin, R.M., Gavin, T., 2014. Molecular mechanisms of aldehyde toxicity: a chemical perspective. Chem. Res. Toxicol. 27, 1081–1091.
Mehling, A., Eriksson, T., Eltze, T., Kolle, S., Ramirez, T., Teubner, W., et al., 2012. Non-animal test methods for predicting skin sensitization potentials. Arch. of Toxicol. 86 (8), 1273–1295.
Mosmann, T., 1983. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. J. Immunol. Methods 65, 55–63.
Natsch, A., Bauch, C., Foertsch, L.M., Gerberick, G.F., Norman, K., Hilberer, H., Inglis, H., Landsiedel, R., Onken, S., Reuter, R., 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., Emter, R., 2008. Skin sensitizers induce antioxidant response element dependent genes: application to the in vitro testing of the sensitization potential of chemicals. Toxicol. Sci. 102, 110–119.
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 (11), 1337–1352.
Natsch, A. et al., submitted for publication. Predicting skin sensitizer potency based on in vitro data from KeratinoSens and kinetic peptide binding: global versus domain-based assessment. Toxicol. Sci.
Nukada, Y., Ashikaga, T., Miyazawa, M., Hirota, M., Sakaguchi, H., Sasa, H., Nishiyama, N., 2012. Prediction of skin sensitization potency of chemicals by human cell line activation test (h-CLAT) and an attempt at classifying skin sensitization potency. Toxicol. In Vitro 26, 1150–1160.
Nukada, Y., Miyazawa, M., Kazutoshi, S., Sakaguchi, H., Nishiyama, N., 2013. Data integration of non-animal tests for the development of a test battery to predict the skin sensitizing potential and potency of chemicals. Toxicol. In Vitro 27, 609–618.
Oesch, F., Fabian, E., Guth, K., Landsiedel, R., 2014. Xenobiotic-metabolizing enzymes in the skin of rat, mouse, pig, guinea pig, man, and in human skin models. Arch. toxicol. 88 (12), 2135–2190.
Patlewicz, G., Basketter, D.A., Smith, C.K., Hotchkiss, S.A., Roberts, D.W., 2001. Skin-sensitization structure–activity relationships for aldehydes. Contact Dermatitis 44, 331–336.
Patlewicz, G., Roberts, D.W., Uriarte, E., 2008. A comparison of reactivity schemes for the prediction skin sensitization potential. Chem. Res. Toxicol. 21, 521–541.
Ramirez, T., Mehling, A., Kolle, S.N., Wruck, C.J., Teubner, W., Eltze, T., Aumann, A., Urbisch, D., Ravenzwaay, B.V., 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 (8), 1482–1497.
Raunio, H., 2011. In silico toxicology – non-testing methods. Front. Pharmacol. 2, 33.
Roberts, D.W., Aptula, A.O., 2014. Electrophilic reactivity and skin sensitization potency of SNAr electrophiles. Chem. Res. Toxicol. 27, 240–246.
Roberts, D.W., Aptula, A.O., Patlewicz, G., 2007a. Electrophilic chemistry related to skin sensitization. Reaction mechanistic applicability domain classification for a published dataset of 106 chemicals tested in the mouse local lymph node assay. Chem. Res. Toxicol. 20, 44–60.
Roberts, D.W., Patlewicz, G., Kern, P.S., Gerberick, F., Kimber, I., Dearman, R.J., Ryan, C.A., Basketter, D.A., Aptula, A.O., 2007b. Mechanistic applicability domain classification of a local lymph node assay dataset for skin sensitization. Chem. Res. Toxicol. 20, 1019–1030.
Rovida, C., Alépée, N., Api, A.M., Basketter, D.A., Bois, F.Y., Caloni, F., et al., 2014. Integrated Testing Strategies (ITS) for safety assessment. ALTEX.
Rudback, J., Hagvall, L., Borje, A., Nilsson, U., Karlberg, A.T., 2014. Characterization of skin sensitizers from autoxidized citronellol – impact of the terpene structure on the autoxidation process. Contact Dermatitis 70, 329–339.
Schmidt, M., Raghavan, B., Muller, V., Vogl, T., Fejer, G., Tchaptchet, S., Keck, S., Kalis, C., Nielsen, P.J., Galanos, C., Roth, J., Skerra, A., Martin, S.F., Freudenberg, M.A., Goebeler, M., 2010. Crucial role for human Toll-like receptor 4 in the development of contact allergy to nickel. Nat. Immunol. 11, 814–819.
Schneider, K., Akkan, Z., 2004. Quantitative relationship between the local lymph node assay and human skin sensitization assays. Regul. Toxicol. Pharmacol. 39, 245–255.
Schnuch, A., Mildau, G., Kratz, E.M., Uter, W., 2011. Risk of sensitization to preservatives estimated on the basis of patch test data and exposure, according to a sample of 3541 leave-on products. Contact Dermatitis 65, 167–174.
Teubner, W., Mehling, A., Schuster, P.X., Guth, K., Worth, A., Burton, J., et al., 2013. Computer models versus reality: How well do *in silico* models currently predict the sensitization potential of a substance. Regul. Toxicol. Pharm. 67 (3), 468–485.
Thyssen, J.P., Linneberg, A., Johansen, J.D., 2007. The epidemiology of contact allergy in the general population – prevalence and main findings. Contact Dermatitis 57, 287–299.
Tsujita-Inoue, K., Hirota, M., Ashikaga, T., Atobe, T., Kouzuki, H., Aiba, S., 2014. Skin sensitization risk assessment model using artificial neural network analysis of data from multiple in vitro assays. Toxicol. In Vitro 28, 626–639.
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 (3), 371–379.

4.3 Assessment of pre- and pro-haptens using non-animal test methods for skin sensitization

Daniel Urbisch, Naveed Honarvar, Matthias Becker, Susanne N. Kolle, Annette Mehling, Britta Wareing, Wera Teubner, Robert Landsiedel

Summary of the publication: This publication analyses, if pro-haptens but also pre-haptens, which require molecular transformations to gain peptide reactivity, can be identified using non-animal methods. Differences in the metabolic capacities of the single non-animal methods are identified and activation mechanism of the pre- and pro-haptens are proposed based on structure elucidations of reaction products. Finally, the sensitivity of the single methods to predict direct-acting haptens vs. pre- and pro-haptens is compared.

Journal: Chemical Research in Toxicology (ACS Publications)

Impact factor: 3.025 (Journal Citation Reports® by Thomson Reuters, 2016)

Type of authorship: first author

Status of publication: published in 2016

My contribution to the publication: approximately 60%

- Identification of pre- and pro-haptens within a comprehensive dataset
- Performance of peptide analyses (DPRA) with subsequent analytics to elucidate the molecular structures of reaction products (e.g. LC-MS/MS, de-novo sequencing)
- Interpretation of mass spectra
- Calculation of Cooper statistics
- Detailed discussion of ‘true false positive’ results (= artefacts) and false negative results of the non-animal methods
- Estimation of the utility of the non-animal methods to identify pre- and pro-haptens
- Preparation of the manuscript including all figures and tables

Chemical Research in Toxicology

Article

pubs.acs.org/crt

Assessment of Pre- and Pro-haptens Using Nonanimal Test Methods for Skin Sensitization

Daniel Urbisch,[†] Matthias Becker,[†] Naveed Honarvar,[†] Susanne Noreen Kolle,[†] Annette Mehling,[‡] Wera Teubner,[§] Britta Wareing,[†] and Robert Landsiedel*[,†]

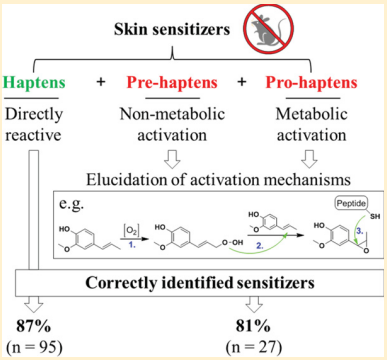
[†]Experimental Toxicology and Ecology, BASF SE, Ludwigshafen, Germany

[‡]BASF Personal Care and Nutrition GmbH, Düsseldorf, Germany

[§]BASF Schweiz AG, Basel, Switzerland

Supporting Information

ABSTRACT: Because of ethical and regulatory reasons, several nonanimal test methods to assess the skin sensitization potential of chemicals have been developed and validated. In contrast to *in vivo* methods, they lack or provide limited metabolic capacity. For this reason, identification of pro-haptens but also pre-haptens, which require molecular transformations to gain peptide reactivity, is a challenge for these methods. In this study, 27 pre- and pro-haptens were tested using nonanimal test methods. Of these, 18 provided true positive results in the direct peptide reactivity assay (DPRA; sensitivity of 67%), although lacking structural alerts for direct peptide reactivity. The reaction mechanisms leading to peptide depletion in the DPRA were therefore elucidated using mass spectrometry. Hapten−peptide adducts were identified for 13 of the 18 chemicals indicating that these pre-haptens were activated and that peptide binding occurred. Positive results for five of the 18 chemicals can be explained by dipeptide formations or the oxidation of the sulfhydryl group of the peptide. Nine of the 27 chemicals were tested negative in the DPRA. Of these, four yielded true positive results in the keratinocyte and dendritic cell based assays. Likewise, 16 of the 18 chemicals tested positive in the DPRA were also positive in either one or both of the cell-based assays. A combination of DPRA, KeratinoSens, and h-CLAT used in a 2 out of 3 weight of evidence (WoE) approach identified 22 of the 27 pre- and pro-haptens correctly (sensitivity of 81%), exhibiting a similar sensitivity as for directly acting haptens. This analysis shows that the combination of *in chemico* and *in vitro* test methods is suitable to identify pre-haptens and the majority of pro-haptens.



1. INTRODUCTION

The evaluation of skin sensitization potential is an essential step to define adequate safety measures for chemicals which may accidentally or purposely come into contact with the skin. Currently, animal tests are generally used to identify and characterize the sensitization potential since they represent internationally accepted test procedures provided in the form of OECD test guidelines (OECD TG) 406 (i.e., the guinea pig maximization test (GPMT) and Buehler test) and 429, and 442A and B (i.e., the murine local lymph node assay (LLNA) and its modifications), with OECD TG 429 being the default method for European legislations (Registration, Evaluation, Authorisation and Restriction of Chemical (REACH; EC 1907/2006)). Because of legal (e.g., the European Cosmetics Regulation)[1] but also ethical considerations, the use of nonanimal test methods is now becoming increasingly essential to assess the sensitizing potentials of chemicals. Research on the underlying molecular and cellular mechanisms has led to a good understanding of the processes involved, and the main events leading to the adverse outcome pathway (AOP) for skin sensitization have now been formally described by the OECD.[2,3] In the meantime, OECD TGs have been published for the direct peptide reactivity assay (DPRA),[4] an *in chemico* test method representing the first key event of the AOP, i.e., "peptide reactivity", and the "KeratinoSens", a cell-based ARE-Nrf2 luciferase test method representing the second key event of the AOP, i.e., "keratinocyte activation"(OECD TG 442C and OECD TG 442D, respectively).[5,6] This key event is also represented by the LuSens assay, which is in an advanced stage of validation.[7,50] The human cell line activation test (h-CLAT) represents the third AOP key event "dendritic cell activation" and is currently a draft OECD TG.[8,9] In contrast to the DPRA, which does not include metabolic activation, the cell lines used in the other tests do possess a certain degree of metabolic capacity. N-acetyl transferase (NAT-1) and esterase activities were detected in all three cell lines, whereas aldehyde dehydrogenase (ALDH) activities were additionally detected in KeratinoSens and LuSens cells; however, cytochrome P450 (CYP) and flavin-dependent monooxygenase (FMO) activities were below the limit of detection.[10,11] In this context, it should be noted that human skin itself has much lower levels of

Received: February 17, 2016

DOI: 10.1021/acs.chemrestox.6b00055
Chem. Res. Toxicol. XXXX, XXX, XXX−XXX

xenobiotic-metabolizing enzymes than the liver and that CYP enzymes are generally only expressed at low levels.[12,13]

Recently, a comprehensive data compilation which reported a wealth of information on the use of these nonanimal test methods for distinguishing sensitizers from nonsensitizers was published.[14] The accuracies of the individual assays addressing the first three key events of the AOP are already good, but using combinations of these tests results in accuracy in predicting skin sensitization potential in humans which are comparable to or even exceed the accuracy of the LLNA. A simple but very effective prediction model is to use the results of the three nonanimal test methods in a 2 out of 3 weight of evidence (WoE) integrated testing strategy (ITS) approach, a so-called "fixed data interpretation procedure" (DIP) within an "integrated approach to testing and assessment" (IATA).[15] Taking a closer look at the skin sensitizers in this data compilation, the majority of chemicals contains electrophilic moieties in their molecular structures. These are directly acting haptens which covalently bind to cutaneous carrier proteins to form immunogenic hapten−protein complexes (i.e., adducts) without additional molecular transformation. Although the formation of adducts is described as being the molecular initiating event (MIE) in the skin sensitization process,[2] a considerable number of skin sensitizers lack electrophilic moieties in their parent structures and are thus not intrinsically peptide reactive. This implies that a certain form of activation is required to transform parts of the molecular structures to electrophilic moieties capable of binding to nucleophilic residues of cutaneous proteins. In order to differentiate mechanisms of activation, two terms are commonly used: (*i*) pre-haptens; these chemicals require abiotic activation by chemical or physical means like the simple contact with air oxygen, dissolved oxygen, or radiation to become electrophilic and are predominantly activated before coming into contact with an organism;[16,17] (*ii*) pro-haptens; these chemicals require biotic modifications as a result of a contact to specific cutaneous enzymes.[18] Both processes may actually act on the same chemicals in certain cases, and chemicals can be both pre- and pro-haptens.[19,20] Also, a number of chemicals with weak peptide reactivity were characterized, which form more potent haptens as a result of additional biotic or abiotic activation.[21]

This article describes the utility of nonanimal test methods to detect pre- and pro-haptens and postulates potential activation mechanisms. Therefore, a *de novo* sequencing approach was first performed to verify that the sulfhydryl group is the sole nucleophilic center in the structure of the cysteine containing (Cys) peptide of the DPRA. Next, reaction products of chemicals tested positive in the DPRA were analyzed using mass spectrometry (LC-MS) to identify molecular mechanisms leading to peptide depletion. Finally, the performance of the DPRA compared to and in combination with the KeratinoSens assay and h-CLAT, both using partially metabolic competent cell lines, was assessed. The results of the three test methods were also incorporated into the 2 out of 3 WoE approach, and the performance of the set of pre- and pro-haptens was then compared to the performance of the set of directly acting haptens.

2. MATERIAL AND METHODS

2.1. Selection of Chemicals. In a recently published data set, 27 putative pre- and pro-haptens were identified, which were (*i*) lacking structural alerts for direct peptide reactivity and (*ii*) gave positive results in the LLNA and, where data were available, also in human tests. The presence or absence of alerts for peptide reactivity was identified using the QSAR Toolbox and/or TIMES SS.[49] All detailed *in vivo*, *in chemico*, and *in vitro* data are listed in the Supporting Information of Urbisch et al.[14]

2.2. Direct Peptide Reactivity Assay (DPRA). The results of the DPRA have previously been published.[14,22,23] The test procedure is in line with OECD TG 442C (adopted in 2015) and was originally described by Gerberick and co-workers in 2004.[24] Briefly, chemicals or vehicle alone were incubated with model heptapeptides containing lysine (Lys) (Ac-RFAAKAA-COOH) or cysteine (Cys) (Ac-RFAACAA-COOH), and the mean peptide depletion or depletion with Cys alone (Cys-only model) was assessed after 24 h. Peptide concentrations were determined by HPLC. Peptide reactivity was reported as percent depletion based on the decrease in nonreacted peptide concentration in the sample relative to the average concentration measured in the control. Criteria defining a negative response and reactivity are defined by the OECD TG 442C and are listed in Table 1.

Table 1. Reactivity Classes for Peptide Binding As Defined for the *in Chemico* DPRA (OECD 442C)[a]

reactivity class	Cys (1:10)/Lys (1:50) prediction model according to OECD TG 442C [%]	Cys-only (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	$42.47 < D_{Cys/Lys} \leq 100$	$98.24 < D_{Cys} \leq 100$

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

In addition, the selected chemicals (Sigma-Aldrich, Germany) retested in the DPRA were also subjected to LC-MS analyses and *de novo* sequencing.

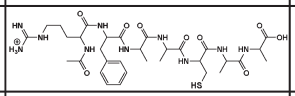
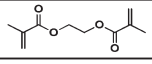
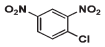
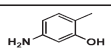
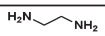
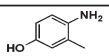
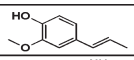
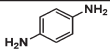
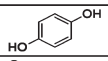
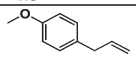
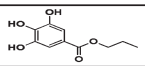
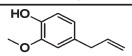
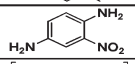
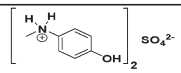
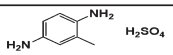
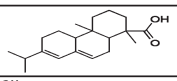
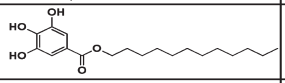
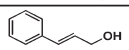
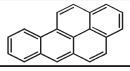
2.3. Mass Spectrometric Analyses. *2.3.1. De novo Sequencing.* *De novo* sequencing by MS is a method to characterize the structure of peptides and proteins and is reported to provide a higher sensitivity and a higher sample throughput than biological methods like Edman degradation.[25] In this approach, the precursor ion is selected and broken into defined fragments, predominantly a series of y- and b-ions, by low energy collision induced dissociation (CID) to determine the sequence of the peptide.[26] Fragments are formed as a result of partial or full side-chain losses from the precursor ion and can occur at each bond depending on the collisional activation. By comparing the defined fragments of the Cys peptide to the fragments of the hapten−peptide adduct, alterations in or losses of defined fragments identify the nucleophilic binding site of the peptide.

For the *de novo* sequencing approach, the samples of the DPRA standard procedure being stored in a freezer (−80 °C) were used undiluted. MS/MS analyses were performed on a TSQ 8000 Evo Triple Quadrupole mass spectrometer by CID between 25 and 60 eV. Spectra of the generated fragment ions were assessed using Excalibur-Software (Thermo Fischer Scientific).

2.3.2. LC-MS Analyses. A liquid chromatography−mass spectrometry (LC-MS) analysis was performed after a test chemical was incubated with the Cys peptide to qualitatively distinguish adducts from peptide oxidation products. This highly sensitive analytical technique combines the physical separation of complex mixtures by HPLC with capabilities of mass spectrometry for the identification of specific chemical structures with defined masses. Using electron spray ionization (ESI), charged molecules or fragments were generated and their mass-to-charge ratios (m/z) were detected. The m/z values of the peptide adducts differ from m/z values of the oxidation products. The DPRA coupled with a subsequent LC-MS analysis is useful to simultaneously detect peptide depletion and identify adducts and peptide oxidation products.[27]

DOI: 10.1021/acs.chemrestox.6b00055
Chem. Res. Toxicol. XXXX, XXX, XXX−XXX

Table 2. Overview of the Pre- and Pro-haptens Being Retested in DPRA According to the Cys-Only Model and Analyzed by Mass Spectrometry[a]

No.	Chemical structure	Test substance	Cas#	Molecular weight [g/mol]	LLNA EC3 [%]	Depletion detected with HPLC UV [%]	Depletion detected with LC-MS [%]
NC		Cys-containing heptapeptide (Ac-RFAACAA-OOH)	-	751.5 [Da]	-	-	-
PC1		Ethylene glycol dimethacrylate (EGDMA)	97-90-5	198.2	28.0	58.8	57.8
PC2		1-Chloro-2,4-dinitrobenzene (Dinitrochlorobenzene, DNCB)	97-00-7	202.6	0.04	100	100
1		5-Amino-2-methylphenol	2835-95-2	123.2	3.4	100	96.6
2		Ethylene diamine	107-15-3	60.1	2.2	18.3	34.6
3		4-Amino-m-cresol	2835-99-6	123.2	1.5	100	100
4		Isoeugenol	97-54-1	164.2	1.8	38.4	100
5		*para*-Phenylendiamine	106-50-3	108.1	0.16	100	89.1
6		Hydroquinone	123-31-9	110.1	0.10	100	100
7		4-Allylanisole	140-67-0	148.2	20.2	22.3	49.9
8		Propyl gallate	121-79-9	212.2	0.32	100	80.2
9		Eugenol	97-53-0	164.2	12.9	26.5	86.8
10		3-Methylcatechol	488-17-5	124.2	0.02	38.9	100
11		2-Nitro-1,4-phenylene diamine	5307-14-2	153.1	0.40	83.4	75.9
12		4-(Methylamino)phenol sulfate (Metol)	55-55-0	344.0	0.78	100	100
13		2,5-Diaminotoluene sulfate	615-50-9	156.2	0.40	100	100
14		Abietic acid	514-10-3	302.5	14.7	100	100
15		Lauryl gallate	1166-52-5	338.4	0.30	100	98.3
16		2-Aminophenol	95-55-6	109.1	0.40	63.9	59.7
17		Cinnamic alcohol	104-54-1	134.2	21.0	31.4	46.2
18		Benzo[a]pyrene	50-32-8	252.4	0.00009	61.3	52.5

[a]NC = negative control. PC = positive control.

For LC-MS analyses, the samples obtained from the standard DPRA procedure were stored in a freezer (−80 °C) and, after thawing, diluted 1:100 with water (HPLC grade) for final analyses. LC-MS analyses were performed using a TSQ 8000 Evo Triple Quadrupole mass spectrometer operated in the ESI(+) mode. An Ascentis C18 (50 × 2.1, 2.7 μm) was used as column for chromatographic separation. Mobile phase A consisted of 950 mL of acetonitrile mixed with 50 mL of water, and mobile phase B consisted of 950 mL of water, 50 mL of acetonitrile, and 0.1 mL of formic acid. The solvent flow rate was adjusted to 500 μL per minute with a gradient starting with 10% mobile phase A at t = 0 min and ending with 90% of the mobile phase A at t = 15 min. The divert valve was for the first minute and then from minutes 14 to 15. Mass spectra were acquired between minute 1 and minute 14 during the chromatographic run and assessed using Excalibur-Software (Thermo Fischer Scientific). In addition, MS/MS detection was performed for eugenol, benzo[a]pyrene, propyl gallate, lauryl gallate, and abietic acid. The theoretical calculations of the mass-to-charge [m/z] ratios of the expected adducts were performed with ChemDraw Prime 15.0.

2.4. Application of the 2 out of 3 Weight of Evidence (WoE) Approach. Experimental data covering the key events peptide reactivity, keratinocyte activation, and dendritic cell activation of the AOP was derived from the DPRA, KeratinoSens, and h-CLAT, respectively, and used in a WoE approach to predict the skin

DOI: 10.1021/acs.chemrestox.6b00055
Chem. Res. Toxicol. XXXX, XXX, XXX−XXX

sensitization potential.[14] According to the 2 out of 3 WoE approach, any two congruent results rule the overall assessment: If at least two results obtained in any of the three tests considered are positive, the chemical is rated to be a skin sensitizer. If at least two of the three tests are negative, the chemical is rated to be a nonsensitizer. In this study, the results for 27 putative pre- and pro-haptens were assessed using this prediction model.[28]

2.5. Statistics. Sensitivities were calculated for the individual *in chemico* and *in vitro* methods, as well as for the 2 out of 3 WoE approach by using Cooper statistics and data obtained from the LLNA (or human data) as the reference.[29] All parameters are based on a 2 × 2 contingency table counting the number of chemicals that are "true positive" (TP), "false positive" (FP), "true negative" (TN), and "false negative" (FN). The sensitivities of the 2 out of 3 WoE approach for a set of pre/pro-haptens and a set of directly acting haptens was also compared using Fisher's exact test.[30]

3. RESULTS

Recently, a comprehensive data compilation consisting of 151 skin sensitizers and 62 nonsensitizers based on LLNA data was published. Among the sensitizers, 27 chemicals were identified which (*i*) lacked structural alerts for direct peptide reactivity, but (*ii*) gave positive results in the LLNA and, where available, also in human tests. These selection criteria imply that peptide reactivity is not inherent but that activation of these chemicals takes place, e.g., after metabolic transformation *in vivo*. The abiotic and metabolic activation of these chemicals is also expected to be different in the nonanimal test methods DPRA, KeratinoSens or LuSens, and h-CLAT, due to the limited metabolic capacities of the test systems. Performances of these test methods as well as probable mechanisms leading to the activation of the pre- and pro-haptens were investigated and described in the following sections.

3.1. Mechanisms of Peptide Depletion in the DPRA.

3.1.1. Peptide Depletion. The DPRA provided positive results for 18 of the 27 selected chemicals lacking electrophilic moieties.[14] When testing according to OECD TG 442C, peptide depletion is detected by HPLC-UV analysis of the model peptides. Conclusions on the molecular structures of the modified peptide cannot be drawn by this read-out since only the loss of the native model peptide is detected. For that reason, the modified peptides were analyzed by LC-MS (Table 2).

For most of the 18 DPRA positives of this data set, Lys-depletion was below the cutoff of 6.38% and thus negative (data not shown). In addition, in all cases, where the Lys-depletion was >6.38%, Cys depletion was even higher; hence, the Cys peptide provided a clearly higher sensitivity (qualitatively and quantitatively) for this data set (of note: also in OECD TG 442C, a Cys-only model is proposed (compare Table 1)). As the Cys peptide is more discriminating, the LC-MS analyses of peptide-adducts were performed with the Cys peptide and are described in more detail in the following section.

For 15 of the 18 of the chemicals in this study-set yielding positive results in the DPRA, as well as for the two positive controls EGDMA (**PC1**) and DNCB (**PC2**), the quantitative results for depletion of the Cys peptide derived by the two detection methods HPLC-UV and LC-MS were comparable. For isoeugenol (**4**), eugenol (**9**), and 3-methylcatechol (**10**), however, peptide depletion detected with LC-MS was higher than depletion detected by HPLC-UV (Figure 1).

3.1.2. LC-MS Analyses of Modified Peptides. For a quantitative read-out of peptide depletion, LC-MS read-out in place of the HPLC-UV read-out was used. On the basis of the

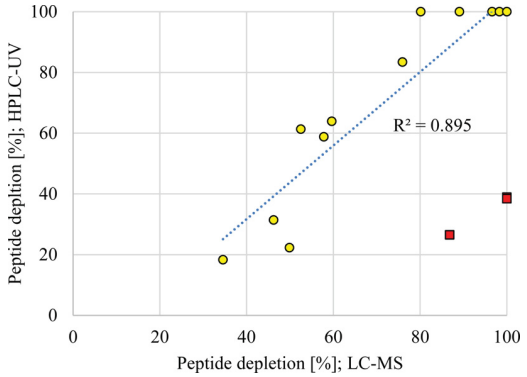


Figure 1. Correlation of Cys peptide depletion detected with HPLC-UV or by LC-MS measurements (open circles; $n = 17$, 5 overlying dots at (100;100)); three outliers (closed squares; **4**, **9**, and **10**) gave higher depletions in LC-MS; detailed values are listed in Table 2.

results, the molecular mechanisms leading to peptide depletion with the 18 sensitizers lacking structural alerts for peptide reactivity are postulated. In principle, two mechanisms leading to peptide depletion can be envisioned. A chemical can be transformed by contact with (air) oxygen to a peptide reactive molecule which covalently binds to the peptide to form an adduct and thus depletes the peptide (illustrated as "**A**" in Figure 2); or the chemical can induce peptide depletion by

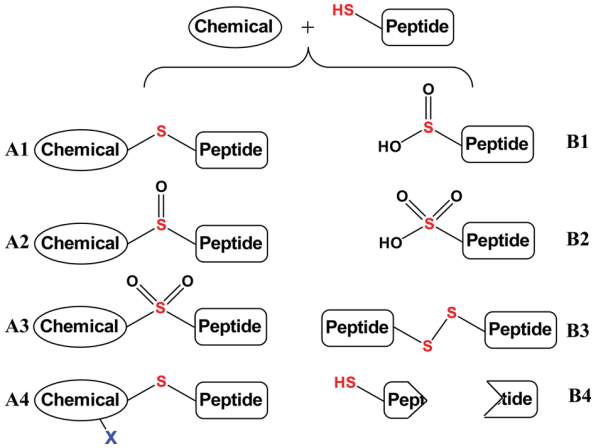


Figure 2. Overview of possible scenarios which could lead to peptide depletions in the DPRA. **A1** illustrates the formation of adducts and **A2** and **A3** further oxidations of such adducts to sulfoxides and sulfones, respectively. **A4** illustrates the formation of adducts as a result of a more complex chemistry. **B1** and **B2** illustrate the transformations of the sulfhydryl group to sulfinic and sulfonic acid residues, respectively. **B3** illustrates the dimer of Cys peptides connected by a disulfide (i.e., cystine bond), whereas **B4** illustrates any kind of fragmentation of the peptide.

inducing fragmentation or oxidation of the peptide without adduct formation taking place (illustrated as "**B**" in Figure 2). A schematic representation of the scenarios (i.e., **A** and **B**) leading to peptide depletion are depicted in Figure 2.

