TOXICOLOGICAL SCIENCES **107(1)**, 106–121 (2009)
doi:10.1093/toxsci/kfn204
Advance Access publication October 1, 2008

Filling the Concept with Data: Integrating Data from Different *In Vitro* and *In Silico* Assays on Skin Sensitizers to Explore the Battery Approach for Animal-Free Skin Sensitization Testing

Andreas Natsch,*[,1] Roger Emter,* and Graham Ellis†

*Givaudan Schweiz AG, CH-8600 Duebendorf, Switzerland; and †Givaudan Suisse SA, 5 chemin de la parfumerie, CH 1214 Vernier, Switzerland

Received June 20, 2008; accepted September 24, 2008

Tests for skin sensitization are required prior to the market launch of new cosmetic ingredients. Significant efforts are made to replace the current animal tests. It is widely recognized that this cannot be accomplished with a single *in vitro* test, but that rather the integration of results from different *in vitro* and *in silico* assays will be needed for the prediction of the skin sensitization potential of chemicals. This has been proposed as a theoretical scheme so far, but no attempts have been made to use experimental data to prove the validity of this concept. Here we thus try for the first time to fill this widely cited concept with data. To this aim, we integrate and report both novel and literature data on 116 chemicals of known skin sensitization potential on the following parameters: (1) peptide reactivity as a surrogate for protein binding, (2) induction of antioxidant/electrophile responsive element dependent luciferase activity as a cell-based assay; (3) Tissue Metabolism Simulator skin sensitization model *in silico* prediction; and (4) calculated octanol-water partition coefficient. The results of the *in vitro* assays were scaled into five classes from 0 to 4 to give an *in vitro* score and compared to the local lymph node assay (LLNA) data, which were also scaled from 0 to 4 (nonsensitizer/weak/moderate/strong/extreme). Different ways of evaluating these data have been assessed to rate the hazard of chemicals (Cooper statistics) and to also scale their potency. With the optimized model an overall accuracy for predicting sensitizers of 87.9% was obtained. There is a linear correlation between the LLNA score and the *in vitro* score. However, the correlation needs further improvement as there is still a relatively high variation in the *in vitro* score between chemicals belonging to the same sensitization potency class.

Key Words: **skin sensitization; *in vitro* testing; battery approach; peptide reactivity; antioxidant response element; electrophile response element.**

The risk of skin sensitization is a critical issue in the development of novel ingredients for cosmetic products. The current skin sensitization testing is based on the local lymph node assay (LLNA) in mice, in which the cellular proliferation in the draining lymph nodes is measured after repeated topical application of the test compound onto the ears (Basketter *et al.*, 2002; Gerberick *et al.*, 2004a, 2007a). Results are expressed as EC3 values indicating the % concentration which induces a threefold increase in cellular proliferation. EC3 values can then be used for risk assessment and for a classification of chemicals into five classes (nonsensitizing/weak/moderate/strong and extreme sensitizers). However, with the forthcoming ban on animal testing for cosmetic ingredients in the European Union and due to the large number of tests needed for the REACH regulation on the registration of existing chemicals in Europe, there is a pressing need for assays which make animal testing obsolete.

Several alternative tests have been proposed and evaluated. The most straightforward approach measures the reactivity of test chemicals with peptides or proteins. This approach has recently been reviewed extensively (Gerberick *et al.*, 2008). It is based on the rationale that the key step in the skin sensitization process is the formation of a covalent adduct between the skin sensitizer and endogenous proteins and/or peptides in the skin. Indeed, using the % depletion of Cys- and Lys-containing heptapeptides after 24 h incubation with a test chemical as a predictor for skin sensitization, an overall accuracy of 89% on a set of 81 chemicals was reported (Gerberick *et al.*, 2007b).

Another approach focuses on the stimulation of dendritic cells, either primary cells (generated from precursor cells obtained from peripheral blood) or cell lines such as THP-1 and U-937. The expression of certain surface markers (especially CD86) was found to be induced by sensitizers in these cell types (Ade *et al.*, 2006; Sakaguchi *et al.*, 2006). A further possibility is to measure the secretion of specific cytokines by dendritic cells and keratinocytes. The cytokine that was most frequently found to be upregulated in different cellular systems is interleukin-8 (Aeby *et al.*, 2004; Bergström *et al.*, 2007; Coquette *et al.*, 2003). All these cell-based assays are very promising, but data sets on large numbers of chemicals have not been published so far. The most extensive dataset is found

[1] To whom correspondence should be addressed at Givaudan Schweiz AG, Ueberlandstrasse 138, CH-8600 Duebendorf, Switzerland. Fax: +41-44-824-29-26. E-mail: andreas.natsch@givaudan.com.

in the recent publication of Sakaguchi *et al.* (in press), reporting data for 21 allergens and 8 nonallergens.

We have recently proposed a further approach which is based on the stimulation of antioxidant response element (ARE) dependent gene activity in a recombinant cell line (Natsch and Emter, 2008). The ARE (also known as electrophile response element) is a DNA element present in many phase II detoxification genes. ARE-regulated genes had been found to be upregulated by electophilic chemicals (Dinkova-Kostova *et al.*, 2005; Wakabayashi *et al.*, 2004). Skin sensitizers in general can be described as electrophilic molecules (reviewed in Smith and Hotchkiss, 2001), and the ARE-based assay offers a straightforward possibility to measure electrophilicity in the cellular context. Moreover, several cellular markers upregulated by sensitizers were shown to be under the control of this regulatory pathway, most notably the expression of IL-8 (Zhang *et al.*, 2005) but also several markers for skin sensitizers identified in the gene-chip and reverse transcription–PCR studies by Ryan *et al.* (2004) and Gildea *et al.* (2006). Thus this assay is functionally linked to the molecular endpoints investigated in several other cell-based assays. Finally, a recent publication of Kim *et al.* (2008) has shown that this signaling pathway is involved in the sensitization reaction *in vivo*, with Nrf2-deficient mice having a reduced (but not abolished) sensitization reaction.

Several *in silico* models have also been explored to develop structure-activity relationships for skin sensitizers, including commercial mechanistically based models such as DEREK (Sanderson and Earnshaw, 1991) and the Tissue Metabolism Simulator (TIMES) software (Dimitrov *et al*, 2005). TIMES comprises a skin sensitization model (TIMES SS) which incorporates skin metabolism and considers the potential of parent chemicals and/or their activated metabolites to react with skin proteins. This model has been recently evaluated in an external validation study which showed good concordance (83%) between experimental and predicted values for 40 new chemicals (Roberts *et al*, 2007a). Additional improvements are still required, however, such developments, which use a mechanistic basis for prediction, offer promising tools to aid in the evaluation of skin sensitization potential.

As summarized above, a plethora of different *in vitro* and *in silico* tests to examine the skin sensitization potential have been proposed, and it is a general view held by many experts that a single test cannot replace the current animal testing for such a complex endpoint. This view has been formalized in a proposal put forward by Jowsey *et al.* (2006) which has been widely cited since then (26 citations by September 2008). According to this proposal, for each chemical three *in vitro* tests would be performed, namely (1) peptide reactivity, (2) stimulation of dendritic cells, and (3) T-cell activation. The results for each test would then be rated on a scale of 0–4. A score of 1 or 2 would then additionally be given to each chemical to score whether it has (1) a structural alert from *in silico* predictions (yes/no) and (2) whether it has low or high bioavailability. The scores of these five separate evaluations would then be multiplied to give a final index of sensitization potential (ISP).

Although widely cited, this model has not yet been filled with data. A routine T-cell activation assay does not yet exist and for each of the dendritic cell activation assays a large dataset has not been published. Nevertheless, based on currently published data and additional data generated in our laboratory recently, a large dataset on peptide reactivity and data from the cell-based assay depending on ARE-regulated luciferase activity have been accumulated. The aim of the current paper is twofold: First to report this full data set on 116 molecules for the two *in vitro* tests along with cLogP and *in silico* predictions according to the TIMES SS model; and then, based on this data compilation, to fill for the first time the "battery approach" proposal of Jowsey *et al.* (2006) with data, in order to test the hypothesis that the results of different assays can be integrated in a scheme such as the one proposed by Jowsey *et al.* (2006). To this aim different ways of calculations are explored for an optimized prediction of the skin sensitization potential of chemicals based on these tests.

MATERIALS AND METHODS

Chemicals and animal data. All fragrance chemicals are commercial qualities obtained from Givaudan Schweiz AG, Geneva, Switzerland. All other test chemicals were purchased from Fluka/Sigma/Aldrich, Buchs, Switzerland. The chemical and trivial names, the structures, along with CAS-numbers and LLNA data of all the test chemicals are summarized in Table I in the supporting information. Many of the chemicals used in this study are moderate to extreme skin sensitizers, and skin contact with these undiluted chemicals should be avoided. LLNA data have all been published previously. The literature references for the original LLNA studies are added to Table SI in the supporting information. The sensitization class in supplementary information Table SI is given based on the scheme of Kimber *et al.* (2003). V. weak/none is indicated for chemicals with EC3 > 30% due to dataset inadequacies (several chemicals considered nonsensitizers have not been tested at > 25% to 50% in the LLNA).

