Toxicology in Vitro 29 (2015) 259–270

Contents lists available at ScienceDirect

Toxicology in Vitro

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

Systematic evaluation of non-animal test methods for skin sensitisation safety assessment

Kerstin Reisinger[a,1,2,*], Sebastian Hoffmann[b,1,2], Nathalie Alépée[c,2], Takao Ashikaga[d,2], Joao Barroso[e,2,3], Cliff Elcombe[f], Nicola Gellatly[g,2], Valentina Galbiati[h], Susan Gibbs[i], Hervé Groux[j], Jalila Hibatallah[k,2], Donald Keller[l], Petra Kern[m,2], Martina Klaric[e,2], Susanne Kolle[n], Jochen Kuehnl[o], Nathalie Lambrechts[p], Malin Lindstedt[q], Marion Millet[r,2], Silvia Martinozzi-Teissier[c,2], Andreas Natsch[s], Dirk Petersohn[a,2], Ian Pike[t], Hitoshi Sakaguchi[u,2], Andreas Schepky[o,2], Magalie Tailhardat[v,2], Marie Templier[r,2], Erwin van Vliet[w,2], Gavin Maxwell[g,2]

[a] Henkel AG & CO KGaA, Henkelstr. 67, 40191 Düsseldorf, Germany
[b] Seh Consulting + Services, Rathausplatz 9, 33098 Paderborn, Germany
[c] L'Oréal Research & Innovation, 1, Avenue Eugene Schueller, 93600 Aulnay sous bois, France
[d] Shiseido Co, Ltd, Quality Assurance Center, Kanagawa, Japan
[e] Cosmetics Europe, Avenue Herrmann Debroux 40, B-1160 Auderghem, Brussels, Belgium
[f] CXR Biosciences Ltd., 2 James Lindsay Place, Dundee Technopole, Dundee DD1 5JJ, UK
[g] Unilever, Safety and Environmental Assurance Centre, Sharnbrook, Bedford MK44 1LQ, UK
[h] Laboratory of Toxicology, University Milan, Via Balzaretti 9, 20133 Milan, Italy
[i] VU University Medical Centre, Department Dermatology, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands
[j] ImmunoSearch, Les Cyclades, Chemin de Camperousse, 06130 Le plan de Grasse, France
[k] Chanel, 135 Avenue Charles de Gaulle, 92521 Neuilly sur Seine, France
[l] Cyprotex, LLC, 4717 Campus Drive, Kalamazoo, MI, USA
[m] Procter & Gamble Technology (Beijing) Co., Ltd., Global Product Stewardship, No. 35 Yu An Road, KongGang Development Area (Area B), Shunyi District, Beijing 101312, PR China
[n] BASF SE, Experimental Toxicology and Ecology, 67056 Ludwigshafen, Germany
[o] Beiersdorf AG, Unnastrasse 48, 20245 Hamburg, Germany
[p] VITO NV, Industriezone Vlasmeer 7, 2400 Mol, Belgium
[q] Department of Immunotechnology, Lund University, Medicon Village (406), 22381 Lund, Sweden
[r] Institut de Recherche Pierre Fabre, CEPC – Bel air de Campans, 81106 Castres Cedex, France
[s] Givaudan Schweiz AG, Ueberlandstrasse 138, 8600 Dübendorf, Switzerland
[t] Proteome Sciences plc, Coveham House, Downside Bridge Road, Cobham KT11 3EP, UK
[u] Kao Corporation, Safety Science Research Laboratories, Ichikai-Machi, Haga-Gun, Tochigi, Japan
[v] LVMH Recherche, 185 avenue de Verdun, 45804 St Jean de Braye Cedex, France
[w] Services & Consultation on Alternative Methods (SeCAM), Via Fontanone 27, 6982 Agno, Switzerland

ARTICLE INFO

Article history:
Received 27 March 2014
Accepted 21 October 2014
Available online 31 October 2014

ABSTRACT

The need for non-animal data to assess skin sensitisation properties of substances, especially cosmetics ingredients, has spawned the development of many in vitro methods. As it is widely believed that no single method can provide a solution, the Cosmetics Europe Skin Tolerance Task Force has defined a three-phase framework for the development of a non-animal testing strategy for skin sensitisation

* Corresponding author. Tel.: +49 211 7972979; fax: +49 211 7977887.

E-mail addresses: kerstin.reisinger@henkel.com (K. Reisinger), sebastian.hoffmann@seh-cs.com (S. Hoffmann), nalepee@rd.loreal.com (N. Alépée), takao.ashikaga@to.shiseido.co.jp (T. Ashikaga), Joao.BARROSO@ec.europa.eu (J. Barroso), CliffElcombe@cxrbiosciences.com (C. Elcombe), Nichola.Gellatly@unilever.com (N. Gellatly), valentina.galbiati@unimi.it (V. Galbiati), S.Gibbs@vumc.nl (S. Gibbs), hgroux@immunosearch.fr (H. Groux), jalila.hibatallah@chanel-corp.com (J. Hibatallah), dkeller@ceetox.com (D. Keller), kern.ps@pg.com (P. Kern), mklaric@cosmeticseurope.eu (M. Klaric), susanne.kolle@basf.com (S. Kolle), jochen.kuehnl@beiersdorf.com (J. Kuehnl), nathalie.lambrechts@vito.be (N. Lambrechts), malin.lindstedt@immun.lth.se (M. Lindstedt), marion.millet@pierre-fabre.com (M. Millet), steissier@rd.loreal.com (S. Martinozzi-Teissier), andreas.natsch@givaudan.com (A. Natsch), ian.pike@proteomics.com (I. Pike), sakaguchi.hitoshi@kao.co.jp (H. Sakaguchi), andreas.schepky@beiersdorf.com (A. Schepky), mtailhardat@research.lvmh-pc.com (M. Tailhardat), marie.templier@pierre-fabre.com (M. Templier), erwin.vanvliet@secam-ce.eu (E. van Vliet), Gavin.Maxwell@unilever.com (G. Maxwell).

[1] Contributed equally.
[2] Members of the Cosmetics Europe Skin Tolerance Task Force, Cosmetics Europe, Avenue Herrmann Debroux 40, B-1160 Auderghem, Brussels, Belgium.
[3] Current address: European Commission, DG Joint Research Centre (JRC), Institute for Health and Consumer Protection (IHCP), via Enrico Fermi 2749, 21027 Ispra (VA), Italy.

http://dx.doi.org/10.1016/j.tiv.2014.10.018
0887-2333/© 2014 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Keywords:
Skin sensitisation
Testing strategy
Safety assessment
Non-animal test methods
Adverse Outcome Pathways

potency prediction. The results of the first phase – systematic evaluation of 16 test methods – are presented here. This evaluation involved generation of data on a common set of ten substances in all methods and systematic collation of information including the level of standardisation, existing test data, potential for throughput, transferability and accessibility in cooperation with the test method developers. A workshop was held with the test method developers to review the outcome of this evaluation and to discuss the results. The evaluation informed the prioritisation of test methods for the next phase of the non-animal testing strategy development framework. Ultimately, the testing strategy – combined with bioavailability and skin metabolism data and exposure consideration – is envisaged to allow establishment of a data integration approach for skin sensitisation safety assessment of cosmetic ingredients.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

1. Introduction

The mechanism behind skin sensitisation and the elicitation of Allergic Contact Dermatitis (ACD) has been investigated for many years and is documented by the OECD as an Adverse Outcome Pathway (AOP) (OECD, 2012). The skin sensitisation AOP captures the impact of skin exposure to sensitising chemicals as a series of biological and chemical key events, which have been reviewed extensively, e.g. by Ainscough et al. (2013), Kimber et al. (2012), Martin et al. (2011), and Toebak et al. (2009). In brief, as a prerequisite, the chemical sensitizer needs to penetrate the stratum corneum as the uppermost layer of the skin in order to become available to the viable cells of the epidermis. It binds covalently to skin proteins of the viable cells (key event 1) to form hapten-protein conjugates, which can be immunogenic. In parallel, keratinocytes become activated and release danger signals e.g. pro-inflammatory cytokines as a response to trauma (key event 2). Next, the phenotype of dendritic cells (DC) changes by the concerted recognition of hapten-protein conjugates by MHC (major histocompatibility complex) molecules and of danger signals (key event 3). The activated DCs mobilise and migrate, after maturational changes, from the skin to the draining lymph node to present the allergen to T cells. After binding to a hapten-peptide specific T cell this clone will expand (key event 4) to elicit the eventual adverse outcome in case of a second exposure with the chemical sensitiser. This level of mechanistic understanding has enabled the development of a multitude of non-animal test methods that each aim to measure the impact of substances on one or more of the AOP key events and therefore to distinguish sensitisers from non-sensitisers or to generate potency information (reviewed previously in Adler et al. (2011)). The complexity of the underlying biology has resulted in the hypothesis that no single measurement will be sufficient to predict sensitiser potency alone (Jowsey et al., 2006). Consequently efforts to apply data from these non-animal test methods for hazard characterisation or risk assessment have focussed upon integration of multiple data types (for example, MacKay et al., 2013; Jaworska et al., 2011; Bauch et al., 2012; Nukada et al., 2012; Natsch et al., 2013). Whilst these approaches continue to show promise, the majority have focused upon integrating non-animal data to predict sensitiser potential. Consequently, one major objective of the Cosmetics Europe Skin Tolerance Task Force has been to identify and evaluate test methods that could allow sensitiser potency prediction without the need for new animal test data, which is of vital importance for the cosmetics industry (Maxwell et al., 2011). This evaluation will inform the development of a non-animal testing strategy for skin sensitisation potency predictions. The resulting strategy will ultimately become an essential part – along with consideration of exposure and other information such as bioavailability or metabolism – of a data integration approach for the skin sensitisation safety assessment of cosmetic ingredients.