A detailed analysis of LC-MS spectra enables postulation of molecular structures of the DPRA reaction products, and the nature of modification can be deduced. These analyses were conducted for the 18 sensitizers lacking structural alerts for peptide reactivity but which (unexpectedly) induced peptide depletion. Table 3 gives an overview of the LC-MS results, the

DOI: 10.1021/acs.chemrestox.6b00055
Chem. Res. Toxicol. XXXX, XXX, XXX−XXX

Table 3. Overview and Interpretations of Adducts and Peptide Oxidations Detected for the Samples with the 18 Pre- and Pro-haptens and the Two Positive Controls[a]

no.	test chemical	signals in LC-MS spectra [m/z]	scenario	interpretation of signals and proposal of product structures
NC	Cys-containing model peptide (Ac-RFAACAA-OOH)	751.6 + 375.8		plausible mass of peptide (z = 1, $[M + H]^+$ + z = 2, $[M + 2H]^{2+}$)
PC1	ethylene glycol dimethacrylate (EGDMA)	949.8 + 475.4	**A1**	plausible adduct (z = 1 + z = 2) after direct MA
PC2	1-chloro-2,4-dinitrobenzene (DNCB)	917.7 + 459.4	**A1**	plausible adduct (z = 1 + z = 2) after direct S_NAr
1	5-amino-2-methylphenol	838.3 + 419.6	**A4**	complex adduct (z = 1 + z = 2); according to the activation of meta-sensitizers suggested by [Aptula et al.[40,45]], the plausible m/z of the adduct would have been 888.40 (z = 1)
		799.1	**B2**	peptide +48, oxidation of sulfhydryl group to sulfonic acid
2	ethylene diamine	no adduct(s)	(**B3**)	no plausible adduct; 2-fold increased amount of dipeptide compared to the control
3	4-amino-m-cresol	872.5 + 436.4	**A1**	plausible adduct (z = 1 + z = 2); *para*-quinone imine formation and subsequent MA
		(*i*) 488.3; (*ii*) 496.5	**A4**	(*i*) complex adduct (z = 2) with *ii* possibly being the sulfonic form of *i* (+ 32)
		1009.1 + 504.9	**A4**	complex adduct (z = 1 + z = 2)
		799.1	**B2**	peptide +48, oxidation of sulfhydryl group to sulfonic acid
		783.6	**B1**	peptide +32, oxidation of sulfhydryl group to sulfinic acid
4	isoeugenol	931.3	**A1**	plausible adduct after autoxidation and epoxidation; proposed mechanism is in the Discussion section
		947.2	**A2**	931.26 + 16; thioether oxidized to sulfoxide
		929.2	**A2**	plausible adduct; *para*-quinone-methide formation and MA; thioether oxidized to sulfoxide
		799.0	**B2; B4**	peptide +48, oxidation of the sulfhydryl group to sulfonic acid; in addition, an intensive signal in the UV spectrum indicating complex fragments
5	*para*-phenylendiamine	858.0 + 428.9	**A1**	plausible adduct (z = 1 + z = 2); *para*-quinone imine formation with subsequent MA
6	hydroquinone	859.1 + 429.0	**A1**	plausible adduct (z = 1 + z = 2); *para*-quinone formation and subsequent MA
		799.0	**B2**	peptide +48, oxidation of sulfhydryl group to sulfonic acid; in addition, 6-fold increased amount of dipeptide
7	4-allylanisole	961.4 + 481.0	**A3**	plausible adduct (z = 1 + z = 2); epoxidation of the vinyl group (proposed mechanism is in the Discussion section), formation of sulfone
		977.6	**A4**	plausible adduct +16, further oxidized
8	propyl gallate	no adduct(s)	(**B3**)	no plausible adduct; obviously complex chemistry[27]; 1.7-fold increased amount of dipeptide compared to the control
9	eugenol	930.5	**A1; B4**	plausible adduct (z = 1) in traces, detected by tandem mass spectrometry (MS-MS); intensive signal in the UV spectrum indicating complex fragments
10	3-methylcatechol	873.2	**A1; B4**	plausible adduct; *ortho*-quinone formation and MA; in addition, intensive signal in the UV spectrum indicating complex fragments
		889.2 + 444.9	**A2**	plausible adduct (z = 1 + z = 2); *ortho*-quinone formation and MA, oxidation of thioether to sulfoxide
		905.2	**A3**	plausible adduct; *ortho*-quinone formation and MA, oxidation of thioether to sulfone
11	2-nitro-1,4-phenylenediamine	902.2 + 451.4	**A1**	plausible adduct (z = 1 + z = 2); *para*-quinone imine formation and subsequent MA
12	4-(methylamino) phenol sulfate (Metol)	436.4	**A1**	plausible adduct; *para*-quinone imine formation and subsequent MA
		443.5	**A2**	plausible adduct +16, further oxidized
		(*i*) 496.0; (*ii*) 1005.2 + 503.0	**A4**	(*i*) 496 (z = 2): complex adduct with further oxidation product ((*ii*) formation of sulfoxide (+16) with z = 1 + z = 2)
		(*i*) 548.6; (*ii*) 556.8; (*iii*) 563.5	**A4**	(*i*) 549 (z = 2): complex adduct with further oxidation products (formation of (*ii*) sulfoxide (+16) and (*iii*) sulfone (+32))
		783.8, 799.8	**B1; B2**	peptide +32 and +48, oxidation of sulfhydryl group to sulfinic and sulfonic acid, respectively
13	2,5-diaminotoluene sulfate	435.9	**A1**	plausible adduct (z = 2); *para*-quinone imine formation and subsequent MA
		799.0	**B2**	peptide +48, oxidation of sulfhydryl group to sulfonic acid; in addition, ~ 80-fold increased amount of dipeptide compared to the control
14	abietic acid	no adduct(s)	(**A4; B4**)	abietic acid is known to form peroxides after contact to air oxygen [Karlberg et al.,[47]; Karlberg,[48]]; obviously complex chemistry leading to peptide depletions
15	lauryl gallate	no adduct(s)	(**A4; B4**)	no plausible adduct; comparable to propyl gallate, obviously complex chemistry[27]
16	2-aminophenol	961.6 + 480.9	**A1**	plausible adduct (z = 1 + z = 2); *ortho*-quinone imine formation and subsequent MA
17	cinnamic alcohol	883.1 + 441.8	**A1**	plausible adduct (z = 1 + z = 2); formation of α,β-unsaturated aldehyde and subsequent MA
		901.7	**A1**	plausible adduct after autoxidation and epoxidation; proposed mechanism is in the Discussion section
18	benzo[*a*]pyrene	no adducts	**B3**	approximately 40% of the peptide depletion can be explained by the dipeptide formation

[a]MA = Michael addition; S_NAr, nucleophilic substitution in aromatic chemicals; m/z = mass-to-charge-ratio; for details on **A1−4** and **B1−4**, see also Figure 2.

DOI: 10.1021/acs.chemrestox.6b00055
Chem. Res. Toxicol. XXXX, XXX, XXX−XXX

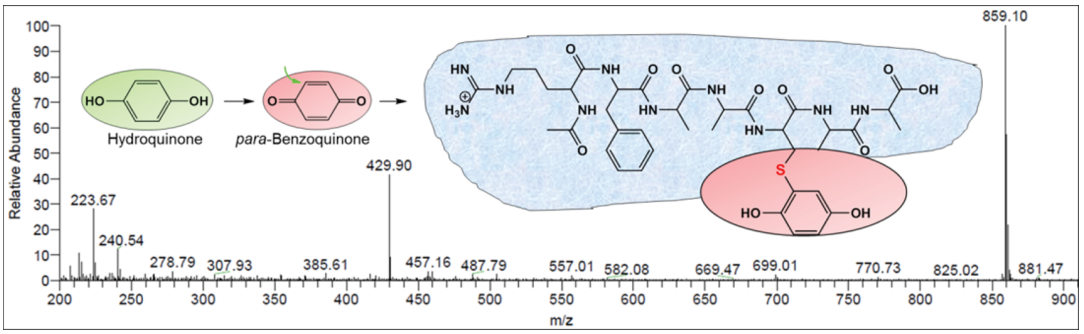


Figure 3. Oxidative activation of the pre-hapten hydroquinone to the highly electrophilic hapten *para*-benzoquinone and corresponding adduct (the blue background illustrates the molecular structure of the Cys peptide). The calculated masses (m/z = 859.4 $[M + H]^+$ and m/z = 430.2 $[M + 2H]^{2+}$) for the complex correspond to the detected masses (m/z = 859.1 $[M + H]^+$ and m/z = 429.9 $[M + 2H]^{2+}$).

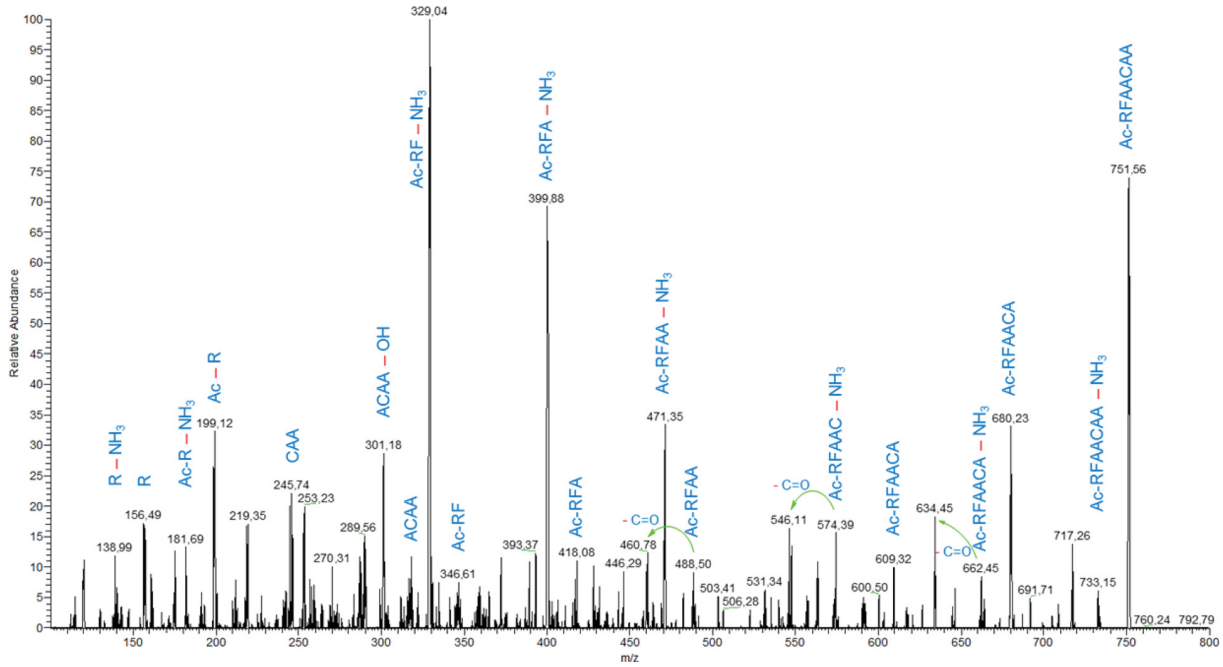


Figure 4. Full fragment spectrum of Ac-RFAACAA (m/z = 751). Ac = acyl, R = arginine, F = phenyl alanine, A = alanine, C = cysteine, NH3 = ammonia, C═O = carbon monoxide.

interpretation of the detected signals, and assigns the reactions scenarios illustrated in Figure 2.

The m/z values were analyzed to detect those belonging to the expected adducts. By the mass spectrometric read-out, chemical structures of plausible adducts were postulated for 13 of the 18 chemicals with positive results in the DPRA. In most of these cases, sensitizers lacking electrophilic moieties were transformed to quinones, quinone imines or quinone methides as reactive intermediates, suggesting autoxidation.[31] The m/z values indicate that the reactive intermediates formed covalent bonds with the peptide following Michael additions (i.e., for **3**, **4**, **5**, **6**, **9**, **10**, **11**, **12**, **13**, **16**, **17**). As an example, the oxidative activation of the pre-hapten hydroquinone (**6**) to the highly electrophilic Michael acceptor *para*-benzoquinone and the molecular structure of its corresponding adduct is shown in Figure 3.

In some cases, the spectrometrically detected m/z values were higher than previously calculated; in other cases, no adducts could be identified, although peptide depletion was detected. Both phenomena can be explained by (additional) oxidation due to contact with (air) oxygen.

In principle, three different types of peptide oxidation reactions were observed during this study (see also Figure 2). (*i*) Some chemicals induced the oxidation of the sulfhydryl group. Chemicals **1**, **3**, **4**, **6**, **12**, and **13** catalyzed the formation of sulfonic acid (m/z of peptide +48), and chemicals **3** and **12** additionally catalyzed the formation of sulfinic acid (m/z of peptide +32). (*ii*) Another form of peptide oxidation is the formation of a dimer of Cys peptides with a mass of 750.5 $[M + 2H]^{2+}$ or rather 1500 $[M + H]^+$ connected by a disulfide (i.e., cystine bond). The dimer is even present in traces (~1% compared to naïve peptide) in the negative control and is normally poorly soluble under the experimental conditions of the DPRA. However, under conditions of the LC-MS experiments (1:100 dilution of the DPRA sample), levels of the dimer were clearly detected. In one example, benzo[*a*]-pyrene (**18**) induced a 39-fold increase in the amount of dimers, which explains ~40% of the observed peptide depletion. In case of metol (**13**), a 79-fold increased amount

DOI: 10.1021/acs.chemrestox.6b00055
Chem. Res. Toxicol. XXXX, XXX, XXX−XXX

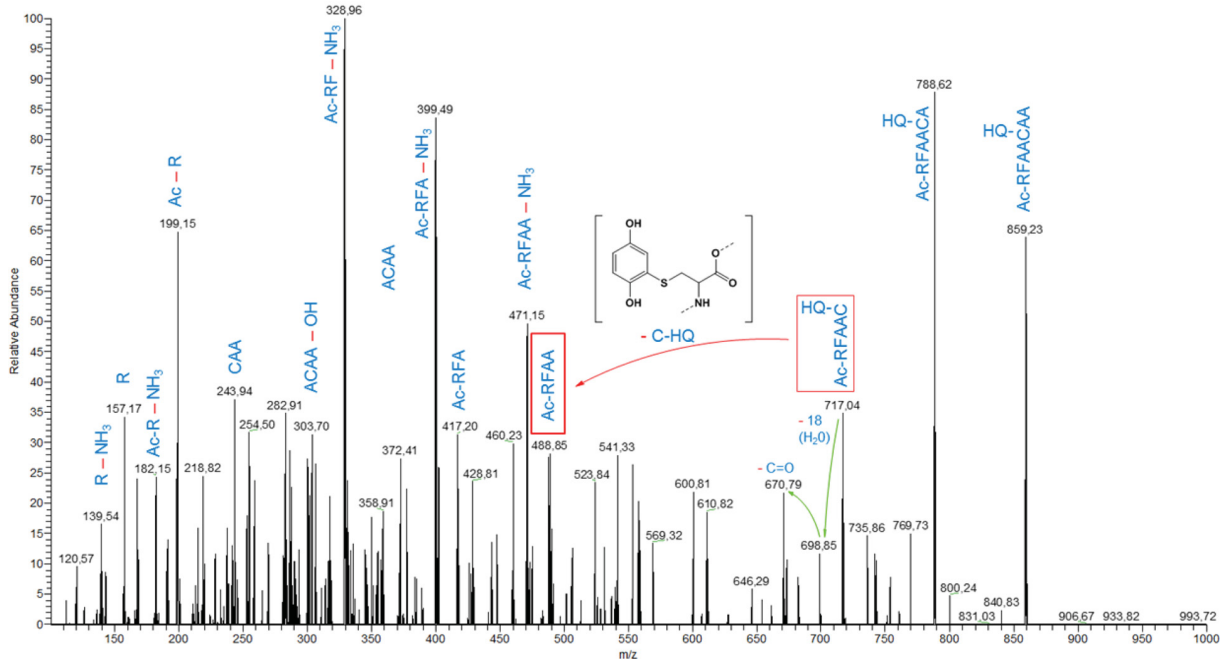


Figure 5. Full fragment spectrum of Ac-RFAACAA after complete reaction with hydroquinone (HQ).

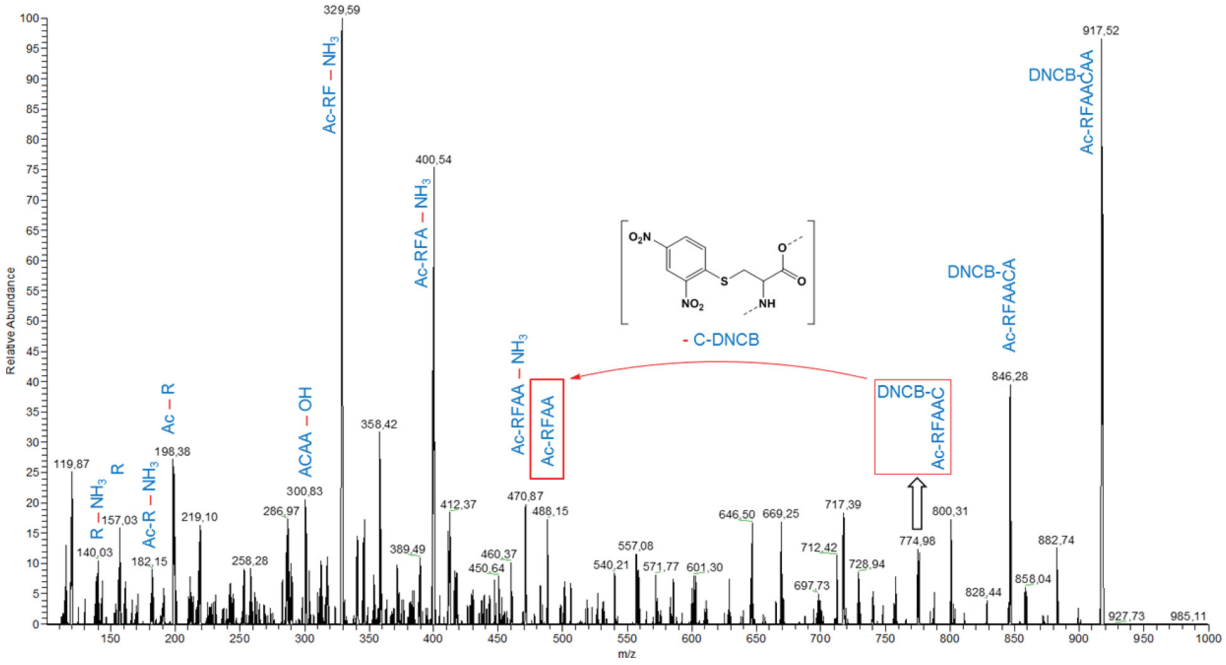


Figure 6. Full fragment spectrum of Ac-RFAACAA after complete reaction with 2,4-dinitrochlorobenzene (DNCB; used as the positive control).

of dimer was detected. (*iii*) A third type of peptide oxidation occurred after the covalent formation of adducts between the test chemical and peptide. The generally oxophilic sulfur was oxidized to a sulfoxide (plausible adduct +16, e.g., **4**, **10**, and **12**) or to a sulfone (plausible adduct +32, e.g., **7**, **10**, and **12**).[32,33]

For another five chemicals (i.e., **1**, **2**, **8**, **14**, and **15**), implausible or no m/z values indicating possible adducts were detected. For **1**, implausible m/z values were detected which were below the m/z value of the calculated adduct but higher than m/z values of the peptide. Any new peak with an observed mass m/z > 751 $[M + H]^+$ which cannot be explained by peptide oxidation is considered to be an adduct.[27] Therefore, the m/z value for **1** is proposed to be based on adduct formation. In addition to plausible adducts, **3**, **4**, and **12** also formed complex adducts for which the structures could not be further clarified. For some of the strong or extreme sensitizers belonging to the class of aromatic amines and phenolic compounds, a more complex pattern is described in the literature, namely, repeated cycles of spontaneous oxidation

DOI: 10.1021/acs.chemrestox.6b00055
Chem. Res. Toxicol. XXXX, XXX, XXX−XXX

Table 4. Overview of the Investigated Pre- and Pro-haptens and the Results of the Single Nonanimal Test Methods as Well as the 2 out of 3 WoE Approach[a]

No.	Name	LLNA final	Human final	DPRA (Cys+Lys) (HPLC)	Cys-Peptide (LC-MS)	Keratino-Sens™	h-CLAT	'2 out of 3' WoE Approach
1	5-Amino-2-methylphenol	+		TP	Adduct	TP	no data	TP
2	Ethylenediamine	+	+	TP	No adduct	TP	TP	TP
3	4-Amino-*m*-cresol	+		TP	Adduct	TP	no data	TP
4	Isoeugenol	+	+	TP	Adduct	TP	FN	TP
5	1,4-Phenylene diamine	+	+	TP	Adduct	TP	TP	TP
6	Hydroquinone	+	+	TP	Adduct	TP	TP	TP
7	4-Allylanisole	+		TP	Adduct	inconclusive	TP	TP
8	Propyl gallate	+	+	TP	No adduct	TP	TP	TP
9	Eugenol	+	+	TP	Adduct	inconclusive	TP	TP
10	3-Methylcatechol	+		TP	Adduct	TP	no data	TP
11	2-Nitro-1,4-phenylendiamine	+	+	TP	Adduct	TP	TP	TP
12	4-(Methylamino) phenol sulfate (Metol)	+	+	TP	Adduct	TP	no data	TP
13	2,5-Diaminotoluene sulfate (PTD)	+	+	TP	Adduct	TP	TP	TP
14	Abietic acid	+	+	TP	No adduct	TP	FN	TP
15	Lauryl gallate	+	+	TP	No adduct	TP	TP	TP
16	2-Aminophenol	+	+	TP	Adduct	TP	TP	TP
17	Cinnamyl Alcohol	+	+	TP	Adduct	TP	TP	TP
18	Benzo(a)pyrene	+		TP	No adduct	TP	TP	TP
19	2-methoxy-4-methylphenol	+		FN		FN	TP	FN
20	Resorcinol	+	+	FN		FN	TP	FN
21	3-Aminophenol	+		FN		FN	TP	FN
22	Geraniol	+	+	FN		TP	TP	TP
23	Diethylenetriamine	+	+	FN		FN	FN	FN
24	Farnesol	+	+	FN		TP	TP	TP
25	3-Dimethylamino propylamine	+	+	FN		TP	TP	TP
26	N,N-Dibutylaniline	+		FN		FN	FN	FN
27	4-Chloroaniline	+		FN		TP	TP	TP
			Sensitivity [%]:	67		80	83	81

[a]TP = true positive, FN = false negative; '+' = positive result *in vivo*; sensitivity = [TP/(TP + FN)].

followed by adduct formation with several nucleophilic residues of Cys peptides leading to the sensitizing reaction product.[27] Unfortunately, since the upper detection limit in the LC-MS approach was m/z = 1500, possible adducts with higher m/z values could not be detected (for instance, in the case of **8**, **14**, or **15**). Peptide depletion observed with **2**, **8**, **14**, or **15** may also be attributable to the formation of insoluble dimers.

The possible mechanisms identified reflect the intrinsic behavior of the chemicals in the *in chemico* tests. The described results are an approximation to the situation *in vivo* and may not exactly demonstrate how these chemicals interact with cutaneous proteins in human skin. Although intensively investigated, actual hapten−protein adducts *in vivo* have not yet been identified.

3.1.3. Binding Site at the Model Peptides. To analyze the molecular structures of the reaction products (or adducts) of the investigated chemicals with the peptide, knowledge of putative binding sites within the model peptide is mandatory. The residue of the amino acid cysteine in the model peptide Ac-RFAACAA-COOH is described to be the sole (nucleophilic) reactive site of the model peptide.[4,17] However, other sites in the structure of the peptide may also conceivably react under certain circumstances. For instance, putative reactions at the side chain of arginine could occur; but also cleavage of the acyl group masking the primary amine in the terminal position with subsequent adduct formation at the amine would both lead to different m/z values compared to adducts at the sulfhydryl group of the Cys-residue.

To exclude that these reaction types had taken place, the actual binding site of the model peptide was analyzed in more detail. The analysis was performed by *de novo* sequencing, a mass spectrometric method which assigns defined fragments to a mass spectrum. First, structural characteristics of the native peptide were examined (Figure 4). Assigning all significant signals in the chromatogram to their assigned respective fragment was possible without observing any unexpected fragments. Next, samples including adducts of the peptide with hydroquinone (**6**, Figure 5) as well as with DNCB (**PC2**, Figure 6) were analyzed using *de novo* sequencing. In the spectrum depicted in Figure 5, it was once again possible to assign all significant signals to their respective fragments. Interestingly, the cleavage of the whole fragment of cysteine bound to hydroquinone was detected and is illustrated by the highlighted mass loss of m/z = 228. Similar to that, Figure 6

DOI: 10.1021/acs.chemrestox.6b00055
Chem. Res. Toxicol. XXXX, XXX, XXX−XXX

illustrates the mass loss of the whole fragment of cysteine bound to dinitrobenzene by a mass loss of m/z = 287 (the chloride substituent was already cleaved during adduct formation according to a nucleophilic substitution (S_NAr)). The signals in Figures 5 and 6 indicate covalent binding of hydroquinone and DNCB only at the sulfhydryl group; reactive positions in the molecular structure of the Cys peptide other than the sulfhydryl group were not detected. These results verify that the sulfhydryl group in the side chain of cysteine is the actual reactive center of the Cys peptide, which is in accordance with the information in the OECD TG.

3.2. Detection of Pre- and Pro-haptens in a Battery of Nonanimal Tests. The predictivity for the identification of pre- and pro-haptens can be increased by taking cell-based nonanimal test methods such as the KeratinoSens, the LuSens, or the h-CLAT into account, as the cell lines used exhibit considerable activity of metabolic enzymes.[10,11,18] Twenty of the 25 (inconclusive results for two chemicals) and 19 of the 24 putative pre/pro-haptens gave positive results in the KeratinoSens and h-CLAT, respectively (results are given in Table 4). A combination of DPRA, KeratinoSens, and h-CLAT within the 2 out of 3 WoE approach correctly identified the sensitizing potential of 22 out of 27 chemicals, resulting in a sensitivity of 81% for this specific set of pre- and pro-haptens (Table 4).

For comparison, sensitivity was also calculated for a set of sensitizers, which contain electrophilic moieties in their parent structures. Here, the 2 out of 3 WoE approach (DPRA, KeratinoSens, and h-CLAT) correctly identified 83 out of the 95 directly acting haptens, which resulted in a sensitivity of 87% (compare Supporting Information (Table S1)). To evaluate statistical significance between the set of pre/pro-haptens and the set of directly acting haptens, a Fisher's exact test was applied.[30] The comparison of the proportion of the sensitivities between both sets indicated no significant difference (p = 0.529).

Among the chemicals evaluated in this study, different subsets can be defined based on characteristics in the molecular structures as well as on the *in chemico* or *in vitro* test results (Table 4). Some of the assignments are in discordance to literature data. For instance, isoeugenol (**4**) and eugenol (**9**) are described to require biotic instead of abiotic activation.[21,34] In the underlying study, adducts were found and characterized for both chemicals in the DPRA, indicating both chemicals to be also susceptible to an abiotic activation.

4. DISCUSSION

In this study, 27 sensitizers lacking structural alerts for peptide reactivity were identified from a comprehensive data set consisting of 213 chemicals and were further analyzed. Eighteen of the 27 putative pre- and pro-haptens were unexpectedly positive in the DPRA and thus interacted in some way with the peptide.

Since the detection of depletion does not give any information on the reactions taking place between the peptide and test chemical, the 18 positives were retested in the DPRA, and the nature of peptide depletion was examined applying additional mass spectrometric detection. The 18 test chemicals were also positive in the second DPRA, indicating a reproducibility of 100% in terms of the overall assay read-out. Quantities of depleted peptides either detected by HPLC-UV or LC-MS were also comparable.