Cell-based ARE assay. AREc32 is a stable cell line derived from the human MCF7 breast carcinoma cell line. The generation of the cell line was described by Wang *et al.* (2006) and the cell line has been licensed from CRX biosciences, Dundee, UK. AREc32 cells were maintained, prepared for the test, treated with chemicals and assayed for luciferase activity exactly as described by Natsch and Emter (2008). The screening on the chemicals not contained in the previous publication was repeated four times, with duplicate analysis for each chemical at each test concentration in each repetition and with six binary dilutions covering the maximal noncytotoxic doses for each test chemical. Based on these experiments, for each test chemical (1) the average maximal induction of gene activity (I_{max}; reported as fold-induction vs. untreated cells) and (2) the average concentration inducing 1.5-fold enhanced gene activity (EC 1.5; reported in μM) were determined. The latter calculations were performed with log-linear extrapolation from the values above and below the induction threshold (as for the EC3 value determination in the LLNA and with the formula described in Gerberick *et al.*, 2007a). A chemical was rated positive, if it induced significantly enhanced gene activity above the threshold at any of the tested concentrations either in all repetitions made or in three out of four repetitions. The literature data are based on three or four repetitions with duplicate analysis in each repetition (Natsch and Emter, 2008).

TABLE 1
Data Compilation of LLNA Results and *In Vitro* and *In Silico* Predictions

	LLNA data			Luciferase induction			TIMES SS modeling	
Test chemicals	EC3 (%)	Class	Cys-depletion (%)	ARE EC 1.5 (μM)	ARE I_{max}	cLogP	Prediction	Training set?
Diphenylcyclopropenone	0.003	Extreme	98.8[a]	1	20.0	3.25	Yes	Yes
Oxazolone	0.003	Extreme	75.8[b]	215	2.3	1.51	Yes	Yes
5-Chloro-2-methyl-4-isothiazolin-3-one	0.009	Extreme	94.8[b]	<< 2	7.2	−0.34	Yes	Yes
p-Benzochinone	0.01	Extreme	91.6[b]	2	19.6	0.25	Yes	Yes
DNCB	0.04	Extreme	100[a]	1	12.3	2.14	Yes	Yes
4-Nitrobenzylbromide	0.05	Extreme	96[b]	4.3	8.9	2.7	Yes	Yes
1,4-Hydrochinone	0.1	Strong	83.3[a]	2	28.6	0.59	Yes	Yes
Glutaraldehyde	0.1	Strong	30[a]	20	17.7	−0.18	Yes	Yes
Fluorescein-5-isothiocyanate	0.14	Strong	100[a]	276	2.2	4.69	Yes	Yes
Phthalic anhydride	0.16	Strong	−1.9[a]	> 1000	1.3	2.07	Yes	Yes
1,4-Phenylenediamine	0.16	Strong	95.2[b]	12	12.7	−0.25	Yes	Yes
Benzyl bromide	0.2	Strong	97.8[c]	68.4	9.6	2.9	Yes	Yes
Lauryl gallate	0.3	Strong	91[a]	5	2.8	6.21	Yes	Yes
Propyl gallate	0.32	Strong	62.2[b]	3	8.3	1.8	Yes	Yes
2-Aminophenol	0.4	Strong	100[b]	2	23.3	0.6	Yes	Yes
Trimellitic anhydride	0.6	Strong	−1.1[a]	> 1000	1.34	2	Yes	Yes
Formaldehyde	0.7	Strong	56.5[b]	34	4.1	0.35	Yes	Yes
Metol	0.78	Strong	100[a]	2	32.4	0.04	Yes	No
Methyldibromo glutaronitrile	1.3	Moderate	75.8[b]	15.0	2.6	1.6	Yes	Yes
1-Phenyl-1,2-propanedione	1.3	Moderate	92.1[b]	114	27.5	1.1	Yes	Yes
1-Naphthol	1.3	Moderate	24.8[a]	16	2.4	2.69	No	Yes
Glyoxal	1.4	Moderate	90.8[b]	192.0	4.2	−1	Yes	Yes
2-Hydroxy-ethyl-acrylate	1.4	Moderate	89.4[c]	6	15.2	−0.25	Yes	Yes
4-Vinyl-pyridine	1.6	Moderate	92[a]	3	13.8	1.7	Yes	No
Pomarose	1.6	Moderate	100[d]	3	38.7	3.3	Yes	No
2-Mercaptobenzothiazol	1.7	Moderate	100[b]	32	10.9	2.86	Yes	Yes
Isoeugenol	1.8	Moderate	98.4[b]	18	60.2	2.1	Yes	Yes
2-Methyl-2H-isothiazolin-3-one	1.9	Moderate	98[a]	3	12.5	−0.486	Yes	Yes
Diethyl maleate	2.1	Moderate	100[b]	2	41.0	2.2	Yes	Yes
Ethylenediamine	2.2	Moderate	15.9[e]	> 1000	1.2	−2	Yes	Yes
3-Dimethyl-amino-1-propylamine	2.2	Moderate	10[a]	156	5.4	−0.45	Yes	Yes
Spirogalbanone	2.2	Moderate	99[d]	25	16.7	5.2	Yes	No
1,2-Benzisothiazolin-3-one	2.3	Moderate	98[a]	2	6.6	0.64	Yes	Yes
Methyl 2-nonynoate	2.5	Moderate	100[a]	1	19.5	3.1	Yes	Yes
trans-2-Decenal	2.5	Moderate	100[d]	56	11.6	3.55	Yes	Yes
trans-Anethole	2.7	Moderate	22[d]	> 1000	0.93	3.39	No	Yes
Benzyl salicylate	2.9	Moderate	0[d]	18	3.6	4	No	Yes
Phenylacetaldehyde	3	Moderate	60.7[a]	48	9.8	1.54	Yes	Yes
Galbanone	3	Moderate	76[d]	24	8.1	4.4	Yes	No
Cinnamic aldehyde	3.1	Moderate	90.5[b]	19	31.6	2.12	Yes	Yes
3-Aminophenol	3.2	Moderate	7[a]	69	5.0	0.18	Yes	Yes
Diethyl sulfate	3.3	Moderate	24[b]	> 1000	1.2	1.14	Yes	Yes
2-Bromotetradecanoic acid	3.4	Moderate	29.3[b]	57.0	3.1	6.2	Yes	Yes
Benzylideneacetone	3.7	Moderate	91.3[b]	9	27.5	2.04	Yes	Yes
α-Methyl cinnamic aldehyde	4.5	Moderate	7.5[d]	47	23.3	2.37	Yes	Yes
Citral	5	Moderate	34.7[b]	64	9.8	3	Yes	Yes
Tetramethylthiuram disulfide	5.2	Moderate	99.4[b]	0.3	4.7	1.7	Yes	Yes
trans-2-Hexenal	5.5	Moderate	97.7[a]	12	29.7	1.8	Yes	Yes
3,4-Dihydrocoumarin	5.6	Moderate	4[d]	> 1000	0.9	0.97	Yes	Yes
Creosol	5.8	Moderate	15.4[b]	> 1000	1.4	1.2	Yes	Yes
Hydratropic aldehyde	6.3	Moderate	48.2[a]	121	2.3	1.96	Yes	Yes
Cyclal C	6.5	Moderate	100[d]	227	8.0	3.1	Yes	Yes
β-Damascone	6.7	Moderate	100[d]	1	26.7	4	Yes	No
Dihydroeugenol	6.8	Moderate	1[d]	46	3.4	2.71	Yes	Yes
Safranal	7.5	Moderate	100[b]	12	32.9	3.2	Yes	Yes