Here we document the first of three phases to develop such a non-animal testing strategy. Sixteen test methods were identified for systematic evaluation, following a review of the available scientific literature. The aim of this evaluation was to gain comparable detailed understanding of the test methods that would allow promising methods to be prioritised for further in-depth evaluation. Therefore, a common set of criteria was assessed involving test method characterisation and standardisation. Such criteria included AOP mapping, ease of transferability, availability and throughput, performance (in terms of reproducibility and predictivity) as well as legal aspects and information. The information was assembled for each test method in collaboration with the developers. In addition, we have compiled data on a set of ten substances for each of the methods to verify publically available data in terms of both sensitiser potential and potency prediction. The resulting analysis forms a comprehensive review of the results obtained, which informed the selection of test methods for the next evaluation phases. Finally, we present our future framework set-up for the development of a non-animal testing strategy for skin sensitisation potency predictions – a data and knowledge gap identified by a previous review of non-animal risk assessment approaches for skin sensitisation (Goebel et al., 2012).

2. Material and methods

2.1. Description of test methods

The following section provides an overview of the 16 test methods, which were analysed during the first phase of the Cosmetics Europe method evaluation process. They are presented according to their alignment to the skin sensitisation AOP (Fig. 1). The description, which covers the status at the beginning of 2013, comprises the test system, read-out parameter, prediction model, and whether the method provides only hazard identification or also includes potency prediction. Finally, the experimental conditions are summarised (including the applied dose range) as this may indicate whether the data obtained have the potential to add information to hazard characterisation beyond the currently used prediction model. As detailed information about each of the test methods is already available in the scientific literature, this is not covered here. The laboratories in which the methods have been developed are indicated and key references are included for further reading.

2.1.1. Protein reactivity test methods

Skin sensitisers show a high diversity in terms of chemical and physiochemical properties. However the AOP considers, chemicals – or in case of pre-/pro-haptens, their respective metabolites – which act as sensitisers due to their ability to react with skin proteins (haptenation). This common characteristic is used in a number of non-animal test methods to differentiate between sensitisers and non-sensitisers. Two *in chemico* assays focus on peptide reactivity using two model peptides as surrogates for cellular proteins. In addition, three cell line assays use the kelch-like ECH-associated protein 1 (Keap1) as an intracellular sensor to

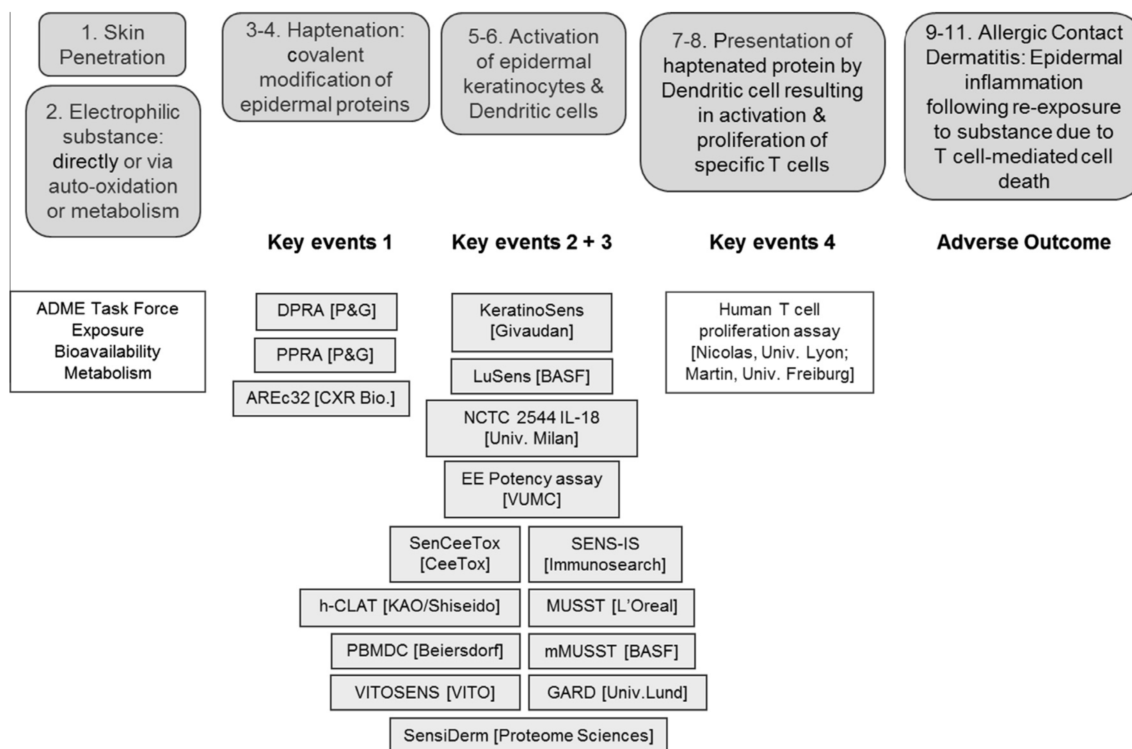

Fig. 1. Non-animal test methods and their alignment to the skin sensitisation AOP. Test methods analysed during of phase I of the Cosmetics Europe method evaluation study (grey boxes). Methods presented in white boxes represent Cosmetics Europe-funded studies to cover the steps in the AOP, which are currently not represented by a non-animal assay.

investigate the reactivity of the test substance. Covalently binding to cysteine residues of Keap 1 causes this repressor protein to delocalize from the transcription factor NF-E2 p45-related factor 2 (Nrf2) which can then bind to and activate antioxidant response element (ARE) containing promoters.

Whilst all five protein reactivity methods reflect the well established importance of interaction between electrophilic haptens and nucleophilic target proteins, the cell line based assays address in addition the induction of cytoprotective mechanisms (referring to AOP key event 2). KeratinoSens™ and LuSens furthermore provide the potential for keratinocyte metabolism of pro-haptens.

2.1.1.1. Direct peptide reactivity assay (DPRA, Procter & Gamble)

The DPRA is a chemistry-based assay that evaluates reactivity of a test compound using two synthetic model peptides including a lysine or cysteine residue. A solution of peptide and test substance in a ratio of 1:10 for cysteine and 1:50 for lysine is incubated for 24 h. After the incubation period, the remaining concentration of the free peptide is measured by high performance liquid chromatography (HPLC) with gradient elution and ultraviolet (UV) detection at 220 nm. Depending on the data obtained from triplicate reactions, averaged peptide depletion of cysteine, lysine or both are used in classification tree models to identify substances as sensitising or non-sensitising. In addition, the prediction model allows the allocation of the protein to the reactivity classes minimal, low, moderate and high (Gerberick et al., 2004, 2007).

2.1.1.2. Peroxidase peptide reactivity assay (PPRA, Procter & Gamble)

The PPRA was developed from the DPRA in order to better identify potential pro- and pre-haptens. Eight concentrations of chemical are tested – instead of one concentration as in the DPRA. The cysteine peptide is incubated for 24 h in the presence and absence of horseradish peroxidase/hydrogen peroxide (HRP/P), whilst the lysine peptide is used only without HRP/P. Following incubation, all concentrations are processed for analysis by HPLC/mass spectrometry (MS)/MS (maximum concentrations of test compound: 5 mM for cysteine, 25 mM for lysine). From the concentration–response peptide depletion data the effective concentration of a test substance that depletes peptides by 25% (i.e., EC25) is estimated by fitting a three-parameter log–logistic model. Substances with an EC25 $\geqslant$ 0.1 mM are considered 'reactive' and those with an EC25 < 0.1 mM are considered 'highly reactive'. Both are therefore classified as 'sensitisers', while substances with less than 15.1% depletion at any concentration are considered 'minimally reactive' and classified as 'non-sensitisers' (Gerberick et al., 2009).

2.1.1.3. AREc32 cell line assay (CXR Bioscience)

The AREc32 cell line assay was the first method exploiting the activation of the Keap1/Nrf2/ARE pathway using a breast cancer cell line (MCF-7), which contains a luciferase gene construct controlled by eight copies of the ARE cis-enhancer element (Wang et al., 2006). The cytotoxicity of the substances is investigated in parallel by measuring adenosine triphosphate (ATP) levels. Luciferase expression at 50% above the vehicle control value is selected as representative of significant induction in any of the applied seven concentrations (max. 100 μM). Hence, test items that induce luciferase expression above this threshold are considered as potential sensitising. More recently, Natsch and Emter proposed to replace the intracellular ATP measurement by the MTT assay (Natsch and Emter, 2008).

2.1.1.4. KeratinoSens™ (Givaudan)

Using the metabolic-competent human keratinocyte HaCaT cell line, the developers of the KeratinoSens™ test method transferred a stable insertion of a luciferase gene under the control of the ARE-element of the human gene AKR1C2, which has been shown to be a key sensitiser-induced gene. These cells are exposed to 12 concentrations of a test substance (max. 2000 mM) for 48 h. Luciferase induction and cytotoxicity as determined with the MTT assay are then evaluated. For luciferase expression the maximal fold-induc-

tion over solvent control (I_{max}) and the concentration needed to reach a 1.5-fold induction (EC1.5) are calculated. For cytotoxicity the IC50, i.e. the concentration inducing 50% of the maximum cytotoxicity, value is derived. A test substance is being identified as sensitiser if the I_{max} shows a >1.5-fold gene induction, this induction is statistically significant above the solvent control value and the EC1.5 value is below 1000 μM in at least two of three repetitions. In addition, at EC1.5, cellular viability needs to be above 70% (Emter et al., 2010; Natsch et al., 2011).