To clarify the underlying mechanisms leading to the activation of pre-haptens, obtained mass spectra were evaluated and interpreted. For 13 of the 18 DPRA positives, adducts were identified by the use of LC-MS indicating the pre-hapten nature of these chemicals. Mechanistically, the initial pre-haptens were probably transformed to quinones, quinone imines, or quinone methides as reactive intermediates in most of the cases to subsequently form covalent bonds with the Cys-peptide following Michael additions (i.e., for **3**, **4**, **5**, **6**, **9**, **10**, **11**, **12**, **13**, **16**, and **17**). Such transformations indicate an activation by contact to dissolved- or air oxygen (i.e., autoxidation). Diradical oxygen activated certain fragrances to increase their potencies in the LLNA after prolonged contact to oxygen; this was recently demonstrated by Karlberg and co-workers.[21] In this context, we analyzed altered peptide reactivity of citral in the DPRA depending on the duration of contact with air oxygen; peptide depletion for the pure chemical was 4- to 5-fold lower than that 11 months after the first exposure to air oxygen (compare Table S2 (Supporting Information)). A similar effect was described for the moderate sensitizer bisabolene, which was only peptide reactive after exposure to oxygen.[35] This can be explained by the formation of primary oxidation products such as hydroperoxides and the subsequent formation secondary oxidation products such as aldehydes or epoxides.[21] A specific mechanism of pre-hapten activation observed in this study is represented by the formation of epoxides. In general, due to high ring tensions, epoxides are in an energetically unfavorable state and relatively unstable. In some cases, epoxides may be quite stable and, depending on their respective residue, only have a low reactivity as electrophiles. Epoxidation is often described as an enzyme-driven reaction (i.e., by CYP enzymes).[36] However, Karlberg and co-workers proposed an epoxide formation of cinnamic alcohol by a peroxy acid under abiotic conditions.[37] Also in our study, only epoxidation (under abiotic conditions) as molecular modification can explain the m/z values of the corresponding adducts for three chemicals (i.e., **4**, **7**, and **17**; of note, for instance in the case of isoeugenol, also adducts based on the formation of a *para*-quinone methide were detected (Table 3)). In Figure 7, a plausible mechanism is proposed for isoeugenol (**4**). Consequently, for specific chemicals with structural characteristics like allylic positions, the duration of air exposure between first opening the boxed sample and actually performing test should be considered.

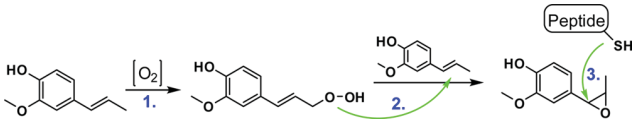


Figure 7. Possible formation of an epoxide as a secondary oxidation product under abiotic and nonmetabolical conditions. In the proposed mechanism, a hydroperoxide could be formed in allylic position in the first step. In the second step, the hydroperoxide could oxidize the unsaturated double bond of another molecule to an epoxide. In the third step, the sulfhydryl group of the peptide could form a covalent bond probably at the benzylic position, leading to a ring opening of the epoxide with a hydroxyl group in β-position to the benzene ring according to an S_N2 mechanism.

In the set of the 18 DPRA positives, no adducts were detected for five chemicals. This could possibly be explained by the formation of large adducts being above the detection limit of the MS or, more likely, by the formation of dimers of the model peptide as observed for benzo[*a*]pyrene. For this

DOI: 10.1021/acs.chemrestox.6b00055
Chem. Res. Toxicol. XXXX, XXX, XXX−XXX

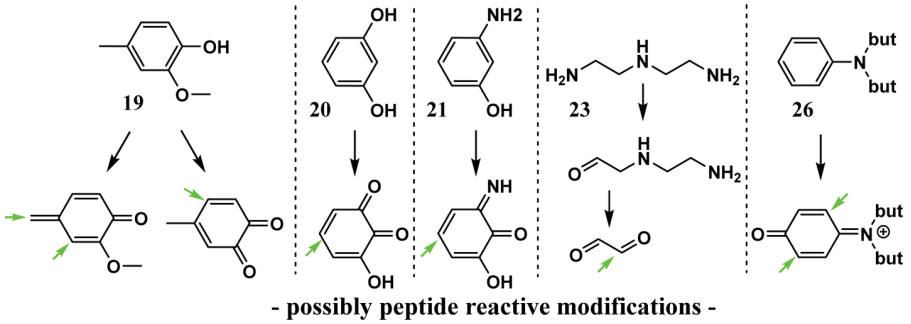


Figure 8. Possible reactive intermediates of the chemicals being false negative in the 2 out of 3 WoE approach. The green arrows indicate highly electrophilic positions in the molecular structures where a nucleophilic attack would probably occur.

particular molecule, the observed peptide depletion is probably not attributable to a peptide binding; thus, the result for this molecule is considered as a "false TP" as classification was still correct. When considering adduct formation instead of peptide depletion as the criterion for a true positive results, the sensitivity of the DPRA would be slightly decreased for the underlying data set. Moving from peptide depletion to adduct formation as a read-out of the DPRA was already described for a set of chemicals predominantly consisting of nonelectrophilic nonsensitizers and electrophilic sensitizers: While the sensitivity was lower, the specificity was considerably higher compared to the depletion read-out.[27,35] However, it should be noted that the assessment of adduct formation on a routine basis is not entirely feasible due to the complexity of molecular structures of the tested chemicals. Furthermore, the biological role of test compound induced alterations of protein structure (such as dimerization as described above) cannot be addressed.

As discussed above, the DPRA was able to detect pre-haptens. Nine of the 27 sensitizing chemicals gave negative results in the DPRA, most likely requiring metabolic activation and would thus be termed pro-haptens. Seven of these nine chemicals yielded positive results in the cell-based assays (KeratinoSens and/or h-CLAT). Thus, a combination of DPRA with methods addressing subsequent key events and, at the same time, providing metabolic competence is beneficial.

For all of the 27 chemicals in this data set, at least two results for the three nonanimal tests DPRA, KeratinoSens, and h-CLAT were present, and the 2 out of 3 WoE approach could be applied. This prediction model correctly identified 22 of the 27 chemicals. Five chemicals were, however, misclassified for various reasons: 2-methoxy-4-methylphenol (**19**), resorcinol (**20**), 3-aminophenol (**21**), diethylenetriamine (**23**), and *N*,*N*-dibutylaniline (**26**).

While the aliphatic amine diethylenetriamine (**23**) was misclassified, two other aliphatic amines of the data set were correctly predicted by the 2 out of 3 WoE approach. *In vivo*, **23** was a moderate sensitizer in the LLNA (EC3 = 3.3% according to the ICCVAM database) and also positive in a human maximization test.[38,39] Assuming a conversion of the primary amine into an aldehyde (Figure 8), **23** can be assigned to the mechanistic domain of pro-Schiff base formers, as would the other two aliphatic amines **2** and **25** of this data set.[40] All three primary amines are hydrophilic, and **23** has the lowest vapor pressure (~37 Pa, calculated with TIMES SS); therefore, low solubility in and evaporation from the DPRA reaction medium can be excluded as reasons for the false negative result. However, it should be kept in mind that test systems using aqueous solutions were previously described to be unsuitable to detect SB formers. The potency of SB formers has been found to be correlated with a combination of an electrophilicity parameter and logP, the latter parameter implying that the MIE occurs in a lipid environment rather than in an aqueous one.[41]

Another chemical with negative results in all three nonanimal tests was the aromatic tertiary amine *N*,*N*-dibutylaniline (**26**), which was weakly sensitizing in the LLNA (EC3 = 19.6%).[42] Aniline as a possible impurity can be excluded as the cause for the positive response *in vivo* since aniline is a much weaker sensitizer in the LLNA. Mechanistically, the formation of a quarternized quinone-imine is likely (Figure 8); several *para*-quinone imines (e.g., **3**, **5**, **11**, or **16**) were positive in the LLNA. However, the addition of S9 mix to the KeratinoSens induced its activation so that **26** was positive in the "KeratinoSens-S9".[43]

2-Methoxy-4-methylphenol (**19**) was negative in the DPRA and KeratinoSens but positive in the h-CLAT.[14] **19** is a moderate sensitizer in the LLNA (EC3 = 5.8)[44] and positive in a HRIPT (compare RIFM database) and can possibly form an *ortho*-quinone or a *para*-quinone methide (Figure 8). In contrast to the formations of quinones and quinone imines, a demethylation of the methoxy group or the formation of a quinone methide might not occur under the conditions of the DPRA. **19** also gave a negative response in KeratinoSens. However, a positive KeratinoSens result for the structurally highly similar dihydroeugenol (data published elsewhere)[14] indicates no general lack in detecting chemicals with similar reaction mechanisms as **19**. The true positive result for **19** in the h-CLAT was due to mechanisms other than those detected by DPRA and KeratinoSens, and may indeed be linked to the specific key event addressed by this assay.

The h-CLAT was also positive for resorcinol (**20**) and 3-aminophenol (**21**), whereas DPRA and KeratinoSens results were negative. These "meta-sensitizers" cannot be oxidized directly to quinones or quinone imines since the amino and hydroxyl groups of the benzene ring are located in the *meta*-position. An activation of such chemicals could take place by an attack of a radical species on the carbon atom located in the *ortho*-position between the two residues.[45] Under the assay conditions of the DPRA and KeratinoSens, this type of activation obviously did not occur. The addition of S9 mix to the KeratinoSens also failed to activate **20** and **21** as indicated by negative results in the "KeratinoSens-S9".[43]

The sensitivity of the 2 out of 3 WoE approach was similar (statistically not different) for the pre- and pro-haptens and the directly acting haptens (compare Supporting Information (Table S1)), indicating that the activation of pre- and pro-haptens is not critical for the performance of this approach. It

DOI: 10.1021/acs.chemrestox.6b00055
Chem. Res. Toxicol. XXXX, XXX, XXX−XXX

should also be kept in mind that the LLNA predicted human data with a limited accuracy of 82%, whereas the nonanimal test methods predicted 90% of the human data correctly.[14] Therefore, it cannot be ruled out that some of the above-mentioned chemicals (e.g., **26**) are true negative in the nonanimal tests and false-positive in the LLNA.

5. CONCLUSIONS

This study was performed on 27 *in vivo* sensitizers, which lacked electrophilic moieties in their parent structures and were therefore not expected to bind to peptides. Eighteen of these 27 putative pre- or pro-haptens gave, yet, positive responses in the DPRA.

The DPRA uses two model peptides containing cysteine and lysine, respectively. The depletion by hapten-binding was generally higher with the cysteine peptide. *De novo* sequencing proved the sulfhydryl group of the cysteine to be the only reacting nucleophilic center of this heptapeptide. This justified the use of the cysteine model peptide (Cys-only model) in this study.

Peptide binding was investigated by conducting LC-MS measurements in addition to the routinely performed HPLC-UV read outs. The peptide depletions measured by the two methods correlated well. In addition, the LC-MS analysis revealed the nature of the altered peptides, which caused the 18 positive DPRA results: 13 of these chemicals were non-metabolically activated and formed peptide-adducts, indicative of them being pre-haptens. However, the remaining five chemicals did not form peptide adducts but induced the formation of different kinds of peptide oxidation products like peptide dimers or sulfonic or sulfinic acid. The biological consequences of those peptide modifications are, however, not clear. Nonetheless, the DPRA is capable of detecting pre-haptens with most of them forming peptide adducts.

The information originating from the cell based assays KeratinoSens (keratinocytes) and h-CLAT (dendritic cell-like cell line) in addition to the DPRA using a 2 out of 3 WoE approach clearly increased the sensitivity in detecting pre- and pro-haptens. Although the metabolic competence of these cell lines may not represent the full metabolic capacity of the native human skin, the actually available metabolic capacity and the and the spontaneously occurring chemical modifications are sufficient to correctly identify most of the pre- and pro-haptens. The addition of an activating system, e.g., hydrogen peroxide and horseradish peroxidase in the DPRA (then called PPRA) or S9 mix to the cell based assays, is, however, relevant for specific pro-haptens (e.g., disubstituted aromatics in *meta*-position or specific aromatic amines).

Overall, the sensitivity of the 2 out of 3 WoE approach in detecting pre- and pro-haptens is actually as high as its sensitivity in detecting directly acting haptens. Recently published findings confirm this.[46]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.chemrestox.6b00055.

Table titles (PDF)

Table S1 represents a list of chemicals (n = 95) with alerts for peptide reactivity in their parent structures and positive results in the *in vivo* LLNA. The chemicals of this set were used to compare the sensitivity of the 2 out of 3 WoE approach model for the set of directly acting haptens (S1) to the set of pre- and pro-haptens, the latter being analyzed in this article (XLSX)

Table S2 lists peptide depletions of citral at different time points to show the substantial impact of air exposure to specific chemicals (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +49 621 60 56203. Fax: +49 621 60 58134. E-mail: robert.landsiedel@basf.com.

Funding

This work was funded by BASF SE, Ludwigshafen, Germany.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Nina Deppermann and Dr. Eric Fabian (both BASF SE) for helpful discussions.

■ ABBREVIATIONS

A, alanine; AOP, adverse outcome pathway; ARE, antioxidant response element; C, cysteine; Cys, cysteine; CID, low energy collision induced dissociation; DPRA, direct peptide reactivity assay; F, phenylalanine; FN, false negative; FP, false positive; h-CLAT, human cell line activation test; HPLC, high performance liquid chromatography; IATA, integrated testing approaches and assessments; K, lysine; LLNA, local lymph node assay; Lys, lysine; MA, Michael acceptor; MIE, molecular initiating event; n, number of chemicals; OECD, Organization for Economic Co-operation and Development; OECD TG, OECD test guideline number; QP, quinone precursor; R, arginine; SB, Schiff base former; S_NAr, aromatics reacting by nucleophilic substitutions; TN, true negative; TP, true positive; WoE, weight of evidence

■ REFERENCES

(1) (2003) *Directive 2003/15/EC* of the European Parliament and the Council of 27 February 2003 amending *Council Directive 76/768/EEC* on the approximations of laws of the Member States relating to cosmetic products. *Off. J. Eur. Union L66*, 26−35.

(2) OECD (2012) The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins; Part 1: Scientific Evidence. Organisation for Economic Co-operation and Development (OECD), Paris, http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part1&doclanguage=en.

(3) OECD (2012) 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. Organisation for Economic Co-operation and Development (OECD), Paris, http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2012)10/part2&doclanguage=en.

(4) OECD (2015) OECD TG 442C: In chemico skin sensitization: Direct Peptide Reactivity Assay (DPRA). Organisation for Economic Co-operation and Development (OECD), Paris, https://ntp.niehs.nih.gov/iccvam/suppdocs/feddocs/oecd/oecd-tg442c-508.pdf.

(5) OECD (2015) OECD TG 442D: *In vitro* skin sensitization: ARE-Nrf2 Luciferase Test Methods. Organisation for Economic Co-operation and Development (OECD), Paris, https://ntp.niehs.nih.gov/iccvam/suppdocs/feddocs/oecd/oecd-tg442d-508.pdf.

DOI: 10.1021/acs.chemrestox.6b00055
Chem. Res. Toxicol. XXXX, XXX, XXX−XXX

(6) Emter, R., Ellis, G., and Natsch, A. (2010) Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers in vitro. *Toxicol. Appl. Pharmacol. 245*, 281−290.

(7) Ramirez, T., Mehling, A., Kolle, S. N., Wruck, C. J., Teubner, W., Eltze, T., Aumann, A., Urbisch, D., van Ravenzwaay, B., and 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.

(8) Ashikaga, T., Yoshida, Y., Hirota, M., Yoneyama, K., Itagaki, H., Sakaguchi, H., Miyazawa, M., Ito, Y., Suzuki, H., and 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.

(9) EURL-ECVAM (2015) EURL ECVAM recommendation on the human Cell Line Activation Test (h-CLAT) for skin sensitisation testing. Publications Office of the European Union, Luxembourg, https://eurl-ecvam.jrc.ec.europa.eu/news/news_docs/eurl-ecvam-recommendation-on-the-human-cell-line-activation-test-h-clat-for-skin-sensitisation-testing.

(10) Fabian, E., Vogel, D., Blatz, V., Ramirez, T., Kolle, S., Eltze, T., van Ravenzwaay, B., Oesch, F., and Landsiedel, R. (2013) Xenobiotic metabolizing enzyme activities in cells used for testing skin sensitization in vitro. *Arch. Toxicol. 87*, 1683−1696.

(11) Oesch, F., Fabian, E., Guth, K., and Landsiedel, R. (2014) Xenobiotic-metabolizing enzymes in the skin of rat, mouse, pig, guinea pig, man, and in human skin models. *Arch. Toxicol. 88*, 2135−2190.

(12) Gotz, C., Pfeiffer, R., Tigges, J., Ruwiedel, K., Hubenthal, U., Merk, H. F., Krutmann, J., Edwards, R. J., Abel, J., Pease, C., Goebel, C., Hewitt, N., and Fritsche, E. (2012) Xenobiotic metabolism capacities of human skin in comparison with a 3D-epidermis model and keratinocyte-based cell culture as in vitro alternatives for chemical testing: phase II enzymes. *Exp. Dermatol. 21*, 364−369.

(13) Gotz, C., Pfeiffer, R., Tigges, J., Blatz, V., Jackh, C., Freytag, E. M., Fabian, E., Landsiedel, R., Merk, H. F., Krutmann, J., Edwards, R. J., Pease, C., Goebel, C., Hewitt, N., and Fritsche, E. (2012) Xenobiotic metabolism capacities of human skin in comparison with a 3D epidermis model and keratinocyte-based cell culture as in vitro alternatives for chemical testing: activating enzymes (Phase I). *Exp. Dermatol. 21*, 358−363.

(14) 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., and Sakaguchi, H. (2015) Assessing skin sensitization hazard in mice and men using non-animal test methods. *Regul. Toxicol. Pharmacol. 71*, 337−351.

(15) Basketter, D., Ashikaga, T., Casati, S., Hubesch, B., Jaworska, J., de Knecht, J., Landsiedel, R., Manou, I., Mehling, A., Petersohn, D., Rorije, E., Rossi, L. H., Steiling, W., Teissier, S., and Worth, A. (2015) Alternatives for skin sensitisation: Hazard identification and potency categorisation: Report from an EPAA/CEFIC LRI/Cosmetics Europe cross sector workshop, ECHA Helsinki, April 23rd and 24th 2015. *Regul. Toxicol. Pharmacol. 73*, 660−666.

(16) Lepoittevin, J. P. (2006) Metabolism versus chemical transformation or pro- versus prehaptens? *Contact Dermatitis 54*, 73−74.

(17) Gerberick, F., Aleksic, M., Basketter, D., Casati, S., Karlberg, A. T., Kern, P., Kimber, I., Lepoittevin, J. P., Natsch, A., Ovigne, J. M., Rovida, C., Sakaguchi, H., and Schultz, T. (2008) Chemical reactivity measurement and the predictive identification of skin sensitisers. *Altern. Lab. Anim. 36*, 215−242.

(18) Jaeckh, C., Blatz, V., Fabian, E., Guth, K., van Ravenzwaay, B., Reisinger, K., and Landsiedel, R. (2011) Characterization of enzyme activities of cytochrome P450 enzymes, flavin-dependent monooxygenases, N-acetyltransferases and UDP-glucuronyltransferases in human reconstructed epidermis and full-thickness skin models. *Toxicol. In Vitro 25*, 1209−1214.

(19) Fabian, E. (2014) Are Xenobiotic Metabolizing Enzyme Activities in Cells Relevant for Skin Sensitization? Oral presentation held at DGPT congress in Hannover/Germany.

(20) Karlberg, A. T., Bergström, M. A., Börje, A., Luthman, K., and Nilsson, J. L. (2008) Allergic contact dermatitis - Formation, structural requirements, and reactivity of skin sensitizers. *Chem. Res. Toxicol. 21*, 53−69.

(21) Karlberg, A. T., Borje, A., Duus Johansen, J., Liden, C., Rastogi, S., Roberts, D., Uter, W., and White, I. R. (2013) Activation of non-sensitizing or low-sensitizing fragrance substances into potent sensitizers - prehaptens and prohaptens. *Contact Dermatitis 69*, 323−334.

(22) 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*, 489−504.

(23) 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. 33*, 1337−1352.

(24) Gerberick, G. F., Vassallo, J. D., Bailey, R. E., Chaney, J. G., Morrall, S. W., and Lepoittevin, J. P. (2004) Development of a peptide reactivity assay for screening contact allergens. *Toxicol. Sci. 81*, 332−343.

(25) Standing, K. G. (2003) Peptide and protein de novo sequencing by mass spectrometry. *Curr. Opin. Struct. Biol. 13*, 595−601.

(26) Medzihradszky, K. F., and Chalkley, R. J. (2015) Lessons in de novo peptide sequencing by tandem mass spectrometry. *Mass Spectrom. Rev. 34*, 43−63.

(27) Natsch, A., and 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.

(28) Basketter, D., Ashikaga, T., Casati, S., Hubesch, B., Jaworska, J., de Knecht, J., Landsiedel, R., Manou, I., Mehling, A., Petersohn, D., Rorije, E., Rossi, L. H., Steiling, W., Teissier, S., and Worth, A. (2015) Alternatives for skin sensitisation: Hazard identification and potency categorisation: Report from an EPAA/CEFIC LRI/Cosmetics Europe cross sector workshop, ECHA Helsinki, April 23rd and 24th 2015. *Regul. Toxicol. Pharmacol. 73*, 660−666.

(29) Cooper, J. A., Saracci, R., and Cole, P. (1979) Describing the validity of carcinogen screening-tests. *Br. J. Cancer 39*, 87−89.

(30) Siegel, S. (1956) *Non Parametric Statistics for Behavioural Sciences* McGraw Hill, New York.

(31) Gerberick, G. F., Troutman, J. A., Foertsch, L. M., Vassallo, J. D., Quijano, M., Dobson, R. L. M., Goebel, C., and Lepoittevin, J. P. (2009) Investigation of peptide reactivity of pro-hapten skin sensitizers using a peroxidase-peroxide oxidation system. *Toxicol. Sci. 112*, 164−174.

(32) Roberts, D. W., and 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.

(33) Aleksic, M., Thain, E., Roger, D., Saib, O., Davies, M., Li, J., Aptula, A., and Zazzeroni, R. (2009) Reactivity profiling: covalent modification of single nucleophile peptides for skin sensitization risk assessment. *Toxicol. Sci. 108*, 401−411.

(34) Scholes, E. W., Pendlington, R. U., Sharma, R. K., and Basketter, D. A. (1994) Skin metabolism of contact allergens. *Toxicol. In Vitro 8*, 551−553.

(35) Natsch, A., Gfeller, H., Rothaupt, M., and Ellis, G. (2007) Utility and limitations of a peptide reactivity assay to predict fragrance allergens in vitro. *Toxicol. In Vitro 21*, 1220−1226.

(36) Bergstrom, M. A., Ott, H., Carlsson, A., Neis, M., Zwadlo-Klarwasser, G., Jonsson, C. A. M., Merk, H. F., Karlberg, A. T., and Baron, J. M. (2007) A skin-like cytochrome P450 cocktail cctivates prohaptens to contact allergenic metabolites. *J. Invest. Dermatol. 127*, 1145−1153.

(37) Niklasson, I. B., Delaine, T., Islam, M. N., Karlsson, R., Luthman, K., and Karlberg, A. T. (2013) Cinnamyl alcohol oxidizes rapidly upon air exposure. *Contact Dermatitis 68*, 129−138.

(38) Schneider, K., and Akkan, Z. (2004) Quantitative relationship between the local lymph node assay and human skin sensitization assays. *Regul. Toxicol. Pharmacol. 39*, 245−255.

DOI: 10.1021/acs.chemrestox.6b00055
Chem. Res. Toxicol. XXXX, XXX, XXX−XXX

(39) Kligman, A. M. (1966) Identification of contact allergens by human assay 0.3. Maximization test - A procedure for screening and rating contact sensitizers. *J. Invest. Dermatol. 47*, 393.
(40) Roberts, D. W., Patlewicz, G., Kern, P. S., Gerberick, F., Kimber, I., Dearman, R. J., Ryan, C. A., Basketter, D. A., and Aptula, A. O. (2007) Mechanistic applicability domain classification of a local lymph node assay dataset for skin sensitization. *Chem. Res. Toxicol. 20*, 1019−1030.
(41) Roberts, D. W., Aptula, A. O., and 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.
(42) Kern, P. S., Gerberick, G. F., Ryan, C. A., Kimber, I., Aptula, A., and Basketter, D. A. (2010) Local lymph node data for the evaluation of skin sensitization alternatives: a second compilation. *Dermatitis 21*, 8−32.
(43) Natsch, A., and Haupt, T. (2013) Utility of rat liver S9 fractions to study skin-sensitizing prohaptens in a modified KeratinoSens assay. *Toxicol. Sci. 135*, 356−368.
(44) Basketter, D. A., Smith Pease, C. K., and Patlewicz, G. Y. (2003) Contact allergy: the local lymph node assay for the prediction of hazard and risk. *Clin. Exp. Dermatol. 28*, 218−221.
(45) Aptula, A. O., Enoch, S. J., and Roberts, D. W. (2009) Chemical mechanisms for skin sensitization by aromatic compounds with hydroxy and amino groups. *Chem. Res. Toxicol. 22*, 1541−1547.
(46) Casati, S., Aschberger, K., Asturiol, D., Basketter, D., Dimitrov, S., Dumont, C., Karlberg, A.-T., Lepoittevin, J.-P., Patlewicz, G., Roberts, D. W., and Worth, A. (2016) Ability of non-animal methods for skin sensitisation to detect pre- and pro-haptens, *JRC Technical Report of EURL ECVAM Expert Meeting*, Publications Office of the European Union, Luxembourg, https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-status-reports/pre-prohapten-workshop-report.
(47) Karlberg, A. T., Boman, A., and Wahlberg, J. E. (1980) Allergenic potential of abietic acid, colophony and pine resin-HA. Clinical and experimental studies. *Contact Dermatitis 6*, 481−487.
(48) Karlberg, A. T. (1989) Pure abietic acid is not allergenic. *Contact Dermatitis 21*, 282−285.
(49) Urbisch, D., Honarvar, N., Kolle, S. N., Mehling, A., Ramirez, T., Teubner, W., and Landsiedel, R. (2016) Peptide reactivity associated with skin sensitization: The QSAR Toolbox and TIMES compared to the DPRA. *Toxicol. In Vitro*, in press.
(50) Ramirez, T., Stein, N., Aumann, A., Remus, T., Edwards, A., Norman, K. G., Ryan, C., Bader, J. E., Fehr, M., Burleson, F., Foertsch, L., Wang, X., Gerberick, F., Beilstein, P., Hoffmann, S., Mehling, A., van, R. B., and Landsiedel, R. (2016) Intra- and inter-laboratory reproducibility and accuracy of the LuSens assay: A reporter gene-cell line to detect keratinocyte activation by skin sensitizers. *Toxicol. In Vitro 32*, 278−286.

DOI: 10.1021/acs.chemrestox.6b00055
Chem. Res. Toxicol. XXXX, XXX, XXX−XXX

4.4 Prediction of skin sensitization potency using peptide reactivity data

Britta Wareing, Daniel Urbisch, Naveed Honarvar, Susanne N. Kolle, Robert Landsiedel

Summary of the publication: This publication analyses the utility of peptide-based methods to predict skin sensitization potency. The DPRA performed according to its OECD test guideline cannot predict potency, while a modification of this assay measuring kinetics shows promising results.

Journal: Toxicology In Vitro (Elsevier)

Impact factor: 2.866 (Clarivate Analytics, 2017)

Type of authorship: co-author

Status of publication: submitted and under revision

My contribution to the publication: approximately 40%

- Estimation of the utility of the DPRA to predict skin sensitization potency
- Performance of kinetic experiments and definition of the prediction model of this new assay protocol
- Estimation of the utility of the kinetic experiments to predict skin sensitization potency
- Calculation of Cooper statistics
- Detailed discussion of false results
- Estimation of the utility of the non-animal methods to identify pre- and pro-haptens
- Preparation of parts of the manuscript referring to the DPRA and the kinetic experiments including all figures and tables

Elsevier Editorial System(tm) for Toxicology in Vitro

Manuscript Draft

Manuscript Number:

Title: Prediction of skin sensitization potency sub-categories using peptide reactivity data

Article Type: Research paper

Section/Category: New Methods and Models

Keywords: skin sensitisation; peptide reactivity; potency; GHS sub-categories

Corresponding Author: Dr. Robert Landsiedel, Dr. rer. nat., ERT, DABT, Fellow ATS

Corresponding Author's Institution: BASF SE

First Author: Britta Wareing, B.Sc.

