

State of the art in non-animal approaches for skin sensitization testing: from individual test methods towards testing strategies

Janine Ezendam¹  · Hedwig M. Braakhuis¹ · Rob J. Vandebriel¹

Received: 18 July 2016 / Accepted: 29 August 2016 / Published online: 14 September 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract The hazard assessment of skin sensitizers relies mainly on animal testing, but much progress is made in the development, validation and regulatory acceptance and implementation of non-animal predictive approaches. In this review, we provide an update on the available computational tools and animal-free test methods for the prediction of skin sensitization hazard. These individual test methods address mostly one mechanistic step of the process of skin sensitization induction. The adverse outcome pathway (AOP) for skin sensitization describes the key events (KEs) that lead to skin sensitization. In our review, we have clustered the available test methods according to the KE they inform: the molecular initiating event (MIE/KE1)—protein binding, KE2—keratinocyte activation, KE3—dendritic cell activation and KE4—T cell activation and proliferation. In recent years, most progress has been made in the development and validation of in vitro assays that address KE2 and KE3. No standardized in vitro assays for T cell activation are available; thus, KE4 cannot be measured in vitro. Three non-animal test methods, addressing either the MIE, KE2 or KE3, are accepted as OECD test guidelines, and this has accelerated the development of integrated or defined approaches for testing and assessment (e.g. testing strategies). The majority of these approaches are mechanism-based, since they combine results from multiple test methods and/or computational tools that address different KEs of the AOP to estimate skin sensitization potential and sometimes potency. Other approaches

are based on statistical tools. Until now, eleven different testing strategies have been published, the majority using the same individual information sources. Our review shows that some of the defined approaches to testing and assessment are able to accurately predict skin sensitization hazard, sometimes even more accurate than the currently used animal test. A few defined approaches are developed to provide an estimate of the potency sub-category of a skin sensitizer as well, but these approaches need further independent evaluation with a new dataset of chemicals. To conclude, this update shows that the field of non-animal approaches for skin sensitization has evolved greatly in recent years and that it is possible to predict skin sensitization hazard without animal testing.

Keywords Skin sensitization · Alternatives to animal testing · Testing strategies · Hazard assessment · Adverse outcome pathway

Introduction

In this review, we provide an update of our previous review on the state of the art of non-animal predictive methods to identify sensitizers (Vandebriel and van Loveren 2010). It is of interest to explore the current state of animal-free approaches for skin sensitization testing, in the light of the prognosis made in expert meetings organized in 2010 by the European Commission (Adler et al. 2011). Different expert groups analysed the status and prospects of alternative methods for full replacement of animal testing for cosmetics testing for five toxicological areas, including skin sensitization. The prospects to assess the skin sensitization potential were optimistic, and it was expected that with a combination of non-animal test methods that address

✉ Janine Ezendam
Janine.Ezendam@rivm.nl

¹ Department of Innovative Testing Strategies, Center for Health Protection, National Institute for Public Health and the Environment (RIVM), PO Box 1, 3720 BA Bilthoven, The Netherlands

different mechanistic steps involved in skin sensitization, hazard identification would become possible without animal testing in 2017–2019. Full replacement for skin-sensitizing potency assessment, however, would not be feasible in 2013, according to these experts.

A skin sensitizer is a chemical that is capable of inducing a T cell-mediated delayed-type hypersensitivity response after dermal contact and penetration into the epidermis. The clinical consequence, allergic contact dermatitis (ACD), occurs after renewed skin exposure to this chemical. ACD is an important consumer and occupational health problem (Kimber et al. 2011). Information on this toxicological endpoint is required in different legal frameworks, including the European regulation of Registration, Evaluation, Authorization and Restriction of Chemical substances (REACH; EC 1907/2006) and under the United Nations Globally Harmonized System for Classification, Labeling and Packaging (GHS/CLP). In addition, in the safety assessment of cosmetic ingredients under the EU Cosmetics Directive (EC 1223/2009) information on skin sensitization is required. Within these legislative frameworks, information is needed on the skin sensitization *hazard potential*, e.g. distinguishing skin sensitizers from non-sensitizers. Furthermore, quantitative information that allows *hazard assessment* is needed as well to meet the regulatory requirements to sub-categorize skin sensitizers according to their potency class. Skin-sensitizing potency is a measure for the intrinsic strength of a skin sensitizer. Strong skin sensitizers will be able to induce skin sensitization at lower concentrations compared to weak skin sensitizers. Quantitative risk characterization that includes exposure and hazard assessment is not done for regulatory purposes. Such a quantitative approach is developed and used by the fragrance industry and uses the potency of a skin sensitizer as a point of departure (Api et al. 2008).

Until now, these legislative frameworks rely on information from animal models. Regulatory accepted models are the guinea pig models described in OECD TG406 (OECD 1992), the guinea pig maximization test (GPMT) and Buehler test, and the local lymph node assay (LLNA) described in OECD TG429 (OECD 2010a) and non-radioactive variants of the LLNA in OECD TG 442A and 442B (OECD 2010b, c). From an animal welfare perspective, the LLNA is a refinement and reduction method compared to the classical guinea pig methods. Another advantage of the LLNA is that it generates dose–response data on lymphocyte proliferation. These quantitative data are used to derive the EC3 value, which is the effective concentration that induces a threefold increase in lymphocyte proliferation. The EC3 correlates relatively well with known human thresholds for contact allergy (Basketter et al. 2005; Griem et al. 2003; Schneider and Akkan 2004). The LLNA EC3 value is used for sub-categorization of skin sensitizers, e.g. in the GHS

sub-categories 1A ($EC3 \leq 2\%$; strong skin sensitizers) and 1B ($EC3 > 2\%$; other skin sensitizers). Clearly, novel predictive test methods or strategies for skin sensitization need to provide information that allow both hazard identification and potency sub-categorization of skin sensitizers.

At the OECD, three non-animal test methods for skin sensitization are accepted as test guidelines: (1) the Direct Peptide Reactivity Assay (DPRA) (OECD TG442C), which measures the *in chemico* peptide reactivity of chemicals (OECD 2015a); (2) the ARE-Nrf2 Luciferase Test Method—KeratiNoSens™—which measures *in vitro* keratinocyte activation (OECD 2015b); and (3) the human cell line activation test (h-CLAT), which measures *in vitro* dendritic cell activation (OECD 2016c). Besides these regulatory accepted methods, many other tests are in the pipeline. Some are still in an early stage of development, whereas others are being validated in ring trials and/or entered the peer-review process at EURL-ECVAM (European Union Reference Laboratory for Alternatives to Animal Testing) (EURL-ECVAM 2015c).

Non-animal test methods for skin sensitization address one of the key events of the adverse outcome pathway (AOP) of skin sensitization, developed by the OECD (OECD 2012), which is recently published in the OECD AOP wiki (Fig. 1). An AOP describes a sequence of key events (KEs) triggered by a chemical. Each AOP starts with a molecular initiating event (MIE), which is the first KE that represents the event that initiates the perturbation caused by the chemical, followed by several KEs at the cellular, organ or organism level, eventually resulting in the adverse outcome. For skin sensitization, the MIE is covalent binding of the chemical or its metabolites to skin proteins, resulting in the formation of an immunogenic conjugate. This triggers two cellular events, KE2 and KE3. The electrophilic properties of skin sensitizers cause an activation of keratinocytes (KE2), resulting in the release of pro-inflammatory cytokines and chemokines and activation of antioxidant responses. These mediators are essential components for the generation of an adaptive immune response (Martin et al. 2011; Van der Veen et al. 2011). The immunogenic conjugate is taken up by dermal dendritic cells (DCs) and/or Langerhans cells, resulting in activation and migration of these cells to the draining lymph nodes (KE3). The dendritic cells present epitopes of the conjugate to T cells, resulting in activation and proliferation of T cells (KE4), and acquisition immunological memory, resulting in sensitization of the immune system (adverse outcome). The AOP is a simplification of a complex biological process, which offers the opportunity to anchor existing test methods or *in silico* tools that measure the MIE or other key events. In addition, it gives insight, which specific test methods need to be developed to cover all key events (Van der Veen et al. 2014b). The AOP can be used to mechanistically underpin

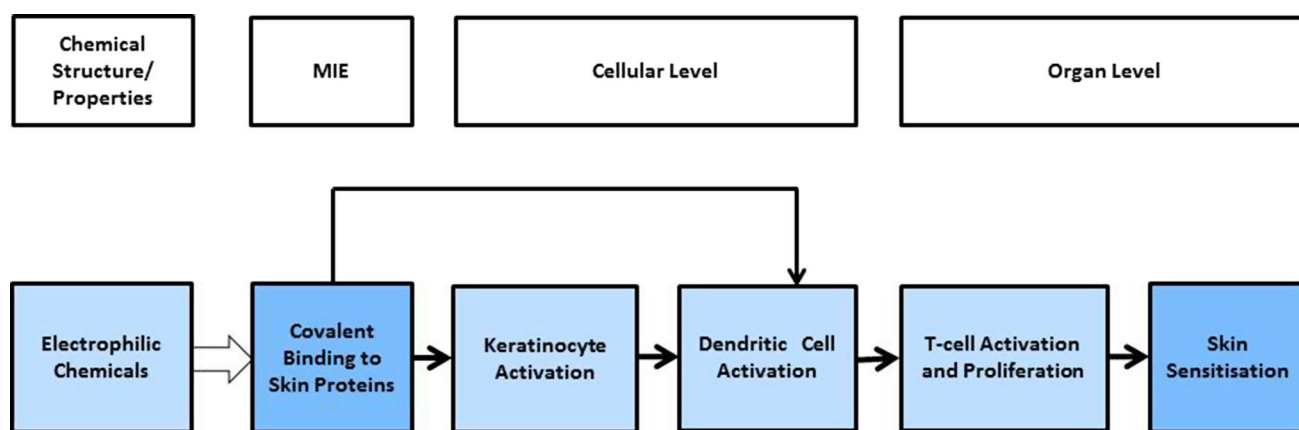


Fig. 1 AOP for skin sensitization. The AOP describes the molecular initiating event (MIE) and subsequent key events (KEs) that lead to the adverse outcome skin sensitization. Covalent binding to skin proteins is the first key event: the MIE. Two cellular events are linked to

the MIE: keratinocyte activation (KE2) and dendritic cell activation (KE3). At the organ level, KE4 takes place, which is T cell activation and proliferation (figure is taken from the OECD AOP wiki: <https://aopwiki.org/Aop:40>. Last accessed on 27 June 2016)

the development of strategies for the testing and assessment of chemicals for a specific regulatory need (MacKay et al. 2013; Patlewicz et al. 2014).

In silico, in chemico and in vitro test methods for skin sensitization testing

This chapter provides an overview of the individual test methods for skin sensitization, including computational tools. The AOP is used to cluster these test methods according to the MIE or KE they address. Notably, the AOP describes the key events that occur once a substance has become bioavailable and does not include dermal bioavailability. The reason is that there is no direct relationship between skin permeability and skin sensitization (Alves et al. 2015); hence, not all substances that are able to penetrate the skin will induce sensitization. In the risk assessment process, however, information on dermal bioavailability of skin sensitizers is an important parameter that needs to be considered (Maxwell et al. 2014). Therefore, methods that can be used to determine dermal bioavailability are described in this chapter as well.

Dermal bioavailability: skin absorption and metabolism

For skin sensitization, the amount that penetrates the skin barrier and reaches the epidermis is the bioavailable concentration. The most important factors that determine skin permeation are lipophilicity and molecular weight (chemicals above 800 Mw do not penetrate the stratum corneum). Other factors that influence skin permeation are volatility, melting point, aqueous solubility of the chemical, contact

time at the skin and the total applied dose (expressed in dose per unit area).

Most of the available methods to assess dermal absorption are designed to determine the systemic concentration after dermal exposure, rather than the epidermal concentration which is the relevant dose metric for skin sensitization (Basketter et al. 2007; MacKay et al. 2013; Maxwell et al. 2014). There are in silico models providing direct estimates of dermal permeability based on physical–chemical properties [QSAR-based Dermal Permeability Coefficient Program (DERMWIN)] (Alves et al. 2015). Physiologically based pharmacokinetic (PBPK) models can be used as well to estimate dermal absorption (Mitrugotri et al. 2011). For actual measurements of skin penetration, OECD test guidelines are available, including an in vitro method (TG428) and an in vivo method (TG 427) for skin absorption. A modification of the in vitro skin absorption assay has been proposed that includes multiple time points and multi-compartmental kinetic modelling to be able to determine epidermal disposition (Davies et al. 2011; Pendlington et al. 2008). Additionally, kinetic modelling approaches that take into account evaporation, protein binding, and clearance to predict the bioavailable concentration in the epidermis have been recently proposed as well (Dancik et al. 2013; Ibrahim et al. 2012).

Although progress is made in this area, challenges remain, including the inclusion of metabolic clearance and activation (MacKay et al. 2013). Skin sensitizers can be divided into haptens, prehaptens and prohaptens. Haptens do not require any form of activation to induce sensitization, whereas prehaptens need to be activated by abiotic transformation via auto-oxidation or hydrolysis and prohaptens require enzymatic conversion (Karlberg et al. 2008). The skin itself is metabolically competent and able

to convert prohaptens into reactive haptens. Several computational tools have simulators to predict skin metabolism, e.g. the OECD QSAR Toolbox, TIMES-SS and Meteor Nexus. The OECD QSAR Toolbox has simulators for auto-oxidation and metabolism (Dumont et al. 2015).

In silico tools for skin sensitization

For the identification of skin sensitization hazard, (quantitative) structure–activity relationship (QSAR) models and expert systems are described in the literature and free and commercial software tools are available. An overview of available computational tools that may be useful to predict toxicological endpoints for regulatory purposes, including skin sensitization, has been compiled by the OECD (2014).