TABLE 1—*Continued*

	LLNA data			Luciferase induction			TIMES SS modeling	
Test chemicals	EC3 (%)	Class	Cys-depletion (%)	ARE EC 1.5 (μM)	ARE I_{max}	cLogP	Prediction	Training set?
Perillaaldehyde	8.1	Moderate	31.9[a]	37	42.1	3.34	Yes	Yes
Silvial	9	Moderate	100[d]	185	3.1	4.4	Yes	No
$NiSO_4$	False-neg.	Weak	35.5[c]	310.2	5.3	n.a.[f]	n.a.	n.a.
Diacetyl	11	Weak	79.9[b]	125	4.8	−0.5	Yes	Yes
Butyl Quinoleine sec	11.2	Weak	6[d]	56	2.0	4.4	Yes	No
Farnesal	11.7	Weak	36.7[b]	108	2.9	5.78	Yes	Yes
α-Hexyl cinnamic aldehyde	12	Weak	−0.3[a]	36	3.9	5.3	Yes	Yes
Eugenol	13	Weak	54[d]	38	2.3	2	Yes	Yes
1-Chlorooctadecane	16	Weak	3.4[b]	> 1000	1.37	9.44	Yes	Yes
Cosmone	16.4	Weak	6[d]	> 1000	1.1	5.6	No	No
Toscanol	16.7	Weak	0[d]	> 1000	0.94	3.8	No	No
Benzyl benzoate	17	Weak	0.2[a]	26	1.96	4	Yes	Yes
Lyral	17.1	Weak	95[d]	98	4.6	2.1	Yes	Yes
Phenyl benzoate	17.1	Weak	38.5[b]	193	3.6	3.04	Yes	Yes
Benzyl cinnamate	18.4	Weak	0[d]	27	8.8	3.65	Yes	Yes
Lilial	18.7	Weak	14[a]	> 1000	1.1	4.2	Yes	Yes
Super muguet	19.3	Weak	0[d]	> 1000	0.95	3.6	Yes	No
Estragole	20.2	Weak	38.5[b]	298	2.4	3.4	Yes	Yes
Cinnamic alcohol	21	Weak	11[d]	309	2.1	1.7	Yes	Yes
Cyclamen aldehyde	22.3	Weak	59.9[b]	56	1.8	4	Yes	Yes
Hydroxycitronellal	23	Weak	46.7[b]	> 1000	0.9	1.54	Yes	Yes
Imidazolidinyl urea	24	Weak	46.1[b]	> 1000	1.3	−8	Yes	Yes
5-Methyl-2,3-hexanedione	26	Weak	25.8[a]	494.0	2.8	0.06	Yes	Yes
Serenolide	27.8	Weak	11[d]	> 1000	1.1	3.8	No	No
Evernyl	28.1	Weak	100[d]	> 1000	1.5[e]	2.1	No	No
Ambrettolide	28.8	Weak	8[d]	> 1000	0.8	6	No	No
Penicillin G	30	Weak	18.5[b]	> 1000	1.4	1.8	Yes	No
Butyl glycidyl ether	31	Weak	84.8[b]	71.8	12.6	0.63	Yes	Yes
Benzocaine	> 50	Weak	11.6[b]	51	3.0	1.8	Yes	Yes
Linalool	30	V.Weak/None	2[d]	> 1000	1.1	3.3	Yes	Yes
Superfix	43.3	V.Weak/None	13[d]	> 1000	1.2	5.91	No	No
Isopropyl myristate	44	V.Weak/None	0.8[a]	> 1000	1.0	> 6	No	Yes
Cydrane	54.8	V.Weak/None	9	> 1000	1.1	4.24	No	No
Geraniol	57	V.Weak/None	0[d]	> 1000	1.3	3.3	Yes	Yes
trans-2-Hexenol	60	V.Weak/None	18[d]	294	1.51	3.9	Yes	No
Pyridine	72	V.Weak/None	10[d]	> 1000	1.1	0.7	Yes	Yes
Benzalkonium chloride	False-pos.	Nonsensitizer	−6.8[c]	1.7	1.7	n.a.	n.a.	n.a.
Diethyl phthalate	> 100	V.Weak/None	0.8[a]	614	1.9	2.47	No	Yes
Propylene glycol	> 100	V.Weak/None	−0.9[a]	> 1000	1.0	−0.92	No	Yes
Glycerol	> 100	V.Weak/None	−3.8[a]	> 1000	1.1	−1.5	No	Yes
Methyl salicylate	> 20	V.Weak/None	0.3[a]	> 1000	1.1	2.2	No	Yes
Benzoic acid	> 20	V.Weak/None	0[b]	> 1000	1.0	1.9	No	No
1-Butanol	> 20	V.Weak/None	2.2[b]	> 1000	1.1	0.84	No	Yes
6-Methyl-coumarin	> 25	V.Weak/None	3.6[b]	69	4.4	2.06	No	Yes
4-Hydroxybenzoic acid	> 25	V.Weak/None	0[b]	> 1000	1.2	1.39	No	Yes
Lactic acid	> 25	V.Weak/None	2.5[b]	> 1000	1.1	−0.65	No	Yes
Salicylic acid	> 25	V.Weak/None	4[a]	> 1000	1.1	2.26	No	Yes
Sulphanilic acid	> 25	V.Weak/None	3.3[b]	> 1000	1.0	−2	Yes	Yes
Benzaldehyde	> 25	V.Weak/None	7[b]	> 1000	1.1	1.48	No	Yes
Ethylene brassylate	> 30	V.Weak/None	0[d]	> 1000	1.1	4.7	No	No
Calone	> 30	V.Weak/None	4[d]	> 1000	1.3	1.2	No	No
Hedione	> 40	V.Weak/None	4[d]	> 1000	1.1	3.1	Yes	Yes
Methyl paraben	> 50	V.Weak/None	3.6[a]	234	2.5	1.7	No	Yes
Propylparaben	> 50	Nonsensitizer	8[a]	35	2.4	2.98	No	Yes
Octanoic acid	> 50	V.Weak/None	0.04[b]	> 1000	1.2	3	No	Yes
Isopropanol	> 50	V.Weak/None	0.3[b]	> 1000	1.1	0.05	No	Yes
Vanillin	> 50	V.Weak/None	3.2[a]	> 1000	1.2	0.1	Yes	Yes
Phenyl ethyl alcohol	> 50	V.Weak/None	4[d]	> 1000	1.1	1.36	No	Yes

TABLE 1—*Continued*

	LLNA data			Luciferase induction			TIMES SS modeling	
Test chemicals	EC3 (%)	Class	Cys-depletion (%)	ARE EC 1.5 (μM)	ARE I_{max}	cLogP	Prediction	Training set?
Benzyl alcohol	> 50	V.Weak/None	6[d]	> 1000	1.0	1.03	No	Yes
Benzenesulfonic acid	n.a.	Nonsensitizer	0[b]	> 1000	1.2	−1.17	No	No
SDS	False-pos.	Nonsensitizer	7.2[c]	> 1000	1.2	1.69	No	Yes

Note. DNCB, 2,4,dinitro-chloro-benzene; SDS, sodium dodecyl sulfate.
[a]Literature data from Gerberick *et al*. (2007b).
[b]Internal unpublished data, assay as in Gerberick *et al*. (2007b).
[c]Internal data, fluorescent detection of peptide depletion used.
[d]Internal data published in Natsch *et al*. (2007), with modified method.
[e]Nonsignificant result, rated zero in Table 3.
[f]Not applicable.

Peptide reactivity assay. Peptide reactivity with the Cys-containing peptide Ac-RFAACAA was determined as described by Gerberick *et al*. (2004b). Further peptide reactivity data were taken from the publications of Gerberick *et al*. (2007b) and from Natsch *et al*. (2007). The data in this latter publication were obtained with a slightly modified method (0.25mM test peptide instead of 0.5mM to reduce peptide precipitation), however, results on the same chemicals obtained under these conditions and the original conditions are very comparable, and thus were used for this combined analysis. All the data (both our own and literature data) are based on triplicate analysis (Gerberick *et al*., 2007b).

Computer modeling and statistics. cLogP values were obtained either from internal data (measured values according to OECD guideline 117) or calculated using KOWWIN V.1.67 obtained from the United States Environmental Protection Agency (U.S. EPA) web site. The TIMES SS software (V.2.25.7) was obtained from OASIS Laboratory of Mathematical Chemistry, Bourgas, Bulgaria, and run using the skin sensitization metabolism activated toxicity model. Regression analysis and plotting of Box plots were performed with the Minitab statistical software (Minitab Inc., version 15.1.1.0, Coventry, UK).

RESULTS

Compilation of the In Vitro *and* In Silico *Data*

Table 1 lists the following parameters for each test chemical: (1) LLNA value, (2) Cys-peptide depletion in %, (3) EC 1.5 from the ARE assay, indicating the concentration for 1.5 fold stimulation of gene activity, (4) I_{max} indicating maximal induction of gene activity in the ARE assay, (5), cLogP value, (6) prediction (yes/no) from the TIMES SS software, and (7) whether the compound was in the training set of TIMES SS. More detailed results with the structural domain assignments and the detailed prediction from TIMES SS are given in the supplementary information Table SII. Peptide reactivity data are only listed as reactivity toward a Cys-containing peptide. In the original assay (Gerberick *et al*., 2004b, 2007b) reactivity with a Lys-peptide at a high pH is also scored. This certainly gives additional information, however, in the published dataset there are only two compounds (phthalic anhydride and trimellitic anhydride) exclusively reactive with the Lys-peptide. Because the Lys-peptide assay involves a large concentration of test compound (25mM) there are solubility issues with many of the hydrophobic compounds included in our current study and not all chemicals were measured with the Lys-assay. We therefore limit the information in this paper to Cys-peptide reactivity.

Classification of the Test Results into Scores

The data in Table 1 were transformed to scores as proposed by Jowsey *et al*. (2006). A score from 0 to 4 was given for Cys-reactivity, ARE EC 1.5, ARE I_{max} and LLNA EC3 according to the thresholds and boundaries set in Table 2. The thresholds for rating a molecule as > 0 in an *in vitro* test are derived from the values needed for significant results in the tests. Thus at least 15% peptide depletion is needed to be significant in most cases. Similarly in the ARE assay, the stimulation must be > 1.5-fold and the EC 1.5 below 1000μM (Natsch and Emter, 2008) in order to rate a chemical positive. The thresholds for the scores 1–4 were then assigned in order to span the whole dynamic range of the test results (see Tables 1 and 2). The scores for the LLNA class are based on the classification of Kimber *et al*. (2003). The scores resulting from this transformation and a simple qualitative discussion for the individual chemicals are summarized in Table 3. We had shown, that both the I_{max} and the EC 1.5 value from the ARE assay are correlated to the potency of a chemical (Natsch and Emter, 2008). To summarize the data from the two ARE-based measures, an average of the score for I_{max} and EC 1.5 was therefore also calculated and included in Table 3.