Order of Authors: Britta Wareing, B.Sc.; Daniel Urbisch, M.Sc.; Annette Mehling, Ph.D.; Naveed Honarvar, Ph.D.; Susanne N Kolle, Dr. sc. hum.; Robert Landsiedel, Dr. rer. nat., ERT, DABT, Fellow ATS

Abstract: While the skin sensitization hazard of substances can already be identified using non-animal methods, the classification of potency sub-categories GHS-1A and 1B is still challenging. Potency can be measured by the dose at which an effect is observed; since the protein-adduct formation is determining the dose of the allergen in the skin, peptide reactivity was used to assess the potency.
The Direct Peptide Reactivity Assay (DPRA; one concentration and reaction-time) provided an adduct yield which did not sufficiently discriminate between sub-categories 1A and 1B (56% accuracy compared to 123 LLNA data). An augmented protocol termed 'quantitative DPRA' (several concentrations and one reaction-time), discriminated sub-categories GHS-1A and 1B with an accuracy of 81% or 57% compared to 36 LLNA or 14 human data, respectively. The analysis of the Cys-adducts was already sufficient; addition of Lys-adducts did not improve the predictivity. An additional modification, the 'kinetic DPRA' (several concentrations and reaction-times) was used to approximate the rate constant of Cys-peptide-adduct formation. 19 of 20 substances were correctly assigned to the potency sub-categories (LLNA data), and the predictivity for 14 human data was equally high.
These results warrant the kinetic DPRA for further validation in order to fully replace in vivo testing for assessing skin sensitization including potency sub-classification.

Suggested Reviewers: Ellen Fritsche Prof. Dr. med.
Professor, Heinrich Heine University Düsseldorf
ellen.fritsche@uni-duesseldorf.de
substantial expertise on skin sensitisation and in vitro methods

Andrew Worth Ph.D.

EU Joint Research Center
Andrew.Worth@ec.europa.eu
significant expertise on in vitro approaches and their regulatory implications

Rodger Curren Ph.D.
President, Institute of In Vitro Science
rcurren@iivs.org
outstanding expertise in establisching and validating in vitro methods, especially on skin and eye effects - as well as their application in product safety.

David Basketter D.Sc.
DABMEB Consultancy Ltd.
dabmebconsultancyltd@me.com
outstanding expertise on skin sensitisation methods.

Jean-Pierre Lepoittevin Prof.
Professor, Institute of Chemistry, Uni Strasbourg
jplepoit@unistra.fr
outstanding expert in protein reactivity of haptens.

Prediction of skin sensitization potency sub-categories using peptide reactivity data

Britta Wareing[a], Daniel Urbisch[a], Annette Mehling[b], Naveed Honarvar[a], Susanne Noreen Kolle[a], Robert Landsiedel[a,*]

[a] BASF SE, Experimental Toxicology and Ecology, Ludwigshafen, Germany

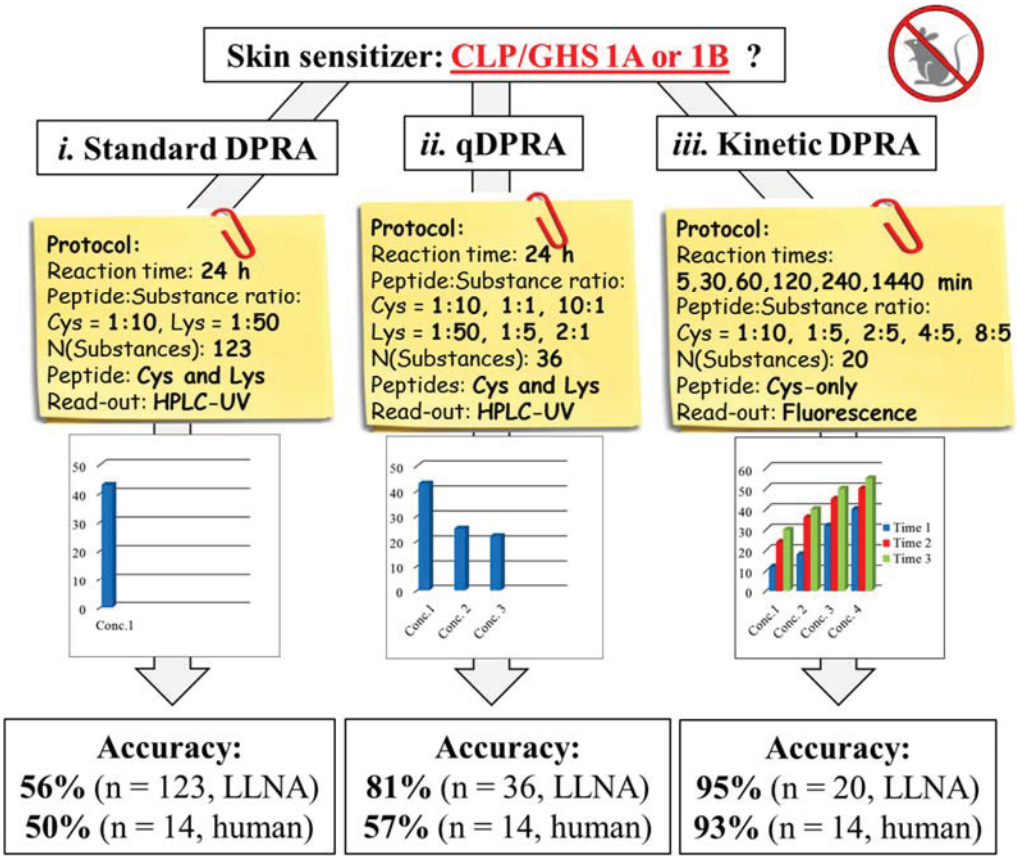
[b] BASF Personal Care and Nutrition GmbH, Duesseldorf, Germany

*Corresponding author; Phone: +49 621 60 56203; Fax: +49 621 60 58134; E-mail address: robert.landsiedel@basf.com

Keywords

Regulatory toxicology, hazard identification, risk assessment, globally harmonized system of classification and labeling of substances (CLP/GHS), protein-binding, allergic contact dermatitis

Graphical abstract



Abstract

While the skin sensitization hazard of substances can already be identified using non-animal methods, the classification of potency sub-categories GHS-1A and 1B is still challenging. Potency can be measured by the dose at which an effect is observed; since the protein-adduct formation is determining the dose of the allergen in the skin, peptide reactivity was used to assess the potency.

The Direct Peptide Reactivity Assay (DPRA; one concentration and reaction-time) provided an adduct yield which did not sufficiently discriminate between sub-categories 1A and 1B (56% accuracy compared to 123 LLNA data). An augmented protocol termed 'quantitative DPRA' (several concentrations and one reaction-time), discriminated sub-categories GHS-1A and 1B with an accuracy of 81% or 57% compared to 36 LLNA or 14 human data, respectively. The analysis of the Cys-adducts was already sufficient; addition of Lys-adducts did not improve the predictivity. An additional modification, the 'kinetic DPRA' (several concentrations and reaction-times) was used to approximate the rate constant of Cys-peptide-adduct formation. 19 of 20 substances were correctly assigned to the potency sub-categories (LLNA data), and the predictivity for 14 human data was equally high.

These results warrant the kinetic DPRA for further validation in order to fully replace *in vivo* testing for assessing skin sensitization including potency sub-classification.

[1] **Abbreviation list**

A, alanine; AOP, adverse outcome pathway; C, cysteine; Cys, cysteine; DNCB, 2,4-Dinotrochlorobenzene; DPRA, direct peptide reactivity assay; EGDMA, Ethylene glycol dimethacrylate; F, phenylalanine; FN, false negative; FP, false positive; h-CLAT, human cell-line activation test; HPLC, high performance liquid chromatography; HRIPT, human repeated insult patch test; IATA, integrated testing approaches and assessments; K, lysine; LLNA, local lymph node assay; logP, octanol-water partition coefficient (logarithmic form); Lys, lysine; MA, Michael acceptor; MIE, molecular initiating event; n, number of chemicals; OECD, Organization for Economic Co-operation and Development; OECD TG, OECD test guideline; qDPRA, quantitative DPRA; QP, quinone precursor; R, arginine; RC, reactive carbonyl; SB, Schiff 'base former; S_NAr, Aromatics reacting by nucleophilic substitutions; TN, true negative; TP, true positive.

1. Introduction

It has been estimated that 15-20% of the general population suffers from ACD (Bruckner *et al.* 2000). Accordingly, the identification of skin sensitizing properties of substances forms an important pillar of substance hazard identification and risk assessment. To identify hazards and to determine the need for risk management measures, the outcomes of toxicity tests are translated into hazard categories. Traditionally, hazard and potency are determined by animal studies (OECD 2010;OECD TG 406 1992).

In the EU, the provisions for hazard classification are laid down in *Regulation (EC) No 1272/2008 on the classification, labelling and packaging of substances and mixtures* (CLP; European Parliament 2008). The CLP Regulation has implemented most of the *United Nation's Globally Harmonized System of Classification and Labelling of Chemicals* (GHS; United Nations 2013). Specifically for skin sensitization, the CLP/GHS system prescribes discriminating between category 1 (sensitizers) and no category (non-sensitizers). Sub-classification of category 1 sensitizers by potency (sub-category 1A/1B) only has to be performed if sufficient data are available. Also under the REACH Regulation (Council Directive 1907/2006/EEC 2006), information on skin sensitization potency is not required if the data are not sufficient for sub-categorization (in which case sensitizers are assigned to category 1). In general, this is based on data generated by the murine local lymph node assay (LLNA; OECD TG 429; OECD 2010), although guidance is also given on how data from other sources, e.g. human or guinea pig data, can be used. The LLNA is preferably used as it includes dose-response assessments, which can be used to determine estimated test substance concentrations that lead to a three-fold increase in the stimulation index (= EC3 value). According to the ECHA guidance (https://echa.europa.eu/documents/10162/13562/clp_en.pdf), which also describes how to use data derived from human or guinea pigs, substances that yield the EC3 value at concentrations ≤ 2% should be evaluated as strong sensitizers, i.e. sub-category 1A; sensitizers with EC3 values above 2% are then classified into sub-category 1B (ICCVAM 2011). Between weak and strong sensitizers, the relative skin sensitizing potency may vary by up to five orders of magnitude (Basketter *et al.* 2007).

In the EU, a number of regulations limit the use of animal testing for regulatory purposes. *Regulation (EC) No 1907/2006 on the Registration, Evaluation, Authorization and Restriction of Chemicals* (REACH;

Council Directive 1907/2006/EEC 2006) requires that animal testing for the generation of new data on a toxicological endpoint may only be undertaken as a last resort. The recast *Regulation (EC) No 1223/2009 on Cosmetic Products* (EU 1223/2009 2009) implemented an animal testing ban with a concomitant marketing ban that came into full force in March 2013 for cosmetic ingredients and products tested on animals after this date. Finally, *Directive 2010/63/EU on the protection of animal used for scientific purposes* (The European Parliament 2010) has implemented the 3Rs principle to replace, reduce and refine animal testing (Russell and Burch 1959). These and other legal and ethical requirements have fostered the development of non-animal methods for regulatory skin sensitization testing and much progress has been made over the past few years (Mehling *et al.* 2012). Three non-animal methods have now been successfully validated and adopted as Organization for Economic Co-operation and Development (OECD) test guidelines (TG) (OECD 2015a;OECD 2015b;OECD 2016b). Test methods adopted as OECD TG can generally be used for regulatory purposes, i.e. for the hazard identification and risk assessment of substances in an occupational setting or that are intended to be marketed. The revision of Annex VII of the REACH legislation now mandates that non-animal testing has to be conducted and that justification must be given for *in vivo* testing in order to generate new data to assess skin sensitization potentials (European Commission 2016).

Currently, a single non-animal method cannot assess the skin sensitization potential of a substance. Instead, different non-animal methods need to be integrated into testing strategies, which preferably address key events of the adverse outcome pathway (AOP) for skin sensitization (OECD 2016a). AOPs describe the sequence of substance-induced pathophysiological events, beginning with a specific molecular initiating event (MIE) that initiates a sequence of early cellular events that ultimately result in an observable (toxic) effect (Ankley *et al.* 2010). Thereby, AOPs provide a framework to describe the mechanisms of toxicity that are relevant for a given toxicological endpoint. The four key events (KE) of the AOP for skin sensitization are: KE 1: The covalent binding of electrophilic substances to nucleophilic centers in skin proteins (also considered to be the MIE); KE 2: Events in keratinocytes (e.g. inflammatory responses); KE 3: Events in dendritic cells (DC; e.g. DC maturation); and KE 4: Events in lymph nodes (e.g. T cell priming

and proliferation) (Basketter *et al.* 2013;Basketter *et al.* 2015;OECD 2012a;OECD 2012b). Together, these four key events reflect the induction or sensitization phase of substance-induced allergic contact dermatitis (ACD) that may become clinically manifest upon secondary exposure to the sensitizing (allergenic) substance (i.e. upon completion of the elicitation or challenge phase) (Mehling *et al.* 2012). The MIE, KE 1, in the AOP for skin sensitization (i.e. the covalent interaction with skin proteins) can be assessed in the *in chemico* Direct Peptide Reactivity Assay (DPRA) that was originally described by Gerberick and coworkers (Gerberick *et al.* 2004;Gerberick *et al.* 2007) and that has been adopted as OECD TG 442C (OECD 2015a). In the DPRA, the covalent interaction with proteins is determined by quantifying the peptide reactivity of a substance towards model synthetic heptapeptides that contain either lysine or cysteine (in the following: Lys-peptide and Cys-peptide). Peptide reactivity may depend on a variety of factors including the test substance's electrophilicity, nucleophilicity, the reaction rate and concurring reactions or the stability of resulting conjugates/adducts (Gerberick *et al.* 2008) as well the presence or absence of abiotic or metabolic activation (Urbisch *et al.* 2016a). The *in chemico* DPRA is generally used to assess skin sensitization potential for hazard identification, a threshold of 6.38% mean Cys- and Lys-peptide depletion is used to discriminate between skin sensitizers and non-sensitizers in the $DPRA_{Cys\&Lys}$ (OECD 2015a). OECD TG 442C also describes a prediction model (PM) that is based upon Cys-peptide depletion alone. In this PM, a threshold of 13.89% Cys-peptide depletion has been laid down to discriminate sensitizers from non-sensitizers in the $DPRA_{Cys\text{-}only}$ (OECD 2015a). Moreover, the OECD TG 442C provides a PM to quantify peptide reactivity by assigning sensitizing substances to one of three 'reactivity classes', i.e. low, moderate and high reactivity. The use of DPRA and other *in vitro* methods to determine skin sensitization for REACH has been critically discussed (Sauer *et al.* 2016).

In the context of skin sensitization, potency can be defined as the amount of an allergen that is needed to sensitize a naïve individual (Kimber *et al.* 2003). Although other factors, e.g. genetic predisposition or whether specific T-cells are present in the lymph node, will also play a role (not all protein conjugates are sensitizers), the more protein-adducts a potentially sensitizing low-molecular weight substance is forming in the skin, the more potent it may be as sensitizer (Friedmann 2007). This appears plausible, since low-

molecular weight substances are generally not allergenic on their own and their peptide reactivity determines the amount of the antigen that will be formed in the skin proteins. Hence, skin sensitization potency is related to the peptide reactivity of a substance. Since the DPRA quantifies peptide reactivity, obtained data can be linked to the determination of skin sensitization potency (Gerberick *et al.* 2007;OECD 2015a).

While *in vivo* methods to assess skin sensitization potency are available, non-animal methods to predict the potency of skin sensitizers are still under development (e.g., (Hirota *et al.* 2013;Jaworska *et al.* 2015;Natsch *et al.* 2015;Takenouchi *et al.* 2015)). The present study was initiated to assess the utility of peptide reactivity data to distinguish between two levels of skin sensitization potency, i.e. to discriminate between CLP/GHS sub-categories 1A and 1B. Specifically, the 'standard DPRA' as laid down in OECD TG 442C was compared to a quantitative DPRA (qDPRA) that allows establishing concentration-response relationships (i.e. the concentration dependence of peptide reactivity) and a kinetic DPRA, a modification of the assay published by Roberts and Natsch (Roberts and Natsch 2009) that allows investigating the reaction kinetics of peptide reactivity (i.e. whether the concentration dependence of peptide reactivity changes with increasing or decreasing reaction times).

2. Materials and Methods

In this study, the standard DPRA and variants thereof were assessed for their ability to discriminate potency classes according to GHS. Figure 1 gives an overview of the chemical selection criteria and the main characteristics of the test methods used.

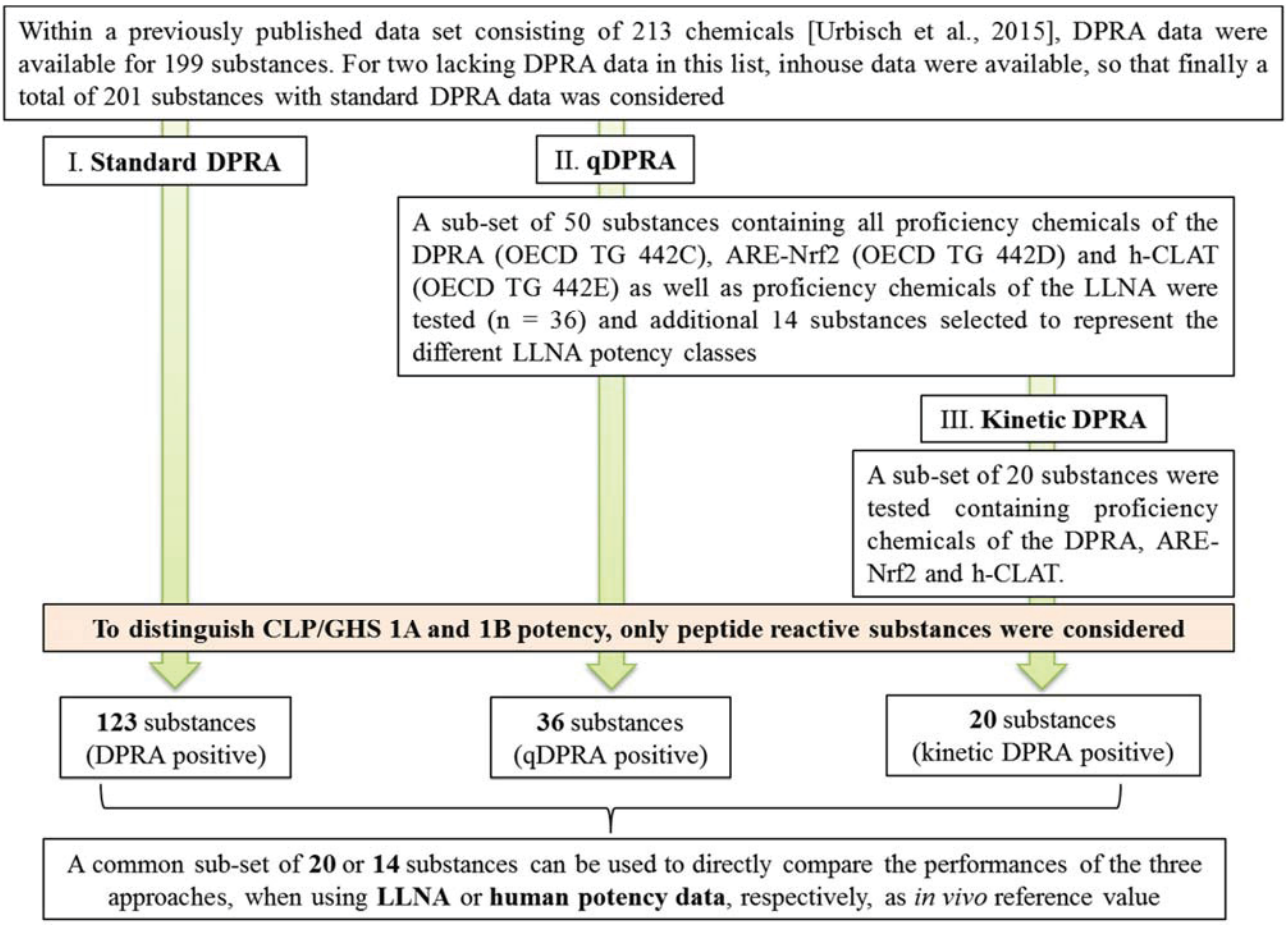


Figure 1: Overview of the selection of substances for the standard DPRA, the qDPRA and the kinetic DPRA.

Albeit high lipophilicity and high molecular weight do not present a general limit, the molecular weight and logP of a substance influence their skin penetration (Magnusson *et al.* 2004;Roberts *et al.* 2016a;Roberts *et al.* 2016b) and hence their dose in the skin and skin sensitization potential. A potential correlation of these two substance properties to the *in vivo* skin sensitization potency was investigated. No correlation of potency and logP or molecular weight could be observed (data not shown; also previously observed by Natsch and coworkers for a slightly different data set (Natsch *et al.* 2013)). Hence, the data set illustrated in Figure 1 and used in this study is not biased by these properties.

2.1 Materials

For newly generated data obtained from the qDPRA and kinetic DPRA conducted in this study, synthetic heptapeptides containing cysteine (Ac-RFAACAA-COOH; MW = 751.9 g/mol; purity >98%) or lysine (Ac-RFAAKAA-COOH; MW = 776.2 g/mol; purity >95%) were obtained from Genscript Inc. (USA) or RS Synthesis (USA). All test substances used in these assays as well as the test reagents acetonitrile, sodium phosphate- and ammonium acetate-buffers, and the fluorescence dye monobromobimane were supplied by Sigma Aldrich (Germany). Substances tested and substance information can be found in Supplementary Tables S1 and S2. The substances tested in the qDPRA and kinetic DPRA can be found in the following result section (Tables 7 and 8, respectively).

For the *in chemico – in vivo* comparisons, previously published *in vivo* reference data were retrieved from Urbisch *et al.* 2015. The data derived from this paper include the comparative data used for the standard DPRA (i.e. OECD TG 442) and LLNA (OECD TG 429). Sub-sets of data were then selected for the three peptide reactivity methods (standard DPRA, qDPRA and kinetic DPRA) according to the criteria given in Figure 1. The selection criteria for the substances tested in the qDPRA includes that the substances are described as performance standards in the test guidelines for the LLNA (OECD TG 429) or as proficiency chemicals in the DPRA (OECD TG 442C), the 'ARE-Nrf2 Luciferase Test Method' (OECD TG 442D) and the 'Human Cell Line Activation Test' (h-CLAT, OECD TG 442E), and that the CLP/GHS sub-categories 1A and 1B were represented. The subset covering the proficiency chemicals of OECD TGs 442C/D/E were then further tested in the kinetic DPRA (see Table 7) to assess whether this method provides further information that can be used in potency assessments.

2.2 Methods

In this study, three different approaches were evaluated for their use in assigning a substance to its respective CLP/GHS sub-category. Table 1 gives an overview of the three approaches, the respective prediction model and the number substances being analyzed.

Table 1: Overview of the test protocols applied in the standard DPRA, the qDPRA, and the kinetic DPRA.

Approach	Peptide used	Peptide: substance ratio	Reaction time	No. of sub-stances	Output for respective PM	Origin of data
Standard DPRA: $DPRA_{Cys\&Lys}$	Cys	1:10	24 h	123	Mean Cys- and Lys-peptide depletion (22.62%)	[a]Literature
	Lys	1:50				
Standard DPRA: $DPRA_{Cys\text{-}only}$	Cys	1:10	24 h	108[c]	Cys depletion (23.09%)	[a]Literature
Quantitative DPRA: $qDPRA_{Cys\&Lys}$	Cys	1:10, 1:1, 10:1	24 h	36	EC6.38	Present study
	Lys	1:50, 1:5, 2:1				
Quantitative DPRA: $qDPRA_{Cys\text{-}only}$	Cys	1:10, 1:1, 10:1	24 h	32[c]	EC13.89	Present study
Kinetic DPRA: [b] Kinetic $DPRA_{Cys\text{-}only}$	Cys	1:10, 1:5, 2:5, 4:5, 8:5	30, 120 min & 24 h (optional 5, 60 & 240 min)	20	log *k* of highest reactivity	Present study

[a] data were obtained from a previously published dataset (Urbisch *et al.* 2015); Cys/Lys = cysteine-/lysine-containing model peptide; *k* = reaction rate constant; [b] standard DPRA and qDPRA were tested with Cys- and Lys-peptide (i.e. $DPRA_{Cys\&Lys}$ and $qDPRA_{Cys\&Lys}$) and Cys-peptide only (i.e. $DPRA_{Cys\text{-}only}$ and $qDPRA_{Cys\text{-}only}$). Kinetic DPRA was only tested with the Cys-peptide (i.e. kientik $DPRA_{Cys\text{-}only}$); [c] subsets of the 123 and 36 substances evaluated in the standard and qDPRA that were Cys-reactive.

2.2.1 Standard DPRA

The standard DPRA was conducted in accordance with OECD TG 442C. Substance classification was the conducted according to the criteria defined in the OECD TG (Table 2).

Table 2: Peptide reactivity classes as defined in OECD TG 442C.

Prediction model (peptide:test substance ratio)	**Cysteine 1:10 / lysine 1:50; $DPRA_{Cys\&Lys}$**	**Cysteine 1:10; $DPRA_{Cys\text{-}only}$**
Reactivity class	**Mean depletion of Cys- & Lys-peptide [%]**	**Cys-peptide depletion [%]**
No or minimal reactivity	[depletion] ≤ 6.38	[depletion] ≤ 13.89
Low reactivity	6.38 < [depletion] ≤ 22.62	13.89 < [depletion] ≤ 23.09
Moderate reactivity	22.62 < [depletion] ≤ 42.47	23.09 < [depletion] ≤ 98.24
High reactivity	42.47 < [depletion] ≤ 100	98.24 < [depletion] ≤ 100

2.2.2 Quantitative DPRA (qDPRA)

The qDPRA was generally performed according to the procedure described in OECD TG 442C; however, in addition to the standard test-substance concentration of 100 mM, substances were also tested at 1 mM and 10 mM (stock concentrations; ratios 'peptide:test substance' were 2:1, 1:5 and 1:50 for lysine-peptide or 10:1, 1:1 and 1:10 for cysteine-peptide) by dissolving substances in a suitable solvent (e.g. acetonitrile or water). All substances were tested in triplicate for each concentration within the same test run (with the exception of EGDMA (ethylene glycol dimethacrylate) and citral, which were tested in 18 or ten test runs, respectively). The estimated test substance concentrations (referring to the stock concentration) required to cause a mean of 6.38% Cys- and Lys-peptide depletion (EC6.38; cysteine 1:10 / lysine 1:50 prediction model) in the $qDPRA_{Cys\&Lys}$ or 13.89% Cys-peptide depletion (EC13.89; cysteine 1:10 prediction model) in the $qDPRA_{Cys\text{-}only}$ were calculated by linear regression using Microsoft Excel 2013. The data obtained were submitted to receiver operator characteristic (ROC) analysis (see Section 2.3.2) to determine the best-fit threshold value for each prediction model that would allow discriminating sub-category 1A from 1B sensitizers.

2.2.3 Kinetic DPRA

The kinetic DPRA was performed as a modification of the assay described by Roberts and Natsch (Roberts and Natsch 2009). All substances were first dissolved in acetonitrile to yield solutions of 20 mM, and then dilution series of 20, 10, 5, 2.5 and 1.25 mM were prepared. The kinetic DPRA consisted of the following steps: First, 120 µL of 0.667 mM Cys-peptide solution in phosphate buffer (pH 7.5) were added to each well of a black 96-well plate. Next, 40 µL of the respective substance solution was added to each well. This yielded 0.5 mM peptide concentrations in each well. Substance dilutions were added leading to concentrations of 5, 2.5, 1.25, 0.625 and 0.3125 mM (final ratios of peptide and test substance = 1:10, 1:5, 2:5, 4:5, 8:5). All substances were tested in triplicate within the same run.

The plates were sealed with impermeable foil directly after application of the substance and incubated in the dark at 25°C. Unless noted otherwise (see Table 7), all substances were incubated for 30, 60, 120, 240 and 1440 min. In addition, seven substances of higher reactivity were incubated for a shorter time periods, i.e. 5 min, in a second test run. After the respective reaction time, each test run was stopped by the addition of 3 mM of the fluorescence dye, i.e. monobromobimane solution (diluted in acetonitrile). Highly reactive and non-fluorescent monobromobimane rapidly reacts with unbound cysteine moieties of the model peptide to form a fluorescent complex. The higher the intensity in fluorescence, the more cysteine moieties were unbound after the respective reaction time and the less peptide-reactive the test chemical.