2.1.1.5. LuSens (BASF)

The LuSens assay uses a keratinocyte-derived cell line, to which a luciferase gene under the control of an ARE promoter (from the NADPH:quinone oxidoreductase 1 rat gene) was inserted (Bauch et al., 2012). In a range finding experiment the cytotoxicity of 12 test substance concentrations is evaluated by determination of a CV75 using the MTT assay. In the main experiments, a minimum of six concentrations are applied, up to a maximum of 1.2 × CV75 of the range finding experiment. For non-toxic substances the maximum concentration (2000 μM) is selected to start the concentration range. Luciferase activity and cytotoxicity is measured after 48 h of treatment. A test substance is considered to exhibit a keratinocyte activating potential if the luciferase activity exceeds 1.5-fold induction with respect to the vehicle control, at a concentration that does not reduce a viability to below 70%.

2.1.2. Keratinocyte based test methods associated to inflammatory-related read-outs

Keratinocytes are relevant to the manifestation of inflammatory effects in the skin in response to haptens, which is important for the activation of hapten-presenting dendritic cells (DC) and for inducing their migration to adjacent lymph nodes. There is growing evidence that the induction of the innate immune system by so-called 'danger signals' is mediated by the same pathways as first described for microbial pathogens (Martin et al., 2011). Danger signals created by pathogen invasion or chemical penetration through the stratum corneum activate keratinocytes to produce inflammatory mediators, such as IL-18 and IL-1β, which in turn activate DCs during the sensitisation process. These processes relate to key event 2 in the skin sensitisation AOP.

2.1.2.1. NCTC 2544 IL-18 assay (Università degli Studi di Milano)

The NCTC 2544 assay is based on the detection of intracellular IL-18 expression by the keratinocyte cell line NCTC 2544. In a dose range-finding experiment, 12 concentrations are used to determine the concentration resulting in a cell viability of 80% (CV80), as assessed by the MTT assay. The CV80 then defines the highest of four concentrations used in the main experiment. If at least one non-cytotoxic concentration induces a 1.2-fold increase in intracellular IL-18, and this increase in IL-18 is statistically significant compared to vehicle treated cells (Dunnett multiple comparisons test), in at least two out of three independent experiments, the substance is classified as a sensitiser. Otherwise it is considered non-sensitising (Corsini et al., 2009; Galbiati et al., 2011).

2.1.2.2. Epidermal equivalent (EE) potency assay (VU University medical Centre)

The epidermal equivalent (EE) potency assay aims to classify sensitiser potency using epidermal equivalents, which requires prior identification of a substance as a sensitiser. In the literature, the NCTC 2544 IL-18 assay has been used to provide this information. Substances (spread on filter papers) are topically applied to the EE at a range of 12 concentrations for 24 h. The effective chemical concentration required to reduce cell viability by 50% relative to vehicle-exposed culture (EE-EC50) is calculated using the MTT assay. The EE-EC50 are then assigned to a potency category using a prediction model correlating previous results with local lymph node assay (LLNA) data (dos Santos et al., 2011; Gibbs et al., 2013).

2.1.3. Test methods using dermal dendritic cell surrogates

The following five assays use primary cells or cell lines as surrogates for dermal DC. Protein expression of cell surface markers which are considered prerequisites for the migration of dermal DC to the adjacent lymph nodes (key event 3 in the skin sensitisation AOP) such as CD54 and/or CD86 or gene expression of e.g. chemokine receptor (CCR)2 are used as measurements of cell activation.

2.1.3.1. Human cell line activation test (h-CLAT, Kao and Shiseido)

The h-CLAT assay uses THP-1 cells (a human monocytic leukemia cell line) as a surrogate for dermal dendritic cells. The THP-1 cells are treated with eight different concentrations of a test substance for 24 h. After removing the test substance, expression of CD86 and CD54 is measured by flow cytometry. Relative fluorescence intensity (RFI) compared to vehicle-only treated control cells is used as an indicator of CD86 and CD54 induction. A test substance is considered a skin sensitiser in case the RFI of either CD86 or CD54 reaches defined thresholds (CD86 ⩾ 150% and/or CD54 ⩾ 200%), in at least two of three independent measurements at any concentration. Concentrations exceeding 50% cytotoxicity, measured with propidium iodide (PI), are excluded from analysis (Ashikaga et al., 2010).

2.1.3.2. Myeloid U937 skin sensitisation test (MUSST, L'Oréal)

The MUSST assay, which uses the U937 cell line (a human histiocytic leukemia cell line) is designed to evaluate the capacity of a substance to induce dendritic cell activation. To achieve this, CD86 expression is assessed by flow cytometry, following a 45 h incubation with the test substance in at least four different concentrations up to a maximum of 200 μg/mL. Concentrations exceeding 30% cytotoxicity, measured with PI, are excluded from analysis. A substance inducing an increase in CD86 protein expression of ⩾150% with evidence of a dose response in at least two concordant experiments is considered to be a sensitiser. If the CD86 positive threshold is not reached and no perturbations are observed in at least two concordant experiments, the substance is considered to be a non-sensitiser. In the other cases, rules based on CD86 expression or cell viabilities are used in order to classify the chemical as sensitising or non-sensitising (Ade et al., 2006).

2.1.3.3. Modified myeloid U937 skin sensitisation test (mMUSST, BASF)

The mMUSST also uses the U937 cell line measuring CD86 by flow cytometry. Five concentrations, chosen based on preliminary PI cytotoxicity assays, are applied for 48 h. The highest tested concentration in the main experiment is two times the concentration causing a cytotoxicity of 25% (CV75). A test substance is predicted to have a dendritic cell line activating potential when CD86 induction exceeds the threshold of 1.2 with respect to vehicle treated cells at any tested concentration showing sufficient cell viability (⩾70%) in at least two independent experiments (Bauch et al., 2012).

2.1.3.4. Peripheral blood monocyte-derived dendritic cell assay (PBMDC, Beiersdorf)

In contrast to the above cell line-based assays, the PBMDC assay uses human peripheral blood monocyte-derived dendritic cells isolated from the fresh buffy coats of five different donors. CD1a negative/CD14 positive monocytes are selected and differentiated by culturing with GM-CSF and IL-4. Cells are then exposed to at least six concentrations of the test substance. The second highest concentration should correspond to a viability of at least 80%. Noncytoxic substances are applied up to 1000 μM or the highest solu-

ble concentration. Following 48 h of stimulation, CD86 expression is determined by flow cytometry. Dead cells are detected using 7-Aminoactinomycin (7-AAD) staining. If a test substance induces on average ⩾20% increase in CD86-positive cells compared to non-treated cells it is considered as a skin sensitiser. The acceptable relative cytotoxicity range is limited to ⩽20% (Reuter et al., 2011).

2.1.3.5. VITOSENS (VITO)

The VITOSENS assay uses differentiated $CD34^+$ progenitor cells derived from human cord blood as surrogate for DC. The response to test substance exposure is evaluated by comparing the fold change in the expression of CCR2 (C–C chemokine receptor type 2) and the transcription factor cAMP responsive element modulator (CREM) compared to solvent-exposed cells (Hooyberghs et al., 2008). In a concentration range-finding experiment using cells from one donor, the concentration that yields around 20% cell death (IC20) at 24 h is determined using PI staining and flow cytometry. Next, the cells are exposed to a dilution series including the IC20 concentration or, in case of a non-cytotoxic substance, with the highest soluble concentration. After 6 h, 0.5 million cells are collected for later RNA extraction and subsequent qPCR of CREM and CCR2 to analyse their relative gene expression. After 24 h, the remainder of the cells is collected and the cell viability is determined using PI. The concentration that is then confirmed to induce 20% cell death in all donors is used for the molecular analysis and prediction of the sensitisation outcome. The experimental set-up is repeated on cell cultures from two different cord blood donors. In case of discordant results, a third donor is tested. The resulting fold changes are combined by a weighted average to predict whether the substance is sensitising or non-sensitising. Furthermore, the fold changes of CREM and CCR2 can be combined with the IC20-value in a tiered approach for potency prediction (is Lambrechts et al., 2010, 2011).

2.1.4. Other test methods

The methods described previously use one or two read-out parameters to provide information on the sensitising potential or potency of a test compound. The following methods were allocated to this section as they investigate a set of 10–200 parameters and so may have the ability to provide further insight into the mechanism by which a specific compound induces skin sensitisation. Note that both GARD and SensiDerm™ use surrogates of dendritic cells (see Section 2.1.3) and Sens-IS and SenCeeTox expose 3D epidermal skin tissues addressing substance activation by keratinocytes as well as the cytotoxicity of a substance (see Sections 2.1.1 and 2.1.2).

2.1.4.1. Sens-IS (Immunosearch)

The Sens-IS method classifies sensitisers according to potency categories based on the expression profiles of 65 genes, which are grouped in one gene set for irritancy and two (SENS-IS and ARE) for sensitisation (Cottrez, 2011). EpiSkin tissues (Skinethic, France) are exposed with the test substances using the protocol for determining in vitro skin irritation described in OECD Guideline 439 (15 min incubation, then washing and post-washing incubation of 6 h). A test substance is considered to be sensitiser if it increases the expression (compared to the solvent control) of at least 7 genes measured by qPCR in either the "SENS-IS" or the "ARE" gene sets. To take into account non-specific gene overexpression due to cell stress, the induction of more than 20 genes in the irritation gene set, classifies a result as inconclusive and the test substance is re-analysed at a lower concentration. Similarly to the LLNA, potency is classified according to the concentration of test material needed to induce a positive response: positive at 0.1%: extreme; positive at 1%: strong; positive at 10%: moderate; positive at 50%: weak. Sens-IS is considered to mainly address key event 2 from the skin sensitisation AOP, but may, as ARE-activated genes are included, also provide information on protein reactivity of a test chemical.