TABLE 2
Set Boundaries for Assigning Scores to the Chemicals

Score	LLNA (EC3 in %)[a]	Peptide reactivity (% depletion)	EC 1.5 in ARE assay (μM)	I_{max} in ARE assay (fold stimulation)
0	> 30	< 15	> 1000	< 1.5
1	10–30	15–40	100–1000	1.5–3.0
2	1–10	40–65	25–100	3.0–6.0
3	0.1–1	65–90	6.25–25	6.0–12.0
4	< 0.1	> 90	< 6.25	> 12

Note. SDS, sodium dodecyl sulfate.
[a]For the chemicals benzocaine, SDS, benezensulfonic acid, the ranking was performed based on human experience or guinea pig tests.

From an inspection of Table 3 and the discussion in the last column, it is obvious, that for many chemicals the Cys-reactivity and the cell-based assay give congruent results, which is also supported by the TIMES SS prediction, but there are also chemicals which are rated positive by only one of the *in vitro* tests, or which are correctly predicted either by the *in vitro* or the *in silico* prediction only.

Jowsey *et al*. (2006) proposed that an additional score of 1 or 2 is given depending on the bioavailability. As discussed below, there is no good published model to rate bioavailability form a hydrophobic test vehicle such as the one used in the LLNA. Therefore a very conservative approach was taken, simplifying bioavailability by giving a score of 2 to any chemical having a cLogP between −2 and 5, and a score of 1 for chemicals outside of this range. (Human *in vitro* skin absorption experiments with a series of unrelated chemicals have determined that the optimum cLogP for maximum absorption is 2, with rapid loss of absorption at Log P value either side; Smith and Hotchkiss, 2001.)

For the *in silico* prediction a score of 1 or 2 (also as proposed by Jowsey *et al*., 2006) for a predicted nonsensitizer or sensitizer, respectively, was given according to the prediction from TIMES SS (Table 3). It should be kept in mind, that these latter data are somewhat biased, as a large fraction of the chemicals were in the training set, and for any *in silico* prediction, the true validity can only be determined with external test data.

Cooper Statistics for the Individual Tests

Before integration of the data, Cooper statistics (Cooper *et al*., 1979) were calculated for this complete dataset for the two individual *in vitro* tests. With peptide depletion > 15% as a positive result, the sensitivity of the Cys-peptide depletion assay is 73.8%, the specificity is 96.9%, positive predictivity is 98.4%, negative predictivity is 58.5%, and the overall accuracy is 80.2% for these 116 chemicals. With the threshold of 1.5-fold luciferase induction in the ARE assay, the sensitivity of the cell-based assay is 78.9%, the specificity is 81.3%, the positive predictivity is 91.7%, the negative predictivity is 59.1%, and the accuracy is 79.3%.

Data Integration According to the Scheme Proposed by Jowsey et al.: *Multiplication of Scores*

In a first attempt of data integration we directly multiplied the scores from the individual tests to calculate a combined score according to Jowsey *et al*. (2006). This is called the ISP (index of sensitization) in the original proposal, and as proposed we multiplied the scores from the two *in vitro* tests with the *in silico* and bioavailability score. The difference to the original proposal is that here we have only one cell-based assay available and not yet a battery of two assays and of course the ARE-based assay is not really a dendritic cell activation assay. A Box plot of the resulting scores is shown in Figure 1, split up for the different sensitization classes. The individual data are included in the column "*multiplied scores*" of Table 3. The Cooper statistics are given in Table 4. It is evident, that this combination of the data yields a very high positive predictivity and a high specificity (96.9%): In the original proposal, a chemical needs to be positive in all the *in vitro* assays to be rated as a sensitizer, failure to give a detectable signal in any of three (here two) *in vitro* assays would give a rating as a nonsensitizer (ISP = 0). The downside of the high specificity of this integration scheme is also shown by this data compilation: For several chemicals only one assay yields a positive result, and therefore the sensitivity, negative predictivity, and thus the overall accuracy are not satisfying with this scheme (Table 4).

In terms of predicting potency, there is clearly a relationship between higher scores and higher sensitization potential of the chemicals as seen in Figure 1. However, calculating the R^2 for a linear correlation for all the 116 single chemicals gives only a value of 0.423, and from Figure 1 it is obvious that the data within one sensitization class are very much scattered. The Spearman rank correlation is highly significant with a coefficient δ for this data evaluation of 0.70.

Data Integration with an Alternative Proposal: Average Scores from Different In Vitro *Tests*

Calculating the product of the scores from the individual assessments as in the original proposal is one option to combine the data. A very simple and intuitive alternative is to take the average of the scores from the different *in vitro* tests and directly relate these average scores to the scores from the LLNA. This was first done only with the two *in vitro* assays, peptide reactivity and ARE induction (taking the average score for EC 1.5 and I_{max}). The results from this evaluation are summarized in Figure 2 and the individual data are included in Table 3 in the column "*average scores*."

Potency prediction. Indeed, as illustrated in Figure 2, there is quite a good linear relationship between the median of the scores for the chemicals within one class and the score of that particular LLNA class (compare the median values in the Box plots to the LLNA scores of the corresponding class). However, as illustrated in this Box plot, there is still a quite broad distribution of the data with a large interquartile range, especially for the moderate sensitizers covering a broad range of the plot. If calculating the linear correlation for all the single values, the R^2 is 0.518. The Spearman rank correlation coefficient δ is 0.758. Alternatively, the same calculations were made with only the EC 1.5 score from the ARE assay (instead of the average score between EC 1.5 and I_{max}) and very similar results are obtained (the linear correlation coefficient was 0.516 instead 0.518) and the Spearman rank correlation is 0.755, indicating that EC 1.5 maybe a sufficient indicator from the ARE assay.

TABLE 3
Data from Table 1 Transformed to Scores for the Single Chemicals and Interpretation of the Results

Generic name	LLNA	Cys Dpl.	ARE induction			Prediction 2/1[b]	CLogP 2/1[c]	Average scores[d]	Multiplied score[e]	Observation[f]
			EC 1.5	I_{max}	Avg[a]					
Extreme sensitizers										
Diphenylcyclopropenone	4	4	4	4	4	2	2	4	64	Peptide, ARE, and TIMES SS concur
Oxazolone	4	3	1	1	1	2	2	2	12	Peptide, ARE, and TIMES SS concur
5-Chloro-2-methyl-4-isothiazolin-3-one	4	4	4	3	3.5	2	2	3.75	56	Peptide, ARE, and TIMES SS concur
p-Benzochinone	4	4	4	4	4	2	2	4	64	Peptide, ARE, and TIMES SS concur
DNCB	4	4	4	4	4	2	2	4	64	Peptide, ARE, and TIMES SS concur
4-nitrobenzylbromide	4	4	4	3	3.5	2	2	3.75	56	Peptide, ARE, and TIMES SS concur
Strong sensitizers										
1,4-Hydrochinone	3	3	4	4	4	2	2	3.5	48	Peptide, ARE, and TIMES SS concur
Glutaraldehyde	3	1	3	4	3.5	2	2	2.25	14	Peptide, ARE, and TIMES SS concur; Lys-peptide would be better predictor
Fluorescein-5-isothiocyanate	3	4	1	1	1	2	2	2.5	16	Peptide, ARE, and TIMES SS concur
Phthalic anhydride	3	0	0	0	0	2	2	0	0	ARE and peptide false negative, Lys-peptide would be correct predictor. TIMES SS correct.
1,4-Phenylenediamine	3	4	3	4	3.5	2	2	3.75	56	Peptide, ARE, and TIMES SS concur
Benzyl bromide	3	4	2	3	2.5	2	2	3.25	40	Peptide, ARE, and TIMES SS concur
Lauryl gallate	3	4	4	1	2.5	2	1	3.25	20	Peptide, ARE, and TIMES SS concur
Propyl gallate	3	2	4	3	3.5	2	2	2.75	28	Peptide, ARE, and TIMES SS concur
2-Aminophenol	3	4	4	4	4	2	2	4	64	Peptide, ARE, and TIMES SS concur
Trimellitic anhydride	3	0	0	0	0	2	2	0	0	ARE and peptide false negative. TIMES SS correct.
Formaldehyde	3	2	2	2	2	2	2	2	16	Peptide, ARE, and TIMES SS concur
Metol	3	4	4	4	4	2	2	4	64	Peptide, ARE, and TIMES SS concur
Moderate sensitizers										
Methyldibromo glutaronitrile	2	3	3	1	2	2	2	2.5	24	Peptide, ARE, and TIMES SS concur
1-Phenyl-1,2-propanedione	2	4	1	4	2.5	2	2	3.25	40	Peptide, ARE, and TIMES SS concur
1-Naphthol	2	1	3	1	2	1	2	1.5	4	Peptide and ARE concur. TIMES SS false negative.
Glyoxal	2	4	1	2	1.5	2	2	2.75	24	Peptide, ARE, and TIMES SS concur
2-Hydroxy-ethyl-acrylate	2	3	4	4	4	2	2	3.5	48	Peptide, ARE, and TIMES SS concur
4-Vinyl-pyridine	2	4	4	4	4	2	2	4	64	Peptide, ARE, and TIMES SS concur
Pomarose	2	4	4	4	4	2	2	4	64	Peptide, ARE, and TIMES SS concur
2-Mercaptobenzothiazol	2	4	2	3	2.5	2	2	3.25	40	Peptide, ARE, and TIMES SS concur
Isoeugenol	2	4	3	4	3.5	2	2	3.75	56	Peptide, ARE, and TIMES SS concur
2-Methyl-2*H*-isothiazolin-3-one	2	4	4	4	4	2	2	4	64	Peptide, ARE, and TIMES SS concur
Diethyl maleate	2	4	4	4	4	2	2	4	64	Peptide, ARE, and TIMES SS concur
Ethylenediamine	2	0	0	0	0	2	2	0	0	ARE and Peptide false negative—requires metabolism?
3-Dimethyl-amino-1-propylamine	2	0	1	2	1.5	2	2	0.75	0	ARE and TIMES SS concur. Peptide false negative—requires metabolism?
Spirogalbanone	2	4	2	4	3	2	1	3.5	24	Peptide, ARE, and TIMES SS concur
1,2-Benzisothiazolin-3-one	2	4	4	3	3.5	2	2	3.75	56	Peptide, ARE, and TIMES SS concur
Methyl 2-nonynoate	2	4	4	4	4	2	2	4	64	Peptide, ARE, and TIMES SS concur
trans-2-Decenal	2	4	2	3	2.5	2	2	3.25	40	Peptide, ARE, and TIMES SS concur