The substance solutions containing monobromobimane were then further incubated for 5 min in the dark on a plate shaker, and fluorescence then detected using a TriStar Multimode reader LB 942 (Berthold, Germany). Fluorescence intensities were normalized relative to the substance without the peptide, as well as the phosphate buffer and acetonitrile (background fluorescence). Peptide depletion was expressed as percent decrease in peptide concentration, which was calculated based on a calibration series. For each incubation time, the remaining (non-reacted) amount of Cys-peptide was determined and plotted against the respective substance concentration. The obtained slope was divided by the incubation time to determine the second order reaction rate constants *k* (Roberts and Natsch 2009). For each of the 20 substances, individual *k* values (i.e. reaction rate constants) were calculated for each incubation time. The maximum *k* value recorded for a given substance (at any concentration or point in time) was used for the further evaluation (in logarithmic form (log k)). The data obtained in the kinetic DPRA were submitted to ROC analysis (see Section 2.3.2) to determine the best-fit threshold value that would allow discriminating CLP/GHS sub-category 1A from sub-category 1B sensitizers.

Peptide reactivity was measured in the kinetic DPRA using a test protocol that is amenable to a high throughput setting and that allows applying exact incubation times. However, highly reactive monobromobimane can only detect sulfhydryl groups of the Cys-peptide. Therefore, only Cys-peptide depletion was measured in the kinetic DPRA.

2.2.3 Mechanistic domains

Correlations with the available LLNA data were established for the total set of 123 substances as well as for seven specific sup-groups of this data set that were established from a chemical perspective to reflect different putative reaction protein-binding mechanisms, as previously described (Urbisch *et al.* 2015). The substances considered for this approach and their probable mechanistic domain adherence are listed in the supplementary part (see Supplementary Table S1). Figure 2 provides the graphical representations of how members of the respective groups generally react with peptides.

Specifically, these were the groups of:

(i) Acylating agents e.g. phthalic anhydride (15 substances: 8 CLP/GHS Cat 1A, 7 CLP/GHS Cat 1B based on LLNA data): Acylating agents contain an activated acyl group where the nucleophile (i.e. the compound that donates a lone pair of electrons) attacks at the carbonyl carbon displacing a leaving group. The rate of acylation of a nucleophile depends on its measure of acidity (Wong SS. 1991).

(ii) Michael acceptors e.g. diethyl maleate (30 substances: 5 CLP/GHS Cat 1A, 24 CLP/GHS Cat 1B, 1 non-sensitizer based on LLNA data): The Michael reaction is the nucleophilic addition of an enolate of a ketone or aldehyde to an α,β-unsaturated carbonyl compound (Michael A. 1887). Michael acceptors are capable of forming irreversible bonds with proteins or DNA, but they may also act reversibly with nucleophiles (Polepally *et al.* 2014).

(iii) Quinone precursors e.g. isoeugenol (14 substances: 12 CLP/GHS Cat 1A, 1 CLP/GHS Cat 1B based on LLNA data): Quinones are cyclic organic compounds containing two carbonyl groups, either adjacent or separated by a vinyl group in a six-membered unsaturated ring. Quinones are usually prepared by oxidation of aromatic amines, polyhydric phenols and polynuclear hydrocarbons (adapted from: Encyclopaedia Britannica; available at https://www.britannica.com/science/quinone). After abiotic or metabolic activation, the resulting quinone reacts with the peptide in a Michael addition.

(iv) Reactive carbonyls (including 1,2-dicarbonyls, 1,3-dicarbonyls, di-substituted α,β-unsaturated carbonyls and aldehydes); in the literature often defined as Schiff base (or imine) formers, i.e. compounds in which the C=O double bond is replaced by a C=N double bond. (16 substances: 12 CLP/GHS Cat 1A, 4 CLP/GHS Cat 1B based on LLNA data): Reactive carbonyl species may damage biological structures, e.g. by disrupting the structure and function of proteins (Semchyshyn 2014).

(v) Substances reacting in nucleophilic substitutions type 2 e.g. 2-mercaptobenzothiazole (18 substances: 8 CLP/GHS Cat 1A, 7 CLP/GHS Cat 1B, 3 non-sensitizer based on LLNA data): Nucleophilic substitution is the reaction of a nucleophile with an electrophile (that accepts the electron pair). In the type 2 nucleophilic substitution, 2 molecules are involved in the actual transition state (http://www.organic-chemistry.org/namedreactions/nucleophilic-substitution-sn1-sn2.shtm).

(vi) Aromatics reacting in nucleophilic substitutions e.g. 2,4-Dinotrochlorobenzene (DNCB) (5 substances: 4 CLP/GHS Cat 1A, 1 CLP/GHS Cat 1B based on LLNA data): In aromatic nucleophilic substitutions, the nucleophile replaces, e.g., halides on an aromatic ring (Senger *et al.* 2012).

(vii) Substances that were devoid of structural alerts, e.g. sodium dodecyl sulfate (15 substances: 1 CLP/GHS Cat 1A, 5 CLP/GHS Cat 1B, 9 non-sensitizer based on LLNA data): this domain contains a diversity of substances, which have a lack for direct peptide reactivity in common. The non-sensitizing substances of this domain are not peptide reactive. However, some of the substances of this domain are sensitizers; therefore, a modification of the molecular structure for instance by metabolic enzymes is probable.

Eight substances could not be assigned to one of the mechanistic domains unambiguously e.g., oxazolone (4 CLP/GHS Cat 1A, 1 CLP/GHS Cat 1B, 3 non-sensitizer based on LLNA data), or they contained metals, i.e. cobalt chloride (CLP/GHS Cat 1A based on LLNA data) and nickel chloride (based on LLNA data non-sensitizer), so that they generally form coordination bonds between metal ions and electron-rich atoms (mainly heteroatoms, such as nitrogen or oxygen). Cobalt (II) complexes are characterized by a tetrahedral

arrangement, whereas nickel (II) is characterized by a square planar tetra-coordinated arrangement. The ligands and the geometry of the complexes determine whether the metal ions are sensitizing or not (Gerberick *et al.* 2008).

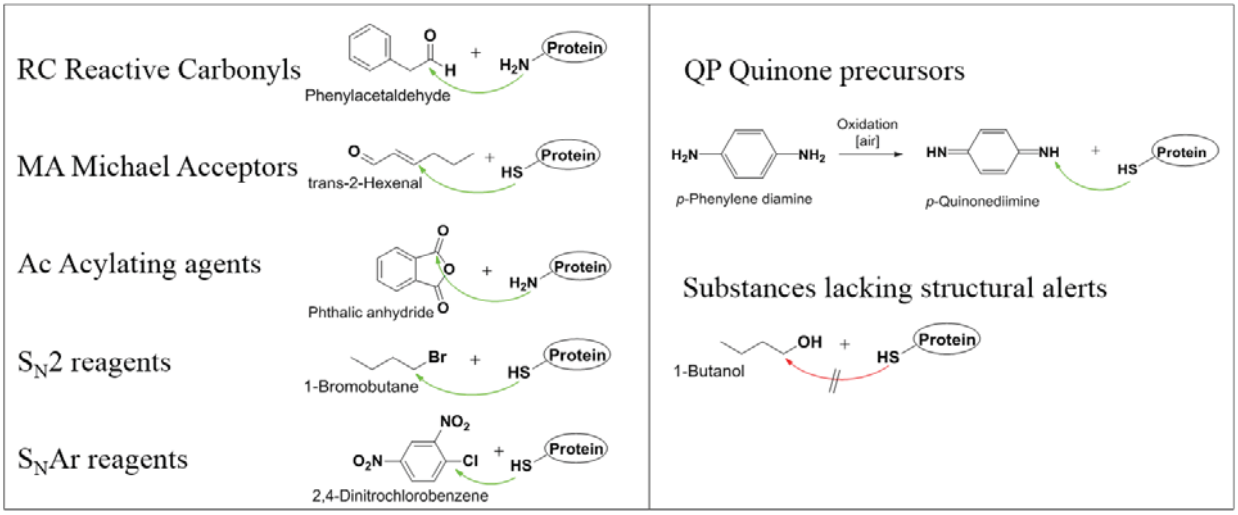


Figure 2: Examples of protein-substance interactions in the different chemical sub-groups (i.e. mechanistic domains).

Of note, the analysis by different reaction mechanisms (seven sub-groups) was only performed on results of the standard DPRA (123 test substances). The numbers of substances submitted to either the qDPRA or the kinetic DPRA was not sufficiently large to allow a similar sub-group analysis. Nevertheless, also for the qDPRA and the kinetic DPRA, information on the respective sub-groups of substances to was used for the discussion of false negative (FN) or false positive (FP) test results.

2.3 Statistics

2.3.1 Cooper statistics and correlation analyses

The performance of each variant of the DPRA was assessed using Cooper statistics (Cooper *et al.* 1979). For this purpose, all peptide reactivity data were compared to the available LLNA data, and the peptide reactivity data of the 14 substances, for which also human skin sensitization potency data were available,

were also compared to these human data. Specifically, contingency tables were generated to identify the number of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) assignments. Sensitivities, specificities and accuracies were calculated by the following equations:

- Sensitivity [%] = TP/(TP+FN)*100
- Specificity [%] = TN/(TN+FP)*100
- Accuracy [%] = (TP+TN)/(TP+FP+TN+FN)*100

2.3.2 Receiver Operator Characteristic (ROC) analyses

ROC curves were generated to assess the quality of the prediction of skin sensitizing potencies using the respective prediction models (PMs) applied in the standard DPRA, the qDPRA and the kinetic DPRA. For each PM, the threshold value leading to the highest accuracy was established. ROC curves illustrate the performance of binary classification systems as their threshold value is varied (Noe 1983). For the generation and evaluation of the ROC curves, the KNIME$^{(R)}$ analytics platform version 2.12.0 (https://www.knime.org/) was used (Berthold 2007), and statistical analysis was conducted using the pROC package version 1.5 of the statistical software R version 2.15.2 (https://www.r-project.org/) (R Core Team 2015;Robin *et al.* 2011). In the ROC analyses, only those substances were considered that were tested positive in the respective DPRA and the LLNA (i.e. not the FP as compared to the LLNA). These were 106 of the 123 substances for which standard DPRA data were available and 35 of the 36 substances evaluated in the qDPRA as well as all 20 substances tested in the kinetic DPRA. The results of the ROC analyses are shown in Supplementary Table S3.

3 Results

Of the 123 substances used to evaluate the standard DPRA-based potency predictions, 48 substances were categorized as strong sensitizers (sub-category 1A in the LLNA), 58 substances as weak to moderate sensitizers (sub-category 1B in the LLNA) and a further 17 substances were classified as non-sensitizers in accordance with the LLNA-based hazard sub-categorizations. The 36 substances evaluated in the qDPRA

can be subdivided into 16 strong and 19 weak sensitizers as well as one non-sensitizer in accordance with the LLNA-based hazard sub-categorizations. Of the subset evaluated in the kinetic DPRA, the 20 substances can be subdivided into 8 strong and 12 weak sensitizers in accordance with the LLNA-based hazard sub-categorizations.

In addition to LLNA data, available human potency data for 14 substances derived from human repeated insult patch tests (HRIPT) or human maximization tests (HMT) were available as dose per skin area (DSA_{05}) values (ICCVAM 2011). In accordance with the human potency data, four substances were assigned as sub-category 1A and ten substances as sub-category 1B. For one substance, the LLNA-based prediction of sub-category 1A was not confirmed by the human data (by which it was assigned to sub-category 1B).

3.1 Standard DPRA

A total of 123 substances (out of the data set of Urbisch et al. 2015) with positive outcomes in the standard DPRA (the test substance selection is depicted in Figure 1) was used to assess the predictivity in assigning substances as either sub-category 1A or 1B, since only these positive data were amenable to quantitative evaluation of sensitizing potency. To compare the standard DPRA-based potency predictions with the LLNA-based hazard sub-categorizations, the standard DPRA data were evaluated using the above-mentioned 'peptide reactivity class' systems laid down in OECD TG 442C for Cys- and Lys-peptide depletion-based and Cys-peptide depletion-based assessments (Table 1). However, the 'peptide reactivity class systems' encompass three classes of potency (determined by low, moderate or high peptide reactivity), whereas the CLP/GHS system that is legally required for hazard identification only comprises the two-level distinction between sub-category 1A and 1B. To account for this difference, the DPRA reactivity class 'low' was interpreted as sub-category 1B, and the DPRA reactivity classes 'moderate' and 'high' were jointly interpreted as sub-category 1A.

Accordingly, for the Cys- and Lys-peptide depletion-based PM ($DPRA_{Cys\&Lys}$), the threshold value of 22.62% that was set in OECD TG 442C to discriminate low reactivity from moderate reactivity was applied

to distinguish between weak and strong sensitizers (Table 1). For the Cys-peptide depletion-based PM ($DPRA_{Cys\text{-}only}$), the corresponding threshold value of 23.09% was applied to distinguish between weak and strong sensitizers (Table 1).

3.1.1 Results of the standard DPRA for all substances

The Cooper statistics for the skin sensitizing potency compared to the available LLNA potency data using the standard DPRA can be found in Tables 3a and 3b. Applying the Cys- and Lys-peptide depletion-based threshold (22.62%) to the 123 substances resulted in an almost perfect sensitivity (98%), but a low specificity (29%) and a low overall accuracy (56%) (Table 3a).

Table 3a: Standard DPRA; sensitivity, specificity, and accuracy of the Cys- and Lys-peptide depletion-based peptide reactivity prediction model ($DPRA_{Cys\&Lys}$). Substances not classified (NC) according to CLP/GHS criteria which tested positive in the DPRA were counted to the group of false positives.

$DPRA_{Cys\&Lys}$ (Threshold = 22.62%)		**DPRA prediction**		
		CLP/GHS 1A	**CLP/GHS 1B**	∑
LNA EC3 prediction	CLP/GHS 1A	47	1	48
	CLP/GHS 1B	36	22	58
	NC	7	10	17
∑		90	33	123

Sensitivity:	98%
Specificity:	29%
Accuracy:	56%

Using ROC analyses, the best-fit threshold values for a standard DPRA Cys- and Lys-peptide depletion-based PM was calculated to be 22.87% (Supplementary Table S3). This value is very close to the OECD TG 442C threshold of 22.62%. Therefore, the ROC analyses of the standard DPRA results did not serve to improve those threshold value already specified in the OECD TG 442C.

By applying the Cys-only peptide depletion-based threshold described by OECD TG 442C (23.09%) for the 108 sensitizers of this set, a high sensitivity (96%), but a very low specificity (15%) and a low overall accuracy (51%) were achieved (Table 3b). A ROC analysis was not performed for the $DPRA_{Cys\text{-}only}$.

Table 3b: Standard DPRA; sensitivity, specificity, and accuracy of the Cys-peptide depletion-based peptide reactivity prediction model ($DPRA_{Cys\text{-}only}$). CLP/GHS NC substances were counted to the group of false positives.

$DPRA_{Cys\text{-}only}$ (moderate = 1A; threshold = 23.09%)		**DPRA prediction**		
		CLP/GHS 1A	**CLP/GHS 1B**	∑
LLNA EC3 prediction	CLP/GHS 1A	46	2	48
	CLP/GHS 1B	43	9	52
	NC	6	2	8
∑		95	13	108

Sensitivity:	96%
Specificity:	15%
Accuracy:	51%

3.1.2 Results of the standard DPRA for substances with common mechanism of peptide reactivity

The specific predictivities of skin sensitization potency that the standard $DPRA_{Cys\&Lys}$ yielded when test substances were assigned to probable mechanism of the peptide binding: The highest accuracies were achieved for aromatic compounds reacting in nucleophilic substitutions (80%; n=5), quinone precursors (93%; n=14) and acylating agents (80%, n=15). For all other sub-groups, accuracies were recorded in a range between 20 and 63% (Table 4).

Table 4: Overview of the predictivity of skin sensitization potency for the standard DPRA Cys&Lys, when substances were assigned to their mechanistic domains.

Mechanistic domain		**n**	**Sensitivity** [%]	**Specificity** [%]	**Accuracy [%]**
Acylating agents	Ac	15	100	57	80
Michael acceptor	MA	30	100	28	40
No alert	No alert	15	100	14	20
Quinone precursors	QP	14	100	50	93
Reactive carbonyls	RC	16	100	50	63
Agents reacting in nucleophilic aromatic substitution	S_N2	18	88	10	44
Agents reacting in nucleophilic substitution type 2	S_NAr	5	100	50	80

3.2 qDPRA

In the qDPRA, the EC6.38 values (i.e. the estimated concentrations resulting in 6.38% peptide depletion) were calculated by linear regression for the 36 that tested positive in the qDPRA of the altogether 50 test substances. Of the 14 substances that tested negative, 11 were TN as compared to the corresponding LLNA data, while 3 substances were FN (sub-category 1B in accordance with the LLNA data (see Supplemental Table S2)). Applying ROC analyses, an EC6.38 value of 2.55 mM (referring to the stock concentration) was calculated as best-fit threshold value to discriminate sub-category 1A from 1B sensitizers in the $qDPRA_{Cys\&Lys}$ (Table 5a). Applying this threshold to the 36 substances yielded a sensitivity of 75%, a specificity of 85% and an accuracy of 81% (seven of the 36 substances were misclassified) (Table 5a).

Table 5a: qDPRA; sensitivity, specificity, and accuracy of the Cys- & Lys peptide depletion-based peptide reactivity prediction model ($qDPRA_{Cys\&Lys}$). The CLP/GHS NC substance was counted to the group of false positives.

$qDPRA_{Cys\&Lys}$ (threshold = EC6.38 value of 2.55 mM)		**qDPRA prediction**		
		CLP/GHS 1A	**CLP/GHS 1B**	∑
LLNA EC3 prediction	CLP/GHS 1A	12	4	16
	CLP/GHS 1B	2	17	19
	NC	0	1	1
∑		14	22	36

Sensitivity:	75%
Specificity:	85%
Accuracy:	81%

Applying the $qDPRA_{Cys\text{-}only}$, a mean peptide depletion of ≤ 13.89% was recorded at 100 mM stock concentration for 18 of the 50 substances tested in the qDPRA. Hence, these 18 substances were assessed as non-sensitizers and they were excluded from the potency evaluation. Of note, however, of these 18 substances that were tested negative, 12 were TN when compared to the corresponding LLNA data, whereas six substances were FN (one sub-category 1A substance and five sub-category 1B substances in accordance with the LLNA data. For the remaining 32 substances that were tested positive in the qDPRA, applying the Cys-peptide depletion-based PM 3, EC13.89 values were calculated. Using ROC analyses, 7.96 mM

(referring to stock concentration) was calculated as best-fit EC13.89 threshold to discriminate sub-category 1A from 1B sensitizers. This threshold was used to assign 26 of the 32 substances to either sub-category 1A or 1B and then compared to the corresponding LLNA data. Applying Cooper statistics, a sensitivity of 93%, a specificity of 71% and an accuracy of 81% was calculated (Table 5b) and all detailed test results are summarized in Table 6.

Table 5b: qDPRA; sensitivity, specificity, and accuracy of the Cys- peptide depletion-based peptide reactivity prediction model.

$qDPRA_{Cys\text{-}only}$ (threshold = EC13.89 value of 7.96 mM)		**qDPRA prediction**		
		CLP/GHS 1A	**CLP/GHS 1B**	∑
LLNA EC3 prediction	CLP/GHS 1A	14	1	15
	CLP/GHS 1B	5	12	17
∑		19	13	32

Sensitivity:	93%
Specificity:	71%
Accuracy:	81%

Table 6: Overview of the 36 substances that tested positive in the qDPRA (considering mean peptide depletion at 100 mM stock concentration) as well as Cys-peptide, Lys-peptide, and mean Cys- and Lys-peptide depletions as well as calculated EC6.38 and EC13.89 values. Negative depletion values were considered to be zero for calculation of mean peptide depletion. For further information on the substances see Supplementary Table S2.

Substance	**LLNA EC3 [%]**	**CLP/ GHS**	**Cys [% peptide depletion]**			**Lys [% peptide depletion]**			**mean [% peptide depletion]**			**EC 6.38 [mM]**	**EC 13.89 [mM]**
			1 mM	**10 mM**	**100 mM**	**1 mM**	**10 mM**	**100 mM**	**1 mM**	**10 mM**	**100 mM**		
Vanillin	> 50	NC	0.91	1.09	4.54	-1.00	-0.92	44.7	0.45	0.55	24.62	**31.81**	**n/a**
Methyl Methacrylate	90	1B	1.08	6.32	47.7	-0.55	-5.78	6.06	0.54	3.16	26.88	**22.22**	**26.46**
EGDMA	28	1B	2.02	16.82	82.41	0.15	3.93	25.64	1.21	10.37	53.96	**6.10**	**8.22**
Imidazolidinyl urea	25	1B	1.72	14.99	48.59	2.22	4.17	16.37	1.97	9.58	32.48	**6.22**	**9.25**
Hydroxycitronellal	23	1B	3.01	5.89	35.64	0.59	3.15	31.84	1.80	4.52	33.74	**15.73**	**34.20**
Cinnamic alcohol	21	1B	0.38	2.63	17.11	0.09	0.82	2.43	0.24	1.72	9.77	**62.08**	**79.99**
Phenyl benzoate	17.1	1B	0.73	7.83	39.97	-1.37	0.41	4.90	0.37	4.12	22.44	**21.11**	**26.97**
Sodium dodecyl sulfate	14	1B	2.33	2.52	-3.63	0.35	0.32	87.48	1.34	1.42	43.74	**20.55**	**n/a**
Eugenol	12.9	1B	1.67	5.11	20.25	-0.96	1.41	6.02	0.83	3.26	13.13	**38.44**	**62.19**
Farnesal	12	1B	0.95	4.41	14.61	-0.44	-2.77	0.10	0.48	2.20	7.35	**83.01**	**93.65**
α-Hexylcinnamicaldehyde	12	1B	1.31	9.77	100	0.53	1.57	3.13	0.92	5.67	51.57	**11.39**	**14.11**
2,3-Butanedione	11.3	1B	2.64	22.53	79.01	0.52	6.62	30.24	1.58	14.57	54.63	**4.32**	**6.09**

Citral	5.3 - 13	1B	0.78	9.11	64.35	0.43	3.00	12.04	0.67	5.96	37.64	**14.59**	**17.79**
Benzylidenacetone	3.7	1B	4.39	33.33	92.90	0.41	-0.04	1.36	2.40	16.67	47.13	**3.51**	**3.95**
3-Propylidenephthalide	3.7	1B	-0.15	0.53	4.83	-0.04	1.84	24.3	0.00	1.18	14.56	**44.96**	**n/a**
Tetramethylthiuram disulfide	3.1	1B	31.50	100	100	0.74	2.12	1.95	16.12	51.06	50.98	**< 1**	**< 1**
Cinnamic aldehyde	3.1	1B	6.43	18.59	67.02	2.65	6.56	I	4.54	12.57	n/a	**3.06**	**6.52**
Phenylacetaldehyde	3 / 4.7	1B	1.27	13.60	62.67	2.37	7.50	16.6	1.82	10.55	39.63	**5.70**	**10.53**
Ethylenediamine	2.2	1B	1.55	3.42	30.79	-1.77	-1.76	-2.13	0.77	1.71	15.39	**40.72**	**44.43**
Diethylmaleate	2.1	1B	11.36	88.12	100	0.57	15.76	100	5.97	51.94	100	**1.08**	**1.30**
2,4-Dinitrobenzenesulfonic acid	2	1A	7.09	46.84	100	0.99	29.45	I	4.04	38.15	n/a	**1.62**	**2.54**
Isoeugenol	1.8	1A	1.14	7.48	50.49	-0.64	3.77	21.93	0.57	5.62	36.21	**12.23**	**23.41**
2-Mercaptobenzothiazole	1.7	1A	4.89	16.96	100	-0.32	-0.59	-0.58	2.45	8.48	50.00	**6.87**	**7.71**
Methyldibromo glutaronitrile	0.9	1A	28.24	100	100	-0.91	2.36	I	14.12	51.18	n/a	**< 1**	**< 1**
4-Methylaminophenol sulfate	0.8	1A	27.47	97.82	I	7.22	I	I	17.34	n/a	n/a	**< 1**	**< 1**
Formaldehyde	0.70	1A	5.00	19.54	56.91	1.01	0.65	2.12	3.00	10.1	29.52	**5.28**	**6.50**
Cobalt chloride	0.57	1A	7.62	16.72	20.23	5.85	25.29	28.36	6.74	21.00	24.30	**< 1**	**7.20**
Chloramin T	0.4	1A	24.58	100	100	1.02	19.59	I	12.8	59.8	n/a	**< 1**	**< 1**
Propyl gallate	0.32	1A	17.29	37.81	99.05	9.14	40.74	I	13.21	39.27	n/a	**< 1**	**< 1**
Phthalic anhydride	0.16	1A	1.46	-2.93	-3.94	1.5	10.83	32.06	1.48	5.41	16.03	**18.22**	**n/a**
p-Phenylenediamine	0.16	1A	7.84	22.12	96.55	10.71	28.71	25.42	9.27	25.42	60.98	**< 1**	**4.81**
1-Chloro-2,4-Dinitrobenzene	0.04	1A	16.29	100	100	0.74	10.86	37.05	8.52	55.43	68.52	**< 1**	**< 1**
Dinitrofluorobenzene	0.03	1A	11.89	100	100	4.68	36.25	100	8.29	68.13	100	**< 1**	**1.20**
p-Benzoquinone	0.0099	1A	100	100	100	7.85	54.53	92.58	53.93	77.27	96.29	**< 1**	**< 1**
Kathon CG	0.005	1A	83.33	100	100	0.56	7.22	I	41.94	53.61	n/a	**< 1**	**< 1**
Oxazolone	0.003	1A	2.49	19.67	74.11	4.84	35.08	56.87	3.67	27.38	65.49	**2.03**	**6.97**

EGDMA = ethylene glycol dimethacrylate; n/a = not applicable; I = interference of test substance and peptide at HPLC analysis (no value for depletion available)

3.1.3 Kinetic DPRA

The Cys-peptide depletion-based maximum *k* values recorded for the 20 substances tested in the kinetic DPRA are presented in Table 7. A threshold value of log *k* = -1.73 was the best fit to discriminate sub-categories 1A and 1B substances. The kinetic $DPRA_{Cys\text{-}only}$ (Table 8) correctly assigned 19 of the 20 substances; only 2,3-butadione (EC3 = 11.3) was over-predicted.

Table 7: Overview of the 20 substances tested in the kinetic DPRA with calculated rate reaction constants for different time points. For further information on the substances see Supplementary Table S2.

Chemical			**Reaction rate constant [s^{-1} M^{-1}]**	

	LLNA-EC3 [%]	CLP/GHS	k 5	k 30	k 60	k 120	k 240	k highest reactivity	log *k* highest reactivity
EGDMA	28	1B	n.d.	0.0009	0.0011	0.0010	0.0008	0.0011	**-2.94**
Cinnamic alcohol	21	1B	n.d.	0.0016	n.d.	0.0009	n.d.	0.0016	**-2.79**
Pentachlorphenol	20	1B	n.d.	0.0053	n.d.	0.0010	n.d.	0.0053	**-2.28**
Phenyl benzoate	17.1	1B	n.d.	0.0007	0.0015	0.0005	0.0001	0.0015	**-2.82**
Eugenol	12.9	1B	n.d.	0.0081	n.d.	0.0030	n.d.	0.0081	**-2.09**
Farnesal	12	1B	n.d.	0.0028	n.d.	0.0012	n.d.	0.0028	**-2.55**
α-Hexylcinnamicaldehyde	12	1B	n.d.	0.0072	0.0069	0.0086	0.0039	0.0086	**-2.07**
2,3-Butanedione	11.3	1B	0.0229	0.0072	0.0028	0.0016	0.0010	0.0229	**-1.64**
Citral	7.3	1B	n.d.	0.0067	n.d.	0.0040	n.d.	0.0067	**-2.17**
Benzylidenacetone	3.7	1B	n.d.	0.0042	n.d.	0.0077	n.d.	0.0077	**-2.12**
Tetramethylthiuram disulfide	3.1	1B	0.0061	0.0006	n.d.	0.0071	n.d.	0.0061	**-2.22**
Diethylmaleate	2.1	1B	n.d.	0.0116	0.0121	0.0144	0.0172	0.0172	**-1.76**
2-Mercaptobenzothiazole	1.7	1A	n.d.	0.1779	0.0878	0.0385	0.0190	0.1779	**-0.75**
Isoeugenol	1.8	1A	0.0200	0.0119	0.0078	0.0032	0.0005	0.0200	**-1.70**
Formaldehyde	0.7	1A	0.1225	0.0227	0.0110	0.0038	0.0023	0.1225	**-0.91**
Phthalic anhydride	0.16	1A	0.2122	0.0349	0.0120	0.0049	0.0015	0.2122	**-0.67**
p-Phenylenediamine	0.16	1A	0.0500	0.0107	0.0055	0.0030	0.0016	0.0500	**-1.30**
1-Chloro-2,4-Dinitrobenzene	0.04	1A	n.d.	0.0616	0.0512	0.0474	0.0423	0.0616	**-1.21**
p-Benzoquinone	0.0099	1A	2.6725	too reactive	too reactive	too reactive	too reactive	2.6725	**0.43**
Oxazolone	0.003	1A	n.d.	0.1627	0.1102	0.0464	0.0173	0.1627	**-0.79**

EGDMA = ethylene glycol dimethacrylate; *k* = Reaction rate constant [s^{-1} M^{-1}]; k 5, 30, 60, 120 or 240 = reaction rate constant at t = 5, 30, 60, 120 or 240 minutes; n.d. = not determined; too reactive = complete peptide depletion at any concentration

Table 8: Kinetic DPRA; sensitivity, specificity, and accuracy of the Cys- and Lys-peptide depletion-based peptide reactivity prediction model.