2.1.4.2. SenCeeTox (Cyprotex, LLC)

The SenCeeTox method is a test battery of three independent assays addressing several key events to provide information on the skin sensitisation potential of substances and to assign them to a certain subset of potency categories (McKim et al., 2012). Protein reactivity is evaluated in a cell-free manner by measurement of the concentration of free glutathione (GSH) after incubation with the test substance for 24 h at room temperature. The amount of free GSH is determined by a colorimetric assay with 5,5′-Dithio-bis(2-nitrobenzoic acid (DTNB) in relation to the vehicle control. An epidermal skin equivalent (EpiDerm™, MatTek, MA) is used for gene expression analysis and cytotoxicity determination. Viability of skin tissues is measured by assaying for lactose dehydrogenase (LDH) activity. Expression of four housekeeping and seven target genes (NADPH-quinone oxidoreductase 1, Aldoketoreductase 1C2, Interleukin 8, Cytochrome P450 1A1, Aldehyde dehydrogenase 3A, Heme-oxygenase 1, Glutamate cysteine ligase catalytic subunit C) is monitored after topical exposure of the model skin tissues to the test substances at a range of six concentrations (0.1, 5, 100, 250, 500, and 2500 μM) for 24 h. Concentrations, which result in cell viability of less than 50% compared to the vehicle control, are disregarded for the determination of the sensitising potential/potency. Finally, a gated algorithm is used to transform the viability, gene induction and glutathione reactivity data into a toxicity index for each substance. This method covers key event 1 (in terms of protein reactivity) and 2 (in terms of keratinocyte activation) in the skin sensitisation AOP.

2.1.4.3. Genomic allergen rapid detection assay (GARD assay, Lund University)

The GARD assay uses proliferating MUTZ-3 cells (a human myeloid leukemia-derived cell line) to measure gene expression induced by test substances. In this cell type, a biomarker signature was identified by analysing the transcriptional profiles of immature cells challenged with a panel of 38 reference substances (18 sensitisers and 20 non-sensitisers). The biomarker signature of 200 genes with the most discriminatory power to separate between skin sensitisers and non-sensitisers was obtained by employing an algorithm for backward elimination (Johansson et al., 2011). To test a substance, cells are treated for 24 h with a maximum concentration of 500 μM for highly soluble non-toxic substances or a concentration yielding 90% viability for toxic substances as measured with PI. Following cell stimulation, the transcriptional levels of the 200 genes, collectively termed the predictive biomarker signature, is evaluated using a whole genome array (Johansson et al., 2013). Classifications of unknown compounds as sensitisers or non-sensitisers are performed with a support vector machine (SVM) model, trained on the 38 reference chemicals used for GARD development, and the output is a decision value as compared to the classification threshold. Key event 3 is covered with this test method.

2.1.4.4. SensiDerm™ TMT-SRM 10-Plex (Proteomics)

SensiDerm™ aims to discriminate sensitisers and non-sensitisers based on pathway-specific biomarker proteins induced in the MUTZ-3 cell line. The biomarker panel comprises the following ten proteins which have been shown to be differentially expressed in MUTZ-3 cells in response to sensitisers compared to non-sensitisers during the assay development: glucose-6-phosphate-1-dehydrogenase, 6-phosphoglucote dehydrogenase, heat shock protein A8, myeloperoxidase (light/heavy chain), S100A4 protein, S100A8 protein, S100A9 protein, 4F2 cell surface antigen heavy chain, superoxide dismutase, thymosin beta-4-like protein.

MUTZ-3 cells are exposed to non-toxic concentrations (>80% viability) of the test substance for 24 h with a maximum concentration of 100 μg/mL. The cellular proteins are then extracted and analysed by mass spectrometry procedure based on selective reaction monitoring. The results of the tests are provided as a ratio of protein expression between the exposed cells and cells grown in a control medium, which is then subjected to a polynomial model that provides a score with a threshold to discriminate sensitisers from non-sensitisers (Thierse et al., 2011). This method addresses key event 3 in the skin sensitisation AOP.

2.2. Method evaluation

2.2.1. Testing of ten substances

In order to obtain a common data set for all test methods, ten substances were selected (see Table 2 for identities). The chemicals were purchased from Sigma–Aldrich with at least 95% purity, with the exception of Lactic acid (approx. 90%), then coded and distributed to the test method developers by Cosmetics Europe. They comprised three non-sensitisers including SLS, which is positive in the LLNA, and seven sensitisers covering all sensitiser potency classes as defined by the LLNA (1 weak, 3 moderate, 2 strong, 1 extreme) including the poorly water-soluble lauryl gallate as a specifically challenging substance. Test methods developed by member companies of Cosmetics Europe (i.e. DPRA, h-CLAT, MUSST and PBMDC) provided existing data on these ten substances from non-blinded testing, if available. Cosmetics Europe collected, decoded and evaluated the respective results. As a minimum, test developers were asked to complete a checklist including the results but also e.g. information on timing or protocol adherence. If provided or available, further supplementary information including the test protocol, publications or raw test data were collected.

2.2.2. Evaluation process and criteria

Information on 15 of the 16 test methods was compiled systematically to enable evaluation on the basis of criteria that were defined by the Cosmetics Europe Skin Sensitisation Task Force. The PPRA is not included in this compilation because its standardisation was finalised only after evaluation had commenced. Twenty evaluation criteria addressing various aspects of interest were considered. For clarity, these were grouped under the headings ‘General points’, ‘Standard Operation Procedure (SOP) and prediction model’, ‘data’, ‘ease of transfer’ and ‘throughput’ (Table 1). Each test method was also mapped onto the skin sensitisation AOP (Fig. 1).

2.2.3. Data analysis and method evaluation

The data analysis focused on the test results for the ten substances. These were available for all 16 methods. The completeness of results and their concordance with the pre-defined reference results based on LLNA EC3 values (and human data for SLS) was evaluated. If data on other substances were available, overall sensitivity, specificity and concordance were calculated. For the 15 test methods differentiating non-sensitising and sensitising substances, the reference results were derived from both LLNA EC3 values distinguishing five potency categories and in parallel from human data using six classes (Basketter et al., 2014). Both result in the same potential, for nine substances (no EC3 value: non-sensitiser; EC3 value: sensitiser; human potency classes 5 and 6: non-sensitiser; human potency classes 1–4: sensitiser). As SLS is false positive in the LLNA compared to human, it was considered as a non-sensitiser. The seven test methods attempting to predict skin sensitisation potency results used method-specific potency categories that did not necessarily correspond to those of the reference results. Therefore, the potency prediction results are described only, without detailed predictivity analysis.

Table 1
Evaluation criteria.

General points	Does the method address hazard (S/NS), GHS classification (1A, 1B, NS) or risk assessment (potency)? Is it clear what AOP step or key event the method is aiming to predict? (i.e. mechanistic relevance of method) Description of test system (how close to in vivo situation) Evidence-based approach (mechanisms) versus unspecific biomarker
SOP/PM	Is a SOP available? Is a prediction model available? How many chemicals have been used to define the PM? Do the methods provide dose response information?
Data	How many data are available with additional chemicals apart from the 10 common ones? Sensitivity Specificity Overall concordance For how many of the 10 chemicals could a definite call not be provided?
Ease of transfer	Is the method covered by IP rights? Has the method already been transferred to other labs? If so, assessment of the transfer Has the method been investigated in a pre-validation? If so, assessment the prevalidation Is it available at a CRO?
Throughput	How many chemicals can be subjected to one experiment? How much time is needed for one experiment (not considering concentration-range finding)? How many valid experiments are needed for a call?

The focus of the method evaluation exercise was to establish a harmonised knowledge base for each of the test methods in order to prioritise methods for further consideration. This evaluation was carried out in close collaboration with the test method developers, whose review concluded the evaluation process. The method developers were invited to a two-day workshop with the Cosmetics Europe Skin Tolerance Task Force held in Brussels on December 3rd and 4th 2012 to discuss their methods and results, the requirements of the cosmetics industry and the strategy of the task force to meet these needs. It should be noted that the data presented here basically reflect the development status of the methods as presented at the workshop. For some methods discussions were extended into early 2013.

3. Results

3.1. Common set of ten substances

3.1.1. Skin sensitisation potential

Fifteen of the evaluated methods reported skin sensitisation potential predictions for the ten substances. These predictions are summarised in a harmonised way as non-sensitiser (NS) and sensitiser (S) (Table 2) alongside the reference results. While all ten substances were tested in all methods, for one method (SensiDerm) inconclusive data were reported because timing constraints did not allow completion of the necessary repeat experiments to reach a final prediction. With one exception, all test methods misclassified a maximum of two substances. The three sensitisers 4-nitrobenzylbromide, cinnamal and tetramethyl thiuram disulphide were correctly identified by all test methods, whereas the sensitisers methyldibromoglutaronitrile, 2-mercaptobenzothiazole and lauryl gallate (selected as challenging due to its poor water solubility), were not classified in up to two test methods. Most challenging was phenyl benzoate, which was misclassified as a non-sensitiser by six test methods. Of the three non-sensitisers, salicylic acid and lactic acid were mis-classified as sensitising by one test method each, while SLS, which is false positive in LLNA

Table 2
Hazard classification by 15 test methods for the ten chemical set. The EE potency assay is not reported here, as this method provided only potency.