The *in silico* tools can be divided into global and local models. Global models are often empirical QSARs that are developed by applying statistical models to biological data or structural descriptors. These models aim to cover all skin sensitizers, e.g. VEGA, CASE Ultra and TOPKAT. Local models aim to predict skin sensitization based on mechanistic knowledge and are applicable to a specific chemical class that belongs to a specific mechanism applicability domain (Aptula and Roberts 2006). Five reaction mechanism applicability domains have been identified: (a) Michael-type acceptors (b) SNAr electrophiles, (c) SN2 electrophiles, (d) Schiff base formers and (e) acylating agents (Aptula and Roberts 2006). Examples of mechanistic computational models are Toxtree, Derek Nexus and OECD QSAR Toolbox.

Table 1 shows a non-exhaustive list of freely and commercially available software tools for skin sensitization. In the literature, more QSARs can be found, but most of these are not yet available as software tools. The *in silico* tools have their advantages and limitations. It is beyond the scope of this review to describe those in detail. Briefly, Rorije et al. (2013) evaluated the performance of *in silico* tools (TOPKAT, Derek Nexus, TIMES-SS, SMART rules (used in the OECD QSAR Toolbox), MultiCASE and VEGA. This evaluation demonstrated for a dataset of 522 chemicals that the individual models have a poor predictivity compared to *in vivo* and *in vitro* models. Their evaluation showed that the predictivity of TOPKAT was low, and sensitivity and specificity of TOPKAT to predict LLNA outcomes were 67 % and 56 %, respectively. MultiCASE had a good sensitivity (88 %), but a poor specificity (11 %). A battery of the best performing *in silico* tools (DEREK, SMARTS and TIMES-SS) based on Bayesian statistics improved predictivity, but only 280 of the 522 substances could be predicted, for the other substances the *in silico* models were unable to provide a prediction (Rorije et al. 2013). Teubner et al. (2013) came to a similar conclusion after evaluating the performance of different *in silico* tools

to predict skin sensitizers according to the GHS criteria. In this study, the performance of Derek Nexus, TIMES-SS, TOPKAT, Toxtree, VEGA, CASE Ultra and the OECD QSAR Toolbox was assessed for a dataset of 45 skin sensitizers and 55 non-sensitizers. This analysis showed also that each individual model had a poor performance. Performance improved when TOPKAT and CASE Ultra were combined, but similar to Rorije et al. (2013), not for all chemicals a prediction could be made. It was only possible to get a prediction for 87 of the 100 chemicals. For these 87 chemicals, the combination of models could correctly predict 69 % of the skin sensitizers and 79 % of the non-sensitizers. In general, this evaluation showed that the mechanism-based *in silico* models had a better performance than the statistical models. The best performance was achieved by TIMES-SS, but only to predict skin sensitization hazard. The *in silico* models were not able to provide an accurate prediction of potency (Teubner et al. 2013).

In chemico assays that address the MIE: protein binding

The ability of substances to bind covalently to proteins is the MIE of the AOP for skin sensitization. The majority of haptens is electrophilic or will become electrophilic after conversion and will react strongly with nucleophiles such as glutathione or the amino acids cysteine and lysine. As described in “Dermal bioavailability: skin absorption and metabolism” section, some *in silico* tools can predict protein reactivity. Next to these *in silico* tools, haptenization can also be measured using *in chemico* peptide reactivity assays. Different methods to measure reactivity exist; some measure the kinetics of reactivity, whereas others quantify the extent of peptide depletion after a fixed incubation time or measure adduct formation (reviewed in Roberts et al. (2008).

The DPRA assay is the most advanced test method to assess protein reactivity in terms of validation and regulatory acceptance. The DPRA measures depletion of cysteine and lysine with high-performance liquid chromatography (HPLC). The per cent depletion of cysteine and lysine is used in a prediction model to discriminate between skin sensitizers and non-sensitizers. The DPRA categorizes chemicals in four reactivity categories. Although this information cannot be used in itself for potency prediction, DPRA data may contribute to assessment of potency when it is used in combination with other test methods (OECD 2015a). The EURL-ECVAM-coordinated validation demonstrated that this assay was transferable to other laboratories experienced in analytical chemistry (EURL-ECVAM 2013a). The accuracy to discriminate skin sensitizers from non-sensitizers was 80 % for a dataset of 157 substances, with a sensitivity of 80 % and specificity of 77 % when

Table 1 Overview of a selection of in silico methods to identify skin sensitizers

Tool	Description	References
Toxtree (free)	This software classifies a substance into its mechanistic reactivity domains based on structural alerts. The tool allows a user to group chemicals according to their mechanistic category. Data can be used for read across to fill any data gap within the category. Toxtree is freely available http://toxtree.sourceforge.net/skinsensitisation.html	(Aptula and Roberts 2006; Enoch et al. 2008)
OECD QSAR Toolbox (free)	This toolbox predicts skin sensitization potential and potency via read across. Substances are grouped according to chemical categories based on similar structural alerts using 10 profilers. The formation of a chemical category is a key step in this toolbox. The toolbox consists of different modules for skin sensitization, including protein-binding profilers that are based on the SMART rules of Enoch et al. (2008). Databases with existing skin sensitization data, mainly from the LLNA and GPMT, are included, and it is possible to fill the data gaps for a chemical when there are measured data for one or more chemicals within the same category. The Toolbox is freely available: http://www.qsartoolbox.org/	(Enoch et al. 2008; OECD 2007)
VEGA (free)	Global QSAR model developed within EU-funded CAESAR project. This statistical tool incorporates a model developed using an Adaptive Fuzzy Partition (AFP) algorithm based on eight descriptors. The software uses atom-centred fragments as descriptors in a multivariate statistical model. The prediction of the model is active or inactive, together with applicability domain information. Substances outside the applicability domain cannot be predicted. Open-source software is available at http://www.vega-qsar.eu	(Chaudhry et al. 2010)
CASE Ultra (commercial)	The CASE models are based on substructure fragments linked to biological activity. Probability assessment is used to determine whether a structural fragment is associated with toxicity. There are 20 models (CASE Ultra) for skin sensitization. Two of those models are commercially available at www.multicase.com/case-ultra-models . One of the skin sensitization CASE models is included in the (Q)SAR database of the Danish Environmental Protection Agency (http://qsar.food.dtu.dk/)	(Gealy et al. 1996; Johnson et al. 1997; Klopman et al. 2005)
Derek Nexus (commercial)	Derek Nexus is a knowledge-based expert system that has an extensive rule base able to identify skin sensitizers. The skin sensitization knowledge base in this tool was initially developed using a historical database of GPMT data. Derek underwent extensive improvements as more skin sensitization (LLNA) data became available. The most recent version contains about seventy alerts for skin sensitization and some alerts for photoallergenicity. The model only identifies skin sensitizers and is not able to identify non-sensitizers. Derek nexus predicts the likelihood that a substance is a skin sensitizer. The alerts are: certain, probable, plausible or equivocal. This system is commercially available at http://www.lhasalimited.org/products/derek-nexus.htm	(Barratt et al. 1994a; Barratt et al. 1994b; Langton et al. 2006)
TIMES-SS (commercial)	TIMES (Tissue Metabolism Simulator) is a hybrid expert system, which has a skin sensitization module (SS). TIMES-SS encodes structure–toxicity and structure–skin metabolism relationships through a number of transformations. The commercial software is available at http://oasis-lmc.org/products/software/times.aspx	(Dimitrov et al. 2005)
TOPKAT (commercial)	This software tool is a suite of two models: one to distinguish skin sensitizers from non-sensitizers and one to sub-categorize the skin sensitizers according to their potency. The tool is based on data from guinea pig tests. Only chemicals that are within the applicability domain of the model can be predicted. The first model calculates the probability of a chemical structure to be a skin sensitizer. Substances with a probability of ≥ 0.7 are considered skin sensitizers and those with a probability of ≤ 0.3 non-sensitizers. In the second model, the potency is estimated; skin sensitizers with a probability of ≥ 0.7 are categorized as strong sensitizers and those with a probability of ≤ 0.3 as weak or moderate ones. Probability values between 0.3 and 0.7 are considered equivocal. The commercial tool is available at http://accelrys.com/solutions/scientific-need/predictive-toxicology.html	(Enslein et al. 1997)

compared to LLNA results (Gerberick et al. 2004; Gerberick et al. 2007). The DPRA was accepted as an OECD test guideline (TG442C) in 2015 (OECD 2015a). An important limitation of the DPRA is that it has no metabolic system and therefore cannot measure the reactivity of prohaptens. To overcome this, the DPRA was modified by adding a peroxidase as metabolizing enzyme, the Peroxidase Peptide Reactivity Assay (PPRA). This assay turned out to be

a promising assay to measure peptide reactivity of prohaptens (Gerberick et al. 2009), but has not been further validated yet.

Since our previous review (Vandebriel and van Loveren 2010), only a few new methods have been published that address protein binding. The assays already described in our previous review have not been further standardized or validated. It is of interest to mention briefly the

Cor1C420 assay, because this assay is used in one of the testing strategies (Natsch et al. 2015) that are described in our current review. This assay is a modification of the DPRA that uses liquid chromatography–mass spectrometry (LC–MS) analysis instead of HPLC. This enables measurement of adduct formation, making it possible to differentiate between actual adduct formation and auto-dimerization, and limits the number of false positives. An additional time point (150 min) was included to be able to assess the kinetics of peptide binding. The Cor1C420 provides a more detailed picture of peptide reactivity, resulting in a slightly improved predictivity compared to the DPRA (Natsch and Gfeller 2008).

The DPRA is performed with high concentrations of test chemicals, and this can be problematic for hydrophobic test chemicals. To solve this problem, the Amino acid Derivative Reactivity Assay Dilution Method (ADRA-DM) was developed that allows the use of low concentrations of test chemicals to measure reactivity to lysine or cysteine by HPLC. The performance of this assay was evaluated with a dataset consisting of 53 skin sensitizers and 29 non-sensitizers. The accuracy, sensitivity and specificity of the ADRA-DM were 90, 87 and 97 % compared to the LLNA (Yamamoto et al. 2015). A novel high-throughput screening (HTS) method addressed protein binding by measuring adduct formation rather than quantifying unreacted peptides. The HTS-DCYA (Dansyl CYsteamine Assay) quantifies adduct formations between the test chemical and a fluorescent cysteamine derivative. The predictive performance was assessed by testing 28 skin sensitizers and 8 non-sensitizers, based on the LLNA. The accuracy, sensitivity and specificity of this assay were 82, 78 and 90 %, respectively. The accuracy, sensitivity and specificity of the DPRA for the same dataset were 93, 100 and 70 % (Avonto et al. 2015). The allergen–protein interaction assay (APIA) uses proteomic technologies to obtain quantitative information on peptide depletion and qualitative mass spectrometry data on cysteine modifications. Only two skin sensitizers and one non-sensitizer were tested in this assay. It was shown that the skin sensitizers induced an exclusive thiol-dependent cysteine modification, whereas the non-sensitizer did not interact with any amino acid (Dietz et al. 2013).

In vitro assays that address KE2: keratinocyte activation

The second key event in the AOP of skin sensitization is activation of keratinocytes that results in inflammatory responses and activation of antioxidant response element (ARE)-dependent pathways (OECD 2012). In vitro keratinocyte assays for skin sensitization have different read-outs. The inflammatory cytokine IL-18 is often measured, because it is induced in keratinocytes after exposure

to skin sensitizers (Galbiati et al. 2011; Van Och et al. 2005). Measurement of activation of ARE-dependent pathways, such as the Keap1 (Kelch-associated protein 1)-Nrf2 (nuclear factor erythroid 2-related factor 2) pathway, is often used as well. Keap1-Nrf2 is the major regulator of cyto-protective responses to oxidative and electrophilic stress and is activated after exposure to skin sensitizers (Natsch 2010). Both human keratinocyte cell lines and reconstructed human epidermis (RhE) models are used to develop predictive skin sensitization assays. RhE models have the advantage that low solubility and hydrophobic substances can be applied, and finished products like creams and lotions can also be tested. RhE models are commercially available, e.g. SkinEthic (RhE, Lyon, France) and MatTek (Epiderm[®], Ashland, Massachusetts), but in-house models are used as well (Gibbs et al. 2013).

Reporter assays measuring Keap1-Nrf2 pathway activation

Both the KeratinoSensTM assay and the LuSens assay are based on activation of the Keap1-Nrf2 pathway. In the KeratinoSensTM assay, the cell line contains a luciferase gene under the transcriptional control of a constitutive promotor fused with an ARE element from the human AKR1C2 gene that is upregulated by skin sensitizers, resulting in activation of the luciferase gene. Subsequently, the luciferase signal can be quantitatively measured using luminescence detection. Test chemicals are considered positive if they induce a statistically significant induction of the luciferase activity >1.5-fold compared to the control at a concentration below 1000 μ M at a cellular viability >70 % (OECD 2015c). EURL-ECVAM concluded that the KeratinoSensTM is transferable to other laboratories and has a within- and between-laboratory reproducibility of 85 % (EURL-ECVAM 2013b). In 2015, the KeratinoSensTM assay was accepted as an OECD test guideline (TG 442D) (OECD 2015c). The accuracy of the KeratinoSensTM assay to identify correctly skin sensitizers has been tested thoroughly, showing an accuracy that varies from 75 to 96 %, dependent on the dataset used (Table 2). Overall, the accuracy of the KeratinoSensTM is higher when compared to human data than compared to LLNA data (Bauch et al. 2012; Urbisch et al. 2015).