Kinetic DPRA$_{Cys\text{-}only}$ (threshold = log *k* = -1.73)		**Kinetic DPRA prediction**		
		CLP/GHS 1A	**CLP/GHS 1B**	∑
LLNA EC3 prediction	CLP/GHS 1A	8	0	8
	CLP/GHS 1B	1	11	12
∑		9	11	20

Sensitivity:	100%
Specificity:	92%
Accuracy:	95%

3.2 Summary of test results

3.2.1 Comparison of performances of the standard DPRA, the qDPRA and the kinetic DPRA

Table 9 provides an overview of performances of the standard DPRA, the qDPRA and the kinetic DPRA in predicting sub-categories 1A and 1B skin sensitization potency as determined using Cooper statistics. Potency assessment using the kinetic DPRA yielded the best accuracy (95%), albeit only using the limited set of 20 substances. Potency assessment using the qDPRA yielded lower accuracies of 81% in both, the $qDPRA_{Cys\&Lys}$ (36 substances) and the $qDPRA_{Cys\text{-}only}$ (32 substances). Potency assessment using the standard DPRA yielded the lowest accuracies of 56% for the $DPRA_{Cys\&Lys}$ (123 substances) and (similarly) 51% for the $qDPRA_{Cys\text{-}only}$ (108 substances). In all three variants of the DPRA, sensitivity was higher than the specificity: The sensitivity in predicting sub-categories 1A or 1B skin sensitizing potency ranged from 100% for the kinetic DPRA to 75% for the qDPRA applying the Cys- and Lys-peptide depletion-based PM. The specificity in predicting sub-categories 1A or 1B skin sensitizing potency ranged from 92% for the kinetic DPRA to 15% for the $DPRA_{Cys\text{-}only}$. Hence, the kinetic DPRA yielded not only the best accuracy, but also the highest sensitivity and specificity.

Table 9: Overview of the performances of the different prediction models as defined for the DPRA, qDPRA and kinetic DPRA compared to LLNA data.

Approach	Peptide used	n	Sens [%]	Spec [%]	Acc [%]
standard DPRA	Cys & Lys	123	98	29	56
	Cys-only	108	96	15	51
qDPRA	Cys & Lys	36	75	85	81
	Cys-only	32	93	71	81
Kinetic DPRA	Cys-only	20	100	92	95

Sens = sensitivity; Spec = specificity; Acc = accuracy; n = number of analyzed substances.

3.2.2 Performances of the standard DPRA, the qDPRA and the kinetic DPRA for a common set of substances

The 20 substances tested in the kinetic DPRA were also tested with the qDPRA and standard DPRA protocols. The performances of the three approaches considering the most capable PMs were calculated for this common sub-set in comparison to LLNA EC3 as well as human DSA_{05} values as illustrated in Table 10. Based on the same set of substances, the kinetic DPRA was superior to the other approaches providing 100% sensitivity and an overall high accuracy to predict skin sensitization potency in the LLNA and humans alike.

Table 10: Performances of the three different DPRA methods for a common set of 20 substances.

Approach	vs. LLNA data (EC3; n = 20)			vs. LLNA data (EC3; n = 14)			vs. Human data (DSA05; n = 14)		
	Sens [%]	Spec [%]	Acc [%]	Sens [%]	Spec [%]	Acc [%]	Sens [%]	Spec [%]	Acc [%]
$DPRA_{Cys\&Lys}$	89	36	60	100	33	57	100	30	50
$qDPRA_{Cys\&Lys}$	44	82	65	17	75	50	20	78	57
Kinetic $DPRA_{Cys\text{-}only}$	100	92	95	100	89	93	100	90	93

Sens = sensitivity; Spec = specificity; Acc = accuracy

4 Discussion

Non-animal methods (alternative methods) for the identification of sensitization testing are now gaining acceptance within the regulatory community. This is evidenced by the meanwhile mandatory use thereof for REACH registrations and the current OECD activities on reporting defined approaches and the methods used. Simple combinations, e.g. the '2 out of 3' approach using e.g. the DPRA , the KeratinoSens™ / LuSens and the h-CLAT give rise to accuracies exceeding those of the LLNA for predicting human sensitization potentials (Bauch *et al.* 2012;Natsch *et al.* 2013;Urbisch *et al.* 2015). Unfortunately, although these methods are to be used, there is reticence to use negative predictions and they are considered to be

insufficient for potency assessments which in turn would trigger subsequent animal testing by necessitating the use of an LLNA for classification according to sub-category 1A (strong sensitizers) or 1B (weak to moderate sensitizers) according to CLP/GHS (question and answer session of ECHA webinar (ECHA 2016)). As currently non-animal methods may only be used in combination, an additional drawback of the use of alternatives in terms of regulatory acceptance is the impending discussions on classification should one of the methods of a test strategy deliver positive results. This could also lead to the requirement to conduct an animal test. A similar problem arose when genotoxicity testing foresaw animal testing if one of the prescribed non-animal methods yielded positive results. Interestingly, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) assessment of the use of the LLNA, the current "gold standard" animal test, for potency assessment in terms of GHS classification and labelling purposes indicated that the value of the results for this purpose is far from adequate (http://iccvam.niehs.nih.gov/methods/immunotox/LLNA-pot/TMER.htm). Similarly, Dumont et al. (2016) found for 28 out of 50 sub-category 1B test substances (56%) discordant LLNA studies and 19 out of 41 (46%) for sub-category 1A test substances (Dumont *et al.* 2016).

A number of studies have now been published, which utilize information obtained from test strategies used to not only the identify skin sensitization hazard of a substance but also to give an indication of the potency of the sensitizer. In the following only a few examples are given: Nukada et al. (2012), evaluated the use of the combination of the DPRA (now OECD 442C) and the h-CLAT (now OECD TG 442E). Their studies using 106 substances demonstrated an 80% accuracy for GHS classification when compared to the LLNA (Nukada *et al.* 2012). Much effort is also being put into the development of *in silico* tools to predict hazard and potency. Jaworska et al. (2015) employed a strategy based on Bayesian networks to define potency based on information derived from the DPRA, KeratinoSens™ and h-CLAT combined with additional *in silico* data. The accuracies using this approach exceeded 90% for potency predictions when comparing to the LLNA (Jaworska *et al.* 2015). Natsch et al. assessed over 300 substances in a method combining data obtained from the KeratinoSens™ and a kinetic peptide binding method to assess potency based on peptide

reactivity (Natsch *et al.* 2015). The extensive analyses conducted by this group actually indicated the use of all three methods instead of only two decreased the predictivity, indicating that more is not always better.

The DPRA addresses the first key event of the adverse outcome pathway, i.e. the covalent binding that is essential for haptenation. The DPRA is also a method that does not require a cell culture laboratory possibly making it more accessible to laboratories. As not all substances can be readily characterized in terms of chemical structure, e.g. substances of unknown or variable composition (UVCBs) such as plant extracts, *in silico* tools are not applicable in all cases (strengths and limitations described in (Urbisch *et al.* 2016b)). Since UVCBs were not the focus of this study as new methods are first evaluated using defined substances, testing methods will still be necessary. In this study, the standard DPRA along with two methods introducing modifications to the DPRA were assessed, namely the quantitative DPRA (qDPRA) and kinetic DPRA, for the predictive properties for potency classifications according to CLP/GHS.

First, the standard DPRA, performed according to OECD TG 442C, was evaluated for its ability to discriminate two potency classes. In the standard procedure, one test-substance stock concentration (100 mM) and one time point (24 h) are used to detect the depletion of Cys- and Lys-peptide. Overall, the performance in predicting CLP/GHS potency sub-categories was poor (accuracy = 56%) and an improvement by adjusting the respective threshold values used for sub-categorization was not successful. Obviously, measuring the yield of peptide-adducts at one test substance concentration after one reaction time is not sufficient.

Hence the standard DPRA was extended by using different test substance concentrations in a so called qDPRA. The test procedure using the highest concentration in the qDPRA was identical to the standard DPRA and thus in accordance to OECD TG 442C; for this reason, potency information in the qDPRA would be available with little additional effort to the regulatory accepted protocol. Compared to the standard DPRA, the sensitivity of the qDPRA was slightly lower (75%), whereas the specificity (85%) and accuracy (81%) was higher. Of the 36 substances considered for this approach, four were falsely assigned to CLP/GHS 1B instead of 1A and thus under-predicted. These substances were phthalic anhydride,

isoeugenol, 2-mercaptobenzothiazole and formaldehyde. Two of the 36 substances were over-predicted and falsely assigned to CLP/GHS 1A instead of 1B, namely diethyl maleate (EC3 = 2.1, Michael acceptor) and tetramethylthiuram disulfide (EC3 = 3.1, nucleophilic substitution type 2). When applying the Cys-only protocol, (EC13.89 value instead of an EC6.38) an increased sensitivity was attained, a decreased specificity and an equivalent accuracy compared to the $qDPRA_{Cys\&Lys}$. However, these changes in the performance of $qDPRA_{Cys\text{-}only}$ were not significantly different to the performance of the $qDPRA_{Cys\&Lys}$. Thus, as in the standard DPRA, the results obtained from the Lys peptide do not necessarily contribute to the determination of potency in the qDPRA. The predictive accuracy was 81% for a substance set of 36 substances compared to the results of LLNAs. For 14 substances human data were available and the predictive accuracy was 57% for this smaller set of substances. While the qDPRA does provide an acceptable predictive accuracy towards LLNA data, it appears to be less sufficient when compared to human data - albeit this is based on a limited number of substances. The accuracy in the expanded set (n = 36) of test substances was clearly higher than that obtained with a smaller sub-set of 14 test substances. It happened that the substance selection in the smaller sub-set contained nearly all substances with false predictions of the larger set of 36 substances. The accuracy calculated from the smaller sub-set may therefore not be representative but was calculated in order to compare the three methods to human data on skin sensitization potency based on a common, however small, set of test substances.

Since neither one test substance concentration at one reaction time (standard DPRA) nor several test substance concentration at one reaction time (qDPRA) seemed to be sufficient to discriminate the potency classes 1A and 1B, a third method, using several test substance concentration at several reaction times (kinetic DPRA) was used. In contrast to the standard DPRA and qDPRA, which use HPLC methods, the kinetic DPRA uses a multiwall plate fluorimetric read-out, in order to facilitate measurements of the large number of samples. The fluorimetric assay is only available for the Cys-peptide as no comparable dye in terms of reactivity is available for the Lys-peptide. The consideration of Cys-only protocol seems justified since results of the standard DPRA or qDPRA indicated that Lys-peptide depletion did not significantly contribute to the discrimination of potency classes. While a Cys-only prediction model has also been

described in OECD TG 442C, data from more substances would be needed to strengthen the hypothesis that Lys-peptide depletion does not significantly contribute to potency assessment.

An advantage of analyzing different reaction times is that highly reactive substances like 2,4-dinitrochlorobenzene (DNCB; reacting by nucleophilic aromatic substitution) or *para*-benzoquinone (reacting in a Michael addition), that completely depleted the peptide at early time points, can be differentiated from ~~s~~substances with lower reactivity that induced complete depletion only after the full incubation period of 24 hours. Another benefit of determining peptide reactivity at early reaction times is the identification of competing reactions to peptide binding like hydrolysis as seen for phthalic anhydride. Using the kinetic DPRA, phthalic anhydride could be correctly assigned to CLP/GHS sub-category 1A, whereas the standard DPRA predicts this substance as false negative when focusing on the Cys-peptide.

The kinetic DPRA, however, does not yet stand in line with the currently adopted OECD TG 442C: While the TG foresees a Cys-only prediction model, it does not mention fluorimetric read-out; rather only HPLC analysis is part of the validated method. The kinetic DPRA would require a full validation with more substances and analysis of inter- and intra-laboratory variability to be regulatory accepted.

The kinetic DPRA data resulted in correct assignments of 19 of 20 test substances to the respective CLP/GHS sub-categories 1A and 1B based on LLNA data (accuracy = 95%; see Table 7); likewise 13 of 14 test substances were correctly sub-classified compared to human data (accuracy = 93%; see Table 7). The predictive accuracy of the LLNA for this set of 14 test substances is also 93%.

The kinetic DPRA is the most promising method to discriminate sub-categories 1A and 1B among skin sensitizing substances. It could be combined with a testing strategy to identify skin sensitizing substances. The straightforward approach is a '2 out of 3' integrated testing strategy approach (i.e. DPRA, LuSens/KeratinoSens™, h-CLAT) followed by a kinetic DPRA in case the substance is predicted as a skin sensitizer by a '2 out of 3' prediction model. If the flourimetric kinetic DPRA was, however, validated, its result could substitute the standard DPRA and hence an integrated testing strategy with kinetic DPRA,

LuSens/KeratinoSens™ and h-CLAT could be used to determine the skin sensitization potential by a '2 out of 3' prediction model and discriminate the skin sensitizing substances in sub-categories 1A and 1B.

5 Conclusions

In this study, a number of approaches based on peptide reactivity were evaluated for their utility to assign skin sensitizers into CLP/GHS potency sub-categories 1A or 1B (based on a LLNA EC3 threshold of 2%). In the first approach - the standard DPRA (OECD TG 442C) using one test substance concentration and one reaction time - was assessed. The standard DPRA did not generate data that was sufficiently accurately discriminate between the potency sub-categories 1A and 1B. Attempts to improve the predictivity by adjusting the classification thresholds of the prediction model did not further improve the predictivity.

In a second approach, the quantitative DPRA (qDPRA), data obtained using three test substance concentrations at one reaction time were used and an improvement of the predictivities were observed. The accuracy of the qDPRA compared to LLNA data was 83% (n=36) but in a smaller subset with human potency data (n=14) only 56% compared to human and 50 compared to LLNA data.

In the third approach, different test substance concentrations and reaction times (kinetic DPRA) were used and this approach resulted in a high accuracy of 95% compared to LLNA data (n=20) and, in a smaller subset with human potency data (n=14), 93% compared to human and 93% compared to LLNA data. When compared to human data, the LLNA itself sub-classified 13 of these 14 substances correctly (accuracy = 93%). Although the number of tested substances is small and further validation with an extended set of substances is needed, the kinetic DPRA could well prove to be a practical tool to sub-classify skin sensitizing substances according to CLP/GHS.

In contrast to the kinetic DPRA, the standard DPRA and the qDPRA procedures fulfill the criteria described in the OECD TG 442C. The kinetic DPRA, however, uses a relatively accessible detection system based on a fluorimetric read-out of the Cys-adducts only (instead of analyzing both Cys- and Lys-peptide depletion by HPLC). Analysis of Lys-depletion in the standard and qDPRA revealed that the Lys-peptide does not

substantially contribute to the discrimination of sub-categories 1A and 1B for the substances assessed here. The kinetic DPRA can be employed in a sequential or integrated testing strategy to identify skin sensitizing substances and sub-classify those to CLP/GHS 1A and 1B. Such a strategy would facilitate the full replacement of *in vivo* studies for REACH and CLP classification data requirements. Whereas currently, the potency assessment may still require LLNA data.

Acknowledgements

This work was funded by the Alternative Methods Program of BASF SE, Ludwigshafen, Germany.

The authors alone are responsible for the writing and contents of the manuscript. They declare no competing financial interest. Ursula G. Sauer (Scientific Consultancy – Animal Welfare, Neubiberg, Germany) assisted in preparing this documents by critically reviewing the content and language.

Supporting information

Supplementary Tables S1 and S2 represent the full list of tested and assessed substances with all additionally obtained data used in the underlying study. Table S3 contains graphical representations being used as part of the ROC analyses.

This material is available free of charge via the Internet at: http://www.journals.elsevier.com/toxicology-in-vitro.

6 Reference list

Ankley, G. T., Bennett, R. S., Erickson, R. J., Hoff, D. J., Hornung, M. W., Johnson, R. D., Mount, D. R., Nichols, J. W., Russom, C. L., Schmieder, P. K., Serrrano, J. A., Tietge, J. E., and Villeneuve, D. L. (2010). Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ.Toxicol.Chem. 29, 730-741.

Basketter, D., Alepee, N., Casati, S., Crozier, J., Eigler, D., Griem, P., Hubesch, B., de, Knecht J., Landsiedel, R., Louekari, K., Manou, I., Maxwell, G., Mehling, A., Netzeva, T., Petry, T., and Rossi, L. H.

(2013). Skin sensitisation--moving forward with non-animal testing strategies for regulatory purposes in the EU. Regul.Toxicol.Pharmacol. 67, 531-535.

Basketter, D., Ashikaga, T., Casati, S., Hubesch, B., Jaworska, J., de, Knecht J., Landsiedel, R., Manou, I., Mehling, A., Petersohn, D., Rorije, E., Rossi, L. H., Steiling, W., Teissier, S., and Worth, A. (2015). Alternatives for skin sensitisation: Hazard identification and potency categorisation: Report from an EPAA/CEFIC LRI/Cosmetics Europe cross sector workshop, ECHA Helsinki, April 23rd and 24th 2015. Regul.Toxicol.Pharmacol. 73, 660-666.

Basketter, D. A., Gerberick, F., and Kimber, I. (2007). The local lymph node assay and the assessment of relative potency: status of validation. Contact Dermatitis 57, 70-75.

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. Regulatory Toxicology and Pharmacology 63, 489-504.

Berthold, M. (2007). KNIME: The Konstanz Information Miner. In 'Studies in Classification, Data Analysis, and Knowledge Organization'. Springer, Berlin.

Bruckner, A. L., Weston, W. L., and Morelli, J. G. (2000). Does sensitization to contact allergens begin in infancy? Pediatrics 105, e3.

Cooper, J. A., Saracci, R., and Cole, P. (1979). Describing the validity of carcinogen screening-tests. British Journal of Cancer 39, 87-89.

Council Directive 1907/2006/EEC (2006). EC: Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorization 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 Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Off J Eur Union, L 396/1.

Dumont, C., Barroso, J., Matys, I., Worth, A., and Casati, S. (2016). Analysis of the Local Lymph Node Assay (LLNA) variability for assessing the prediction of skin sensitisation potential and potency of chemicals with non-animal approaches. Toxicol.In Vitro 34, 220-228.

ECHA. Use of alternative methods to animal testing in your REACH registration (webinar). 2016. https://echa.europa.eu/view-webinar/-/journal_content/56_INSTANCE_DdN5/title/use-of-alternative-methods-to-animal-testing-in-your-reach-registration (accessed 29.11.2016).

EU 1223/2009 (2009). Regulation (EC) No 1223/2009 of the European Parliament and of the Council. Official Journal of the European Union L342, 59-209.

European Commission (2016). Amendmend of Annex VII to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) as regards skin sensitisation. Off J Eur Union, L 255/14.

European Parliament. REGULATION (EC) No 1272/2008 on the classification, labelling and packaging of substances and mixtures. 2008.

Friedmann, P. S. (2007). The relationships between exposure dose and response in induction and elicitation of contact hypersensitivity in humans. Br.J.Dermatol. 157, 1093-1102.

Gerberick, F., Aleksic, M., Basketter, D., Casati, S., Karlberg, A. T., Kern, P., Kimber, I., Lepoittevin, J. P., Natsch, A., Ovigne, J. M., Rovida, C., Sakaguchi, H., and Schultz, T. (2008). Chemical reactivity measurement and the predictive identification of skin sensitisers. Atla-Alternatives to Laboratory Animals 36, 215-242.

Gerberick, G. F., Vassallo, J. D., Bailey, R. E., Chaney, J. G., Morrall, S. W., and Lepoittevin, J. P. (2004). Development of a peptide reactivity assay for screening contact allergens. Toxicol.Sci. 81, 332-343.

Gerberick, G. Frank, Vassallo, Jeffrey D., Foertsch, Leslie M., Price, Brad B., Chaney, Joel G., and Lepoittevin, Jean Pierre (2007). Quantification of chemical peptide reactivity for screening contact allergens: A classification tree model approach. Toxicol.Sci. 97, 417-427.

Hirota, M., Kouzuki, H., Ashikaga, T., Sono, S., Tsujita, K., Sasa, H., and Aiba, S. (2013). Artificial neural network analysis of data from multiple in vitro assays for prediction of skin sensitization potency of chemicals. Toxicol.In Vitro 27, 1233-1246.

ICCVAM. ICCVAM LLNA Potency Evaluation Report. NIH Publication Number 11-7709. 2011.

Jaworska, J. S., Natsch, A., Ryan, C., Strickland, J., Ashikaga, T., and 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.

Kimber, I., Basketter, D. A., Butler, M., Gamer, A., Garrigue, J. L., Gerberick, G. F., Newsome, C., Steiling, W., and Vohr, H. W. (2003). Classification of contact allergens according to potency: Proposals. Food Chem. Toxicol. 41, 1799-1809.

Magnusson, B. M., Anissimov, Y. G., Cross, S. E., and Roberts, M. S. (2004). Molecular size as the main determinant of solute maximum flux across the skin. J.Invest Dermatol. 122, 993-999.

Mehling, A., Eriksson, T., Eltze, T., Kolle, S. N., Ramirez, T., Teubner, W., van Ravenzwaay, B., and Landsiedel, R. (2012). Non-animal test methods for predicting skin sensitization potentials. Arch. Toxicol. 86, 1273-1295.

Michael A. (1887). Über die Addition von Natriumacetesig- und Natriummalonsäureäthern zu den Aethern ungesättigter Säuren. J Praktische Chemie 35, 349-356.

Natsch, A., Emter, R., Gfeller, H., Haupt, T., and 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., 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. 33, 1337-52.

Noe, D. A. (1983). Selecting a diagnostic study's cutoff value by using its receiver operating characteristic curve. Clin.Chem. 29, 571-572.

Nukada, Y., Ashikaga, T., Miyazawa, M., Hirota, M., Sakaguchi, H., Sasa, H., and Nishiyama, N. (2012). Prediction of skin sensitization potency of chemicals by human Cell Line Activation Test (h-CLAT) and an attempt at classifying skin sensitization potency. Toxicol. in Vitro 26, 1150-1160.

OECD. Guideline for Testing of Chemicals No. 429/429A/429B, ("Skin Sensitization: Local Lymph Node Assay"). 2010.

OECD. 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. 2012a.

OECD. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins; Part 1: Scientific Evidence. 2012b.

OECD. OECD TG 442C: In chemico skin sensitization: Direct Peptide Reactivity Assay (DPRA). 2015a.

OECD. OECD TG 442D: In vitro skin sensitization: ARE-Nrf2 Luciferase Test Methods. 2015b.

OECD. Guidance document on the reporting of defined approaches and individual information sources to be used within integrated approaches to testing and assessment (IATA) for skin sensitization. ENV/JM/HA(2016)11. 2016a.

OECD. OECD TG 442E: In vitro human Cell Line Activation Test (h-CLAT). 2016b.

OECD TG 406. Guideline for Testing of Chemicals No. 406: Skin Sensitization. 1992.

Polepally, P. R., Huben, K., Vardy, E., Setola, V., Mosier, P. D., Roth, B. L., and Zjawiony, J. K. (2014). Michael acceptor approach to the design of new salvinorin A-based high affinity ligands for the kappa-opioid receptor. Eur.J.Med.Chem. 85, 818-829.

R Core Team. R: A Language and Environment for Statistical Computing. 2015.

Roberts, D. W., Api, A. M., and Aptula, A. O. (2016a). Chemical applicability domain of the Local Lymph Node Assay (LLNA) for skin sensitisation potency. Part 2. The biological variability of the murine Local Lymph Node Assay (LLNA) for skin sensitisation. Regul.Toxicol.Pharmacol. 80, 255-259.

Roberts, D. W., Api, A. M., Patlewicz, G., and Schultz, T. W. (2016b). Chemical applicability domain of the Local Lymph Node Assay (LLNA) for skin sensitization potency. Part 1. Underlying physical organic chemistry principles and the extent to which they are represented in the LLNA validation dataset. Regul.Toxicol.Pharmacol. 80, 247-254.

Roberts, D. W. and 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.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., and Muller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC. Bioinformatics. 12, 77.

Russell, W. M. S. and Burch, R. L. (1959). The principles of humane experimental technique. (Methuen, Co. LTD., London.

Sauer, U. G., Hill, E. H., Curren, R. D., Raabe, H. A., Kolle, S. N., Teubner, W., Mehling, A., and Landsiedel, R. (2016). Local tolerance testing under REACH: Accepted non-animal methods are not on equal footing with animal tests. Altern.Lab Anim 44, 281-299.

Semchyshyn, H. M. (2014). Reactive carbonyl species in vivo: generation and dual biological effects. ScientificWorldJournal. 2014, 417842.

Senger, N. A., Bo, B., Cheng, Q., Keeffe, J. R., Gronert, S., and Wu, W. (2012). The element effect revisited: factors determining leaving group ability in activated nucleophilic aromatic substitution reactions. J Org.Chem. 77, 9535-9540.

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., and 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.

The European Parliament (2010). Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. Official Journal of the European Union L276, 20.10.2010, 33–79.

United Nations. Globally Harmonized System of Classification and Labelling of Chemicals (GHS). 2013.

Urbisch, D., Becker, M., Honarvar, N., Kolle, S. N., Mehling, A., Teubner, W., Wareing, B., and Landsiedel, R. (2016a). Assessment of pre- and pro-haptens using non-animal test methods for skin sensitization. Chem.Res.Toxicol. 29, 901-13.

Urbisch, D., Honarvar, N., Kolle, S. N., Mehling, A., Ramirez, T., Teubner, W., and Landsiedel, R. (2016b). Peptide reactivity associated with skin sensitization: The QSAR Toolbox and TIMES compared to the DPRA. Toxicol.In Vitro. 34, 194-203.

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., and Sakaguchi, H. (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods. Regul.Toxicol.Pharmacol. 71, 337-351.

Wong SS. (1991). Chemistry of Protein Conjugation and Cross-Linking. ISBN 9780849358869, CRC Press, 340 pp.

HIGHLIGHTS

- Three approaches based on peptide reactivity were evaluated for their utility to assign skin sensitizers into CLP/GHS potency sub-categories 1A or 1B.
- The standard DPRA did not generate data that was sufficiently accurately discriminate between the potency sub-categories 1A and 1B.
- A second approach, the quantitative DPRA (qDPRA, three test substance concentrations at one reaction time) achieved an accuracy of 83% (n=36) compared to LLNA data.
- A third approach, kinetic DPRA (multiple test substance concentrations and reaction times resulted in a high accuracy of 95% compared to LLNA data (n=20) and, in a smaller subset with human potency data (n=14), 93% compared to human and 93% compared to LLNA data.
- The kinetic DPRA can be employed in a sequential or integrated testing strategy to identify skin sensitizing substances and sub-classify those to CLP/GHS 1A and 1B. Such a strategy would facilitate the full replacement of *in vivo* studies for REACH and CLP classification data requirements.

End of scientific publications

End of scientific publications

5 Discussion and conclusions

This work studied skin sensitization as first complex toxicological endpoint, for which a complete adverse outcome pathway (AOP) is defined. The first three key events (KE) of this AOP can be addressed using non-animal methods: *in silico* models based on (Q)SAR and the *in chemico* DPRA reflect the peptide reactivity of a substance (KE 1 and molecular initiating event (MIE)), the cell-based KeratinoSens^{TM} or LuSens assays reflect keratinocyte activation (KE 2) and the cell-based h-CLAT reflects dendritic cell activation (KE 3). The aim of this work was to analyze, how and in which cases existing non-animal methods can be used to adequately replace animal testing for skin sensitization and how the methods and underlying prediction models could be improved. In the following part, the five questions which guided the research of this thesis (see **Section 3**: Aim of this thesis) are discussed and referred to. The outcome of this research to address these questions describes a straightforward use of non-animal methods to assess skin sensitization hazard and potency.