trans-Anethole	2	1	0	0	0	1	2	0.5	0	Peptide concurs with LLNA. ARE and TIMES SS false negative
Benzyl salicylate	2	0	3	2	2.5	1	2	1.25	0	ARE concurs with LLNA. TIMES SS and peptide false negative.
Phenylacetaldehyde	2	2	2	3	2.5	2	2	2.25	20	Peptide, ARE, and TIMES SS concur
Galbanone	2	3	3	3	3	2	2	3	36	Peptide, ARE, and TIMES SS concur
Cinnamic aldehyde	2	4	3	4	3.5	2	2	3.75	56	Peptide, ARE, and TIMES SS concur
3-Aminophenol	2	0	2	2	2	2	2	1	0	ARE and TIMES SS concur. Peptide false negative—requires metabolism?
Diethyl sulfate	2	1	0	0	0	2	2	0.5	0	ARE false negative, Peptide and TIMES SS concur
2-Bromotetradecanoic acid	2	1	2	2	2	2	1	1.5	4	Peptide, ARE, and TIMES SS concur
Benzylideneacetone	2	4	3	4	3.5	2	2	3.75	56	Peptide, ARE, and TIMES SS concur
α-Methyl cinnamic aldehyde	2	0	2	4	3	2	2	1.5	0	ARE and TIMES SS concur. Peptide false negative.
Citral	2	1	2	3	2.5	2	2	1.75	10	Peptide, ARE and TIMES SS concur
Tetramethylthiuram disulfide	2	4	4	2	3	2	2	3.5	48	Peptide, ARE and TIMES SS concur
trans-2-Hexenal	2	4	3	4	3.5	2	2	3.75	56	Peptide, ARE and TIMES SS concur
3,4-Dihydrocoumarin	2	0	0	0	0	2	2	0	0	ARE and peptide false negative. TIMES SS correct—requires metabolism?
Creosol	2	1	0	0	0	2	2	0.5	0	ARE false negative, peptide borderline. TIMES SS concurs. Requires metabolism?
Hydratropic aldehyde	2	2	1	1	1	2	2	1.5	8	Peptide, ARE, and TIMES SS concur
Cyclal C	2	4	1	3	2	2	2	3	32	Peptide, ARE, and TIMES SS concur
β-Damascone	2	4	4	4	4	2	2	4	64	Peptide, ARE, and TIMES SS concur
Dihydroeugenol	2	0	2	2	2	2	2	1	0	ARE and TIMES SS concur. Peptide false negative—requires metabolism?
Safranal	2	4	3	4	3.5	2	2	3.75	56	Peptide, ARE, and TIMES SS concur
Perillaaldehyde	2	1	2	4	3	2	2	2	12	Peptide, ARE, and TIMES SS concur
Silvial	2	4	1	2	1.5	2	2	2.75	24	Peptide, ARE, and TIMES SS concur
Weak sensitizers										
$NiSO_4$	1	1	1	2	1.5		2	1.25	3	Peptide and ARE concur. Peptide precipitates
Diacetyl	1	3	1	2	1.5	2	2	2.25	18	Peptide, ARE, and TIMES SS concur
Butyl Quinoleine sec	1	0	2	1	1.5	2	2	0.75	0	ARE and TIMES SS concur. Peptide false negative—requires metabolism?
Farnesal	1	1	1	1	1	2	1	1	2	Peptide, ARE, and TIMES SS concur
α-Hexyl cinnamic aldehyde	1	0	2	2	2	2	1	1	0	ARE and TIMES SS concur. Peptide false negative
Eugenol	1	2	2	1	1.5	2	2	1.75	12	Peptide, ARE, and TIMES SS concur
1-Chlorooctadecane	1	0	0	0	0	2	1	0	0	ARE and peptide false negative—requires metabolism or due to poor solubility?
Cosmone	1	0	0	0	0	1	1	0	0	ARE, peptide, and TIMES SS all predict nonsensitizer. False positive in LLNA?
Toscanol	1	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS all predict nonsensitizer. False positive in LLNA?
Benzyl benzoate	1	0	2	1	1.5	2	2	0.75	0	ARE and TIMES SS concur. Peptide false negative.
Lyral	1	4	2	2	2	2	2	3	32	Peptide, ARE, and TIMES SS concur
Phenyl benzoate	1	1	1	2	1.5	2	2	1.25	6	Peptide, ARE, and TIMES SS concur.
Benzyl cinnamate	1	0	2	3	2.5	2	2	1.25	0	ARE and TIMES SS concur. Peptide false negative.

TABLE 3—*Continues*

Generic name	LLNA	Cys Dpl.	ARE induction			Prediction 2/1[b]	CLogP 2/1[c]	Average scores[d]	Multiplied score[e]	Observation[f]
			EC 1.5	I_{max}	Avg[a]					
Lilial	1	0	0	0	0	2	2	0	0	ARE and peptide false negative
Super muguet	1	0	0	0	0	2	2	0	0	Peptide and ARE negative. False positive in LLNA or requires metabolism?
Estragole	1	1	1	1	1	2	2	1	4	Peptide, ARE, and TIMES SS concur
Cinnamic alcohol	1	0	1	1	1	2	2	0.5	0	ARE and TIMES SS concur. Peptide false negative.
Cyclamen aldehyde	1	2	2	1	1.5	2	2	1.75	12	Peptide, ARE, and TIMES SS concur
Hydroxycitronellal	1	2	0	0	0	2	2	1	0	Peptide and TIMES SS concur. ARE false negative.
Imidazolidinyl urea	1	2	0	0	0	2	1	1	0	Peptide and TIMES SS concur. ARE false negative.
5-Methyl-2,3-hexanedione	1	1	1	1	1	2	2	1	4	Peptide, ARE, and TIMES SS concur
Serenolide	1	0	0	0	0	1	2	0	0	ARE, peptide and TIMES SS all predict nonsensitizer. False positive in LLNA?
Evernyl	1	4	0	0	0	1	2	2.25	4	Peptide concurs with LLNA. ARE and TIMES SS false negative
Ambrettolide	1	0	0	0	0	1	1	0	0	ARE, peptide, and TIMES SS all predict nonsensitizer. False positive in LLNA?
Penicillin G	1	1	0	0	0	2	2	0.5	0	Peptide and TIMES SS concur, ARE false negative, peptide borderline
Butyl glycidyl ether	1	3	2	4	3	2	2	3	36	Peptide, ARE, and TIMES SS concur
Benzocaine	1	0	2	2	2	2	2	1	0	ARE and TIMES SS concur. Peptide false negative. LLNA known to be variable/false negative.
V. weak/nonsensitizers										
Linalool	0	0	0	0	0	2	2	0	0	Peptide and ARE concur. TIMES incorrect?
Superfix	0	0	0	0	0	1	1	0	0	ARE, peptide, and TIMES SS concur
Isopropyl myristate	0	0	0	0	0	1	1	0	0	ARE, peptide, and TIMES SS concur
Cydrane	0	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS concur
Geraniol	0	0	0	0	0	2	2	0	0	ARE and peptide concur. TIMES false positive
trans-2-Hexenol	0	1	1	1	1	2	2	1	4	Peptide and ARE false positive or LLNA false negative due to very high volatility
Pyridine	0	0	0	0	0	2	2	0	0	ARE and peptide concur. TIMES false positive
Benzalkonium chloride	0	0	4	1	2.5	1	1	1.25	0	Peptide and TIMES SS concur. ARE false positive with unclear dose-response
Diethyl phthalate	0	0	1	1	1	1	2	0.5	0	Peptide and TIMES SS concur. ARE false positive (borderline)
Propylene glycol	0	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS concur
Glycerol	0	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS concur
Methyl salicylate	0	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS concur
Benzoic acid	0	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS concur
1-Butanol	0	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS concur
6-Methyl-coumarin	0	0	2	2	2	1	2	1	0	Peptide and TIMES SS concur. ARE false positive
4-Hydroxybenzoic acid	0	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS concur
Lactic acid	0	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS concur
Salicylic acid	0	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS concur

Sulphanilic acid	0	0	0	0	0	2	2	0	0	ARE and peptide concur. TIMES false positive
Benzaldehyde	0	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS concur
Ethylene brassylate	0	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS concur
Calone	0	0	0	0	0	1	2	0	0	ARE, peptide, and Prediction concur.
Hedione	0	0	0	0	0	2	2	0	0	ARE and peptide concur. TIMES false positive
Methyl paraben	0	0	1	1	1	1	2	0.5	0	ARE concurs with human data, peptide and TIMES SS with LLNA
Propylparaben	0	0	2	1	1.5	1	2	0.75	0	ARE concurs with human data, peptide and TIMES SS with LLNA
Octanoic acid	0	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS concur
Isopropanol	0	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS concur
Vanillin	0	0	0	0	0	2	2	0	0	ARE and peptide concur. TIMES false positive
Phenyl ethyl alcohol	0	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS concur
Benzyl alcohol	0	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS concur
Benzenesulfonic acid	0	0	0	0	0	1	2	0	0	ARE, peptide, and TIMES SS concur
SDS	0	0	0	0	0	1	2	0	0	ARE, peptide, and Prediction concur—known false positive in LLNA

Note. DNCB, 2,4,dinitro-chloro-benzene.
[a]Indicates the average for the scores calculated separately for EC 1.5 and I_{max}.
[b]Indicates whether TIMES SS predicts it as a sensitizer (2) or not (1).
[c]Indicates whether cLogP is between −2 and 5 (2) or not (1).
[d]Average of the Cys-depletion score and the average ARE-score.
[e]Multiplied scores from Cys-depletion score, average ARE-score, cLogP score and TIMES SS score.
[f]Indicates whether *in vitro* or *in silico* concur with each other and with LLNA result.