In the LuSens assay, the luciferase gene is under the control of an ARE element from the rat NADPH Quinone oxidoreductase (NQO1), instead of the human AKR1C2 in the KeratinoSensTM assay. In the LuSens, a test compound is considered to have sensitizing potential when the luciferase induction is statistically significant $\geq$ 1.5-fold compared to the vehicle control in at least 2 of the tested concentrations. The cell viability at these concentrations needs to be above 70 %. The accuracy of the LuSens assay is similar to the KeratinoSensTM and ranges between 71 and 85 %

Table 2 Predictivity of in vitro assays that address KE2 to identify skin sensitization potential

Type of read-outs	Assay	Compared to	Sensitivity	Specificity	Accuracy (%)	N	References
Keap1-Nrf2 reporter cell lines	KeratinoSens™	Human	86 %	73 %	80	54	(Bauch et al. 2012)
		LLNA	83 %	76 %	81	54	
		Human	82 %	84 %	82	102	(Urbisch et al. 2015)
		LLNA	74 %	73 %	74	103	
		LLNA	79 %	80 %	80	127	(EURL-ECVAM 2014)
		LLNA	84–95 %	78–100 %	86–96	28	(Natsch et al. 2011)
		LLNA	86 %	83 %	85	67	
		LLNA	79 %	72 %	77	145	(Natsch et al. 2013)
	LuSens	LLNA	83 %	88 %	85	20	(Ramirez et al. 2016)
		Human	78 %	78 %	79	60	(Urbisch et al. 2015)
		LLNA	73 %	70 %	71	62	
		LLNA	73 %	74 %	74	74	(Ramirez et al. 2014)
		Human	83 %	81 %	83	69	
		LLNA	92 %	75 %	85	20	(Ramirez et al. 2016)
		Human	89 %	77 %	84	54	(Bauch et al. 2012)
		LLNA	81 %	71 %	77	54	
Gene expression	HaCaT	Human	75 %	71 %	74	19	(Van der Veen et al. 2013)
		Human	100 %	71 %	86	41	(Van der Veen et al. 2014a)
		Human	100 %	71 %	86	41	(McKim et al. 2010)
		LLNA	81 %	92 %	84	58	
	RhE	Not specified	100 %	100 %	100	7	(McKim et al. 2012)
		LLNA and human	–	–	95	19	(Gibbs et al. 2013)
		LLNA	72 %	100 %	80	10	(Coleman et al. 2015)
		LLNA	100 %	100 %	100	10	(Cottrez et al. 2015)
		LLNA	98 %	95 %	97	150	(Cottrez et al. 2016)
		Human	96 %	97 %	96	130	
LLNA	–	–	88–100	16	(Saito et al. 2013)		
IL-18 production	NCTC2544	LLNA	100 %	94 %	97	33	(Corsini et al. 2013)
	HaCaT	Human	23 %	92 %	47	34	(Van der Veen et al. 2014a, b)
	RhE	LLNA and human	–	–	95	19	(Gibbs et al. 2013)

(see Table 2). The inter-laboratory validation of the LuSens was performed in five laboratories. The transferability of the assay to other laboratories was >75 %, while the reproducibility both within and between laboratories was 100 % (Ramirez et al. 2016). The validation study of the LuSens assay has been evaluated by EURL-ECVAM, who considered that ‘this me-too assay would qualify to progress into the independent peer-review process once the test submitter provides the clarifications requested on some aspects of the submission’ (EURL-ECVAM 2015b).

Limitations of the KeratinoSens™ and LuSens assay are related to solubility and stability of the test substance in the vehicle. For example, hydrophobic substances do not dissolve and hydrolysis can occur, resulting in the cleavage of chemical bonds by water, which excludes this type of substances for testing in the KeratinoSens™ and LuSens assays.

In vitro keratinocyte assays measuring gene expression

The response of keratinocytes after contact with skin sensitizers can be measured at the gene expression level as well, to elucidate molecular pathways that are regulated by skin sensitizers and to identify predictive biomarker genes. Van der Veen et al. (2013) used the results of a DNA microarray to find a gene signature of 10 genes that could distinguish skin sensitizers from non-sensitizers with a high accuracy, using the human keratinocyte cell line HaCaT (Van der Veen et al. 2013). This gene signature includes genes that play a role in the oxidative stress response, including Keap1-Nrf2 genes, and genes related to the inflammatory response. The predictivity of this gene signature has been evaluated using an independent dataset of 41 chemicals, including substances that were misclassified in the LLNA. The results showed the gene signature had an accuracy of

86 % compared to human data, whereas the LLNA had an accuracy of 78 % for the same dataset (Van der Veen et al. 2014a). Although promising, the performance is based on a relatively small dataset and more chemicals need to be tested to evaluate the predictive capacity of this gene signature.

Another study evaluated the predictivity to identify skin sensitizers by measuring 11 genes in the HaCaT cell line (McKim et al. 2010). The genes are based on three signaling pathways associated with sensitization: genes under the control of Keap1-Nrf2, CYP1A and metallothionein 1 and 2 were measured. They developed an algorithm based on viability, reactivity and gene expression to identify skin sensitizers. Using 41 test substances, this assay had an accuracy, sensitivity and specificity of 84, 100 and 71 %, respectively, compared to human data (McKim et al. 2010).

Gene expression has been used as a read-out as well in RhE models. The EpiSensA assay uses the MatTek EpiDerm® RhE model and measures the expression of five genes to identify skin sensitizers (Saito et al. 2013). Of these five genes, three genes showed a high predictive accuracy of 100 % (ATF3), 94 % (DNAJB4) and 88 % (GCLM) compared to the LLNA. Coleman et al. (2015) used the MatTek EpiDerm® model to determine the skin sensitization potential of chemical substances. They developed an algorithm based on glutathione reactivity, viability and fold change in gene expression induced by chemical exposure. Expression of 15 different genes controlled by the Keap1-Nrf2 pathway was measured. The dataset of ten substances is too low to provide reliable information on the accuracy of this assay (Coleman et al. 2015). McKim et al. (2012) used both the MatTek EpiDerm® and the SkinEthic RhE to identify skin sensitizers by measuring 11 target genes involved in the oxidative stress, metal and xenobiotic response. They tested seven chemicals that are used in medical devices, and all were classified correctly as either sensitizer or non-sensitizer (McKim et al. 2012). Again, the dataset is too small to conclude on the predictivity of this assay.

The SENS-IS assay determines both the sensitizing potential and potency of chemicals using a panel of genes on an EpiSkin® RhE model. During development of this assay, a panel of 24 genes was selected from the initial set of 200 genes involved in the antioxidant and inflammatory responses. Using the SENS-IS assay, 50 substances were all correctly classified as either sensitizer or non-sensitizer (Cottrez et al. 2015). In a subsequent test, two panels of genes were used to predict the skin-sensitizing potential and potency of chemicals: 17 genes that contain Keap1-Nrf2- genes (REDOX group) and 21 genes involved in inflammation, danger signals and cell migration (SENS-IS group). Using this prediction model, 125 out of 130 substances were correctly classified (Cottrez et al. 2016). In

a ring study in three laboratories using 19 substances, the reproducibility of the SENS-IS assay was assessed and the assay had a within-laboratory reproducibility of 95 % and a between-laboratory reproducibility of 100 %. The SENS-IS assay is currently under evaluation at EURL-ECVAM and EURL-ECVAM recommended to explore the impact of the patent that covers this assay and possible consequences for future development as an OECD TG (EURL-ECVAM 2015b).

In vitro keratinocyte assays measuring IL-18 production

Both the keratinocyte cell line NCTC2544 and primary keratinocytes are able to produce IL-18 in a dose-dependent fashion after exposure to skin sensitizers, whereas irritants and respiratory sensitizers did not (Corsini et al. 2009). Next to NCTC2544, the keratinocyte cell lines HPKII and HaCaT selectively produce IL-18 in response to skin sensitizers (Galbiati et al. 2011). However, the IL-18 assay in NCTC2544 cells could not easily be reproduced in different laboratories, as shown in a transferability study (Teunis et al. 2013). This was supported by a study, demonstrating that IL-18 production by HaCaT cells was not an accurate read-out for identifying skin sensitizers; in HaCaT cells, only five out of the 22 tested skin sensitizers were correctly classified (Van der Veen et al. 2014a). IL-18 production in RhE models can also be used to discriminate between skin sensitizers and non-sensitizers (Gibbs et al. 2013). The accuracy of IL-18 production to identify sensitizers is with 95–97 % quite high, but based on a small dataset of 19 substances (Corsini et al. 2013; Gibbs et al. 2013). Table 2 shows the performance of keratinocyte assays measuring IL-18 production to predict skin sensitization.

In vitro keratinocyte assays that predict skin sensitization potency

Some of the keratinocyte-based models described above claim to be able to predict both the skin-sensitizing potential and potency of substances (Table 3). McKim et al. (2012) tested seven chemicals for their sensitizing potency by determining the magnitude of gene induction in MatTek EpiDerm® and the SkinEthic RhE. They could correctly predict the sensitizing potency category (extreme, strong, moderate and non-sensitizer) for five out of seven of the chemicals (accuracy of 71 %) (McKim et al. 2012). Coleman et al. (2015) developed an algorithm based on re-analysis of the data from McKim et al. (2010) and combined this with information on viability, GSH depletion and in silico data to determine sensitization potency. This method proved to provide a poor prediction of potency categories. The accuracy for the ten tested substances was only 40 % compared to LLNA EC3 data (Coleman et al. 2015). In the

Table 3 Accuracy of in vitro assays that address KE2 to identify skin sensitization potency

Model	Read-out	Compared to	Accuracy (%)	N	References
MatTek Epiderm [®] and SkinEthic [®] RhE	Gene expression	Not specified	71	7	McKim et al. (2012)
MatTek Epiderm [®]	Gene expression	LLNA	40	10	Coleman et al. (2015)
Episkin [®]	Gene expression	LLNA	96	50	Cottrez et al. (2015)
RhE in-house model	Viability and IL-1 α production	LLNA	70–72	12	Dos Santos et al. (2011)
		Human	62–71	12	

SENS-IS assay, the number of regulated genes was used to predict sensitization potency: skin sensitizers were categorized in four potency categories (extreme, strong, moderate and weak). The assay could predict the potency categories with an accuracy of 96 % compared to LLNA data using 50 substances (Cottrez et al. 2015).

In the epidermal equivalent assay, using an in-house RhE model based on primary keratinocytes the concentration that induces 50 % reduction in viability correlated with sensitizer potency. In addition, the concentration that induced a tenfold increase in IL-1 α release also correlated with potency. The potency predicted in this epidermal equivalent assay correlated well with LLNA EC3 values for seven of the 10 tested skin sensitizers (Dos Santos et al. 2011). Limitations of the epidermal equivalent potency assay are that it can only be used for sensitizers that have cytotoxic properties. Furthermore, the assay itself can only be used for potency prediction, but it cannot discriminate between sensitizers and non-sensitizers. It was proposed to use this assay in a tiered strategy to estimate the potency of skin sensitizers identified in another in vitro assay (Teunis et al. 2013). The assay was successfully transferable between laboratories (Teunis et al. 2013).

In vitro assays that address KE3: dendritic cell activation

Dendritic and Langerhans cell maturation and migration is a key step in the skin immune response to skin sensitizers. Because Langerhans cells are only available in low numbers and sometimes spontaneously mature, DCs of human or sometimes mouse origin are used instead (Casati et al. 2005). Human DCs are cultured either from CD14⁺ peripheral monocytes (moDCs) or from CD34⁺ cord blood cells (CD34⁺ DCs). Maturation is measured by expression of cell surface markers such as CD40, CD80, CD83, CD86, and human leucocyte antigen (HLA)-DR, and production of cytokines such as IL-6, IL-8, IL-12p40, and TNF- α . In addition to cell surface marker expression and cytokine production, signal transduction is a parameter that is changed upon sensitizer exposure. The use of these primary human cells has its limitations, such as limited reproducibility within and between laboratories due to inter-donor

variability and variations in cell isolation and culture techniques (Casati et al. 2005). To circumvent this inter-donor variability, cell lines such as THP-1, U937, KG-1, and MUTZ-3 have been used as surrogate models for DCs (Galvão dos Santos et al. 2009).

In vitro DC assays measuring cell surface markers

The h-CLAT is based on CD54 and CD86 expression by the human THP-1 monocytic cell line (Ashikaga et al. 2010; Sakaguchi et al. 2006; Sakaguchi et al. 2010). This assay has been validated by EURL-ECVAM and proved transferable to other laboratories experienced in cell culture techniques and flow cytometry analysis. The reproducibility of within and between laboratories was around 80 % and did not meet the predefined reproducibility target of 85 % for within-laboratory reproducibility (EURL-ECVAM 2015a). In the OECD test guideline, criteria are included that enable monitoring of the reproducibility of this assay within laboratories performing it routinely, including demonstration of technical proficiency and maintenance of a historical database of results.

The performance of the h-CLAT was assessed in several studies. Using a set of 66 chemicals, the h-CLAT had an accuracy of 83 %, a sensitivity of 83 % and a specificity of 67 % compared to human data (Nukada et al. 2011). Chemicals with a low-to-moderate skin sensitization potency more often gave false-negative predictions in the h-CLAT than chemicals showing high skin sensitization potency. In addition, water-insoluble chemicals tended to produce false-negative results. Using a set of 100 chemicals, the h-CLAT had an accuracy, sensitivity and specificity of 84, 88 and 75 %, respectively, compared to the LLNA. Eight out of the nine false negatives were water-insoluble chemicals (Ashikaga et al. 2010). A recent study confirmed that chemicals with an octanol–water partition coefficient (log Kow) greater than 3.5 tend to produce false-negative results. The h-CLAT had an accuracy of 80 %, a sensitivity of 83 % and a specificity of 71 % for a dataset of 143 chemicals. When all substances with a log Kow greater than 3.5 were excluded, the performance of the h-CLAT improved. Accuracy, sensitivity and specificity were 88, 94 and 74 %, respectively (Takenouchi et al. 2013).

A similar DC assay that measures (only) CD86 expression and employs the U937 cell line is the U-SENS, formerly known as the MUSST (myeloid U937 skin sensitization test). This assay uses a cut-off of 1.5-fold induction of CD86 expression to identify skin sensitizers. A multicentre study with 24 chemicals showed a predictivity of 94 % (Alépée et al. 2015). Using a set of 175 chemicals, the assay showed 88 % accuracy, 90 % sensitivity and 79 % specificity compared to human and LLNA data (Pirouard et al. 2015). Using a subset of 92 substances for which both human and LLNA data were available, the predictive capacity was comparable to the set of 175 chemicals, with 85 % accuracy, 90 % sensitivity and 60 % specificity with regard to human data and 84 % accuracy, 92 % sensitivity and 55 % specificity with regard to the LLNA. The results of the validation study were submitted to EURL-ECVAM, and the method is currently being peer-reviewed. Furthermore, a project proposal for the development of a Test Guideline on the U-SENS™ test method was submitted by France to the OECD and is included in the OECD work program in 2015 (EURL-ECVAM 2015c).