1. Can the molecular initiating event (MIE) be addressed using *in chemico* and *in silico* tools?

The MIE was investigated by two approaches either using computational (namely the *in silico* tools TIMES SS and the QSAR Toolbox) or experimental methods (namely the *in chemico* DPRA).

To date, the *in chemico* DPRA is the only regulatory accepted method addressing the MIE [78] and experimental DPRA results are more consistent with human than with LLNA data [80]. However, the applicability of the DPRA shows limitations when testing insoluble substances, substances co-eluting with the model peptide or pro-haptens.

The *in silico* tools TIMES SS and the QSAR Toolbox are in general applicable to these substances. Since results from *in silico* tools were more consistent with LLNA data than with human data, their algorithms should be refined by *i.* considering available human data in the training sets, *ii.* giving a greater weight to such human data, and *iii.* also include available external LLNA data not yet considered in the respective algorithms. This would lead to an increase in their chemical space and to a higher predictivity in identifying human sensitizers and non-sensitizers.

Care should be taken by using the 'DPRA profilers' within the QSAR Toolbox (also see **Section 4.1**). When compared to experimental DPRA data, the current version of these profilers faces severe limitations:

i. the predictivity is low, *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 on peptide reactivity are already implemented within the same tool. For this reason, predictions of the DPRA profilers were not further considered in this work.

Overall, the selected *in silico* tools have the potential to address the MIE of skin sensitization as accurately as *in chemico* methods. Both approaches can be used complementarily, as each has its own limitations and advantages.

2. **How predictive are the non-animal methods - either individually or in combination?**

For this study, a comprehensive dataset of 213 substances was compiled, for which human and/or LLNA data were available in the literature to be used as reference values (also see **Section 4.2**). For most of these substances, results from non-animal methods were available, while data gaps were closed by additional testing during this thesis work. When comparing the results of the non-animal methods to the respective reference values it is shown that the DPRA, KeratinoSens™ or LuSens, and h-CLAT can predict the respective AOP key event with a high accuracy. Interestingly, the murine local lymph node assay (LLNA) itself (as first-choice animal test) predicts human sensitizers and non-sensitizers approximately as reliable as the single non-animal methods.

Since no single non-animal method can cover the complexity of the given endpoint [117], methods addressing the first three key events of the adverse outcome pathway (AOP) are combined within testing strategies [65;137]. The testing strategy providing the highest predictivity is the '2 out of 3' integrated testing strategy (ITS), which predicts both, human sensitizers and non-sensitizers, with a high accuracy of 90%.
Notably, this testing strategy discriminates human skin sensitizers from non-sensitizers with a higher accuracy than the LLNA itself (accuracy of 82%).

3. **How can the applicability domain of these methods be defined?**

In addition to factors like the physical state or the lipophilicity of a test substance (see also limitations of the applicability of the DPRA in **Section 4.1**), the applicability of the non-animal methods can be classified by the chemical reaction mechanism underlying the MIE (see also **Section 4.2**). *In silico* tools can be used to assign a test substance to its respective mechanistic group. This allows a more accurate estimation of the predictivity in the experimental non-animal methods, since false predictions associated with a specific molecular functionality can be uncovered.
This is for example proven for acylating agents in this work, which are not applicable in the LuSens and KeratinoSens™. For the hazard identification of acylating agents, more weight should be given to the dendritic cell-based assays and especially the Lysine reactivity in the DPRA, as a conclusion. In contrast, Michael acceptors, quinone precursors or substances reacting in nucleophilic substitutions could be reliably predicted in all investigated non-animal methods.

4. **Can the individual methods or combinations identify pre-haptens and pro-haptens?**

For this study, a set of 27 pre- and pro-haptens was investigated in the DPRA. Eighteen of these (surprisingly) gave positive responses in the DPRA despite the absence of any metabolic activity in this assay. The altered peptides resulting from the incubation of these 18 substances in the DPRA were analyzed by LC-MS. This enabled a postulation of molecular structures of the DPRA-reaction products so that the nature of modification could be deduced. Thirteen of these substances were non-metabolically activated and actually formed peptide

adducts. This indicates their pre-hapten nature, since oxygen is sufficient to transform an unreactive parent into a peptide-reactive intermediate – rather than a metabolic activation of pro-haptens. The remaining five substances were positive due to artefacts, which are discussed in more detail in **Section 4.3**. As a conclusion, the DPRA is capable of detecting pre-haptens, but lacks identifying pro-haptens, which may, however, be detected in the PPRA representing a modification of the DPRA using hydrogen peroxide and horse-radish peroxidase as surrogate for cutaneous enzymes [138;139].

Since the cell-based alternative methods KeratinoSens^{TM} (keratinocytes) and h-CLAT (dendritic cell-like cell line) provide a specific metabolic capacity, their sensitivity in detecting pre- and pro-haptens was clearly higher compared to the sensitivity of the DPRA. Although the metabolic competence of these cell lines may not represent the full metabolic capacity of the native human skin, the data show that they are sufficient to correctly identify a number of pro-haptens. Combined in the '2 out of 3' ITS, the sensitivity in detecting pre- and pro-haptens (i.e. 81%) is almost as high as its sensitivity in detecting directly acting haptens (i.e. 87%).

5. Can these methods evaluate the potency of sensitizers?

So far, the identification of skin sensitization hazard has been addressed in this work answering the first four questions. In some applications, e.g. cosmetics, not all substances identified as sensitizers can be banned from the market, since a number cannot be substituted due to essential properties as seen in case of specific preservatives, hair color ingredients or fragrances. For these substances, the identification of skin sensitization potency is important to enable actual risk assessments. Also for industrial chemicals, the CLP/GHS system prescribes discriminating a sub-classification of category 1 sensitizers by their potency into two sub-categories: 1A for strong and 1B for weak sensitizers. [4].

Relative potency (GHS 1A or 1B) can be investigated with the kinetic DPRA representing a modified protocol of the DPRA: testing different test substance concentrations and reaction time points enables the calculation of reaction rate constants for peptide binding. The reaction rate constants correlate with potency to a high degree in the kinetic DPRA. When applying a prediction model to transform the continuous read-out of the kinetic DPRA to dichotomous potency sub-categorization as required for chemicals (i.e. CLP/GHS 1A or 1B), 19 out of 20 test substances could be correctly assigned to their respective potency sub-category (see **Section 4.4**).

Risk assessments for cosmetics require continuous potency information rather than merely dichotomous sub-categories. The kinetic DPRA provided continuous potency values of 20 test substances which concurred with those of the LLNA ($R^2 = 0.71$, **Figure 5.1)**.

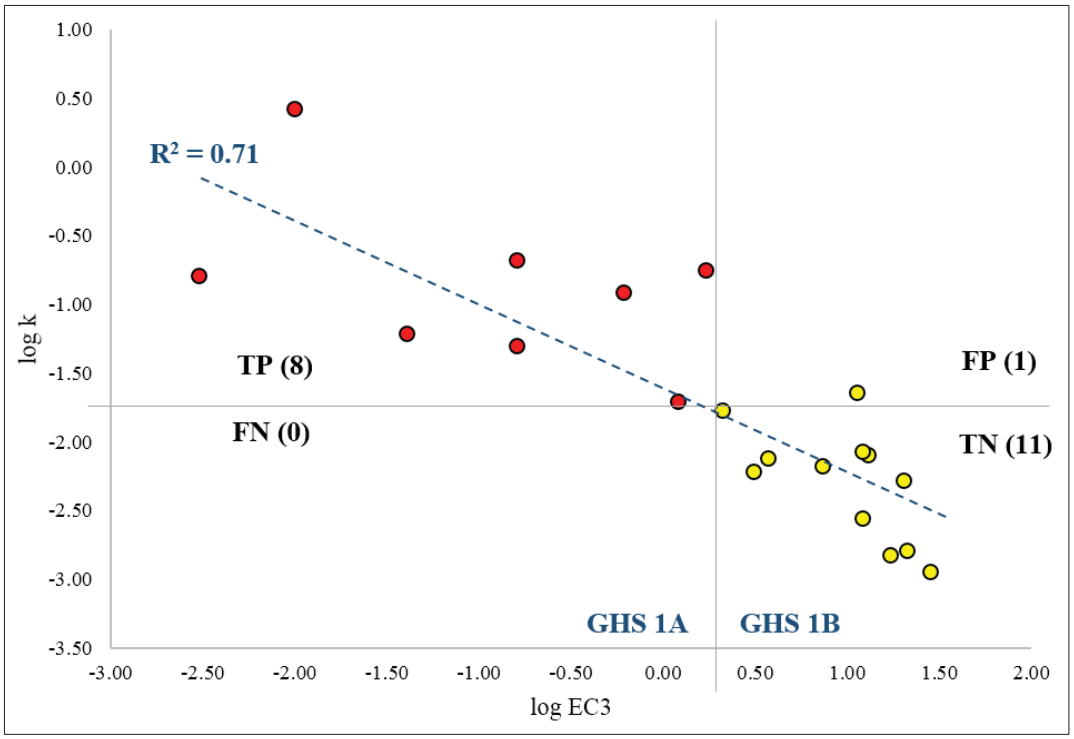


Figure 5.1: Reaction rate constants (*log* k) calculated from kinetic DPRA results plotted vs. LLNA-EC3 values (log EC3). The horizontal solid line represents the cut-off value of *log* k = -1.7 for assigning a test chemical to GHS 1A (red dots) or 1B (yellow dots), while the vertical solid line represents the cut-off value derived from LLNA (log EC3 = 0.30). The resulting four quadrants illustrate false negative (FN), false positive (FP), true negative (TN) and true positive (TP) results from kinetic DPRA when compared to LLNA results. Only 2,3-butadione was falsely predicted as strong sensitizer, even though being close (*log* k = -1.6) to the cut-off value. When considering numerical potency values instead of sub-categories, a high concordance is indicated by the regression line (dashed, blue) with a degree of certainty of $R^2 = 0.71$.

The obtained data proof the kinetic DPRA to be a practical tool to (at least) sub-classify substances as weak (CLP/GHS 1B) or strong sensitizers (1A). This good predictivity of the kinetic DPRA supports the hypothesis that sensitization potency of a hapten increases with its fast and efficient binding to dermal proteins.

A further advantage of the kinetic DPRA is the analysis of different reaction times so that highly reactive substances like 2,4-dinitrochlorobenzene (DNCB; reacting by nucleophilic aromatic substitution) or para-benzoquinone (reacting in a Michael addition), which completely depleted the peptide at early time points, can be differentiated from substances with lower reactivity that induced complete depletion only after an incubation period of 24 hours (as used by the standard DPRA). Another benefit of determining peptide reactivity at early reaction times is the identification of competing reactions to peptide binding like hydrolysis as seen for phthalic anhydride.

Overall conclusions

The current issues in replacing animal testing by non-animal methods to evaluate skin sensitization are defined as five questions that guided the research of this thesis. **Figure 5.2** highlights the conclusions of the different studies referring to these questions.

The '2 out of 3' integrated testing strategy provides reliable results in the identification of direct-acting haptens, pre-haptens and most of the pro-haptens and distinguishes human sensitizers from non-sensitizers with a higher predictivity than the local lymph node assay (LLNA) as first choice animal test. The results of this thesis contributed to the regulatory acceptance of animal-free methods for hazard identification of skin sensitizers in the framework of REACH.

Currently, this testing strategy only provides a binary answer (i.e. sensitizer? yes/no). In this context, the kinetic variant of the DPRA provides promising results to distinguish weak from strong sensitizers.

Overall, the kinetic DPRA can be employed in the '2 out of 3' integrated testing strategy, which can then be used *i.* to identify skin sensitizers and *ii.* to sub-classify those as weak or strong sensitizers. Such a testing strategy would facilitate a full replacement of *in vivo* studies for REACH and CLP classification data requirements.

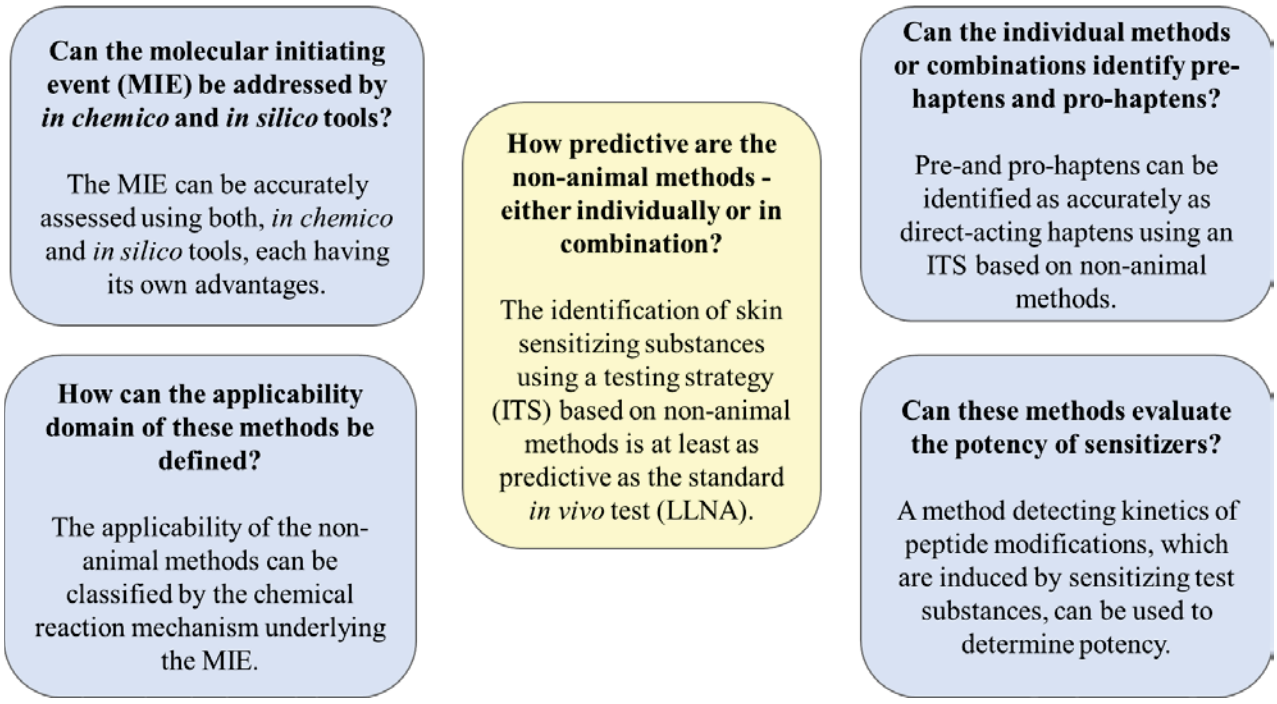


Figure 5.2: Overview of the five specific questions, which guided the research of the thesis work, and the corresponding answers stated as brief theses.

6 Outlook

The work described in this thesis contributes to new approaches to reliably assess the skin sensitization hazard and potency of substances. It is the first complex toxicological endpoint to be assessed without performing animal tests. This thesis confirms the effectiveness of non-animal methods for sensitization hazard identification and potency sub-classification. Parts of this work were considered and cited in the new ECHA guidance document for the evaluation of chemicals in the light of REACH [57].

While the approaches presented here give an example how to address complex toxicological effects with new approaches, there are still unsolved problems and future challenges:

- ***Solubility***

Most of the substances considered in this work were readily soluble in the aqueous media of the non-animal methods. However, if a substance *not* being readily soluble is negative in the DPRA [78], no firm conclusion on the lack of reactivity could be drawn. In the cell-based assays, substances with a *log* K_{OW} value >3.5 or >5 are outside the known applicability domain of the h-CLAT [96] or KeratinoSens™ [83], respectively, and tend to produce false negative results [140]. Hence, further progress may be required in the investigation of insoluble solid and highly lipophilic liquid substances.
One option to gain further information on peptide reactivity may be the inclusion of QSAR tools [80]. For instance, a negative QSAR prediction may confirm a negative DPRA result for a substance causing solubility issues.

Regarding *in vitro* assays, 3D skin tissue models consisting of human-derived epidermal keratinocytes were proposed for testing substances with solubility issues [141]. The pure test substance or a respective suspension can be topically applied on the outermost cell layer. After an incubation period, cytotoxicity [141], the release of specific enzymes or cytokines [142] or the up-/down-regulation of certain genes [25;143] can be used as read-out parameters. However, the suitability and reliability of such models needs to be confirmed by testing a much larger number of (insoluble) substances.

- ***Pro-haptens***

In this work, it was shown that pre-haptens and most of the investigated pro-haptens could be reliably identified using non-animal methods. However, a challenging sub-group of pro-haptens is represented by the class of substances containing aliphatic amine moieties. Of the four sensitizing substances with aliphatic amine moieties in the data set, only ethylene diamine was correctly identified by all three non-animal methods (of note, the positive result in the enzyme-free DPRA was probably based on an artefact, since no peptide adduct was detected; see **Section 4.3**). The other three substances were false negatives in the DPRA. N,N-Dibutyl aniline and diethylene triamine were even false negatives in all three tests, leading to the conclusion that the metabolic capacity in the investigated cell-lines was not sufficient for the biotic activation of these specific pro-haptens.

One possibility is the inclusion of QSAR tools with their ability to simulate metabolic activation. The putative pro-hapten character of substances with aliphatic amine moieties in their molecular structures could be identified

and their sensitizing property could be predicted *in silico* and used as information source within the framework of an integrated approach to testing and assessment (IATA). More precisely: if the pro-hapten nature would be confirmed, the QSAR prediction on peptide reactivity could be used as DPRA surrogate in the '2 out of 3' ITS, since the DPRA is not applicable to pro-haptens anyway [144].

Another possibility to overcome the limited metabolic capacity of the cell-based assays is the addition of rat liver microsomes (S9-mix) as typically used in the field of *in vitro* toxicology. However, such a surrogate system contains much higher enzyme activities compared to those reported for human skin [95] and the increase in sensitivity of a combination of the KeratinoSens™ with S9-mix was associated with a decrease in the specificity at the same time [93].

- ***Complex tissue models***

Multi-organ chips (MOCs) like co-cultures of human artificial liver microtissues and skin biopsies [145] represent a great improvement in monitoring human organ-level functionality. Such MOCs comprise several advantages such as an air liquid interface simulating the dermal route of exposure, the reproduction of major organ-specific functions (organotypic properties), crosstalk between organs roughly reflecting parts of the human organism and a high metabolic capacity [145;146]. If draw-backs like a lack in reproducibility (due donor variability or limited technical validity [147]) can be addressed successfully, MOCs may increase the accuracy in toxicological risk assessments for skin sensitization especially for 'problem candidates' such as insoluble substances or specific pro-haptens.

- ***Numerical potency values***

The accurate sub-classification of weak and strong sensitizers is proved in this work, which is sufficient in the registration of chemicals according to GHS/CLP. In the sector of cosmetics, however, the generation of continuous numerical potency values using non-animal methods is essential to derive reference values for risk assessments - and still remains a challenge.

Beyond sub-classifying substances as weak or strong sensitizers, the kinetic DPRA may also predict EC3 values of the LLNA with a high accuracy (as illustrated in **Figure 5.1**). The high concordance of numerical potency values derived in the kinetic DPRA with EC3 values of the LLNA should be explored in the future by testing a higher number of substances and by comparing kinetic data with known human potency values.

Another promising testing strategy is considering kinetic data derived by peptide reactivity testing combined with concentration-response data of the KeratinoSens™ [123]. Also, the above described 3D skin tissue models are proposed to be used for evaluating potency, e.g. based on the release of specific cytokines (i.e. IL-18) and cytotoxicity [141]. A further approach is the "Bayesian network ITS" [125], which combines test results and other (computational) data (such as *log* K_{OW} values) on a substance in a probabilistic decision model. However, the algorithms underlying these testing strategies are complex and difficult to comprehend.

- ***Borderline range***

Evaluating continuous read-outs of *in vitro* studies to make draw dichotomous conclusions (e.g. hazard – no hazard) has inherent limitations given by the precision of the method. This is not limited to *in vitro* studies but also applies to *in vivo* methods. It is often ignored that the experimental readouts are *no* error-free numbers [148]. Due to biological and technical variabilities, every test method has a degree of uncertainty and readouts may be close to the classification threshold, which is used to dichotomize continuous data. The limited precision of testing methods and its impact on the toxicological decision making is a matter of recent scientific discussions [148-151] and needs to be further explored.

The borderline range (BR) was previously not considered, probably since the characteristics of each sample have to be considered and evaluated in a comprehensive statistical analysis. In addition, test results falling into the BR are uncertain and should be interpreted with due care; thus, a considerable number of substances could possibly not be classified. One approach to account for limited precision is the definition of a BR around the classification threshold; test results falling into the BR would indicate that no firm conclusion can be drawn and the test result is hence 'inconclusive' [149].

Concluding remarks

Overall, research in future should focus on furthering the experimental models (e.g. include metabolic competence and 3D models), include more substance-specific data (like physico-chemical properties and *in silico* predictions) as well as analyze and consider uncertainties (not only the referring to predictivity, but also to precision) and finally apply comprehensive data interpretation and prediction models (e.g. probabilistic). This will improve the applicability, accuracy and ultimately the acceptance of skin sensitization assessments using non-animal methods. Even more, it will be a bare necessity to address more complex effects of systemic and developmental toxicity in the future.

7 Abbreviations

Ac, acylating agent; Acc, accuracy; ACD, allergic contact dermatitis; ADH, alcohol dehydrogenase; AKR, human aldoketo reductase gene; ALDH, aldehyde dehydrogenase; AOP, adverse outcome pathway; APC, antigen-presenting cell; ARE, antioxidant response elements; AUC, area under the curve;

CD, cluster of differentiation; CYP, cytochrome P450 isoenzymes; Cys, cysteine-containing heptapeptide; DC, dendritic cell;

D$_{Cys}$, Cys depletion; $D_{Cys/Lys}$, mean of Cys and Lys depletion; DIP, fixed data interpretation procedure; DPRA, direct peptide reactivity assay;

EC3, effective concentration [%] inducing a three-fold cell proliferation in the murine lymph node (LLNA); ECHA, the European Chemicals Agency; ECVAM, European Centre for the Validation of Alternative Methods; ESAC, EURL ECVAM Scientific Advisory Committee;

FMO, Flavin-containing monooxygenase; FP, false positive; FN, false negative;

GPMT, guinea pig maximization test;

^{3}**H**, Tritium; h-CLAT, human cell-line activation test; HMT, human maximization test; HPLC-UV, high pressure liquid chromatography with ultra-violet light absorbance detection; HRIPT, human repeated insult patch test;

IATA, Integrated Approach to Testing and Assessment; IL, interleukin; ITS, integrated testing strategy;

k, reaction rate constant; KC, keratinocyte; KE, key event;

LLNA, local lymph node assay; LNCC, local lymph node assay detecting cell counts; *log* k, reaction rate constant; *log* K_{OW}, octanol/water partition coefficient; Lys, lysine-containing heptapeptide;

MA, Michael acceptors; MIE, molecular initiating event; MS, mass spectrometry; MUSST, myeloid U937 skin sensitization test;

n, number of substances analyzed; NAT1, N-acetyl transferase-1; Nrf2-Keap, Nuclear-like 2 Kelch-like ECH associated protein-1;

OECD, Organization for Economic Co-operation and Development;

PM, prediction model; PPD, *p*-phenylene diamine; PPRA, Peroxidase Peptide Reactivity Assay; PPV, positive predictive value;

qDPRA, quantitative DPRA; (Q)SAR, (quantitative) structure activity relationship;

R^2, degree of certainty; 3R, refine, reduce, replace; RC, reactive carbonyls; REACH, registration, evaluation, authorization (and restriction) of chemicals;

S9-mix, supernatant fraction obtained from rat liver homogenate by centrifuging at 9000 g for 20 minutes in a suitable medium; Sens, sensitivity; SI, stimulation index; $S_N1/2$, agents reacting in nucleophilic substitutions of type ½; S_NAr, nucleophilic substitutions occurring in aromatic agents; Spec, specificity;

TCR, T cell receptor; TG, test guideline; TIMES SS, tissue metabolism simulator for skin sensitization; TN, true negative; TNF, tumor necrosis factor; TP, true positive;

UGT, uridine diphosphate-glucuronosyltransferase; UN-GHS/CLP, United Nations Globally Harmonized System of Classification and Labelling of Chemicals; UVCBs, substances of unknown or variable composition;

WoE, weight of evidence;

8 References

1. Russell WMS, Burch RL. The principles of humane experimental technique. Methuen, Co. LTD., London; 1959.

2. Mehling A, Eriksson T, Eltze T, Kolle SN, Ramirez T, Teubner W et al. Non-animal test methods for predicting skin sensitization potentials. Archives of Toxicology 2012; 86:1273-1295.

3. OECD. Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment, Series on Testing and Assessment, No. 34. 2005.

4. Council Directive 1907/2006/EEC. EC: Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorization 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 Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Official Journal of the European Union 2006; L 396/1.

5. 2003/15/EC. Directive 2003/15/EC of the European Parliament and the Council of 27 February 2003 amending Council Directive 76/768/EEC on the approximations of laws of the Member States relating to cosmetic products. Official Journal of the European Union 2003; L66:26-35.

6. OECD. OECD Guidelines for Testing of Chemicals, Section 4 - Health Effects. 2016.

7. Hutchings CV, Shum KW, Gawkrodger DJ. Occupational contact dermatitis has an appreciable impact on quality of life. Contact Dermatitis 2001; 45(1):17-20.

8. Evans CC, Fleming JD. Images in clinical medicine. Allergic contact dermatitis from a henna tattoo. New England Journal of Medicine 2008; 359(6):627.

9. Thyssen JP, Linneberg A, Johansen JD. The epidemiology of contact allergy in the general population-prelevance and main findings. Contact Dermatitis 2007; 57:287-299.

10. Bruckner AL, Weston WL, Morelli JG. Does sensitization to contact allergens begin in infancy? Pediatrics 2000; 105(1):e3.

11. Zug KA, Warshaw EM, Fowler JF, Jr., Maibach HI, Belsito DL, Pratt MD et al. Patch test results of the North American contact dermatitis group 2005-2006. Dermatitis 2009; 20:149-160.

12. Lepoittevin J, Basketter DA, Goossens A, Karlberg AT. Allergic contact dermatits: The molecular basis. Springer, Berlin; 1998.

13. Kimber I, Basketter DA, Gerberick GF, Dearman RJ. Allergic contact dermatitis. International Immunopharmacology 2002; 2:201-211.

14. Saint-Mezard P, Bérard F, Dubois B, Kaiserlian D, Nicolas JF. The role of CD4+ and CD8+ T cells in contact hypersensitivity and allergic contact dermatitis. European Journal of Dermatology 2004; 14:131-138.

15. Bergstrom MA, Ott H, Carlsson A, Neis M, Zwadlo-Klarwasser G, Jonsson CAM et al. A skin-like cytochrome P450 cocktail cctivates prohaptens to contact allergenic metabolites. Journal of Investigative Dermatology 2007; 127:1145-1153.

16. Svensson CK. Biotransformation of drugs in human skin. Drug Metabolism and Disposition 2008; 37:247-253.

17. Pendlington RU, Minter HJ, Stupart L, MacKay C, Roper CS, Sanders DJ et al. Development of a modified in vitro skin absorption method to study the epidermal/dermal disposition of a contact allergen in human skin. Cutaneous and Ocular Toxicology 2008; 27:283-294.

18. Toebak MJ, Gibbs S, Bruynzell DP, Scheper RJ, Rustemeyer T. Dendritic cells: Biology of the skin. Contact Dermatitis 2009; 60:2-20.

19. Dupuis G, Benezra C. Allergic Contact Dermatitis to simple chemicals: A molecular approach. New York and Basel; 1982.

20. Landsteiner K, Jacobs J. Studies on the sensitization of animals with simple chemical compounds. Journal of Experimental Medicine 1936; 64:625-639.

21. Weltzien HU, Moulon C, Martin SF, Padovan E, Hartmann U, Kohler J. T cell immune response to haptens. Structural models for allergic and autoimmune reaction. Toxicology 1996; 107:141-151.

22. Rock KL, Goldberg AL. Degradation of cell proteins and the generation of MHC class I-presented peptides. Immunology 1999; 17:739-779.