Cooper statistics. The Cooper statistics were calculated from the data obtained with average scores from the two *in vitro* tests (Table 5). One possible option is to rate any chemical positive, which has an average *in vitro* score > 0. In this case the underlying rationale is that any compound which is rated positive by either the peptide reactivity assay OR by the ARE assay is a potential sensitizer. This approach gives a relatively high sensitivity (86.9%), as most sensitizers were positive in at least one of the tests. The overall accuracy is 85.3%. (Another option would be to rate any chemical with a score above 0.5 as a sensitizer: In this case the lowest score (1) in one of the test and a negative test result in the second test would be regarded as not sufficient evidence for the rating of a chemical as a sensitizer. But either a moderate score (2) in one of the tests or a weak score (1) in two tests would be considered enough evidence for rating a chemical as a sensitizer. With this specific calculation sensitivity is 81.0%, specificity is 87.5%, positive predictivity is 94.4%, negative predictivity is 63.6%, and accuracy is 82.8%.)

False positives and false negatives. At this stage, it may be appropriate to discuss the false positives and false negatives resulting from this approach with average scores and this is summarized in Table 6. For several compounds, there is substantial evidence that the wrong assignment could be due to a false positive or false negative LLNA result. Thus, if (1) the four compounds ambrettolide, serenolide, cosmone, and toscanol which all have no structural alert and which gave clear irritation reaction in the LLNA (see Table 6) would be rated as false positives in the LLNA, and (2) the human evidence would overrule the LLNA data for the parabens, then the accuracy of the data in Table 5 would be 90.5%. Such corrections *post hoc* are certainly critical, and it will be important for future validation studies to select a large array of test chemicals excluding such ambiguous compounds.

Cooper statistics versus human data. Finally, the Cooper statistics with this model (as in Table 5, average score > 0 indicating sensitizers) were also calculated for the 45 compounds for which unambiguous human data are available (as summarized in Table SV in the supplementary information and mainly taken from the LLNA validation study of Basketter *et al.*, 1999). For this subset of chemicals, the Cooper statistics against human evidence are: Sensitivity, 97.1%, Specificity, 81.8%, positive predictivity, 94.3%, negative predictivity, 90.0%, and accuracy 93.3%.

The Effect of Integrating cLogP as a Surrogate for Bioavailability in the Model with Average Scores

We have further tried to refine the data obtained by the model with average scores by integration of cLogP. Two models were tried:

1. The score for any molecule with a cLogP < −2 or > 5 was reduced by 0.5 to account for the potentially low bioavailability.

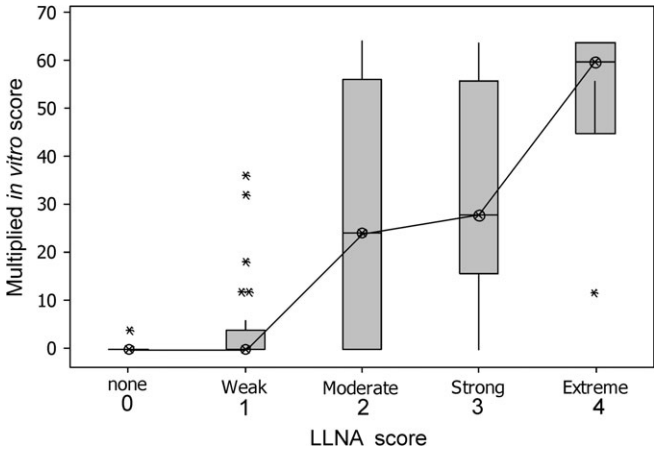

FIG. 1. Chemical classification using the original model of Jowsey *et al.* (2006) with the product of individual scores. A Box plot with the interquartile range is given and the median is indicated and connected with a line.

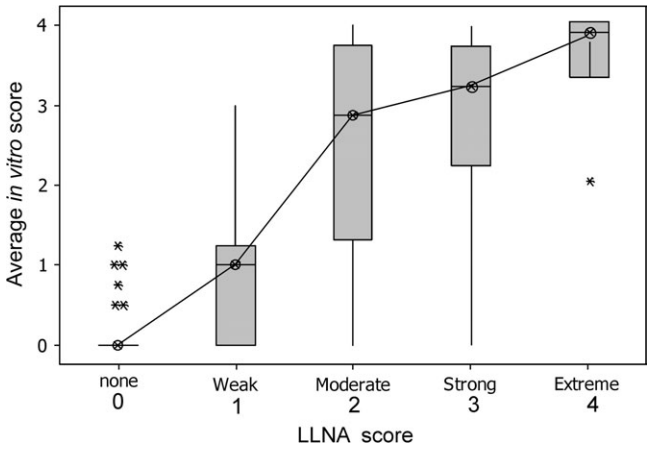

FIG. 2. Chemical classification using the average of the scores from two *in vitro* tests. A Box plot with the interquartile range is given and the median is indicated and connected with a line.

2. As above, and in addition the score for chemicals with a cLogP > −0.5 and < 3 was enhanced by 0.5 to account for the potentially higher bioavailability of these chemicals.

These refined scores were then used again for linear correlation analysis. Interestingly, the correlation coefficient R^2 was reduced from 0.518 without correction to 0.515 with the model (1) and to 0.489 with the model (2). Also the Spearman correlation coefficients were reduced from 0.758 to 0.748/0.696. If correcting the scores even more (by 1 instead of 0.5, and with the same boundaries set), the correlation was further reduced to 0.502 and 0.428 with the two models. The Spearman rank correlations became 0.746/0.686, thus with all this corrections for a cLogP outside the optimal range, the correlation and thus the predictivity of the model is reduced.

Integrating the Predictions from TIMES SS in the Model with Average Scores

In the Jowsey *et al.* (2006) proposal a yes/no scoring for structural alerts was included, increasing the scores for any chemical with a known structural alert by a factor of two. A knowledge-based approach based on established reaction mechanisms for skin sensitizers is certainly valuable to refine the *in vitro* predictions and it may especially help to correct for weak false positives in *in vitro* tests. In a weight-of-evidence based approach, one may suggest, that more convincing *in vitro* evidence is needed for a molecule to be rated positive in the complete absence of structural alerts, whereas a molecule with clear structural alerts will be rated positive even if the *in vitro* evidence is somewhat weaker. We have therefore used the TIMES SS model, and we subtracted a value of 1 from the score of any chemical which is not rated as a sensitizer by the TIMES SS model: Thus any chemical not predicted to be a sensitizer by TIMES SS but rated positive by the *in vitro* tests is put into the next lower class, and a chemical with weak *in vitro* prediction and no alert becomes a nonsensitizer. The

TABLE 4
Summary Results and Cooper Statistics for the Model with Multiplied Scores

		Predicted classification		
		Nonsensitizer	Sensitizer	Total
Chemical data classification	Nonsensitizer	31	1	32
	Sensitizer	28	56	84
	Total	59	57	116
Sensitivity	66.7%			
Specificity	96.9%			
Positive predictivity	98.2%			
Negative predictivity	52.5%			
Accuracy	75.0%			

TABLE 5
Summary Results and Cooper Statistics for the Alternative Model with Average Scores, Rating Each Chemical with a Score > 0 as a Sensitizer

		Predicted classification		
		Nonsensitizer	Sensitizer	Total
Chemical classification	Nonsensitizer	26	6	32
	Sensitizer	11	73	84
	Total	36	80	116
Sensitivity	86.9%			
Specificity	81.3%			
Positive predictivity	92.4%			
Negative predictivity	70.3%			
Accuracy	85.3%			