Test substance	CAS number	Reference result: potential	AREc32	DPRA	GARD	h-CLAT	Keratino-Sens™	Lu-Sens	mMUSST	MUSST	NCTC2544	PBMDC	PPRA[d]	SenCeeTox	SensiDerm	Sens-IS	VITO-SENS
2-Mercaptobenzothiazole	149-30-4	S[a]	NS	S	S	S	S	S	S	S	NS	S	S	S	-	S	S
4-Nitrobenzylbromide	100-11-8	S[b]	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Cinnamal	104-55-2	S[a]	S	S	S	S	S	S	S	S	S	S	S	S	-	S	S
Lactic acid	50-21-5	NS[a]	NS	NS	NS	NS	NS	NS	NS	NS	S	NS	NS	NS	NS	NS	NS
Lauryl gallate	1166-52-5	S[a]	S	S	S	S	S	NS	S	S	S	S	S	S	-	S	S
Methyldibromoglutaronitrile	35691-65-7	S[a]	S	S	S	S	S	S	S	S	NS	NS	S	S	S	S	S
Phenyl benzoate	93-99-2	S[a]	S	S	NS	S	NS	NS	S	S	NS	NS	NS	S	-	S	NS
Salicylic acid	69-72-7	NS[a]	NS	NS	NS	S	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Sodium lauryl sulphate	151-21-3	NS[a,c]	S	NS	NS	NS	NS	NS	S	NS	NS	NS	NS	S	NS	NS	NS
Tetramethyl thiuram disulphide	137-26-8	S[a]	S	S	S	S	S	S	S	S	S	S	Insoluble	S	S	S	S
Concordance			8/10	10/10	9/10	9/10	9/10	8/10	9/10	10/10	6/10	8/10	8/9	9/10	6/6	10/10	9/10

NS: non-sensitiser; S: Sensitiser; -: inconclusive.
[a] Basketter et al. (2014).
[b] Gerberick et al. (2005).

Table 3
Potency classification for the ten substance set by each of seven test methods which predict sensitiser potency.

Test substance	CAS number	Reference results		DPRA	EE	KeratinoSens™	PPRA	SenCeeTox	Sens-IS	VITOSENS
		Potency based on LLNA)	Human potency class[a]							
2-Mercaptobenzothiazole	149-30-4	Moderate[a]	3	High	na (MTT interference)	Weak/moderate	Reactive	Moderate	Moderate	Moderate
4-Nitrobenzylbromide	100-11-8	Extreme[b]	na	High	Extreme	Strong/extreme	Highly reactive	Strong	Extreme	Extreme
Cinnamal	104-55-2	Moderate[a]	2	High	Strong	Moderate	Reactive	Moderate	Strong	Moderate
Lactic acid	50-21-5	NS[a]	6	Minimal	Not tested	NS	Minimally reactive	NS	NS	NS
Lauryl gallate	1166-52-5	Strong[a]	2	high	na (MTT interference)	strong/extreme	highly reactive	NS-weak	moderate	strong
Methyldibromoglutaronitrile	35691-65-7	Strong[a]	2	High	Strong	Strong/extreme	Highly reactive	Strong-extreme	Strong	Strong
Phenyl benzoate	93-99-2	Weak[a]	3	Moderate	Very weak/NS	NS	Minimally reactive	NS-weak	Weak	NS
Salicylic acid	69-72-7	NS[a]	6	Minimal	Not tested	NS	Minimally reactive	NS-weak	NS	NS
Sodium lauryl sulphate (SLS)	151-21-3	NS[a,c]	6	Minimal	Not tested	NS	Minimally reactive	Weak	NS	NS
Tetramethyl thiuram disulphide	137-26-8	Moderate[a]	3	High	na (solubility)	Strong/extreme	na (solubility)	Moderate	Moderate	Extreme

NS: non-sensitiser; na: not available.
[a] Basketter et al. (2014).
[b] Gerberick et al. (2005).
[c] False positive in the LLNA, thus considered as non sensitiser.

but not found to be a sensitiser in humans, was classified as sensitising by three test methods. Interestingly, some differences in prediction were found with similar test methods. The three ARE cell line assays (KeratinoSens™, LuSens, AREc32) showed concordant results for only six of the ten substances. This was also the case for the test methods based on dendritic cell surrogates (h-CLAT, MUSST, mMUSST, PBMDC), which came to the same conclusion for six substances only. The reasons for these differences remain to be discussed, but are most likely due to differences in the test method protocols such as cells or prediction models used.

3.1.2. Skin sensitisation potency

Of the seven test methods predicting skin sensitiser potency, six do not require prior classification of a chemical as sensitising, but the EE potency assay does. Therefore the three non-sensitisers were not tested in this assay. Potency categories are not defined consistently across different test methods. Sens-IS, VITOSENS and the EE potency assay apply the five LLNA categories from non-sensitiser to extreme, whilst KeratinoSens™ and SenCeeTox in addition allow assignment of substance to intermediate categories such as non-weak or strong/extreme. In contrast, DPRA categorises chemical reactivity with peptides as minimal, low, moderate or high, and the PPRA as minimally reactive, reactive or highly reactive. Table 3 summarises the potency predictions of all seven methods together with the reference results as derived from the LLNA and in terms of human potency categories as reported in Basketter et al. (2014).

The EE potency assay only assesses known sensitisers for their sensitising potency and was susceptible to reduction of cell viability (as shown by MTT staining) or solubility problems for three substances. Therefore test results for only four of the seven sensitisers were available (non sensitisers were not tested). The PPRA encountered solubility issues with tetramethyl thiuram disulphide, but test results were obtained for the remaining nine chemicals. Potency predictions for all ten chemicals were obtained from the other five test methods. With the exception of the strong sensitiser lauryl gallate being predicted as 'NS-weak' in SenCeeTox, potency predictions were either correct or differed to the reference result by only one category in all cases for Sens-IS, KeratinoSens™, VitoSens and SenCeeTox. No bias towards under- or over-prediction of potency was observed.

The DPRA and the PPRA use fewer potency categories than the LLNA. The six substances with LLNA reference results of moderate, strong and extreme were all classified by the DPRA as having 'high' reactivity, phenyl benzoate (classified as weak by the LLNA) as 'moderate' and the three non-sensitisers as 'minimal'. The PPRA classified LLNA extreme and strong sensitisers as highly reactive, the LLNA moderate sensitisers as reactive, and the LLNA weak and non-sensitisers as minimally reactive.

Human skin sensitisation data are available for six of the seven sensitising substances, which were all assigned as human potency class '2' and '3' (Basketter et al., 2014). This correlated well with their classification based on LLNA results – which ranged from weak to strong – with only minor differences for cinnamal and phenyl benzoate. Consequently, the potency prediction from the test methods broadly matched the human potency classes in a similar manner as described above for the LLNA.

At the time of the workshop the h-CLAT had already been proposed for potency predictions (Nukada et al., 2012), but it was not proposed by the test developer for this application at the time of evaluation.

3.1.3. Method evaluation according to the pre-defined criteria

The evaluation of all test methods, except the PPRA (because method standardisation was finalised only after evaluation had commenced), was performed according to the criteria detailed above and is presented in Table 4. In summary, the methods were characterised by the test system (cell line – 9 methods; 3D tissue – 3; primary cells – 2; synthetic peptide – 1) and the number of skin sensitisation biomarkers (specific or non-specific) measured. Regarding conduct of the methods and the data analysis, SOP and prediction models were – unless they were considered as confidential – provided by the test developers. As an indicator of the robustness of the prediction model, the number of chemicals used to develop the model was also captured. For most methods prediction models were based on more than 25 substances, which was considered as sufficient. Similarly, the number of test concentrations used was considered as an indicator for the potential generation of concentration–response data. Whilst three test methods use fewer than four concentrations (which is unlikely to allow investigation of concentration–responses), the five test methods which involve more than six test concentrations are likely to be able to provide this type of information.

The number of substances tested – in addition to the ten test substance set – for which data and/or predictions were available for each method, was captured. This number was smaller than 10 for three test methods. More than 40 substances had been tested in the remaining methods, for which the predictive performance in terms of specificity, sensitivity and concordance with the skin sensitisation potential as determined by the LLNA was calculated. While both sensitivity and specificity ranged from approximately 65% to 100%, the concordance was at least 73%. As many factors, especially the identity and number of substances tested, may have a significant impact on these performance parameters, they should be considered with care as they therefore do not lend themselves necessary for comparison.

Information on transferability and throughput that were used to characterise practical aspects of testing were of particular interest to our evaluation. Intellectual property rights protected about half of the methods. While locally restricted rights – as in the case of the h-CLAT – were of little concern, rights constituting an obstacle to wide and non-exclusive availability of methods were of higher concern. Aspects such as previously successful method transfer, pre-validation activities and the availability of test methods at CROs were of interest in this regard. It was found that most methods had already been transferred or a transfer was planned or ongoing. Likewise, most methods are available at a CRO. Obviously, more established methods, such as the DPRA, KeratinoSens™, PBMDC, MUSST or h-CLAT are more likely to have undergone a validation exercise establishing their transferability and reproducibility.

Regarding the throughput, most methods can test at least six substances in parallel in one experiment. However, the duration and minimum number of required valid experiments may differ considerably. As a consequence, the average time to test a substance may be a short as one week (for example in the DPRA), or also as long as three to four weeks (using VITOSENS).