23. Smith-Pease CK. From xenobiotic chemistry and metabolism to better prediction and risk assessment of skin allergy. Toxicology 2003; 192:1-22.

24. Gerberick F, Aleksic M, Basketter D, Casati S, Karlberg AT, Kern P et al. Chemical reactivity measurement and the predictive identification of skin sensitisers. ATLA-Alternatives to Laboratory Animals 2008; 36:215-242.

25. Cottrez F, Boitel E, Auriault C, Aeby P, Groux H. Genes specifically modulated in sensitized skins allow the detection of sensitizers in a reconstructed human skin model. Development of the SENS-IS assay. Toxicology In Vitro 2015; 29(4):787-802.

26. Banchereau J, Steinman RM. Dendritic cells and the control of immunity. Nature 1998; 392:245-252.

27. Ryan CA, Kimber I, Basketter DA, Pallardy M, Gildea LA, Gerberick GF. Dendritic cells and skin sensitization: Biological roles and uses in hazard identification. Toxicology and Applied Pharmacology 2007; 221:384-394.

28. Steinman RM. The denritic cell system and its role in immunogenicity. Annual Reviews Immunology 1991; 9:271-296.

29. Steinman RM. Dendritic cells and the control of immunity: Enhancing the efficiency of antigen presentation. The Mount Sinai Journal of Medicine 2001; 68:160-166.

30. Steinman RM, Mellman IS, Muller WA, Cohn ZA. Endocytosis and the recycling of plasma membrane. The Journal of Cell Biology 1983; 96:1-27.

31. Kimber I, Cumberbatch M. Dendritic cells in cuteanous immune response to chemical allergens. Food and Chemical Toxicology 1992; 117:137-146.

32. Hulette BC, Ryan CA, Gildea LA, Gerberick GF. Elucidating changes in surface marker expression of dendritic cells following chemical allergen treatment. Toxicology and Applied Pharmacology 2002; 182:226-233.

33. Ohl L, Mohaupt M, Czeloth N, Hintzen G, Kiafard Z, Zwirmer J et al. CCR7 governs skin dendritic cell migration under inflammatory and steady-state conditions. Immunity 2004; 21:279-288.

34. Sallusto F, Schaerlie P, Loetscher P, Schaniel C, Lenig D, Mackay CR et al. Rapid and coordinated switch in chemokine receptor expression during dendritic cell maturation. European Journal of Immunology 1998; 28:2760-2769.

35. Aiba S, Terunuma A, Manome H, Tagami H. Dendritic cells differently respond to haptens and irritants by their production of cytokines and expression of co-stimulatory molecules. European Journal of Immunology 1997; 27:3031-3038.

36. Rambukkana A, Bos JD, Irik D, Menko WJ, Kapsenberg ML, Das PK. In situ behavior of human langerhans cells in skin organ culture. Journal of Investigative Dermatology 1995; 109:510-512.

37. Ozawa H, Nakagawa S, Tagamim H, Aiba S. Interleukin-1b and granulocyte macrophage colony-stimulating factor mediate Langerhans cell maturation differently. Journal of Investigative Dermatology 1996; 106:441-445.

38. Roychowdhury S, Svensson CK. Mechanism of drug-induced delayed-type hypersensitivity reactions in the skin. The American Association of Pharmaceutical Scientists (AAPS) Journal 2005; 7:E834-E846.

39. Banchereau J, Briere F, Caus C, Davoust J, Lebecque S, Liu YJ et al. Immunobiology of dendritic cells. Annual Reviews Immunology 2000; 18:767-811.

40. Seidenari S, Mantovani L, Manzini BM, Pignatti M. Cross-sensitizations between azo dyes and para-amino compound. A study of 236 azo-dye-sensitive subjects. Contact Dermatitis 1997; 36(2):91-96.

41. Coopman S, Degreef H, Dooms-Goossens A. Identification of cross-reaction patterns in allergic contact dermatitis from topical corticosteroids. British Journal of Dermatology 1989; 121(1):27-34.

42. Pezutto A, Ulrichs T, Burmester GR. Taschenatlas der Immunologie. Second Edition. 2006.

43. Patlewicz G, Roberts DW, Walker JD. QSARs for the skin sensitization potential of aldehydes and related compounds. QSAR and Combinatorial Science 2003; 22:196-203.

44. Aptula AO, Roberts DW. Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: General principles and application to reactive toxicity. Chemical Research in Toxicology 2006; 19:1097-1105.

45. Patlewicz G, Basketter DA, Smith CK, Hotchkiss SA, Roberts DW. Skin-sensitization structure-activity relationships for aldehydes. Contact Dermatitis 2001; 44(6):331-336.

46. Enoch SJ, Roberts DW. Predicting skin sensitization potency for michael acceptors in the LLNA using quantum mechanics calculations. Chemical Research in Toxicology 2013; 26(5):767-774.

47. Aptula AO, Patlewicz G, Roberts DW. Skin sensitization: reaction mechanistic applicability domains for structure-activity relationships. Chemical Research in Toxicology 2005; 18(9):1420-1426.

48. Roberts DW, Aptula AO. Electrophilic reactivity and skin sensitization potency of SNAr electrophiles. Chemical Research in Toxicology 2014; 27(2):240-246.

49. Thierse HJ, Moulon C, Allespach Y, Zimmermann B, Doetze A, Kuppig S et al. Metal-protein complex-mediated transport and delivery of Ni^{2+} to TCR/MHC contact sites in nickel-specific human T cell activation. Journal of Immunology 2004; 172:1926-1934.

50. Aptula AO, Roberts DW, Pease CK. Haptens, prohaptens and prehaptens, or electrophiles and proelectrophiles. Contact Dermatitis 2007; 56:54-56.

51. Lepoittevin JP. Metabolism versus chemical transformation or pro- versus prehaptens? Contact Dermatitis 2006; 54:73-74.

52. Urbisch D, Becker M, Honarvar N, Kolle SN, Mehling A, Teubner W et al. Assessment of pre- and pro-haptens using non-animal test methods for skin sensitization. Chemical Research in Toxicology 2016.

53. European Parliament. Regulation (EC) No 1278/2008. 2008. http://reach-compliance.eu/english/legislation/docs/launchers/CLP/launch-2008-1272-EC-CLP.html

54. European Chemicals Agency. The Use of Alternatives to Testing on Animals for the REACH Regulation (Second report under Article 117(3) of the REACH Regulation). 2014.

55. Angers-Loustau A, Tosti L, Casati S. The regulatory use of the Local Lymph Node Assay for the notification of new chemicals in Europe. Regulatory Toxicology and Pharmacology 2011; 60(3):300-307.

56. European Chemcial Agency. Guidance on information requirements and Chemical Safety Assessment. 2016.

57. European Commission. Amendmend of Annex VII to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) as regards skin sensitisation. Official Journal of the European Union 2016; L 255/14.

58. European Commission. Amendmend of Annex VII to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) as regards skin sensitisation. Official Journal of the European Union 2016; L 255/14.

59. Regulation (EC) No 1107/2009 of the European Parliament and the Council of the European Union. Regulation of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC; Official Journal of the European Union. 2009.

60. OECD TG 406. Guideline for Testing of Chemicals No. 406: Skin Sensitization. 1992.

61. Buehler EV. Delayed contact hypersensitivity in the guinea pig. Archives in Dermatology 1965; 91:171-177.

62. Magnusson B, Kligman AM. The Identification of contact allergens by animal assay. The guinea pig maximization test. Journal of Investigative Dermatology 1969; 52:268-276.

63. Kimber I, Basketter DA, Butler M, Gamer A, Garrigue JL, Gerberick GF et al. Classification of contact allergens according to potency: Proposals. Food and Chemical Toxicology 2003; 41:1799-1809.

64. Kolle SN, Basketter D, Schrage A, Gamer AO, van RB, Landsiedel R. Further experience with the local lymph node assay using standard radioactive and nonradioactive cell count measurements. Journal of Applied Toxicology 2012; 32(8):597-607.

65. Basketter D, Alepee N, Casati S, Crozier J, Eigler D, Griem P et al. Skin sensitisation--moving forward with non-animal testing strategies for regulatory purposes in the EU. Regulatory Toxicology and Pharmacology 2013; 67(3):531-535.

66. OECD. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins; Part 1: Scientific Evidence. 2012.

67. OECD. 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. 2012.

68. Ankley GT, Bennett RS, Erickson RJ, Hoff DJ, Hornung MW, Johnson RD et al. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environmental Toxicology and Chemistry 2010; 29(3):730-741.

69. Goebel C, Aeby P, Ade N, Alepee N, Aptula A, Araki D et al. Guiding principles for the implementation of non-animal safety assessment approaches for cosmetics: skin sensitisation. Regulatory Toxicology and Pharmacology 2012; 63(1):40-52.

70. Teubner W, Mehling A, Schuster PX, Guth K, Worth A, Burton J et al. Computer models versus reality: how well do in silico models currently predict the sensitization potential of a substance. Regulatory Toxicology and Pharmacology 2013; 67(3):468-485.

71. Devillers J, Mombelli E. Evaluation of the OECD QSAR Application Toolbox and Toxtree for estimating the mutagenicity of chemicals. Part 1. Aromatic amines. SAR and QSAR in Environmental Research 2010; 21(7-8):753-769.

72. Devillers J, Mombelli E. Evaluation of the OECD QSAR Application Toolbox and Toxtree for estimating the mutagenicity of chemicals. Part 2. alpha-beta unsaturated aliphatic aldehydes. SAR and QSAR in Environmental Research 2010; 21(7-8):771-783.

73. Patlewicz G, Dimitrov SD, Low LK, Kern PS, Dimitrova GD, Comber MI et al. 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. Regulatory Toxicology and Pharmacology 2007; 48(2):225-239.

74. Patlewicz G, Kuseva C, Mehmed A, Popova Y, Dimitrova G, Ellis G et al. TIMES-SS--recent refinements resulting from an industrial skin sensitisation consortium. SAR and QSAR in Environmental Research 2014; 25(5):367-391.

75. Aeby P, Ashikaga T, Bessou-Touya S, Schepky A, Gerberick GF, Kern P et al. Identifying and charaterizing chemical sensitizers without animal testing: COLIPA's research and method development program. Toxicology In Vitro 2010; 24:1465-1473 .

76. Dimitrov SD, Low LK, Patlewicz GY, Kern PS, Dimitrova GD, Comber MHI et al. Skin sensitization: Modeling based on skin metabolism simulation and formation of protein conjugates. International Journal of Toxicology 2005; 24:189-204.

77. Gerberick GF, Vassallo JD, Bailey RE, Chaney JG, Morrall SW, Lepoittevin JP. Development of a peptide reactivity assay for screening contact allergens. Toxicological Sciences 2004; 81:332-343.

78. OECD. OECD TG 442C: In chemico skin sensitization: Direct Peptide Reactivity Assay (DPRA). 2015.

79. Gerberick GF, Vassallo JD, Foertsch LM, Price BB, Chaney JG, Lepoittevin JP. Quantification of chemical peptide reactivity for screening contact allergens: A classification tree model approach. Toxicological Sciences 2007; 97:417-427.

80. Urbisch D, Honarvar N, Kolle SN, Mehling A, Ramirez T, Teubner W et al. Peptide reactivity associated with skin sensitization: The QSAR Toolbox and TIMES compared to the DPRA. Toxicology In Vitro 2016.

81. Emter R, Ellis G, Natsch A. Performance of a novel keratinocyte-based reporter cell line to screen skin sensitizers in vitro. Toxicology and Applied Pharmacology 2010; 245:281-290.

82. Ramirez T, Mehling A, Kolle SN, Wruck CJ, Teubner W, Eltze T et al. LuSens: a keratinocyte based ARE reporter gene assay for use in integrated testing strategies for skin sensitization hazard identification. Toxicology In Vitro 2014; 28(8):1482-1497.

83. OECD. OECD TG 442D: *In vitro* skin sensitization: ARE-Nrf2 Luciferase Test Methods. 2015.

84. Ramirez T, Stein N, Aumann A, Remus T, Edwards A, Norman KG et al. Intra- and inter-laboratory reproducibility and accuracy of the LuSens assay: A reporter gene-cell line to detect keratinocyte activation by skin sensitizers. Toxicology In Vitro 2016; 32:278-286.

85. ESAC. ESAC Opinion on the BASF-coordinated Performance Standards-based validation of the LuSens test method for skin sensitisation testing. 2016.

86. Dinkova-Kostova AT, Holtzclaw WD, Kensler TW. The role of Keap1 in cellular protective responses. Chemical Research in Toxicology 2005; 18:1779-1791.

87. Wakabayashi N, Dinkova-Kostova AT, Holtzclaw WD, ang M, obayashi A, amamoto M et al. Protection against electrophile and oxidant stress by induction of the phase 2 response: Fate of cysteines of the

Keap1 sensor modified by inducers. Proceedings of the National Academy of Sciences of the United States of America 2004; 101:2040-2045.

88. Wang W, Jaiswal AK. Nuclear factor Nrf2 and antioxidant response element regulate NRH:quinone oxidoreductase 2 (NQO2) gene expression and antioxidant induction. Free Radical Biology and Medicine Journal 2006; 40:1119-1130.

89. Bauch C, Kolle SN, Ramirez T, Eltze T, Fabian E, Mehling A et al. Putting the parts together: Combining in vitro methods to test for skin sensitizing potentials. Regulatory Toxicology and Pharmacology 2012; 63:489-504.

90. Natsch A, Emter R. Skin sensitizers induce antioxidant response element dependent genes: Application to the in vitro testing of the sensitization potential of chemicals. Toxicological Sciences 2008; 102:110-119.

91. McKim JM, Keller DJ, Gorski JR. A new in vitro method for identifying chemical sensitizers combining peptide binding with ARE/EpRE-mediated gene expression in human skin cells. Cutaneous and Ocular Toxicology 2010; 29:171-192.

92. Fabian E, Vogel D, Blatz V, Ramirez T, Kolle S, Eltze T et al. Xenobiotic metabolizing enzyme activities in cells used for testing skin sensitization in vitro. Archives in Toxicology 2013; 87(9):1683-1696.

93. Natsch A, Haupt T. Utility of rat liver S9 fractions to study skin-sensitizing prohaptens in a modified KeratinoSens assay. Toxicological Sciences 2013; 135(2):356-368.

94. Oesch F, Fabian E, Oesch-Bartlomowicz B, Werner C, Landsiedel R. Drug-metabolizing enzymes in the skin of man, rat, and pig. Drug Metabolism Reviews 2007; 39:659-698.

95. Oesch F, Fabian E, Guth K, Landsiedel R. Xenobiotic-metabolizing enzymes in the skin of rat, mouse, pig, guinea pig, man, and in human skin models. Archives in Toxicology 2014; 88(12):2135-2190.

96. OECD. OECD TG 442E: *In vitro* human Cell Line Activation Test (h-CLAT). 2016.

97. Ade N, Leon F, Pallardy M, Peiffer JL, Kerdine-Romer S, Tissier MH et al. HMOX1 and NQO1 genes are upregulated in response to contact sensitizers in dendritic cells and THP-1 cell line: Role of the Keap1/Nrf2 pathway. Toxicological Sciences 2009; 107:451-460.

98. Ashikaga T, Hoya M, Itagaki H, Kutumura Y, Aiba S. Evaluation of CD86 expression and MHC class II molecular internalization in THP-1 human monocyte cells as predicitive endpoint for contact sensitizers. Toxicol In Vitro 2002; 16:711-716.

99. Sakaguchi H, Ashikaga T, Miyazawa M, Yoshida Y, Ito Y, Yoneyama K et al. Development of an in vitro skin sensitization test using human cell lines; human cell line activation test (h-CLAT) II. An inter-laboratory study of the h-CLAT. Toxicology in Vitro 2006; 20:774-784.

100. Yoshida Y, Sakaguchi H, Ito Y, Okuda M, Suzuki H. Evaluation of the skin senitization potential of chemicals using expression of co-stimulatory molecules CD54 and CD86, on the naive THP-1 cell line. Toxicology In Vitro 2003; 17(221):228.

101. Ashikaga T, Sakaguchi H, Sono S, Kosaka N, Ishikawa M, Nukada Y et al. A comparitive evaluation of in vitro skin sensitization test: The human cell-line activation test (h-CLAT) versus local lumph node assay (LLNA). Alternatives to Laboratory Animals 2010; 38:275-284.

102. Nukada Y, Ashikaga T, Sakaguchi H, Sono S, Mugita N, Hirota M et al. Predictive performance for human skin sensitizing potential of the human cell line activation test (h-CLAT). Contact Dermatitis 2011;1-11.

103. Nukada Y, Ito Y, Miyazawa M, Sakaguchi H, Nishiyama N. The relationship between CD86 and CD54 protein expression and cytotoxicity following stimulation with contact allergen in THP-1 cells. Journal of Toxicological Sciences 2011; 36:313-324.

104. Sakaguchi H, Ryan CA, Ovigne JM, Schroeder KR, Ashikaga T. Predicting skin sensitization potential and interlaboratory reproducibility of a human cell line activation test (h-CLAT) in the European Cosmetics Association (COLIPA) ring trials. Toxicology In Vitro 2010; 24:1810-1820.

105. Schnuch A, Mildau G, Kratz EM, Uter W. Risk of sensitization to preservatives estimated on the basis of patch test data and exposure, according to a sample of 3541 leave-on products. Contact Dermatitis 2011; 65(3):167-174.

106. Python F, Goebel C, Aeby P. Assessment of the U937 cell line for the detection of contact allergens. Toxicology and Applied Pharmacology 2007; 220:113-124.

107. Piroird C, Ovigne JM, Rousset F, Martinozzi-Teissier S, Gomes C, Cotovio J et al. The Myeloid U937 Skin Sensitization Test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome pathway for skin sensitization. Toxicology In Vitro 2015; 29(5):901-916.

108. Dai R, Streilein JW. Naive hapten-specific human T-lymphocytes are primed in vitro with derivatized blood mononuclear cells. Journal of Investigative Dermatology 1998; 110:29-33.

109. Dietz L, Esser PR, Schmucker SS, Goette I, Richter A, Schnölzer M et al. Tracking human contact allergens: From mass spectrometric identification of peptide-bound reactive small chemicals to chemical-specific naive human T cell priming. Toxicological Sciences 2010; 117:336-347.

110. Rustemeyer T, De Ligter S, von Blomberg BM, Frosch PJ, Scheper RJ. Human T lymphocate priming in vitro by haptenated autologous dendritic cells. Clinical and Experimental Immunology 1999; 117:209-216.

111. Krasteva M, Peguet-Navarro J, Moulon C, Courtellemont P, Redziniak G, Schmitt D. In vitro primary sensitization of hapten-specific T cells by cultured human epideraml Langerhans cells - a screening predicitve assay for contact sensitizers. Clinical and Experimental Allergy 1996; 26:563-570.

112. Maxwell G, Aeby P, Ashikaga T, Bessou-Touya S, Diembeck W, Gerberick GF et al. Skin sensitization: the Colipa strategy for developing and evaluating non-animal test methods for risk assessment. ALTEX - Alternatives to Animal Experimentation 2011; 28:50-55.

113. Kimber I, Basketter DA, Gerberick GF, Ryan CA, Dearman RJ. Chemical allergy: Translating Biology into hazard characterizazion. Toxicological Sciences 2011; 120:S238-S268.

114. Moon JJ, Chu H, Pepper M, McSorley SJ, Jameson SC, Kedl R et al. Naive CD4 T cell frequency varies for different epitopes and predicts repertoire and response magnitude. Immunity 2007; 27:203-213.

115. Martin SF. Contact dermatits: from pathomechanism to immunotoxicology. Experimental Dermatology 2012; 21:382-289.

116. Basketter D, Ashikaga T, Casati S, Hubesch B, Jaworska J, de KJ et al. Alternatives for skin sensitisation: Hazard identification and potency categorisation: Report from an EPAA/CEFIC LRI/Cosmetics Europe cross sector workshop, ECHA Helsinki, April 23rd and 24th 2015. Regulatory Toxicology and Pharmacology 2015; 73(2):660-666.

117. Sauer UG, Hill EH, Curren RD, Raabe HA, Kolle SN, Teubner W et al. Local tolerance testing under REACH: Accepted non-animal methods are not on equal footing with animal tests. ATLA - - Alternatives to Laboratory Animals 2016; 44(3):281-299.

118. Urbisch D, Mehling A, Guth K, Ramirez T, Honarvar N, Kolle S et al. Assessing skin sensitization hazard in mice and men using non-animal test methods. Regulatory Toxicology and Pharmacology 2015; 71(2):337-351.

119. Patlewicz G, Kuseva C, Kesova A, Popova I, Zhechev T, Pavlov T et al. Towards AOP application--implementation of an integrated approach to testing and assessment (IATA) into a pipeline tool for skin sensitization. Regulatory Toxicology and Pharmacology 2014; 69(3):529-545.

120. Takenouchi O, Fukui S, Okamoto K, Kurotani S, Imai N, Fujishiro M et al. 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. Journal of Applied Toxicology 2015.

121. Nukada Y, Miyazawa M, Kazutoshi S, Sakaguchi H, Nishiyama N. Data integration of non-animal tests for the development of a test battery to predict the skin sensitizing potential and potency of chemicals. Toxicology In Vitro 2013; 27(2):609-618.

122. van der Veen JW, Rorije E, Emter R, Natsch A, van LH, Ezendam J. Evaluating the performance of integrated approaches for hazard identification of skin sensitizing chemicals. Regulatory Toxicology and Pharmacology 2014; 69(3):371-379.

123. Natsch A, Emter R, Gfeller H, Haupt T, Ellis G. Predicting skin sensitizer potency based on in vitro data from KeratinoSens and kinetic peptide binding: global versus domain-based assessment. Toxicological Sciences 2015; 143(2):319-332.

124. Jaworska J, Dancik Y, Kern P, Gerberick F, Natsch A. Bayesian integrated testing strategy to assess skin sensitization potency: from theory to practice. Journal of Applied Toxicology 2013.

125. Jaworska JS, Natsch A, Ryan C, Strickland J, Ashikaga T, Miyazawa M. Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: a decision support system for quantitative weight of evidence and adaptive testing strategy. Archives in Toxicology 2015; 89(12):2355-2383.

126. Hirota M, Kouzuki H, Ashikaga T, Sono S, Tsujita K, Sasa H et al. Artificial neural network analysis of data from multiple in vitro assays for prediction of skin sensitization potency of chemicals. Toxicology In Vitro 2013; 27(4):1233-1246.

127. Hirota M, Fukui S, Okamoto K, Kurotani S, Imai N, Fujishiro M et al. Evaluation of combinations of in vitro sensitization test descriptors for the artificial neural network-based risk assessment model of skin sensitization. Journal of Applied Toxicology 2015.

128. Tsujita-Inoue K, Hirota M, Ashikaga T, Atobe T, Kouzuki H, Aiba S. Skin sensitization risk assessment model using artificial neural network analysis of data from multiple in vitro assays. Toxicology In Vitro 2014; 28(4):626-639.

129. Silvia Teissier, Nathalie Alépée. Case study 5: L'OREAL approach and decision strategy (http://cefic-lri.org/wp-content/uploads/2014/03/G-Case-Study-3-LOreal-April-2015.pdf). 2015.

130. J Matheson. ICCVAM Integrated Decision Strategy for Skin Sensitization (http://ntp.niehs.nih.gov/iccvam/meetings/sot15/matheson-poster-text-508.pdf). 2015.

131. Basketter D, Ashikaga T, Casati S, Hubesch B, Jaworska J, de KJ et al. Alternatives for skin sensitisation: Hazard identification and potency categorisation: Report from an EPAA/CEFIC LRI/Cosmetics Europe cross sector workshop, ECHA Helsinki, April 23rd and 24th 2015. Regulatory Toxicology and Pharmacology 2015; 73(2):660-666 .

132. OECD. Guidance document on the reporting of defined approaches and individual information sources to be used within integrated approaches to testing and assessment (IATA) for skin sensitization. ENV/JM/HA(2016)11. 2016.

133. Basketter DA, Alepee N, Ashikaga T, Barroso J, Gilmour N, Goebel C et al. Categorization of chemicals according to their relative human skin sensitizing potency. Dermatitis 2014; 25(1):11-21.

134. Natsch A, Ryan CA, Foertsch L, Emter R, Jaworska J, Gerberick F et al. A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. Journal of Applied Toxicology 2013.

135. Kligman AM. Identification of contact allergens by human assay .3. Maximization test - A procedure for screening and rating contact sensitizers. Journal of Investigative Dermatology 1966; 47:393.

136. Basketter DA. The human repeated insult patch test in the 21st century: a commentary. Cutaneous and Ocular Toxicology 2009; 28(2):49-53.

137. De Wever B, Fuchs HW, Gaca M, Krul C, Mikulowski S, Poth A et al. Implementation challenges for designing integrated in vitro testing strategies (ITS) aiming at reducing and replacing animal experimentation. Toxicology In Vitro 2012; 26:526-534.

138. Gerberick GF, Troutman JA, Foertsch LM, Vassallo JD, Quijano M, Dobson RLM et al. Investigation of peptide reactivity of pro-hapten skin sensitizers using a peroxidase-peroxide oxidation system. Toxicological Sciences 2009; 112:164-174.

139. Troutman JA, Foertsch L, Kern PS, Dai HJ, Quijano M, Dobson RLM et al. The incorporation of lysine into the peroxide-peptide reactivity assay for skin sensitization assessment. Toxicological Sciences 2011; 122:422-436.

140. Takenouchi O, Miyazawa M, Saito K, Ashikaga T, Sakaguchi H. Predictive performance of the human Cell Line Activation Test (h-CLAT) for lipophilic chemicals with high octanol-water partition coefficients. Journal of Toxicological Sciences 2013; 38(4):599-609.

141. Gibbs S, Corsini E, Spiekstra SW, Galbiati V, Fuchs HW, Degeorge G et al. An epidermal equivalent assay for identification and ranking potency of contact sensitizers. Toxicology and Applied Pharmacology 2013; 272(2):529-541.

142. Andres E, Barry M, Hundt A, Dini C, Corsini E, Gibbs S et al. Preliminary performance data of the RHE/IL-18 assay performed on SkinEthic RHE for the identification of contact sensitizers. International Journal of Cosmetic Sciences; 39(2):121-132.

143. Cottrez F, Boitel E, Ourlin JC, Peiffer JL, Fabre I, Henaoui IS et al. SENS-IS, a 3D reconstituted epidermis based model for quantifying chemical sensitization potency: Reproducibility and predictivity results from an inter-laboratory study. Toxicology In Vitro 2016; 32:248-260.

144. Gibbs S, Corsini E, Spiekstra SW, Galbiati V, Fuchs HW, Degeorge G et al. An epidermal equivalent assay for identification and ranking potency of contact sensitizers. Toxicology and Applied Pharmacology 2013; 272(2):529-541.

145. Wagner I, Materne EM, Brincker S, Sussbier U, Fradrich C, Busek M et al. A dynamic multi-organ-chip for long-term cultivation and substance testing proven by 3D human liver and skin tissue co-culture. Lab on a Chip 2013; 13(18):3538-3547.

146. Alepee N, Bahinski A, Daneshian M, De WB, Fritsche E, Goldberg A et al. State-of-the-art of 3D cultures (organs-on-a-chip) in safety testing and pathophysiology. ALTEX - Alternatives to Animal Experimentation 2014; 31(4):441-477.

147. Leist M, Efremova L, Karreman C. Food for thought ... considerations and guidelines for basic test method descriptions in toxicology. ALTEX - Alternatives to Animal Experimentation 2010; 27(4):309-317.

148. Dimitrov S, Detroyer A, Piroird C, Gomes C, Eilstein J, Pauloin T et al. Accounting for data variability, a key factor in in vivo/in vitro relationships: application to the skin sensitization potency (in vivo LLNA versus in vitro DPRA) example. Journal of Applied Toxicology. 2016.

149. Leontaridou M, Urbisch D, Kolle SN, Ott K, Mulliner DS, Gabbert S et al. Quantification of the borderline range and implications for evaluating non-animal testing methods' precision. ALTEX - Alternatives to Animal Experimentation. 2017.

150. Hoffmann S. LLNA variability: An essential ingredient for a comprehensive assessment of non-animal skin sensitization test methods and strategies. ALTEX - Alternatives to Animal Experimentation 2015; 32(4):379-383.

151. Hoffmann S, Cole T, Hartung T. Skin irritation: prevalence, variability, and regulatory classification of existing in vivo data from industrial chemicals. Regulatory Toxicology and Pharmacology 2005; 41(3):159-166.