The modified MUSST (mMUSST) uses the similar cell line and read-out, but a different cut-off value. A 20 % increase in CD86 expression is taken as being indicative of sensitization hazard. The mMUSST correctly predicted 43 out of 50 or 39 out of 53 substances, compared to human or LLNA data, respectively. It offered an overall accuracy of 86 or 74 %, a sensitivity of 75 or 64 % and a specificity of 100 or 94 %, when compared to human or LLNA data, respectively (Bauch et al. 2011). Using a set of 105 test chemicals, a common outcome between h-CLAT and MUSST was observed for (only) 76 chemicals (Urbisch et al. 2015). The h-CLAT showed an accuracy of 86 or 80 % and the (m)MUSST of 78 or 75 %, when comparing the predictions to human or LLNA data, respectively.

The SH (thiol) test measures changes in cell surface thiols in THP-1 cells (Suzuki et al. 2009). When as a cut-off an increase or decrease of 15 % in surface thiols was taken, a 90 % concordance was found between the effects on surface thiols and in vivo data (guinea pig maximization test and Buhler test, LLNA, and human maximization test) using a set of 36 sensitizers and 16 non-sensitizers. All sensitizers tested decreased cell surface thiols on moDC (Kagatani et al. 2010). The mechanistic relevance of a decrease in cell surface thiols in skin sensitization induction was demonstrated by treatment with a thiol oxidizer that decreased cell surface thiols. As a result, phosphorylation of p38 MAPK, a process crucial in DC maturation by sensitizers, and DC maturation were induced. The increase in CD86 expression after treatment with a membrane-impermeable thiol blocker provided additional mechanistic evidence for a role of cell surface thiols in DC maturation (Hirota et al. 2009).

Only one assay based on primary monocytes has been validated in a ring study. This assay is based on moDCs; in this model, induction of CD86 expression is measured together with viability assessment. A ring study with five laboratories using five skin sensitizers and two non-sensitizers confirmed the transferability of the protocol. However, SDS, a non-sensitizer, was correctly predicted in only two laboratories. The skin sensitizers eugenol and hexylcinnamaldehyde were not identified as such in all participating laboratories (Reuter et al. 2015).

Besides assays based solely on keratinocytes or dendritic cells, a co-culture model has been described as well. The ‘Loose-fit Coculture-based Sensitization Assay’ (LCSA) consists of a co-culture of primary human keratinocytes (KC) and human moDC. This assay uses CD86 induction as a read-out and could identify the three skin sensitizers tested (Schreiner et al. 2007). Subsequent studies were done to establish whether LCSA data could be used to estimate potency. This was investigated with a small set of ten skin sensitizers, but results were not convincing (Frombach et al. 2016; Sonnenburg et al. 2015).

In vitro dendritic cell assays measuring gene expression

One approach to improving current DC maturation assays is by finding additional maturation markers, and microarray analysis has been used for this purpose. The Genomic Allergen Rapid Detection (GARD) assay is based on gene profiling of the response of the human dendritic MUTZ-3 cell line to allergens (Johansson et al. 2012; Johansson et al. 2011). Discrimination between sensitizers and non-sensitizers is based on a set of 200 genes, and a support vector machine for supervised classification. The markers identified could be linked to immunologically relevant pathways; the ones most highly populated involved the Keap1-Nrf2 pathway, xenobiotic metabolism signalling, protein ubiquitination, LPS/IL-1-mediated inhibition of RXR function, aryl hydrocarbon receptor signalling and protein kinase A signalling. In a follow-up study, it was found that molecular pathways for metabolic processes, cell cycling and oxidative stress responses were regulated by skin sensitizers (Albrekt et al. 2014). It was found that with increasing potency of the sensitizer, the chemical reactivity groups gradually regulated more pathways and more genes per pathway. Accuracy, sensitivity and specificity were found to be 89, 89 and 88 %, respectively, for a dataset of 37 chemicals (Johansson et al. 2014).

A classification model of 13 genes was devised using the data of 10 sensitizers and 11 non-sensitizers tested in CD34⁺ DC (Hooyberghs et al. 2008). Cross-validation showed an accuracy, sensitivity and specificity of 89, 82 and 97 %, respectively. The expression profiles of two genes, cAMP Responsive Element Modulator (CREM)

and CCR2, showed the highest discriminative potential between sensitizers and non-sensitizers. For some of the 13 genes, their functional relevance was established as well as their functional relevance in DC maturation (Lambrechts et al. 2010a, 2011). In an extension of these studies, the use of the classification model was tested for potency evaluation. For a set of 15 skin sensitizers, it was shown that the concentration that caused 20 % cell damage (IC₂₀) had the best correlation with the LLNA EC₃. A linear regression model was built based on the IC₂₀ and expression changes in CREM and CCR2. It was shown that after removal of two outliers, this model had a good correlation for 13 skin sensitizers (Lambrechts et al. 2010b).

In vitro assay measuring IL-8 induction

The IL-8 Luc assay is based on the THP-G8 cell line, a THP-1-derived IL-8 reporter cell line. The cell line contains two luciferase genes, one under the control of the IL-8 promoter and one under the control of the G3PDH promoter. The ratio of induction between the two luciferase genes is measured and divided by that of untreated cells. This ratio was increased by 13 out of 15 sensitizers and 5 out of 7 non-sensitizers tested. Using cut-off values of ≥ 1.4 for induction and ≤ 0.8 for inhibition, the test accuracy was 82 %, the sensitivity 87 % and the specificity 72 % (Takahashi et al. 2011). The IL-8 Luc assay has been validated in a ring study, and the independent peer review coordinated by the Japanese Centre for Validation of Alternative Methods (JacVAM) is finalized. An OECD test guideline is in preparation (EURL-ECVAM 2015c).

In vitro assays that address KE4: T cell activation

The development of T cell-based in vitro assays turned out to be complex and challenging, and no standardized or validated assays are available yet (Martin et al. 2010). One of the reasons is that it is difficult to mimic in vitro the tightly organized process of in vivo T cell priming that involves complex molecular and cellular processes in both the skin and lymph nodes. The most commonly used in vitro method is the human T cell priming assay (hTCPA). In this assay, the test chemical itself or conjugated to a protein carrier is added to a co-culture of immature monocyte-derived DCs with human peripheral blood leucocytes as a source for naïve T cells. After 5 and 10 days, T cell proliferation and cytokine production are measured.

The most important drawbacks of this assay are summarized by Martin et al. (2010). One important drawback is the high variability of the assay caused by the donor-to-donor variations in T cell repertoire between blood donors. This makes it difficult to standardize the assay. On the other hand, improvements in protocol of the assay, such as

the removal of regulatory T cells, increased the sensitivity of the assay, especially for weak skin sensitizers (Vocanson et al. 2014). Recently, efforts were made to further standardize the hTCPA by developing a standard operation procedure (SOP) (Richter et al. 2013). Although some progress is made, standardized in vitro T cell-based assays will not be available in the near future.

Testing strategies for skin sensitization

The concept of strategies that combine existing data with experimental assays and computational tools is not new (Hartung et al. 2013; Rorije et al. 2013; Rovida et al. 2015). Different terminologies are used for approaches that combine different methods, e.g. integrated testing strategies (ITS), sequential testing strategies (STS) and integrated approaches to testing and assessment (IATA). According to the OECD, ITS and STS are so-called defined approaches to testing and assessment, which use a fixed data interpretation procedure (DIP) that is applied to data generated with a defined set of individual information sources. In contrast, an IATA follows an iterative process of information collection and the data gathered targets the generation of new data, when needed. The decision-making in an IATA is based on a weight-of-evidence approach (OECD 2008; Patlewicz et al. 2014; Tollefsen et al. 2014).

Recently, the OECD has developed two guidance documents on defined approaches to testing and assessment. The first is a general document that provides a set of principles for reporting defined approaches to testing and assessment and how these can be used within an IATA (OECD 2016b). The second document specifically focuses on defined approaches and individual information sources that are available for testing and assessment of skin sensitization and presents a number of case studies illustrating several of these defined approaches (OECD 2016a). In case of skin sensitization, the majority of the testing strategies are defined approaches, and up to now, only one IATA has been proposed (Patlewicz et al. 2014). Some of the strategies provide information that can be used for hazard identification, e.g. distinguishing skin sensitizers from non-sensitizers (summarized in Table 4), while other approaches provide information that allow potency sub-categorization as well (summarized in Table 5). The different defined approaches for skin sensitization are constructed either as STS or as ITS (OECD 2016b). In an STS, the information is collected in a fixed stepwise approach and involves interim decision steps. After each step, a prediction model (DIP) is applied to reach a (interim) decision whether more information and subsequent testing is needed. When sufficient information is available at this step, a final prediction is made. A generic example of how an STS can be

Table 4 Summary of defined approaches to testing and assessment and IATA that can be used for hazard identification

Defined approach	Compared to	Sensitivity (%)	Specificity (%)	Accuracy (%)	N	Individual data sources	References
'2 out of 3' ITS	Human	93	95	94	54	DPRA, KeratinoSens™, LuSens, U-SENS, mMUSST, h-CLAT	Bauch et al. (2012)
	LLNA	81	88	83			
	LLNA	81	77	81	145	DPRA, KeratinoSens™, U937-CD86	Natsch et al. (2013)
RIVM STS	Human	90	90	90	101	DPRA, KeratinoSens™, h-CLAT, LuSens, U-SENS, mMUSST	Urbisch et al. (2015)
	LLNA	79	82	77	180		
	Human	96	93	95	41	Bayesian QSAR approach, DPRA, KeratinoSens™, HaCaT gene signature, h-CLAT	Van der Veen et al. (2014a, b)
IDS	LLNA	93	64	83	41		
	LLNA	95	100	96	26	OECD QSAR Toolbox, h-CLAT and physicochemical properties (octanol/water partition coefficient, water solubility, vapour pressure, Mw, melting point and boiling point)	Strickland et al. (2016)
	LLNA	97	96	97	94		
IATA	LLNA and guinea pig	86	89	88	100	Existing data (read across), substance-specific information (physical form, physical–chemical properties), QSARs for skin sensitization, protein reactivity, skin irritation and corrosion, mutagenicity	Patlewicz et al. (2014)
Decision tree ITS	Human	94	49	80	97	Derek Nexus 4.1.0, DPRA, LuSens, KeratinoSens, h-CLAT, U-SENS	Macmillan et al. (2016)
Potency prediction based on MIE and KE2	LLNA	95	54	85	183		
	LLNA	82	78	80	244	TIMES-SS, CorlC420 assay, KeratinoSens™	Natsch et al. (2015)
Kao STS	LLNA	98	51	85	139	DPRA, h-CLAT, DEREK Nexus	Takenouchi et al. (2015)
Score-based ITS	LLNA	89	70	84		DPRA, h-CLAT, DEREK Nexus	Takenouchi et al. (2015)
Artificial neural network	LLNA	100	61.5	90.6	53	SH test, ARE assay, h-CLAT, LogP	Tsujita-Inoue et al. (2014)
Bayesian ITS	LLNA	100	100	100	60	DPRA, h-CLAT, TIMES, KeratinoSens™, bioavailability parameters	Jaworska et al. (2015)
Probabilistic hazard assessment	LLNA	90	91	81	145	TIMES, Dragon, KeratinoSens™, DPRA, U937-CD86, glutathione depletion	Luechtefeld et al. (2015)

Table 5 Summary of defined approaches to testing and assessment that predict potency sub-categories

Defined approach	Compared to	Accuracy (%)	<i>N</i>	Individual data sources	References
Potency prediction based on KE1 and KE2	Human GHS class 1A	83	23	TIMES-SS Cor1C420 assay, KeratinoSens™	Natsch et al. (2015)
	Human GHS class 1B	83	29		
	LLNA GHS 1A	70	54		
	LLNA GHS 1A	77	79		
Kao STS	LLNA EC3 < 1 % (strong)	65	29	DPRA, h-CLAT,	Takenouchi et al. (2015)
	LLNA EC3 > 1 % (moderate/weak)	78	73	DEREK Nexus	
Score-based ITS	LLNA EC3 < 1 % (strong)	52	29	DPRA, h-CLAT,	Takenouchi et al. (2015)
	LLNA EC3 > 1 % (moderate/weak)	77	73	DEREK Nexus	
Bayesian ITS	LLNA negative	92	39	DPRA, h-CLAT, TIMES, KeratinoSens™, bioavailability parameters	Jaworska et al. (2015)
	LLNA weak	82	39		
	LLNA moderate	95	40		
	LLNA strong	72	29		
	LLNA negative	100	14		
	LLNA weak	89	19		
	LLNA moderate	75	12		
	LLNA strong	87	15		
Artificial neural network	LLNA EC3 < 1 % (strong)	92	13	SH test, ARE assay, h-CLAT, LogP	Tsujita-Inoue et al. (2014)
	LLNA EC3 > 1 % (moderate/weak)	72	39		
Probabilistic hazard assessment	LLNA negative	83	145	TIMES, Dragon, KeratinoSens™, DPRA, U937-CD86	Luechtefeld et al. (2015)
	LLNA moderate	55			
	LLNA strong	63			
	LLNA extreme	60			
	LLNA negative	65	89	TIMES, Dragon, KeratinoSens™, DPRA, U937-CD86	
	LLNA moderate	59	64	TIMES, Dragon, KeratinoSens™, DPRA, U937-CD86, glutathione depletion	
	LLNA strong	66			
	LLNA extreme	59			
	LLNA negative	73			
	LLNA moderate	69			
	LLNA strong	65			
	LLNA extreme	54			

constructed is illustrated in Fig. 2. An ITS is an approach that assesses information of multiple information sources at the same time by applying a variety of specific methodologies to convert inputs from the different information sources into a prediction. A variety of specific methodologies can be applied, such as statistical and mathematical models. Figure 3 shows a generic example of how an ITS can be constructed.