Based on the information collected, test methods were prioritised based on voting by the Cosmetics Europe member companies represented in the Skin Tolerance Task Force for further evaluation in a more detailed second evaluation phase. For initial data integration exercises, test methods were chosen, for which substantial information was available. Protocol robustness, proven transferability and reproducibility – generally demonstrated by successful multi-laboratory studies – apparently were important test methods characteristics considered in this process, together with amount of existing data and availability through contract research organisations. The voting resulted in the selection of the DPRA, KeratinoSens™, MUSST and h-CLAT for further evaluation. These methods primarily focus on prediction of hazard potential. Therefore methods that were considered to show potential for prediction of skin sensitisation potency and that use gene regulation or proteomics as biomarkers (GARD, SensiDerm™, Sens-IS, SenCeeTox

Table 4

Results of the evaluation of 15 skin sensitisation methods according to pre-defined criteria.

Criterion		AREc32	DPRA	EE potency	GARD	h-CLAT	Keratino-Sens™	LuSens	mMUSST	MUSST	NCTC2544	PBMDC	SenCee-Tox	SensiDerm	Sens-IS	VITOSENS
Gernal points	Description of test system	Cell line	Synthetic peptides	3D tissue	Cell line	Cell line	Cell line	Cell line	Cell line	Cell line	Cell line	Primary cells	3D tissue	Cell line	3D tissue	Primary cells
	Evidence based approach (mechanisms) versus unspecific biomarker: no. of biomarkers.	1–2 (specific)	1–2 (specific)	1–2 (unspecific)	≥3 (specific)	1–2 (specific)	1–2 (specific)	1–2 (specific)	1–2 (specific)	1–2 (specific)	1–2 (specific)	1–2 (specific)	≥3 (specific)	≥3 (specific)	≥3 (specific)	1–2 (specific)
SOP/PM	Is a SOP available?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	Yes
	Is a Prediction Model available?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
	No. of chemicals used to define the PM?	na	>25	<10	>25	>25	na	>25	>25	>25	>25	>25	na	11–25	>25	11–25
	Do the method provide dose response information: no. of concentration used?	>6	≤3	>6	≤3	>6	>6	>6	4–6	4–6	4–6	4–6	4–6	4–6	4–6	≤3
Data	How many data are available with chemicals (besides 10 common)?	<10	≥40	<10	≥40	≥40	≥40	≥40	≥40	≥40	10–39	≥40	≥40	<10	≥40	10–39
	Sensitivity	na	87.3%	na	na[c]	87.5%	88.0%	77.1%	62.8%	86.2%	86.8%	74,4%	na[b]	na	100%	96.9%
	Specificity	na	87.1%	na	na[c]	75.0%	83.3%	66.7%	97.2%	89.5%	95.0%	77.8%	na[b]	na	93.1%	90.9%
	Overall concordance	8/10[a]	87.3%	na	9/10[a]	84.0%	86.3%	73.6%	73.7%	87.5%	89.7%	77.0%	9/10[a]	6/6[a]	96.9%	94.4%
	For how many of the 10 common chemicals could not a definite call be provided	0	0	≥3	1	0	0	0	0	0	0	0	0	≥3	0	0
Ease of transfer	Is the method covered by intellectual property rights?	Yes	No	No	Yes	Yes (in Japan only)	Yes	Yes	No[d]	No[d]	No	No	Yes	Yes	Yes	Yes
	Has the method already been transferred to other labs? If so, assess it.	No	Yes: Transfer to ≥2 labs and well reprod.	Yes: Transfer to ≥2 labs and well reprod	Planned/ ongoing	Yes: Transfer to ≥2 labs and well reprod	Yes: Transfer to ≥2 labs and well reprod	Planned/ ongoing	No	Yes: Transfer to ≥2 labs	Yes: Transfer to ≥2 labs and well reprod.	Yes: Transfer to ≥2 labs and well reprod	Yes: Transfer to 2 labs and well reprod	Yes (no further information)	Yes: Transfer to 1 lab (no details)	Yes (no further information)
	Has the method been investigated in a pre-validation? If so, assess it.	No	Yes (see above)	Yes (see above)	No	Yes (see above)	Yes (see above)	Planned/ ongoing	No	Planned/ ongoing	Yes	Yes (see above)	Planned/ ongoing	No	No	No
	Is it available at a CRO?	Yes	Yes	Yes	Planned/ ongoing	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes
Throughput	How many chemicals can be subjected to one experiment?	≥6	≥6	3–5	≥6	3–5	≥6	≥6	≥6	≥6	≤3	≥6	3–5	≥6	≥6	≥6
	How many working days are needed for one experiment (not considering concentration-range finding)?	≤3	≤3	≤3	>6	≤3	4–6	≤3	≤3	≤3	≤3	>6	≤3	>6	≤3	>6
	How many valid experiments are needed for a call?	2	2	2	2	2	2	2	2	2	3–4	3–4	1	2	2	3–4

[a] Concordance for 10 test substances, sensitivity and specificity not calculated due to small sample size.
[b] Not calculated as no unambiguous hazard predictions available.
[c] Available data were not considered as these were used to develop or adjust the prediction model.
[d] U937 cell line requires a licence; na: not available.

and VITOSENS) were also selected. In addition, the PPRA as a potential improvement of the DPRA was prioritised.

4. Discussion

Cosmetic Europe's Skin Tolerance Task Force is developing a data integration approach for the skin sensitisation safety assessment of cosmetic ingredients. This requires a non-animal testing strategy, which delivers skin sensitisation potency predictions. It is of utmost importance that the strategy is developed in a way that ensures all stakeholders will have a high level of confidence in the produced results. Confidence will be built by (a) incorporating current mechanistic understanding – guided by the OECD AOP, (b) the amount and quality of data used in strategy construction, (c) transparent and objective strategy composition (Jaworska and Hoffmann, 2010) and (d) satisfactory predictive performance. It will need to offer flexibility to adjust to specific purposes, e.g. for cases requiring only hazard identification not potency estimation, and demands (including applicability domain issues). Therefore, we envisage that the term 'strategy' is used here to collectively describe an array of testing and data integration approaches. It is planned that a default or standard strategy for potency prediction will be developed, that is intended for cases without any relevant a priori information on the substance to be tested. In other cases where a priori information exists, or the purpose is not potency estimation, modifications to this default and/or specifically tailored strategies will be available. A-priori information may include

(i) physico-chemical properties, including molecular weight, the octanol–water partition coefficient and physical form at room temperature,
(ii) potency expectation such as derived by read-across or Q(SAR) approaches,
(iii) intended function (e.g. dye) or origin (e.g. natural extracts),
(iv) information on suspected transformation – whether activation or deactivation by oxidation or metabolic processes (e.g. for pre- or pro-haptens),
(v) skin penetration from data or expectations based on modelling approaches,
(vi) available in vitro test method data (where insufficient for concluding on potency).

As examples, approaches to confirm the expectation of a substance being a non-sensitiser or approaches specially suited for lipophilic substances (which may be difficult to test in an aqueous, cell line based assay) are likely to be required.

The testing strategy is expected to provide an ordinal resolution of the potency spectrum preferably distinguishing five categories (non, weak, moderate, strong, extreme). The references for assessing the strategy's performance will be human (six categories) as proposed by Basketter et al. (2014) and LLNA (EC3, categorised in 5 classes) data. EC3 values will be harvested from existing publications and qualify for inclusion only if certain criteria, including the specification of the vehicle, test concentrations and stimulation indices, are fulfilled. Variability associated with replicate EC3 values will be taken into account.

To reach this aim of developing a non-animal testing strategy for potency prediction, three phases frame our efforts (Fig. 2). The outcome of the first of these phases, the evaluation of existing test methods supporting prioritisation for further consideration is presented here. Sixteen test methods with data in common for a set of 10 substances were considered during this evaluation. With the exception of test methods developed by member companies of Cosmetics Europe (i.e. DPRA, h-CLAT, MUSST and PBMDC) that provided existing data from non-blinded testing, coded substances were tested. However, for calculation of the predictivity of most methods including these four, available data on additional chemicals were considered (in most cases $\geqslant$40 substances, Table 4), so that potential impact of coded versus non-coded testing on predictivity become marginal. With the cooperation of the test method developers, additional information relevant to a pre-defined list of criteria that addressed a number of parameters including the level of standardisation, existing test data, potential for throughput, transferability and accessibility was systematically collated. The outcome of this evaluation was reviewed by each test method developer, discussed at a workshop held with the method developers, and ultimately informed the prioritisation of test methods for phase II of the evaluation process.

Initially, the ten test methods DPRA, GARD, h-CLAT, KeratinoSens™, MUSST, PPRA, SenCeeTox, SensiDerm, Sens-IS and VITOSENS were prioritised based on voting by the Cosmetics Europe member companies represented in the Skin Tolerance Task Force. At a later stage, one test method was dropped because significant optimisation would be needed, while another was stopped due to organisational issues. During phases II and III of the Cosmetics Europe framework new developments of existing or up-coming methods such as the efforts by Teunis et al. (2013, 2014), or van der Veen et al. (2015), will be monitored and considered in case they can be expected to improve the testing strategy.