TABLE 6
Discussion of False Positives and False Negatives in Table 5

Generic name	Discussion
False negatives	
Phthalic anhydride	Negative in both assays, would be rated positive with inclusion of a Lys-containing peptide in reactivity assay
Trimellitic anhydride	Negative in both assays, rapid hydrolysis in any aqueous test system proposed by Mitjans *et al.* (2008)
3,4-Dihydrocoumarin	No explanation
1-Chlorooctadecane	Highly hydrophobic compound, not soluble in assay solution for peptide reactivity and culture media
Cosmone	Putative false positive in LLNA, no structural alert, Slight to moderate persistent erythema seen on ear at top dose in LLNA (25%).
Toscanol	Putative false positive in LLNA, Severe ear swelling and erythema seen at 10% and 100% dose groups. Irritant in rabbits at 100%
Lilial	No explanation
Super muguet	Putative prohapten?
Serenolide	Putative false positive in LLNA, slight ear swelling and erythema seen at 30 and 50% dose groups. Mild irritant in rabbits at 100%
Ambrettolide	Putative false positive in LLNA, EC3 = 28.8%, borderline, moderate swelling and persistent erythema seen at 100%
Ethylenediamine	Putative prohapten?
False positives	
trans-2-Hexenol	Clear structural alert as prohapten, Putative false negative in LLNA due to very high volatility. LLNA would need to be done occluded. Rapid evaporation from LLNA vehicle of *trans*-2-Hexenal, > 95% lost within 10 min (our unpublished data).
Benzalkonium chloride	No clear explanation for observed ARE induction, borderline result, with $I_{max} = 1.7$, close to threshold
Diethyl phthalate[a]	No clear explanation for observed weak ARE induction ($I_{max} = 1.9$) at maximal test concentration (1000μM), borderline result, due to impurity?
6-Methyl-coumarin[a]	No explanation for observed ARE induction
Methyl paraben[a]	Putative false negative in LLNA, sensitization reactions in humans reported
Propylparaben[a]	Putative false negative in LLNA, sensitization reactions in humans reported

[a]For these compounds the false positive result is corrected, if the correction for negative prediction by TIMES-SS is included as discussed below and summarized in Table 7.

resulting data were then corrected by setting any negative value to zero (i.e., a compound not predicted as sensitizer by both TIMES SS and i*n vitro* data keeps its score 0 and does not get a negative value). The linear correlation coefficient for this corrected sensitizations score is now slightly increased form 0.518 to 0.525, and the Spearman coefficient is increased from 0.758 to 0.769. The Cooper statistics after this correction are shown in Table 7. The accuracy is clearly improved to 87.9%, due to the reduction of false positives from *in vitro* tests, as noted in footnote to Table 6. This approach, with results integrated from the two *in vitro* assays and refined with the result from the *in silico* prediction model, give clearly an enhanced accuracy as compared with the accuracy obtained from the single *in vitro* tests.

A Further Alternative Calculation: Regression Analysis

In the calculations above, we have followed a very intuitive and not profoundly mathematical approach of data integration, based on and modified from the widely cited proposal of Jowsey *et al.*: Assigning scores to the individual test results and combining them together with a simple intuitive calculation (which is not based on an empirical mathematical model) to get a final score. However, a very classical mathematical approach of data integration from multiple endpoints is regression analysis. We have therefore further analyzed the data with regression analysis of LLNA class versus the *in vitro* data.

First, regression of LLNA class versus log EC 1.5, log I_{max}, Peptide depletion in % and cLogP was performed. This resulted in the following regression equation and statistics:

$$\text{LLNA Class} = 1.2 - 0.011\,\text{clogP} + 0.053\log I_{max} - 0.383\log\text{EC } 1.5 + 0.010\,\text{CysDpl}. \quad (1)$$

The standard error, and T and p values for the predictors in this equation are

Constant: SE = 0.455, $T = 3.99$, $p < 0.0005$
clogP: SE = 0.036, $T = -0.30$, $p = 0.765$
logI_{max}: SE = 0.287, $T = 0.18$, $p = 0.855$
log EC 1.5: SE = 0.133, $T = -2.87$, $p = 0.005$
CysDpl.: SE = 0.0029, $T = 3.56$, $p = 0.001$
$S = 0.817$, $F = 27.8$, $R^2 = 51.0\%$

logI_{max} apparently had no significant influence, and repeating the analysis without this parameter gave a similar equation with improved statistical parameters (see Equation 1b in the Supplementary Information). cLogP had also no significant influence and removing this parameter resulted in a similar and simplified equation (Equation 1c in the Supplementary Information). (Note. a logarithmic metric also for the peptide reactivity result such as measuring amount of reaction as log (100 - depletion) may appear a more appropriate parameter for regression analysis but did not result in an improved regression model).

Regression analysis was also performed based on the scores attributed to the raw data according to Table 2. The regression line for this analysis was forced through zero, based on the rationale that, without any other evidence from an *in vitro* test, a chemical is rated as zero/nonsensitizer (whereas calculating a regression with an *y*-intercept would automatically attribute to each chemical a certain minimal sensitization class). In this analysis, again, EC 1.5 had a significant influence, but the

TABLE 7
Summary Results and Cooper Statistics for the Alternative Model with Average Scores Corrected for Negative TIMES SS Predictions

		Predicted classification		
		Nonsensitizer	Sensitizer	Total
Chemical classification	Nonsensitizer	30	2	32
	Sensitizer	12	72	84
	Total	41	75	116
Sensitivity	85.7%			
Specificity	93.8%			
Positive predictivity	97.3%			
Negative predictivity	71.4%			
Accuracy	87.9%			

Note. Rating each chemical with a score > 0 as a sensitizer.

effect of I_{max} was low and nonsignificant, underlining the notion that the EC 1.5 value is the key predictor from the ARE assay (see Equation 2b in Supplementary Information). Therefore, only the results for regression with the EC 1.5 score, the Cys-depletion score and cLogP are given below in Equation 2:

$$\text{LLNA Class} = 0.319\,\text{Cys Score} + 0.376\,\text{EC 1.5 Score} + 0.084\,\text{clogP} \quad (2)$$

The standard error, and T and p values for the predictors in this equation are

cLogP: SE = 0.0315, $T = 2.66$, $p = 0.009$
EC 1.5 Score: SE = 0.075, $T = 4.92$, $p < 0.0005$
Cys Score: SE = 0.065, $T = 4.90$, $p < 0.0005$
$S = 0.871$, $F = 132.9$

Repeating this analysis without cLogP yielded the following result:

$$\text{LLNA Class} = 0.357\,\text{Cys Score} + 0.391\,\text{EC 1.5 Score} \quad (3)$$

The standard error, and T and p values for the predictors in this equation are

EC 1.5 Score: SE = 0.071, $T = 5.53$, $p < 0.0005$
Cys Score: SE = 0.064, $T = 5.48$, $p < 0.0005$
$S = 0.896$, $F = 183.76$

Both by directly analyzing the raw data (Equation 1) or by using the scores (Equations 2 and 3), it is clear that both the Cys-depletion and the EC 1.5 similarly (the coefficients are indeed similar in Equations 2 and 3!) and highly significantly contribute to the prediction of the sensitization potential class, but the cLogP has negligible influence in the analysis of the raw data and a small influence in Equation 2. This nicely confirms the more intuitive approach of data evaluation chosen above, which had shown that an arithmetic combination (calculating averages) of the scores form both i*n vitro* tests without integrating cLogP yields the best prediction. Indeed, the intuitive, nonempirical model with average scores can be expressed by Equation 4, which indeed is very similar to the empirically calculated Equation 3:

$$\text{LLNA Class} = 0.5\,\text{Cys Score} + 0.5\,\text{ARE Score} \quad (4)$$

DISCUSSION

Here we have jointly reported data from two *in vitro* assays and one *in silico* prediction model for a large set of chemicals of known skin sensitization potential. These data are also reported in an Excel format in the Supplementary information to allow the scientific community to perform more sophisticated calculations on them. Different ways of data integration were explored and they are discussed below.

Multiplication of Scores versus Average of Scores

Starting from the proposal of Jowsey *et al*. (2006), we have scored the data in classes and performed multiple calculations. The original proposal rates a chemical positive only if it is positive in all individual tests. A logical rationale behind this approach is the assumption that a chemical must be able to react with a protein AND stimulate dendritic cells AND stimulate T-cell proliferation; these are all hurdles that must be overcome to give the sensitization reaction.

Multiplication of the scores gives a high specificity: Indeed all the chemicals positive in both assays are rated as sensitizers by the LLNA, with 2-hexenol as the only false positive. With 2-hexenal being a moderate sensitizer, 2-hexenol may act as prohapten. The negative result in the LLNA may be due to its high volatility, which would explain it as a false negative in the LLNA rather than false positive *in vitro*. On the other hand the sensitivity of the approach with multiplication of scores is low, with several sensitizers only recognized by one of the *in vitro* test. This problem might be further enhanced if several cell-based tests system with a limited sensitivity for weak sensitizers will be combined.

A multiplication of the scores, as in the original proposal, thus did not appear to be the most useful approach to integrate the data. Calculating average scores from different tests is another intuitive possibility and it gives a more transparent view of the data. Whereas the original proposal is based on a mechanistic approach (a chemical must overcome a series of hurdles to sensitize), this data integration is based on a weight-of-evidence approach: If *in vitro* results are sufficiently indicative of sensitization, the chemical is scored positive. This approach already allows for a better yes/no prediction of

the skin sensitization potential, and on the average it gives a good estimate of the sensitization potency (Fig. 2). However, the prediction of potency is not yet satisfying at the level of the individual chemical.