Defined approaches to testing and assessment and IATA to predict skin sensitization potential

The first ITS for skin sensitization was proposed by BASF (Bauch et al. 2012). Their initial approach was developed

using assays that reflect the first three key events of the AOP: the DPRA (KE1), the LuSens assay and the KeratinoSens™ (KE2) and the mMUSST and the h-CLAT (KE3). The preliminary prediction model was based on a dataset of 54 substances. Application of this prediction model resulted in an accuracy of 94 or 83 %, a sensitivity of 93 or 81 % and a specificity of 95 or 88 % compared to human or LLNA data, respectively (Bauch et al. 2012) (Table 4). This ITS further evolved after evaluating it with a larger dataset of 145 substances. This dataset contained 102 skin sensitizers and 43 non-sensitizers according to the LLNA. In this ITS, the DPRA (KE1), the KeratinoSens™ (KE2) and the U937-CD86 assay (KE3) were used. A simplified data interpretation procedure was used compared

Fig. 2 Generic example of a sequential testing strategy (STS). In an STS a substance is tested in a stepwise manner by performing a test method followed by an interim decision step in which it is decided if the information is sufficient for a prediction or that additional information is needed. In the latter case, a next test is performed. This figure is a modification of a figure published in (OECD 2016b)

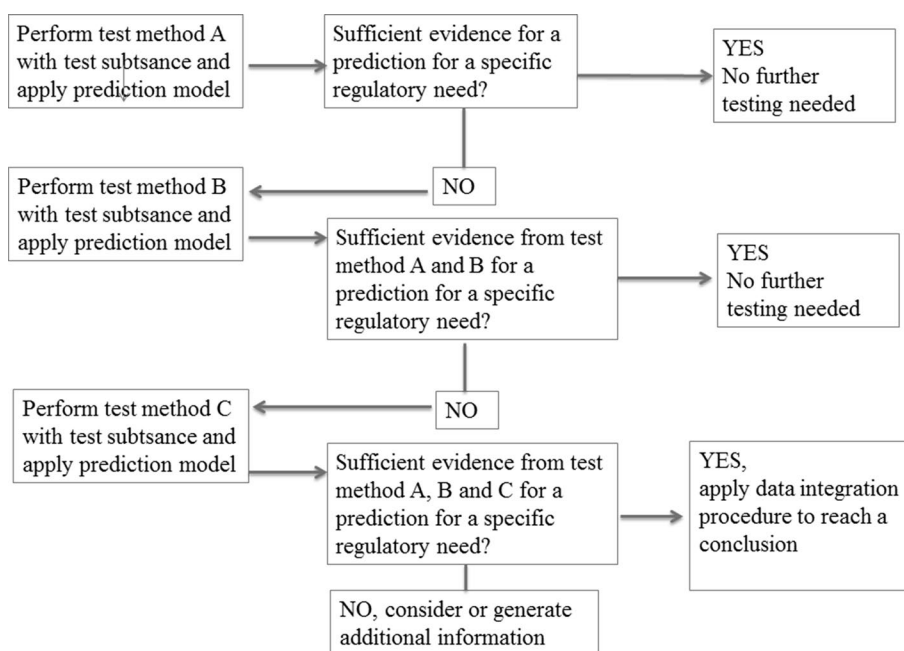
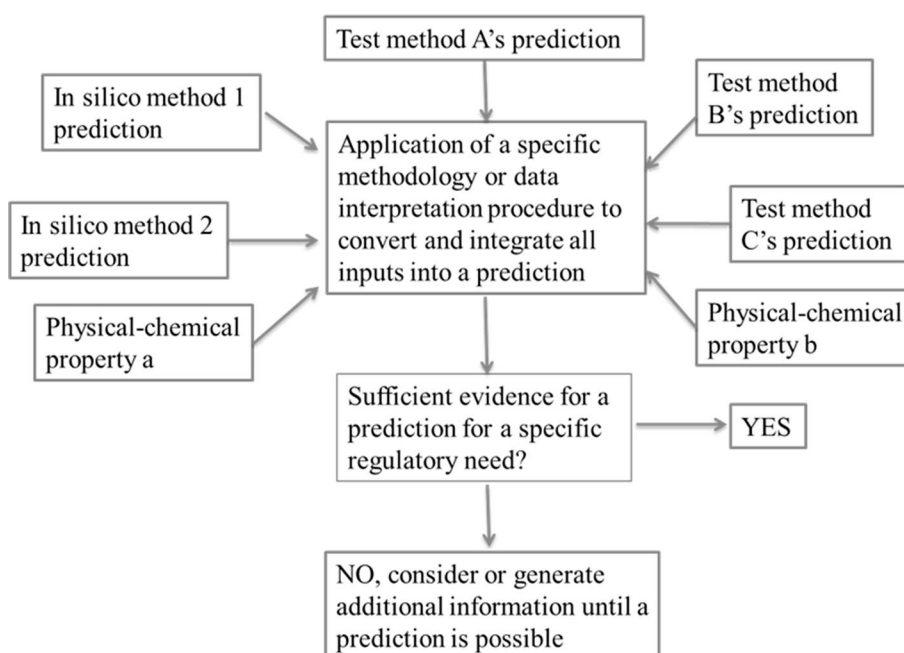


Fig. 3 Generic example of an integrated testing strategy (ITS). In an ITS all input from individual data sources are combined and integrated to reach a decision. Statistical tools or algorithms can be used to convert the input to prediction. This figure is a modification of a figure published in (OECD 2016b)



to the initial study of Bauch et al. (2012). If at least 2 out of 3 tests give a positive prediction, a substance is considered a skin sensitizer. In case at least 2 out of 3 tests give a negative prediction, a substance is considered a non-sensitizer. The accuracy of this ‘2 out of 3’ ITS compared to the LLNA is 81 %, while sensitivity and specificity are 82 and 77 %, respectively (Natsch et al. 2013).

More recently, this ‘2 out of 3’ ITS was further evaluated by testing its performance using a dataset of 213 substances; human data were available for 114 substances (Urbisch et al. 2015). The predictive capacity of the ITS

was evaluated using data from different test methods, including the DPRA (KE1), KeratinoSens™ and the LuSens assay (KE2), the h-CLAT, mMUSST and U-SENS assays (KE3). It is important to note that not all assays tested the complete dataset of 213 substances. The available information allowed a thorough performance analysis and comparison of the individual test methods. In addition, it was possible to analyse the predictive capacities of ITS approaches consisting of different assays. For example, an ITS combining data from the regulatory accepted test methods DPRA, KeratinoSens™ and h-CLAT resulted in

the highest accuracy of 90 %; sensitivity and specificity both were 90 % as well, compared to human data. This predictivity was more accurate compared to predictions based on the LLNA for the same dataset, for which the accuracy, sensitivity and specificity for the same dataset were 82, 91 and 64 %, respectively (Urbisch et al. 2015).

An STS developed by the RIVM was designed as a tiered approach that allows stepwise gathering of information on skin sensitization hazard (Van der Veen et al. 2014a). The RIVM STS includes computational models and experimental assays that inform on the MIE, KE2 and KE3. Decision criteria are included after each tier in order to enable targeted and efficient information gathering and to avoid redundant testing. Tier 1 starts with a Bayesian QSAR approach that combines predictions from four computational tools (MultiCASE, CAESAR, DEREK and OECD QSAR Toolbox) (Rorije et al. 2013). A substance is rated positive, negative or equivocal. Positive and negative substances proceed to Tier 2; equivocal substances are further tested in Tier 1 using the DPRA. Tier 2 consists of keratinocyte-based assays. Substances that are positive in Tier 1 are tested in the KeratinoSens™ assay, which has the lowest level of false-positive results. Negative substances are tested in the HaCaT gene signature (Van der Veen et al. 2013), which has the lowest number of false-negative results. When Tiers 1 and 2 give concordant results, no further testing is needed and the substance is considered a skin sensitizer if the two tiers give concordant positive results and vice versa for non-sensitizers. In case of discordant results in Tiers 1 and 2, further testing in Tier 3 is needed, in which the h-CLAT assay is performed. The final decision is based on the majority voting, i.e. two out of three tests need to give concordant results to reach a prediction. The performance of this STS was evaluated with a dataset of 27 human skin sensitizers and 14 non-sensitizing chemicals. The dataset included five substances that gave false-positive results in the LLNA and two substances that were false-negative results in the LLNA. This STS predicted human skin sensitizers with an accuracy, sensitivity and specificity of 95, 96 and 93 %, respectively (Van der Veen et al. 2014a). The performance of this STS needs to be further evaluated with a larger dataset.

Recently, the Interagency Coordinating Committee for the Validation of Alternative Methods (ICCVAM) published their integrated decision strategy (IDS) based on the AOP using existing data from *in vitro*, *in chemico* and *in silico* methods (Strickland et al. 2016). To develop this strategy, a dataset of 87 skin sensitizers and 33 non-sensitizers based on the LLNA was used. The IDS was developed using existing data from the DPRA, h-CLAT and KeratinoSens™, physicochemical properties that may affect skin sensitization and read-across predictions using the OECD QSAR Toolbox. To determine the most accurate IDS, the

machine learning approach support vector machine (SVM) was applied to a training set of 94 substances and subsequently tested with 26 substances. Data integration using SVM resulted in the highest accuracy when data from the h-CLAT, the OECD QSAR Toolbox and six physicochemical properties that affect skin permeability (octanol/water partition coefficient, water solubility, vapour pressure, melting point, boiling point, molecular weight) were used, both for the training and for the test sets. The accuracy, sensitivity and specificity for the test set were, respectively, 97, 96 and 96 % compared to the LLNA (Table 4).

Lhasa proposed an ITS based on a decision tree that includes their computational tool Derek Nexus combined with either one of two *in chemico*/*in vitro* assays that address MIE (DPRA), KE2 (KeratinoSens™, LuSens) or KE3 (h-CLAT or U-SENS) (Macmillan et al. 2016). A decision tree was constructed to identify skin sensitizers in a stepwise manner. First, a test substance is processed in Derek and assigned as in domain or outside domain of the *in chemico* and *in vitro* assays. Chemicals that have a ClogP below 3.5 and do not need prior activation are considered within domain. Test substances with an alert in Derek Nexus that are outside the domain of the assays are classified as skin sensitizers and not further tested. Test substances that are within domain are tested in one of the *in chemico* or *in vitro* assays. If that assay is concordant with Derek, a test substance is considered a skin sensitizer. In case of discordant results, a second *in vitro*/*in chemico* assay is performed. The overall call is based on the majority result in the two assays and DEREK. Test substances that do not give an alert in Derek that are within domain are tested according to the same procedure as described for those that give an alert. Test substances without an alert in Derek that are outside domain are considered equivocal. The performance of the Lhasa decision tree ITS was evaluated with a dataset of 183 substances for which LLNA data were available and a comparison to human data was made for a dataset of 97 substances. The decision tree ITS had an accuracy of 80 % compared to human data and 85 % compared to LLNA data. Sensitivities were 94 and 95 %, compared to human and LLNA, respectively. The specificity of this ITS was low, 49 % compared to human data and 54 % compared to the LLNA. Furthermore, no prediction is possible for 19 substances that do not have a Derek alert and are outside the domain of the *in chemico* and *in vitro* assays (Macmillan et al. 2016).

Patlewicz et al. (2014) developed an IATA for skin sensitization hazard using a workflow that incorporates physicochemical properties of chemicals, read-across and QSAR models in a weight-of-evidence (WoE) approach. In this IATA, the information sources used are not fixed, but the information that is collected depends on the structural alerts and the physical–chemical information that is

available for a substance. The WoE decisions are dependent on expert opinion. The first steps in this IATA involve information gathering on hazardous and physical chemical properties of the test chemical to determine whether further testing is relevant. If the answer to that is yes, existing skin sensitization (in vivo, in vitro, human) data is gathered and evaluated to see whether new data need to be generated. In case of structural analogues with experimental sensitization data, read across is done, otherwise, (Q)SAR models are applied. For example, TIMES-SS and the OECD Toolbox are used to obtain information on protein-binding alerts. Furthermore, QSARs that provide information on mutagenicity and skin irritation or corrosion are included in this IATA. The workflow of the IATA is quite complex, and it is not completely transparent how the final WoE decision is made. The workflow was transformed into a software tool that is incorporated in the OASIS pipeline of the OECD toolbox. The performance of this IATA was assessed with 45 skin sensitizers and 55 non-sensitizers based on animal (LLNA and guinea pig) data (Teubner et al. 2013). Three substances were flagged as inorganic and were not evaluated further. For the remaining substances, an overall accuracy of 88 %, a sensitivity of 89 % and a specificity of 86 % compared to LLNA data were obtained (Patlewicz et al. 2014).

Defined approaches to testing and assessment to predict skin sensitization potency sub-categories

Table 5 summarizes the defined approaches that estimate the potency sub-category of a skin sensitizer. Skin sensitizers are categorized according to their LLNA EC3 values. Skin sensitizers can be divided into two classes according to the GHS scheme: 1A ($EC3 \leq 2\%$) and 1B ($EC3 > 2\%$) or be divided in 3 classes: strong or extreme ($EC3 \leq 1\%$), moderate $EC3 > 1\text{--}\leq 10\%$, weak ($EC3 > 10\%$).

An approach based on the MIE and KE2 used quantitative dose–response information from the Cor1C420 assay that measures kinetic peptide reactivity and the KeratinoSens™ to predict potency (Natsch et al. 2015). Multiple regression analysis was performed to compare the potency predictions with LLNA or human effect levels. Two models were developed: an approach for chemicals that could be attributed to a specific mechanistic domain and an approach based on a global regression equation for substances outside these chemical domains.

The accuracy, sensitivity and specificity of this approach to distinguish skin sensitizers from non-sensitizers was 80, 82 and 78 %, respectively, for a set of 244 compared to the LLNA. The approach was used to predict the GHS potency sub-categories 1A and B and to derive a point value that was compared to the LLNA EC3 or the human effect level (DSA_{05} : dose per skin area leading to a sensitization

incidence of 5 %). Compared to human data, this approach predicted both GHS classes with an accuracy of approximately 82 % (Table 5) for a dataset of 71 substances. Based on the LLNA EC3, this approach had an accuracy of 70 % for the 1A and 77 % for the 1B GHS categories (Natsch et al. 2015).