The basis for the testing strategy composition will be more than 100 substances, for which both LLNA and human data are available. It is planned that test results from all eight phase II methods for all substance will be available. For each test method the data considered most useful for the testing strategy composition will be defined. This implies that the potential contribution of read-out parameters – instead of currently applied prediction models – to the strategy will be explored, especially for the methods that on hazard assessment. For example, for the DPRA relative cysteine and lysine depletion will be used. It has to be noted that properties of the data of the various methods differ. While the methods of first priority have a few relevant read-outs to be captured, this will be more complex for other methods, such as Sens-IS or GARD that measure an array of genes. Variability of the methods will be accounted for. While for methods that have undergone a validation exercise, the usually derived information on within- and between-laboratory reproducibility is readily available, this will be less straightforward at least for some of the other methods. As a substantial number of the substances have already been tested for some of the methods, it is expected that the remaining data gaps will be filled soon. This will allow re-assessment of already proposed testing strategies, e.g. by Bauch et al. (2012), Gomes et al. (2012), Nukada et al. (2012), Natsch et al. (2013) or Jaworska et al. (2013), with new data. Once the data for the eight test methods will be available, a testing strategy will be composed addressing the specific purposes and needs as described above. Driven by the mechanistic understanding and supported by data analysis specialists, data mining and other statistical tools will be used to combine test method data in an objective and transparent way to obtain a predictive testing strategy that will be made publicly available. Predictive performance will be assessed correlatively/probabilistically against the reference human and LLNA data or a combination thereof. Although efficiency and other factors, such as availability or duration, may – at least at this stage – not be accounted for, it is nevertheless expected that the strategy will comprise a limited number of test methods.

In the third phase of the framework, applicability domain issues specifically relevant for substance used by cosmetic industry will be addressed. It is anticipated that for inherently problematic substance types, such as natural extracts, dyes or polymers, further data may be required in order to provide sufficient evidence that the testing strategy works for these substance types or to optimise the adaptation of the strategy.

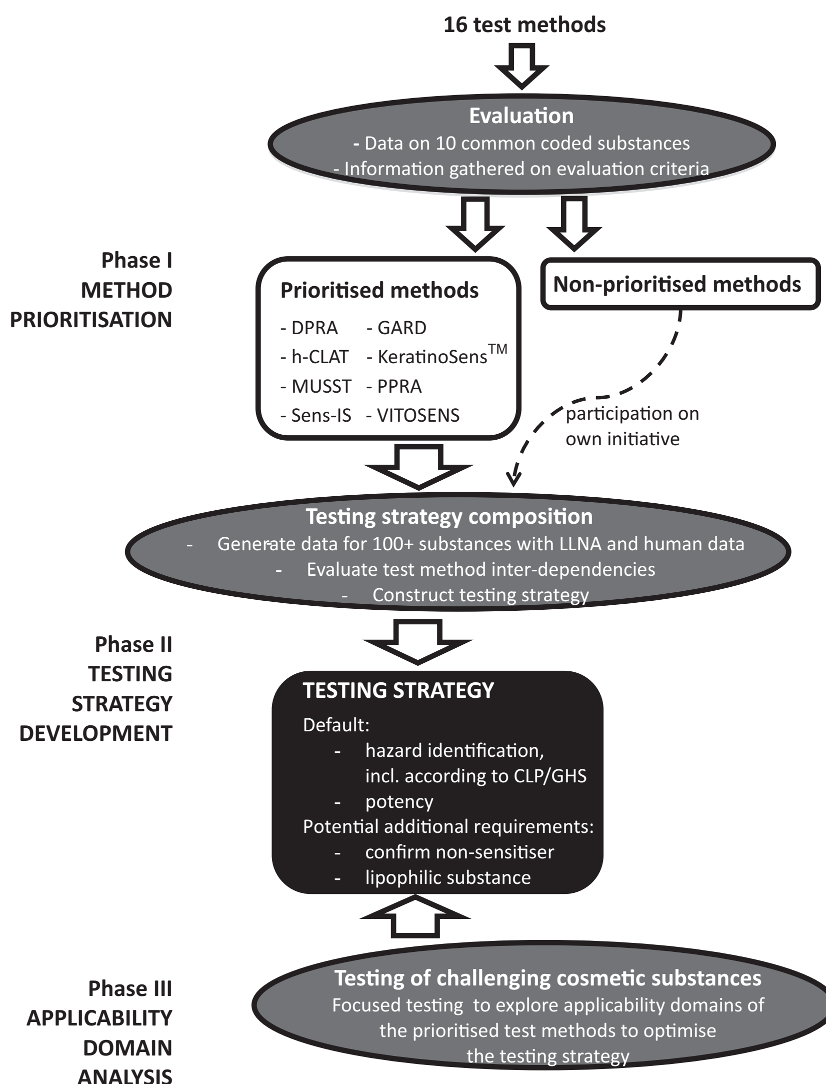

Fig. 2. Cosmetics Europe's three-phase framework for development of a non-animal testing strategy for skin sensitisation potency prediction (CLP: Classification, Labelling and Packaging; GHS: Globally Harmonised System). In phase I 16 non-animal test methods were evaluated based on predefined criteria detailed in Table 1. In phase II an extensive database will be generated comprising more than 100 chemicals, for which human data and LLNA EC3 values are available. This resulting data set will used to develop a test strategy for predicting hazard identification and characterisation (potency) with non-animal test methods only. In phase III the test strategy will be optimised after challenging the prioritized methods with compounds of specific interest for cosmetic industry (e.g. UV-filters, preservatives, hair dyes).

The resulting non-animal testing strategy for skin sensitisation potency predictions will be combined with bioavailability and skin metabolism data, exposure consideration and in exceptional cases with data from T cell activation assays, to satisfy the ultimate goal of a data integration approach for skin sensitisation safety assessment of cosmetic ingredients.

Conflict of Interest

The authors declare that there are no conflicts of interest.

Transparency Document

The Transparency document associated with this article can be found in the online version.

Acknowledgments

We would like to thank Pierre Aeby for his support of the method evaluation leading to the workshop. Furthermore, we would like to thank Silvia Casati (EC, EURL-ECVAM) for her active participation in the workshop.

References

Ade, N., Martinozzi-Teissier, S., Pallardy, M., Rousset, F., 2006. Activation of U937 cells by contact sensitisers: CD86 expression is independent of apoptosis. J. Immunotoxicol. 3, 189–197.
Adler, S., Basketter, D., Creton, S., Pelkonen, O., van Benthem, J., Zuang, V., Andersen, K.E., Angers-Loustau, A., Aptula, A., Bal-Price, A., Benfenati, E., Bernauer, U., Bessems, J., Bois, F.Y., Boobis, A., Brandon, E., Bremer, S., Broschard, T., Casati, S., Coecke, S., Corvi, R., Cronin, M., Daston, G., Dekant, W., Felter, S., Grignard, E., Gundert-Remy, U., Heinonen, T., Kimber, I., Kleinjans, J., Komulainen, H., Kreiling, R., Kreysa, J., Leite, S.B., Loizou, G., Maxwell, G., Mazzatorta, P., Munn, S., Pfuhler, S., Phrakonkham, P., Piersma, A., Poth, A., Prieto, P., Repetto, G., Rogiers, V., Schoeters, G., Schwarz, M., Serafimova, R., Tähti, H., Testai, E., van Delft, J., van Loveren, H., Vinken, M., Worth, A., Zaldivar, J.M., 2011. Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010. Arch. Toxicol. 85, 367–485.
Ainscough, J.S., Gerberick, F.G., Dearman, R.J., Kimber, I., 2013. Danger, intracellular signaling, and the orchestration of dendritic cell function in skin sensitization. J. Immunotoxicol. 10, 223–234.
Ashikaga, T., Sakaguchi, H., Sono, S., Kosaka, N., Ishikawa, M., Nukada, Y., Miyazawa, M., Ito, Y., Nishiyama, N., Itagaki, H., 2010. A comparative evaluation of in vitro

skin sensitisation tests: the human cell-line activation test (h-CLAT) versus the local lymph node assay (LLNA). Altern. Lab. Anim. 38 (4), 275–284.

Basketter, D., Alépée, N., Ashikaga, T., Barroso, J., Gilmour, N., Goebel, C., Hibatallah, J., Hoffmann, S., Kern, P., Martinozzi-Tessier, S., Maxwell, G., Reisinger, K., Sakaguchi, H., Schepky, A., Tailhardat, M., Templier, M., 2014. Categorization of chemicals according to their relative human skin sensitising potency. Dermatitis 25 (1), 11–21.

Bauch, C., Kolle, S.N., Ramirez, T., Eltze, T., Fabian, E., Mehling, A., Teubner, W., van Ravenzwaay, B., Lansiedel, R., 2012. Putting the parts together: combining in vitro methods to test for skin sensitising potentials. Regul. Toxicol. Pharmacol. 63 (3), 489–504.

Corsini, E., Mitjans, M., Galbiati, V., Lucchi, L., Galli, C.L., Marinovich, M., 2009. Use of IL-18 production in a human keratinocyte cell line to discriminate contact sensitisers from irritants and low molecular weight respiratory allergens. Toxicol. In Vitro 23, 789–796.

Cottrez, F., 2011. Development of Sens-IS – an EpiSkin based model for measuring chemical sensitisers potency. In: Poster at the 8th World Congress on Alternatives, Montreal (Canada).

dos Santos, G.G., Spiekstra, S.W., Sampat-Sardjoepersad, S.C., Reinders, J., Scheper, R.J., Gibbs, S., 2011. A potential in vitro epidermal equivalent assay to determine sensitiser potency. Toxicol. In Vitro 25, 347–357.

Emter, R., Ellis, G., Natsch, A., 2010. Performance of a novel keratinocyte-based reporter cell line to screen skin sensitisers in vitro. Toxicol. Appl. Pharmacol. 245, 281–290.

Galbiati, V., Mitjans, M., Lucchi, L., Viviani, B., Galli, C.L., Marinovich, M., Corsini, E., 2011. Further development of the NCTC 2544 IL-18 assay to identify in vitro contact allergens. Toxicol. In Vitro 25 (3), 724–732.

Gerberick, G.F., Vassallo, J.D., Bailey, R.E., Chaney, J.G., Morrall, S.W., Lepoittevin, J.P., 2004. Development of a peptide reactivity assay for screening contact allergens. Toxicol. Sci. 81 (2), 332–343.

Gerberick, G.F., Ryan, C.A., Kern, P.S., Schlatter, H., Dearman, R.J., Kimber, I., Patlewicz, G.Y., Basketter, D.A., 2005. Compilation of historical local lymph node data for evaluation of skin sensitisation alternative methods. Dermatitis 16 (4), 157–202.

Gerberick, G.F., Vassallo, J.D., Foertsch, L.M., Price, B.B., Chaney, J.G., Lepoittevin, J.P., 2007. Quantification of chemical peptide reactivity for screening contact allergens: a classification tree model approach. Toxicol. Sci. 97 (2), 417–427.

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

Gibbs, S., Corsini, E., Spiekstra, S.W., Galbiate, V., Fuchs, H.W., Degeorge, G., Troese, M., Hayden, P., Deng, W., Roggen, E., 2013. An epidermal equivalent assay for identification and ranking potency of contact sensitizers. Toxicol. Appl. Pharmacol. 272, 529–541.

Goebel, C., Aeby, P., Ade, N., Alépée, N., Aptula, A., Araki, D., Dufour, E., Gilmour, N., Hibatallah, J., Keller, D., Kern, P., Kirst, A., Marrec-Fairley, M., Maxwell, G., Rowland, J., Safford, B., Schellauf, F., Schepky, A., Seaman, C., Teichert, T., Tessier, N., Teissier, S., Weltzien, H.U., Winkler, P., Scheel, J., 2012. Guiding principles for the implementation of non-animal safety assessment approaches for cosmetics: skin sensitisation. Regul. Toxicol. Pharmacol. 63 (1), 40–52.

Gomes, C., Nocairi, H., Thomas, M., Inanez, F., Collin, J.F., Saporta, G., 2012. Stacking Prediction for a Binary Outcome. Compstat 2012, Proceedings, pp. 271–282 <http://www.compstat2012.org/Proceedings_COMPSTAT2012.pdf>.

Hooyberghs, J., Schoeters, E., Lambrechts, N., Nelissen, I., Witters, H., Schoeters, G., Van Den Heuvel, R.A., 2008. Cell-based in vitro alternative to identify skin sensitisers by gene expression. Toxicol. Appl. Pharmacol. 231, 103–111.

Jaworska, J., Hoffmann, S., 2010. Integrated Testing Strategy (ITS) – opportunities to better use existing data and guide future testing in toxicology. Altex 27, 231–242.

Jaworska, J., Harol, A., Kern, P., Gerberick, G.F., 2011. Integrating non-animal test information into an adaptive testing strategy–skin sensitisation proof of concept case. Altex 28, 211–225.

Jaworska, J., Dancik, Y., Kern, P., Gerberick, F., Natsch, A., 2013. Bayesian integrated testing strategy to assess skin sensitisation potency: from theory to practice. J. Appl. Toxicol., May 14, Epub.

Johansson, H., Lindstedt, M., Albrekt, A.S., Borrebaeck, C.A., 2011. A genomic biomarker signature can predict skin sensitisers using a cell-based in vitro alternative to animal tests. BMC Genomics 12, 399.

Johansson, H., Albrekt, A.S., Borrebaeck, C.A., Lindstedt, M., 2013. The GARD assay for assessment of chemical skin sensitisers. Toxicol. In Vitro 37, 1163–1169.

Jowsey, I.R., Basketter, D.A., Westmoreland, C., Kimber, L., 2006. A future approach to measuring relative skin sensitising potency: a proposal. J. Appl. Toxicol. 26, 341–350.

Kimber, I., Travis, M.A., Martin, S.F., Dearman, R.J., 2012. Immunoregulation of skin sensitization and regulatory T cells. Contact Dermatitis 67, 179–183.

Lambrechts, N., Vanheel, H., Nelissen, I., Witters, H., Van Den Heuvel, R., Van Tendeloo, V., Schoeters, G., Hooyberghs, J., 2010. Assessment of chemical skin sensitizing potency by an in vitro assay based on human dendritic cells. Toxicol. Sci. 116, 122–129.

Lambrechts, N., Nelissen, I., Van Tendeloo, V., Witters, H., Van Den Heuvel, R., Hooyberghs, J., Schoeters, G., 2011. Functionality and specificity of gene markers for skin sensitisation in dendritic cells. Toxicol. Lett. 203, 106–110.

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.

Martin, S.F., Esser, P.R., Weber, F.C., Jakob, T., Freudenberg, M.A., Schmidt, M., Goebeler, M., 2011. Mechanisms of chemical-induced innate immunity in allergic contact dermatitis. Allergy 66, 1152–1163.

Maxwell, G., Aeby, P., Ashikaga, T., Bessou-Touya, S., Diembeck, W., Gerberick, F., Kern, P., Marrec-Fairley, M., Ovigne, J.M., Sakaguchi, H., Schroeder, K., Tailhardat, M., Teissier, S., Winkler, P., 2011. Skin sensitisation: the Colipa strategy for developing and evaluating non-animal test methods for risk assessment. Altex 28, 50–55.

McKim Jr., J.M., Keller 3rd, D.J., Gorski, J.R., 2012. An in vitro method for detecting chemical sensitisation using human reconstructed skin models and its applicability to cosmetic, pharmaceutical, and medical device safety testing. Cutan. Ocul. Toxicol. 31, 292–305.

Natsch, A., Emter, R., 2008. Skin sensitisers induce antioxidant response element dependent genes: application to the in vitro testing of the sensitisation potential of chemicals. Toxicol. Sci. 102 (1), 110–119.

Natsch, A., Bauch, C., Foertsch, L., Geberick, F., Norman, K., Hilberer, A., Inglis, H., Landsiedel, R., Onken, S., Reuter, H., Schepky, A., Emter, R., 2011. The intra- and inter-laboratory reproducibility and predictivity of the KeratinoSens assay to predict skin sensitisers in vitro: results of a ring-study in five laboratories. Toxicol. In Vitro 25, 733–744.

Natsch, A., Ryan, C.A., Foertsch, L., Emter, R., Jaworska, J., Gerberick, F., Kern, P., 2013. A dataset on 145 chemicals tested in alternative assays for skin sensitisation undergoing prevalidation. J. Appl. Toxicol.. Apr 9, Epub.

Nukada, Y., Miyazawa, M., Kazutoshi, S., Sakaguchi, H., Nishiyama, N., 2012. Data integration of non-animal tests for the development of a test battery to predict the skin sensitising potential and potency of chemicals. Toxicol. In Vitro 27, 609–618.

OECD, 2012. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 2: Use of the AOP to Develop Chemical Categories and Integrated Assessment and Testing Approaches. Series on Testing and Assessment No. 168. ENV/JM/MONO(2012)/PART2.

Reuter, H., Spieker, J., Gerlach, S., Engels, U., Pape, W., Kolbe, L., Schmucker, R., Wenck, H., Diembeck, W., Wittern, K.P., Reisinger, K., Schepky, A.G., 2011. In vitro detection of contact allergens: development of an optimized protocol using human peripheral blood monocyte-derived dendritic cells. Toxicol. In Vitro 25, 315–323.

Teunis, M., Corsini, E., Smits, M., Madsen, C.B., Eltze, T., Ezendam, J., Galbiati, V., Gremmer, E., Krul, C., Landin, A., Landsiedel, R., Pieters, R., Rasmussen, T.F., Reinders, J., Roggen, E., Spiekstra, S., Gibbs, S., 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, 1135–1150.

Teunis, M.A., Spiekstra, S.W., Smits, M., Adriaens, E., Eltze, T., Galbiati, V., Krul, C., Landsiedel, R., Pieters, R., Reinders, J., Roggen, E., Corsini, E., Gibbs, S., 2014. International ring trial of the epidermal equivalent sensitizer potency assay: reproducibility and predictive capacity. Altex 31, 251–268.

Thierse, H.J., Budde, P., Dietz, L., Ohnesorge, S., Eikelmeier, S., Conde, M., Zucht, H.D., Schulz-Knappe, P., 2011. Proteomic identification of allergen-regulated proteins and allergen-protein interaction networks in assisting biomarker and assay development. In: Roggen, E.L., Weltzien, H.U., Helma, H. (Eds.), Progress Towards Novel Testing Strategies for in Vitro Assessment of Allergens. Transworld Research Network, Kerala, India, pp. 145–166.

Toebak, M.J., Gibbs, S., Bruynzeel, D.P., Scheper, R.J., Rustemeyer, T., 2009. Dendritic cells: biology of the skin. Contact Dermatitis 60, 2–20.

van der Veen J.W., Hodemaekers H., Reus A.A., Maas W.J., van Loveren H., Ezendam J., 2015. Human relevance of an in vitro gene signature in HaCaT for skin sensitization. Toxicol. In Vitro 29, 81–84.

Wang, X.J., Hayes, J.D., Wolf, C.R., 2006. Generation of a stable antioxidant response element-driven reporter gene cell line and its use to show redox-dependent activation of nrf2 by cancer chemotherapeutic agents. Cancer Res. 66 (22), 10983–10994.