The company Kao constructed an STS based on the MIE, addressed by the DPRA, and KE3, addressed by the h-CLAT. KE2 is not included in this STS. In their STS, the h-CLAT is first performed and substances that are positive in the h-CLAT are considered a skin sensitizer, and no further testing is needed. The skin-sensitizing potency is based on the minimal induction threshold (MIT) derived from the dose–response curve of the h-CLAT. A substance is considered a strong skin sensitizer when the MIT is $\leq 10\text{ }\mu\text{g/ml}$ and a weak sensitizer if the MIT is >10 and $\leq 5000\text{ }\mu\text{g/ml}$. Substances that are negative in the h-CLAT are subsequently tested in the DPRA. If the DPRA is positive, the substance is considered a weak skin sensitizer. If both assays are negative, the substance is considered a non-sensitizer (Takenouchi et al. 2015).

The performance of this assay has been evaluated in two studies, one with 101 substances (Nukada et al. 2013), and later this dataset was enlarged with 38 additional chemicals (Takenouchi et al. 2015). Since both studies found comparable results, only the last one is discussed here. Skin sensitization potential was predicted with an accuracy, sensitivity and specificity of 81, 90 and 54 %, respectively, compared to the LLNA, demonstrating that this STS has a poor specificity. The dominant assay in this STS is the h-CLAT, and it is known that chemicals with a $\log Kow > 3.5$ tend to give false-negative results (Takenouchi et al. 2013). A new analysis of this STS, in which 10 chemicals with a $\log Kow > 3.5$ were excluded, resulted in an accuracy, sensitivity and specificity of 85, 98 and 51 %, respectively (Takenouchi et al. 2015). Exclusion of these 10 chemicals slightly improved the accuracy and sensitivity of the STS, but the specificity for hazard identification remains low.

The sub-classes predicted with this STS were compared to the sub-categorization based on the LLNA EC3 (extreme, strong, moderate and weak sensitizers), using a cut-off of EC3 1 % to divide substances into two sub-categories (extreme to strong versus weak to moderate). It was shown that the STS categorized 19 of the 29 extreme-to-strong skin sensitizers in the strong sub-category and 57 of the 73 moderate and weak skin sensitizers in the weak class. Hence, the STS especially under-predicts the potency of strong and extreme skin sensitizers (Takenouchi et al. 2015). It is remarkable that the cut-off values for the potency sub-categories were different from those used for the GHS CLP sub-categories 1A and B that use an EC3 of 2 %. Therefore, it is difficult to assess whether this STS

is a valuable tool for sub-categorization under GHS and REACH, and an additional analysis using the EC3 value of 2 % as a cut-off is highly recommendable.

Takenouchi et al. (2015) constructed a score-based system with the same assays and evaluated the system with the same dataset of 139 substances. In this approach, results from the DPRA and h-CLAT were converted in scores from 0 to 3. If a substance is negative in those assays and has no alert in DEREK Nexus, it was assigned a score of 0. Scores 1–3 were assigned using predefined cut-offs for the MIT from the h-CLAT and the percentage cysteine depletion from the DPRA. Substances were considered skin sensitizers if the score was 2 or higher. The total score was used to sub-categorize the skin sensitizers in two sub-categories: strong sensitizer if the score was 7, 8 or 9 and weak sensitizer if the score was between 2 and 6. This score-based system identified skin sensitizers with an accuracy, sensitivity and specificity of 84, 89 and 70 %, respectively, compared to the LLNA. This score-based system has a higher specificity, but a lower sensitivity, compared to the Kao STS, in which the same dataset and assays were used. The ability to predict potency in two sub-categories was again not so accurate for the extreme and strong skin sensitizers ($EC_3 < 1\%$), as only 52 % was correctly categorized as strong sensitizer. Potency prediction for the moderate and weak sensitizers was somewhat better, 78 % were correctly classified (Takenouchi et al. 2015).

An artificial neural network model (ANN) was used to integrate dose–response data from the ARE assay in AREC32 cells (Tsujita-Inoue et al. 2014) and the h-CLAT assay. Also, the water–octanol partition coefficient ($\log K_{ow}$) was included as a parameter that reflects both the ability of chemicals to penetrate into the epidermis and the water solubility (Tsujita-Inoue et al. 2014). ANN analysis is a nonlinear statistical data-modelling tool that can be used to model complex inputs and outputs. The performance of this model was tested with 62 test chemicals based on the LLNA. The model was trained with a set of 53 chemicals and then tested with the remaining 9 chemicals to evaluate the predictive capacity of the model. Skin sensitization hazard was predicted with an accuracy, sensitivity and specificity of, respectively, 91, 100 and 62 % for the test set. The performance of this model to predict potency sub-categories was evaluated as well. Substances were divided into two potency sub-categories, using the LLNA EC_3 value of 1 % as a cut-off to distinguish extreme and strong sensitizers from moderate and weak ones. For the full dataset of 62 chemicals, the model predicted the extreme-to-strong skin sensitizers with an accuracy of 92 % and the moderate-to-weak ones with 72 % (Tsujita-Inoue et al. 2014). Again, sub-categories were distinguished using a cut-off of 1 %, making it not possible to compare them to the GHS CLP sub-categories 1A and 1B. The ANN model was used to predict a point value as well that was compared

to the LLNA EC_3 value. A relatively good correlation (R value 0.838; RMS error 0.628) was shown (Tsujita-Inoue et al. 2014).

A Bayesian network was used to construct and refine a Bayesian ITS (Jaworska et al. 2013; Jaworska and Hoffmann 2010; Jaworska et al. 2015). A Bayesian network formulates a probabilistic hypothesis about the target variable, in this case the LLNA potency. A probabilistic model can be used to represent uncertain knowledge in case it is known which variables are important in the process of interest, but the relationships between the variables are not well characterized (Jaworska and Hoffmann 2010). This Bayesian ITS has been first presented as a proof of concept (Jaworska et al. 2011) and was later revised to improve the potency prediction and transform the concept in a practical tool for hazard assessment: ITS-2 (Jaworska et al. 2013). ITS-2 was distributed in an open source version by collaboration with the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) (Pirone et al. 2014).

To improve further the precision and accuracy for predicting LLNA potency, ITS-2 was revised by improving the representation of both chemistry and biology (Jaworska et al. 2015). ITS-3 uses the following data sources as inputs: bioavailability-related variables, in silico prediction of potency using TIMES and quantitative data from the DPRA, KeratinoSens™ and h-CLAT. The Bayesian network was used to construct this ITS, in which each variable is presented as a node in the network and each node has its own probability distribution conditioned to the other network variables. The ITS-3 estimates skin sensitization potency expressed as the probability distribution of the LLNA EC_3 , spread among four potency sub-categories: non-sensitizers, weak, moderate and strong sensitizers. First, ITS-3 was trained with a training set compiled of 147 substances. This ITS was validated with a test set which contained 60 substances. This evaluation showed that the Bayesian ITS could accurately predict the different potency classes. The ITS did not produce any false positives in the test set. Strong and weak skin sensitizers were predicted with a high accuracy (87–89 %), whereas the prediction of moderate skin sensitizers was somewhat lower, around 75 % (Table 5). The ITS predicted three moderate skin sensitizers as weak ones, which may be explained by the limited dynamic range of the in vitro assays, making it difficult to distinguish weak from moderate skin sensitizers (Jaworska et al. 2013; Jaworska et al. 2015).

A computational approach was used to integrate and optimize information from in silico, *in chemico* and in vitro assays. The initial dataset ($n = 145$) evaluated in this ITS was based on the work by Jaworska et al. (2013), using the same quantitative parameters from the DPRA, KeratinoSens and U937-CD86 assay and the TIMES potency

prediction. These seven *in chemico/in vitro* descriptors were combined with Dragon descriptors, which were used to evaluate molecular structure–activity or structure–property relationships. This first dataset was subdivided in two smaller datasets, for which additional *in vitro* results were available to enlarge the number of *in chemico/in vitro* descriptors to nine and ten, respectively. The algorithm Random Forest was used as a supervised learning model on all available data to predict the class of a chemical. Subsequently, recursive feature elimination was used to show which descriptors are the most important to make an accurate prediction and to remove the features that are least valuable. To be able to sub-categorize substances according to their potency, a dose-informed random forest algorithm combined with the hidden Markov model was used to transform the output into four classes (non-sensitizer, moderate, strong and extreme skin sensitizers). The overall accuracy of this probabilistic approach was 60–70 %. The accuracy to predict the correct potency class differed per dataset, for details see Table 5 (Luechtefeld et al. 2015).

Discussion and conclusions

This update illustrates that since our previous review in 2010 much progress has been made, both on the development, validation and regulatory acceptance of individual predictive test methods and on combining them in defined approaches to testing and assessment or IATA. For individual test methods, progress is made mainly in the area of *in vitro* keratinocyte and dendritic cell assays, with several assays being validated and/or regulatory accepted, whereas non-animal approaches measuring protein binding (MIE) and T cell activation (KE4) have not advanced during this period. The acceptance of the DPRA, KeratinoSensTM and h-CLAT as OECD test guidelines and their implementation in the REACH annex is a first step towards animal-free hazard identification of skin sensitizers. However, actual replacement of the animal tests for regulatory purposes can only be achieved if one or maybe multiple defined approaches to testing and assessment are regulatory accepted and implemented. It is evident from our review that there are multiple ways to combine and analyse data from different information sources. Some approaches are constructed on a mechanistic base using parts of the AOP, e.g. the ‘2 out of 3’ ITS (Urbisch et al. 2015) or the STS of the RIVM (Van der Veen et al. 2014a, b). Others apply statistical approaches to construct their strategy (e.g. the integrated decision strategy of Strickland et al. (2016) or use a combination of both (e.g. the Bayesian ITS of Jaworska et al. (2015)). The differences in predictivity between the different defined approaches are not always that big, although a direct comparison is often not possible due to

differences in datasets used. To progress towards regulatory acceptance, an independent review of the different defined approaches is urgently needed using the same set of chemicals to enable a comparison. This next step is essential to allow skin sensitization testing without animals, which is urgently needed with the EU ban on animal testing for cosmetic ingredients and the REACH deadline of 2018 approaching.

It became evident from our review that many AOP-based testing strategies are based mainly on MIE, KE2 and KE3; hence, they do not cover the complete AOP. Yet they are able to accurately identify the majority of substances and often have a better performance than the LLNA. The question remains whether or coverage of the complete AOP is required and if addition of assays informing on activation of KE4 would improve the prediction. Apparently, information on MIE, KE2 and KE3 is sufficient to predict skin sensitization with a high certainty. Furthermore, the ‘2 out of 3’ approach shows that it is possible to identify skin sensitizers with a high accuracy based on only two KEs. Notably, discordant results in the individual test methods do not imply that a substance does not trigger all KEs. One explanation is that substances are misclassified due to limitations of the assay, e.g. prohaptens will not be detected in the DPRA. Understanding the limitations and applicability domain is thus important to explain why a substance provides a discordant result. Furthermore, several skin sensitization assays are in the pipeline towards regulatory acceptance and implementation and may be a useful component in the available testing strategies. To optimally use these novel assays within a testing strategy, it is of utmost importance to understand their limitations and applicability domains, because this facilitates the selection of the most appropriate test method within a strategy. Although a defined approach follows a fixed procedure, flexibility is needed in the assays that are used within their construction. For example, in the future multiple (me-too) assays are available that inform on KE2 and KE3 and ideally the assay that is best fit-for-purpose needs to be selected to address a certain KE. To be able to select the most optimal assay, it is essential to understand its limitations and applicability domain. Transparent sharing of data from different test methods, as done in the study described by Urbisch et al. (2015), is of major importance to be able to compare the interchangeability of different test methods that address the same KE.

Interestingly, the predictive capacity of several defined approaches described in our review is comparable; for example, the ‘2 out of 3’ ITS and the Bayesian ITS are quite similar in their accuracy, sensitivity and specificity for a large set of chemical. The construction of the Bayesian ITS is quite complex, and the decision is based on integration of multiple information sources, whereas the ‘2 out of 3’ ITS is simple and allows a prediction on only two assays.

This may indicate that the construction of a strategy does not have a major impact on the predictivity of a strategy. In the process towards regulatory acceptance, multiple aspects of a strategy need to be considered, including mechanistic relevance, reliability and reproducibility. Another important aspect in this process is that a defined approach needs to be applicable in practice, i.e. not only by the companies or institutes that have developed the approach but also by SMEs and regulators. In a recent workshop on this topic organized by the European Partnership for Alternatives to Animal Testing (EPAA), it was concluded that there is a need for any new approach to have a sufficient degree of transparency and ‘user-friendliness’. The workshop concluded that a simple STS, such as the one developed by the RIVM, may be the optimal way forward to obtain AOP-informed information for hazard identification of skin sensitizers (Basketter et al. 2015).

Another important hurdle in this field is that it is difficult to estimate potency of skin sensitizers with the newly developed test methods or strategies. In general, the majority of predictive *in vitro* or *in chemico* assays cannot estimate potency on their own. A possible explanation may be that the dynamic range of *in vitro* and *in chemico* assays is too limited to detect the broad range of skin-sensitizing potencies, which span four orders of magnitude. An exception to this is the recently developed SENS-IS assay, which provides an accurate prediction of potency categories (Cottrez et al. 2015). Possibly, RhE models are more suitable to predict potency compared to submerged *in vitro* assays, as has been shown in the epidermal equivalent assay as well (Dos Santos et al. 2011). In addition, several studies have shown that regulation of gene expression correlates with potency (McKim et al. 2012; Van der Veen et al. 2013), so the read-out in the SENS-IS assay may more capable of assigning skin sensitizers to their potency sub-category.

Quantitative results from individual assays have been used as inputs in several defined approaches in order to estimate potency. The majority of the currently published strategies enable sub-categorization in potency classes. An accurate prediction of a point value, such as the LLNA EC₃, is not yet possible or does not correlate well with the available human or LLNA benchmarks. Some of the defined approaches are able to predict the potency sub-categories quite accurately, e.g. the approach that uses MIE and KE2 for potency prediction (Natsch et al. 2015) and the Bayesian ITS (Jaworska et al. 2015). Other approaches are only able to predict a specific sub-category with high accuracy. For example, the ANN approach had a high accuracy for the strong skin sensitizers, whereas moderate and weak skin sensitizers were predicted with less certainty. In contrast, the Kao ITS and score-based ITS have a poor accuracy for the strong skin sensitizers (Takenouchi et al. 2015). The defined approaches described by Takenouchi et al. (2015) are based

on a similar dataset tested in the same test methods. It is evident that the construction and data interpretation model have an impact on the performance of the approach, since the accuracy to predict strong skin sensitizers is somewhat better for the ITS. Clearly, an independent review using a similar dataset of chemicals is needed to evaluate the performance of these defined approaches as well.

In reviewing the strategies that estimate potency, it is essential to consider recent evaluations on the performance of the LLNA to predict potency. Thorough evaluations of an extensive database of LLNA studies provided insight into the variability of the LLNA to predict the potency. Importantly, variability of the LLNA could only partly be explained by the vehicle used, demonstrating that LLNA results are variable, also when the same vehicle is used. Also, potency estimations of weak skin sensitizers were more variable than that of strong ones (Dumont et al. 2016; Hoffmann 2015). These analyses are important, since they provide insight into the variability and uncertainty related to the LLNA. In the development and evaluation of approaches that predict potency, this variability needs to be considered. Furthermore, whenever possible and available, human potency data should be used as a benchmark.

To conclude, we have presented an updated overview of animal-free test methods and defined approaches for testing and assessment for skin sensitization. This review offers the opportunity to look back on the forecast of Adler et al. (2011). These experts expected that full replacement for skin-sensitizing potency assessment would not be feasible in 2013, but that assessment of skin sensitization potential would be possible with a combination of non-animal test methods that address different mechanistic steps involved in skin sensitization in 2017–2019. In our opinion, it should already be possible to determine skin sensitization hazard potential with a defined approach, although independent review using a similar set of chemicals of the most promising defined approaches is an essential step to push these approaches further towards regulatory acceptance.

Acknowledgments We highly appreciate the critical comments and suggestions of Prof. Dr. Aldert Piersma of the RIVM. This review was written within the context of a research project funded by ZonMw (Grant No. 114022001; Systems toxicology supported data infrastructure for human risk assessment). In addition, the Dutch Ministries of Economic Affairs and of Health, Welfare and Sports are acknowledged for supporting this work.

References

- Adler S, Basketter D, Creton S et al (2011) Alternative (non-animal) methods for cosmetics testing: current status and future prospects—2010. Arch Toxicol 85(5):367–485
- Albrekt A-S, Johansson H, Börje A, Borrebaeck C, Lindstedt M (2014) Skin sensitizers differentially regulate signaling pathways in

- MUTZ-3 cells in relation to their individual potency. *BMC Pharmacol Toxicol* 15(1):1–13
- Alépée N, Piroird C, Aujoulat M et al (2015) Prospective multicentre study of the U-SENS test method for skin sensitization testing. *Toxicol in Vitro* 30(1, Part B):373–382
- Alves VM, Muratov E, Fourches D et al (2015) Predicting chemically-induced skin reactions. Part II: QSAR models of skin permeability and the relationships between skin permeability and skin sensitization. *Toxicol Appl Pharm* 284(2):273–280
- Api AM, Basketter DA, Cadby PA et al (2008) Dermal sensitization quantitative risk assessment (QRA) for fragrance ingredients. *Regul Toxicol Pharmacol* 52(1):3–23
- Aptula AO, Roberts DW (2006) Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity. *Chem Res Toxicol* 19(8):1097–1105
- Ashikaga T, Sakaguchi H, Sono S et al (2010) A comparative evaluation of in vitro skin sensitisation tests: the human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA). *Altern Lab Anim* 38(4):275–284
- Avonto C, Chittiboyina AG, Rua D, Khan IA (2015) A fluorescence high throughput screening method for the detection of reactive electrophiles as potential skin sensitizers. *Toxicol Appl Pharmacol* 289(2):177–184
- Barratt MD, Basketter DA, Chamberlain M, Admans GD, Langowski JJ (1994a) An expert system rulebase for identifying contact allergens. *Toxicol In Vitro* 8(5):1053–1060
- Barratt MD, Basketter DA, Chamberlain M, Payne MP, Admans GD, Langowski JJ (1994b) Development of an expert system rulebase for identifying contact allergens. *Toxicol In Vitro* 8(4):837–839
- Basketter DA, Andersen KE, Liden C et al (2005) Evaluation of the skin sensitizing potency of chemicals by using the existing methods and considerations of relevance for elicitation. *Contact Dermat* 52(1):39–43
- Basketter D, Pease C, Kasting G et al (2007) Skin sensitisation and epidermal disposition: the relevance of epidermal disposition for sensitisation hazard identification and risk assessment. The report and recommendations of ECVAM workshop 59. *Altern Lab Anim* 35(1):137–154
- Basketter D, Ashikaga T, Casati S et al (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(2):660–666
- Bauch C, Kolle SN, Fabian E 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 SN, Ramirez T et al (2012) Putting the parts together: combining in vitro methods to test for skin sensitizing potentials. *Regul Toxicol Pharmacol* 63(3):489–504
- Casati S, Aeby P, Basketter DA et al (2005) Dendritic cells as a tool for the predictive identification of skin sensitisation hazard. *Altern Lab Anim* 33(1):47–62
- Chaudhry Q, Piclin N, Cotterill J et al (2010) Global QSAR models of skin sensitizers for regulatory purposes. *Chem Cent J* 4(Suppl 1):S5
- Coleman KP, McNamara LR, Grailer TP et al (2015) Evaluation of an in vitro human dermal sensitization test for use with medical device extracts. *Appl In Vitro Toxicol* 1(2):13
- Corsini E, Mitjans M, Galbiati V, Lucchi L, Galli CL, Marinovich M (2009) Use of IL-18 production in a human keratinocyte cell line to discriminate contact sensitizers from irritants and low molecular weight respiratory allergens. *Toxicol In Vitro* 23(5):789–796
- Corsini E, Galbiati V, Mitjans M, Galli CL, Marinovich M (2013) NCTC 2544 and IL-18 production: a tool for the identification of contact allergens. *Toxicology In Vitro* 27(3):1127–1134
- Cottrez F, Boitel E, Auriault C, Aeby P, Groux H (2015) Genes specifically modulated in sensitized skins allow the detection of sensitizers in a reconstructed human skin model. Development of the SENS-IS assay. *Toxicol In Vitro* 29(4):787–802
- Cottrez F, Boitel E, Ourlin J-C et al (2016) SENS-IS, a 3D reconstituted epidermis based model for quantifying chemical sensitization potency: reproducibility and predictivity results from an inter-laboratory study. *Toxicol In Vitro* 32:248–260
- Dancik Y, Miller MA, Jaworska J, Kasting GB (2013) Design and performance of a spreadsheet-based model for estimating bioavailability of chemicals from dermal exposure. *Adv Drug Deliv Rev* 65(2):221–236
- Davies M, Pendlington RU, Page L et al (2011) Determining epidermal disposition kinetics for use in an integrated nonanimal approach to skin sensitization risk assessment. *Toxicol Sci* 119(2):308–318
- Dietz L, Kinzebach S, Ohnesorge S et al (2013) Proteomic allergen-peptide/protein interaction assay for the identification of human skin sensitizers. *Toxicol In Vitro* 27(3):1157–1162
- Dimitrov SD, Low LK, Patlewicz GY et al (2005) Skin sensitization: modeling based on skin metabolism simulation and formation of protein conjugates. *Int J Toxicol* 24(4):189–204
- Dos Santos GG, Spiekstra SW, Sampat-Sardjoepersad SC, Reinders J, Scheper RJ, Gibbs S (2011) A potential in vitro epidermal equivalent assay to determine sensitizer potency. *Toxicol In Vitro* 25(1):347–357
- Dumont C, Prieto P, Asturiol D, Worth A (2015) Review of the availability of in vitro and in silico methods for assessing dermal bioavailability. *Appl In Vitro Toxicol* 1(2):147–164
- Dumont C, Barroso J, Matys I, Worth A, 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
- Enoch SJ, Madden JC, Cronin MT (2008) Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. *SAR QSAR Environ Res* 19(5–6):555–578
- Enslein K, Gombar VK, Blake BW et al (1997) A quantitative structure-toxicity relationships model for the dermal sensitization guinea pig maximization assay. *Food Chem Toxicol* 35(10–11):1091–1098
- EURL-ECVAM (2013a) EURL ECVAM recommendation on the Direct Peptide Reactivity Assay (DRPA) for skin sensitisation testing European Commission Joint Research Centre. Institute for Health and Consumer Protection. European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)
- EURL-ECVAM (2013b) EURL ECVAM recommendation on the Keratinosens™ assay for skin sensitisation testing European Commission Joint Research Centre. Institute for Health and Consumer Protection. European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)
- EURL-ECVAM (2014) EURL ECVAM recommendation on the Keratinosens™ assay for skin sensitisation testing. European Commission Joint Research Centre. Institute for Health and Consumer Protection. European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)
- EURL-ECVAM (2015a) EURL ECVAM Recommendation on the human Cell Line Activation Test (h-CLAT) for skin sensitisation testing. European Commission, Joint Research Centre, Institute for Health and Consumer Protection
- EURL-ECVAM (2015b) EURL ECVAM Status Report on the Development, Validation and Regulatory Acceptance of Alternative Methods and Approaches. European Commission, Joint Research Centre, Institute for Health and Consumer Protection
- EURL-ECVAM (2015c) EURL ECVAM status report on the development, validation and regulatory acceptance of alternative

- methods and approaches (2015) European Commission, Joint Research Centre, Institute for Health and Consumer Protection
- Frombach J, Sonnenburg A, Krapohl B-D, Zuberbier T, Stahlmann R, Schreiner M (2016) A novel method to generate monocyte-derived dendritic cells during coculture with HaCaT facilitates detection of weak contact allergens in cosmetics. *Arch Toxicol*. doi:10.1007/s00204-016-1722-y
- Galbiati V, Mitjans M, Lucchi L et al (2011) Further development of the NCTC 2544 IL-18 assay to identify in vitro contact allergens. *Toxicol In Vitro* 25(3):724–732
- Galvão dos Santos G, Reinders J, Ouwehand K, Rustemeyer T, Scheper RJ, Gibbs S (2009) Progress on the development of human in vitro dendritic cell based assays for assessment of the sensitizing potential of a compound. *Toxicol Appl Pharmacol* 236(3):372–382
- Gealy R, Graham C, Sussman NB, Macina OT, Rosenkranz HS, Karol MH (1996) Evaluating clinical case report data for SAR modeling of allergic contact dermatitis. *Hum Exp Toxicol* 15(6):489–493
- Gerberick GF, Vassallo JD, Bailey RE, Chaney JG, Morrall SW, Lepoittevin JP (2004) Development of a peptide reactivity assay for screening contact allergens. *Toxicol Sci* 81(2):332–343
- Gerberick GF, Vassallo JD, Foertsch LM, Price BB, Chaney JG, Lepoittevin JP (2007) Quantification of chemical peptide reactivity for screening contact allergens: a classification tree model approach. *Toxicol Sci* 97(2):417–427
- Gerberick GF, Troutman JA, Foertsch LM et al (2009) Investigation of peptide reactivity of pro-hapten skin sensitizers using a peroxidase-peroxide oxidation system. *Toxicol Sci* 112(1):164–174
- Gibbs S, Corsini E, Spiekstra SW et al (2013) An epidermal equivalent assay for identification and ranking potency of contact sensitizers. *Toxicol Appl Pharmacol* 272(2):529–541
- 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(3):269–290
- Hartung T, Luechtefeld T, Maertens A, Kleensang A (2013) Integrated testing strategies for safety assessments. *ALTEX* 30(1):3–18
- Hirota M, Suzuki M, Hagino S et al (2009) Modification of cell-surface thiols elicits activation of human monocytic cell line THP-1: possible involvement in effect of haptens 2,4-dinitrochlorobenzene and nickel sulfate. *J Toxicol Sci* 34(2):139–150
- Hoffmann S (2015) LLNA variability: an essential ingredient for a comprehensive assessment of non-animal skin sensitization test methods and strategies. *ALTEX* 32(4):379–383
- Hooyberghs J, Schoeters E, Lambrechts N et al (2008) A cell-based in vitro alternative to identify skin sensitizers by gene expression. *Toxicol Appl Pharmacol* 231(1):103–111
- Ibrahim R, Nitsche JM, Kasting GB (2012) Dermal clearance model for epidermal bioavailability calculations. *J Pharm Sci* 101(6):2094–2108
- Jaworska J, Hoffmann S (2010) Integrated Testing Strategy (ITS)—opportunities to better use existing data and guide future testing in toxicology. *ALTEX* 27(4):231–242
- Jaworska J, Harol A, Kern PS, Gerberick F (2011) Integrating non-animal test information into an adaptive testing strategy—skin sensitization proof of concept case. *ALTEX* 28(3):211–225
- 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
- Jaworska JS, 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(12):2355–2383
- Johansson H, Lindstedt M, Albrekt AS, Borrebaeck CA (2011) A genomic biomarker signature can predict skin sensitizers using a cell-based in vitro alternative to animal tests. *BMC Genomics* 12:399
- Johansson H, Albrekt AS, Borrebaeck CA, Lindstedt M (2012) The GARD assay for assessment of chemical skin sensitizers. *Toxicol In Vitro* 27(3):1163–1169
- Johansson H, Rydnert F, Kuhn J, Schepky A, Borrebaeck C, Lindstedt M (2014) Genomic allergen rapid detection in-house validation—a proof of concept. *Toxicol Sci* 139(2):362–370
- Johnson R, Macina OT, Graham C, Rosenkranz HS, Cass GR, Karol MH (1997) Prioritizing testing of organic compounds detected as gas phase air pollutants: structure-activity study for human contact allergens. *Environ Health Perspect* 105(9):986–992
- Kagatani S, Sasaki Y, Hirota M et al (2010) Oxidation of cell surface thiol groups by contact sensitizers triggers the maturation of dendritic cells. *J Invest Dermatol* 130(1):175–183
- Karlberg AT, Bergstrom MA, Borje A, Luthman K, Nilsson JL (2008) Allergic contact dermatitis—formation, structural requirements, and reactivity of skin sensitizers. *Chem Res Toxicol* 21(1):53–69
- Kimber I, Basketter DA, Gerberick GF, Ryan CA, Dearman RJ (2011) Chemical allergy: translating biology into hazard characterization. *Toxicol Sci* 120(suppl 1):S238–S268
- Klopman G, Ivanov J, Saikhov R, Chakravarti S (2005) MC4PC—an artificial intelligence approach to the discovery of structure toxic activity relationships (STAR). In: Helma C (ed) *Predictive Toxicology*. CRC Press, Boca Raton, FL, pp 423–457
- Lambrechts N, Vanheel H, Hooyberghs J et al (2010a) Gene markers in dendritic cells unravel pieces of the skin sensitization puzzle. *Toxicol Lett* 196(2):95–103
- Lambrechts N, Vanheel H, Nelissen I et al (2010b) Assessment of chemical skin sensitizing potency by an in vitro assay based on human dendritic cells. *Toxicol Sci* 116(1):122–129
- Lambrechts N, Nelissen I, Van Tendeloo V et al (2011) Functionality and specificity of gene markers for skin sensitization in dendritic cells. *Toxicol Lett* 203(2):106–110
- Langton K, Patlewicz GY, Long A, Marchant CA, Basketter DA (2006) Structure-activity relationships for skin sensitization: recent improvements to Derek for Windows. *Contact Dermat* 55(6):342–347
- Luechtefeld T, Maertens A, McKim JM, Hartung T, Kleensang A, Sá-Rocha V (2015) Probabilistic hazard assessment for skin sensitization potency by dose–response modeling using feature elimination instead of quantitative structure–activity relationships. *J Appl Toxicol* 35(11):1361–1371
- MacKay C, Davies M, Summerfield V, Maxwell G (2013) From pathways to people: applying the adverse outcome pathway (AOP) for skin sensitization to risk assessment. *ALTEX* 30(4):473–486
- Macmillan DS, Canipa SJ, Chilton ML, Williams RV, Barber CG (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
- Martin SF, Esser PR, Schmucker S et al (2010) T-cell recognition of chemicals, protein allergens and drugs: towards the development of in vitro assays. *Cell Mol Life Sci* 67(24):4171–4184
- Martin SF, Esser PR, Weber FC et al (2011) Mechanisms of chemical-induced innate immunity in allergic contact dermatitis. *Allergy* 66(9):1152–1163
- Maxwell G, MacKay C, Cubberley R et al (2014) Applying the skin sensitisation adverse outcome pathway (AOP) to quantitative risk assessment. *Toxicol In Vitro* 28(1):8–12
- McKim JM, Keller DJ, Gorski JR (2010) A new in vitro method for identifying chemical sensitizers combining peptide binding with ARE/EpRE-mediated gene expression in human skin cells. *Cutan Ocul Toxicol* 29(3):171–192
- McKim JM, Jr., Keller DJ, Gorski JR (2012) An in vitro method for detecting chemical sensitization using human reconstructed skin models and its applicability to cosmetic, pharmaceutical, and medical device safety testing. *Cutan Ocul Toxicol* 31(4):292–305

- Mitragotri S, Anissimov YG, Bunge AL et al (2011) Mathematical models of skin permeability: an overview. *Int J Pharm* 418(1):115–129
- Natsch A (2010) The Nrf2-Keap1-ARE toxicity pathway as a cellular sensor for skin sensitizers—functional relevance and a hypothesis on innate reactions to skin sensitizers. *Toxicol Sci* 113(2):284–292
- 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(2):464–478
- Natsch A, Bauch C, Foertsch L et al (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(3):733–744
- Natsch A, Ryan CA, Foertsch L et al (2013) A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. *J Appl Toxicol JAT* 33(11):1337–1352
- 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(2):319–332
- Nukada Y, Ashikaga T, Sakaguchi H et al (2011) Predictive performance for human skin sensitizing potential of the human cell line activation test (h-CLAT). *Contact Dermat* 65(6):343–353
- 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(2):609–618
- OECD (1992) OECD Guidelines for the testing of chemicals, section 4: health effects, Test No. 406 skin sensitization. OECD Publishing, Paris
- OECD (2007) Guidance on grouping of chemicals OECD environmental health and safety publications, series on testing and assessment no 80. OECD Publishing, Paris
- OECD (2008) Workshop on integrated approaches to testing and assessment. In: Publishing O (ed) Series on testing and assessment no 88. OECD Publishing, Paris
- OECD (2010a) Guidelines for the testing of chemicals. Section 4: health effects. Test No. 429. Skin sensitization: local lymph node assay. OECD Publishing, Paris
- OECD (2010b) Guidelines for the testing of chemicals. Section 4: health effects. Test no. 442A skin sensitization: local lymph node assay: DA. OECD Publishing, Paris
- OECD (2010c) OECD guidelines for the testing of chemicals. Section 4: health effects. Test no. 442B skin sensitization: local lymph node assay: BrdU-ELISA. OECD Publishing, Paris
- OECD (2012) The adverse outcome pathway for skin sensitisation initiated by covalent binding for proteins. Part 1. Scientific evidence OECD environment, health and safety publications series on testing and assessment, vol 168. OECD Publishing, Paris, pp 1–46
- OECD (2014) Guidance on grouping of chemicals, 2nd Ed. Environment, health 22 and safety publications, series on testing and assessment no. 194 23 series on testing and assessment, vol ENV/JM/MONO(2014)4. OECD Publishing, Paris
- OECD (2015a) OECD guidelines for the testing of chemicals, Section 4. Test No. 442C: in chemico skin sensitisation: Direct Peptide Reactivity Assay (DPRA). OECD Publishing, Paris
- OECD (2015b) OECD guidelines for the testing of chemicals, Section 4. Test No. 442D: in vitro skin sensitisation: ARE-Nrf2 Luciferase test method. OECD Publishing, Paris
- OECD (2015c) Test No. 442D: in vitro skin sensitisation: ARE-Nrf2 Luciferase test method. OECD Publishing, Paris
- OECD (2016a) 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 sensitisation In: OECD Environment HaSP (ed) Series on testing and assessment, vol ENV/JM/HA(2016)11. OECD Publishing, Paris
- OECD (2016b) Guidance document on the reporting of defined approaches to be used within integrated approaches to testing and assessment In: OECD Environment HaSP (ed) Series on testing and assessment, vol ENV/JM/HA(2016)10. OECD Publishing, Paris
- OECD (2016c) OECD Guidelines for the Testing of Chemicals, Section 4. Test No. 442E: in vitro skin sensitisation: human Cell Line Activation Test (h-CLAT) vol ENV/JM/WRPR(2016)19]. OECD Publishing, Paris
- Patlewicz G, Kuseva C, Kesova A et al (2014) Towards AOP application—implementation of an integrated approach to testing and assessment (IATA) into a pipeline tool for skin sensitization. *Regul Toxicol Pharmacol* 69(3):529–545
- Pendlington RU, Minter HJ, Stupart L et al (2008) Development of a modified in vitro skin absorption method to study the epidermal/dermal disposition of a contact allergen in human skin. *Cutan Ocul Toxicol* 27(4):283–294
- Piroid C, Ovigne J-M, Rousset F et al (2015) The Myeloid U937 Skin Sensitization Test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome pathway for skin sensitization. *Toxicol In Vitro* 29(5):901–916
- Pirone JR, Smith M, Kleinstreuer NC et al (2014) Open source software implementation of an integrated testing strategy for skin sensitization potency based on a Bayesian network. *ALTEX* 31(3):336–340
- Ramirez T, Mehling A, Kolle SN et al (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
- Ramirez T, Stein N, Aumann A et al (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
- Reuter H, Gerlach S, Spieker J et al (2015) Evaluation of an optimized protocol using human peripheral blood monocyte derived dendritic cells for the in vitro detection of sensitizers: results of a ring study in five laboratories. *TiV* 29(5):976–986
- Richter A, Schmucker SS, Esser PR et al (2013) Human T cell priming assay (hTCPA) for the identification of contact allergens based on naive T cells and DC—IFN- γ and TNF- α readout. *Toxicol In Vitro* 27(3):1180–1185
- Roberts DW, Aptula AO, Patlewicz G, Pease C (2008) Chemical reactivity indices and mechanism-based read-across for non-animal based assessment of skin sensitisation potential. *J Appl Toxicol* 28(4):443–454
- Rorije E, Aldenberg T, Buist H, Kroese D, Schuurmann G (2013) The OSIRIS weight of evidence approach: ITS for skin sensitisation. *Regul Toxicol Pharmacol* 67(2):146–156
- Rovida C, Alepee N, Api AM et al (2015) Integrated testing strategies (ITS) for safety assessment. *ALTEX* 32(1):25–40
- Saito K, Nukada Y, Takenouchi O, Miyazawa M, Sakaguchi H, Nishiyama N (2013) Development of a new in vitro skin sensitization assay (Epidermal Sensitization Assay; EpiSensA) using reconstructed human epidermis. *Toxicol In Vitro* 27(8):2213–2224
- Sakaguchi H, Ashikaga T, Miyazawa M et al (2006) 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. *Toxicol In Vitro* 5:774–784
- Sakaguchi H, Ryan C, Ovigne JM, Schroeder KR, Ashikaga T (2010) Predicting skin sensitization potential and inter-laboratory reproducibility of a human Cell Line Activation Test (h-CLAT) in the European Cosmetics Association (COLIPA) ring trials. *Toxicol In Vitro* 24(6):1810–1820

- Schneider K, Akkan Z (2004) Quantitative relationship between the local lymph node assay and human skin sensitization assays. *Regul Toxicol Pharmacol* 39(3):245–255
- Schreiner M, Peiser M, Briechle D, Stahlmann R, Zuberbier T, Waner R (2007) A loose-fit coculture of activated keratinocytes and dendritic cell-related cells for prediction of sensitizing potential. *Allergy* 62(12):1419–1428
- Sonnenburg A, Schreiner M, Stahlmann R (2015) Assessment of the sensitizing potency of preservatives with chance of skin contact by the loose-fit coculture-based sensitization assay (LCSA). *Arch Toxicol* 89(12):2339–2344
- Strickland J, Zang Q, Kleinstreuer N et al (2016) Integrated decision strategies for skin sensitization hazard. *J Appl Toxicol*. doi:10.1002/jat.3281
- Suzuki M, Hirota M, Hagino S, Itagaki H, Aiba S (2009) Evaluation of changes of cell-surface thiols as a new biomarker for in vitro sensitization test. *Toxicol In Vitro* 23(4):687–696
- Takahashi T, Kimura Y, Saito R et al (2011) An in vitro test to screen skin sensitizers using a stable THP-1-derived IL-8 reporter cell line, THP-G8. *Toxicol Sci* 124(2):359–369
- Takenouchi O, Miyazawa M, Saito K, Ashikaga T, Sakaguchi H (2013) Predictive performance of the human Cell Line Activation Test (h-CLAT) for lipophilic chemicals with high octanol-water partition coefficients. *J Toxicol Sci* 38(4):599–609
- Takenouchi O, Fukui S, Okamoto K et al (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(11):1318–1332
- Teubner W, Mehling A, Schuster PX et al (2013) Computer models versus reality: how well do in silico models currently predict the sensitization potential of a substance. *Regul Toxicol Pharmacol* 67(3):468–485
- Teunis M, Corsini E, Smits M et al (2013) Transfer of a two-tiered keratinocyte assay: IL-18 production by NCTC2544 to determine the skin sensitizing capacity and epidermal equivalent assay to determine sensitizer potency. *Toxicol In Vitro* 27(3):1135–1150
- Tollefsen KE, Scholz S, Cronin MT et al (2014) Applying adverse outcome pathways (AOPs) to support integrated approaches to testing and assessment (IATA). *Regul Toxicol Pharmacol* 70(3):629–640
- 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(4):626–639
- Urbisch D, Mehling A, Guth K et al (2015) Assessing skin sensitization hazard in mice and men using non-animal test methods. *Regul Toxicol Pharmacol* 71(2):337–351
- Van der Veen JW, Vandebriel R, Van Loveren H, Ezendam J (2011) Keratinocytes, innate immunity and allergic contact dermatitis—opportunities for the development of in vitro assays to predict the sensitizing potential of chemicals. In: Ro YS (ed) *Contact dermatitis*. Available from: <http://www.intechopen.com/books/contact-dermatitis/keratinocytes-innate-immunity-and-allergic-contact-dermatitis-opportunities-for-the-development-of-i>
- Van der Veen JW, Pronk TE, van Loveren H, Ezendam J (2013) Applicability of a keratinocyte gene signature to predict skin sensitizing potential. *Toxicol In Vitro* 27(1):314–322
- Van der Veen JW, Rorije E, Emter R, Natsch A, van Loveren H, Ezendam J (2014a) Evaluating the performance of integrated approaches for hazard identification of skin sensitizing chemicals. *Regul Toxicol Pharmacol* 69(3):371–379
- Van der Veen JW, Soeteman-Hernandez LG, Ezendam J, Stierum R, Kuper FC, van Loveren H (2014b) Anchoring molecular mechanisms to the adverse outcome pathway for skin sensitization: analysis of existing data. *Crit Rev Toxicol* 44(7):590–599
- Van Och FM, Van Loveren H, Van Wolfswinkel JC, Machielsen AJ, Vandebriel RJ (2005) Assessment of potency of allergenic activity of low molecular weight compounds based on IL-1 α and IL-18 production by a murine and human keratinocyte cell line. *Toxicology* 210(2–3):95–109
- Vandebriel RJ, van Loveren H (2010) Non-animal sensitization testing: state-of-the-art. *Crit Rev Toxicol* 40(5):389–404
- Vocanson M, Achachi A, Mutez V et al (2014) Human T cell priming assay: depletion of peripheral blood lymphocytes in CD25+ cells improves the in vitro detection of weak allergen-specific T cells. In: Martin SF (ed) *T Lymphocytes as tools in diagnostics and immunotoxicology*. Springer, Basel
- Yamamoto Y, Tahara H, Usami R et al (2015) A novel in chemico method to detect skin sensitizers in highly diluted reaction conditions. *J Appl Toxicol* 35(11):1348–1360