Regression Analysis instead of a Intuitive Data Integration

To evaluate the data in a more mathematical way, regression analysis was used. Indeed regression analysis indicated that the scores obtained from the two *in vitro* assays both highly significantly contribute to the prediction of the sensitization class. When performing the regression analysis with the two *in vitro* scores only, the best regression statistics were obtained (Equation 3), and interestingly this regression equation is very close to the intuitive approach with average scores from the tests (expressed by Equation 4). The regression analysis also indicated that both assays similarly contribute to the prediction (similar coefficients in Equations 2 and 3).

The Effect of cLogP

At present it is clear that chemicals with a wide range of physicochemical properties can sufficiently gain access to the viable epidermis in order to induce skin sensitization, with only very hydrophilic or highly hydrophobic compounds having a reduced sensitization potential (recently reviewed by Roberts and Aptula, 2008). Classical data sets from skin penetration studies contain penetration data from aqueous solution, but both LLNA tests and cosmetic usage involve application of chemicals either from an organic solvent or from mixed lipid/surfactant/water systems. Thus classical skin penetration models are of questionable value to predict the sensitization situation. This is an area of much current research (reviewed by Basketter *et al*., 2007). Nevertheless, cLogP has often been shown to be a simple indicator correlating to bioavailability in the skin, and thus attempts were made to integrate cLogP into the prediction models. In the model with average scores, a reduced final score was given for chemicals with cLogP outside the range between −2 and 5, as chemicals outside this range are considered to be less bioavailable. Yet using cLogP with this intuitive approach to account for potentially different bioavailability did reduce the correlation. This finding was also confirmed by regression analysis, with cLogP making only a small contribution to the overall regression equation.

Integrating the In Silico *Prediction*

In the original proposal, the *in vitro* score for any chemical with a structural alert is doubled. In the model with average scores we have taken a more conservative approach, by putting each chemical without a structural alert into the next lower sensitization class. With this knowledge-based approach, absence of structural alerts reduces predicted sensitization potential, but a known structural alert does not further enhance the *in vitro* score: if *in vitro* tests rate a molecule a moderate sensitizer, this classification is not changed by the mere fact that this has also been predicted based on existing knowledge.

TIMES SS model was selected as the preferred model because it incorporates structure-toxicity and structure-metabolism relationships through a number of transformations simulating skin metabolism and interaction of the parent molecule and reactive metabolites with skin proteins (Roberts *et al*., 2007). Other models are available, including statistical quantitative structure activity relationships such as TOPKAT (Accelrys, San Diego, CA) and MCASE (Mulitcase, Inc., Beachwood, OH) or expert systems such a Derek for Windows (LHASA Inc., Leeds, UK) which could also be investigated for use in such a decision process.

Pooled or Separate Analysis of Different Structural Classes

Here we have combined all structural classes together for a simple pooled data analysis. However, it has been argued, that the evaluation of the sensitization potential should always be done for "applicability domains" (Roberts *et al*., 2007b). Indeed, if the data were analyzed for single classes, a more refined result may be obtained. For example Michael acceptors are rated quite high both with the ARE assay and with the Cys-peptide reactivity (see Table 3) and therefore many moderate chemicals in this structural class are rated as strong or extreme with the generalized model, and this could be avoided with a class-wise treatment of the data.

However, many structural classes contain only a low number of chemicals and thus, a separate data interpretation model cannot be developed. Another option is then to use the *in vitro* data as a basis for read-across: The *in vitro* data of the novel compound would directly be compared with the *in vitro* data of the closest structural/mechanistic neighbors in the dataset in order to make an informed rating. Only chemicals with no structural relatives in the data set would be rated solely based on a global model such as the ones presented here using the scores from the *in vitro* data.

Selected Tests and Test Improvements

The peptide reactivity assay and the ARE assay were selected, because for these tests the largest data sets are available and because they give a quantitative read-out. Both tests have proven to be valuable to rate a large number of chemicals and combining the results already gives an improved prediction of the sensitization risk. Yet both tests may still need refinement:

1. We had shown that depletion in the peptide reactivity assay may be due to adduct formation or peptide oxidation (Natsch *et al*., 2007). Although adduct formation with proteins is considered a hallmark of the sensitization process, it is not established whether the capacity of a chemical to catalyze peptide oxidation is a relevant predictor for sensitization. A test to discriminate between oxidizing and adduct forming chemicals will be published shortly.
2. The ARE cell-based assay is based on a breast cancer cell line. This test may already be sufficient, as the Nrf2/Keap1/

ARE regulatory pathway is present in most or all cell types. However, it might be possible to improve the test by using cell types directly involved in the sensitization process.

Furthermore, as more data on the dendritic cell assays with CD86 expression (such as the hCLAT test or the U-937 test; Ade *et al*., 2006; Sakaguchi *et al*., 2006, in press) accumulate, it will be interesting to repeat the current analysis with these data to further analyze the contribution of these additional tests for a more optimized prediction of the sensitization risk.

Finally, both in the current approach and in the original proposal by Jowsey *et al*., no special attention is paid to prohaptens. Without additional tests, prohaptens would need to be metabolized and recognized directly within the cell-based assay(s). Indeed several putative prohaptens (for example dihydroeugenol) are positive in the ARE assay. Whether this is sufficient, or whether a specific test for metabolic activation needs to be added, will also need a careful assessment.

Further Research Directions

Along with refinements of these tests and the inclusion of other tests and *in silico* models, also a refinement of the calculations will be needed. Whether it is useful to use ordinal data for classification or whether a direct regression analysis based on raw data (e.g., inducing concentrations from cell-based assays or kinetic constants from peptide reactivity) might be a better choice needs to be further explored. Also a decision tree approach, weighing the evidence from different *in vitro* tests in a tiered approach, needs to be considered.

It will be important to include large chemical data sets to test the validity of each single test and the validity of integration from different tests and it is of special importance to perform these validations on a standard list of real positives and real negatives: As exemplified in Table 6, the LLNA results for some compounds included in this study are questionable. Because there is no such standard list, we included all our data in the evaluation, and the fact that the dataset contained borderline results certainly did deteriorate the Cooper statistics.

Even if the ultimate standardized tests and all the final data sets are not yet available, we felt it timely to publish this first integrated dataset and apply the calculations of the Jowsey *et al*. proposal to the data in order that the discussion can progress how data can be integrated in the ''battery concept'' for skin sensitization testing.

SUPPLEMENTARY DATA

Supplementary data are available online at http://toxsci.oxfordjournals.org/.

FUNDING

Givaudan Schweiz AG.

ACKNOWLEDGMENTS

We would like to thank M. Rothaupt for HPLC analysis and Boris Schilling for critical discussions.

REFERENCES

Ade, N., Martinozzi-Teissier, S., Pallardy, M., and Rousset, F. (2006). Activation of U937 cells by contact sensitizers: CD86 expression is independent of apoptosis. *J. Immunotoxicol.* **3,** 189–197.

Aeby, P., Wyss, C., Beck, H., Griem, P., Scheffler, H., and Goebel, C. (2004). Characterization of the sensitization potential of chemicals by *in vitro* analysis of dendritic cell activation and skin penetration. *J. Invest. Dermatol.* **122,** 1154–1164.

Basketter, D. A., Evans, P., Fielder, R. J., Gerberick, G. F., Dearman, R. J., and Kimber, I. (2002). Local lymph node assay—Validation and use in practice. *Food Chem. Toxicol.* **40,** 593–598.

Basketter, D. A., Lea, L. J., Cooper, K., Stocks, J., Dickens, A., Pate, I., Dearman, R. J., and Kimber, I. (1999). Threshold for classification as a skin sensitizer in the local lymph node assay: A statistical evaluation. *Food Chem. Toxicol.* **37,** 1167–1174.

Basketter, D., Pease, C., Kasting, G., Kimber, I., Casati, S., Cronin, M., Diembeck, W., Gerberick, G., Hadgraft, J., Hartung, T., *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 59a. *ATLA Altern. Lab. Anim.* **35,** 137–154.

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

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

Coquette, A., Berna, N., Vandenbosch, A., Rosdy, M., De Wever, B., and Poumay, Y. (2003). Analysis of IL-1β and IL-8 expression and release in *in vitro* reconstructed human epidermis for the prediction of *in vivo* skin irritation and/or sensitization. *Toxicol. In Vitro* **17,** 311–321.

Dimitrov, S. D., Low, L. K., Patlewicz, G. Y., Kern, P. S., Dimitrova, G. D., Comber, M. H. I., Phillips, R. D., Niemela, J., Bailey, P. T., and Mekenyan, O. G. (2005). Skin sensitization: Modelling based on skin metabolism simulation and formation of protein conjugates. *Int. J. Toxicol.* **24,** 189–204.

Dinkova-Kostova, A. T., Holtzclaw, W. D., and Kensler, T. W. (2005). The role of Keap1 in cellular protective responses. *Chem. Res. Toxicol.* **18,** 1779–1791.

Gerberick, G. F., Ryan, C., Dearman, R. J., and Kimber, I. (2007a). Local lymph node assay (LLNA) for detection of sensitization capacity of chemicals. *Methods* **41,** 54–60.

Gerberick, G. F., Ryan, C. A., Kern, P. S., Dearman, R. J., Kimber, I., Patlewicz, G. Y., and Basketter, D. (2004a). A chemical dataset for evaluation of alternative approaches to skin-sensitization testing. *Contact Dermatitis* **50,** 274–288.

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

Gerberick, G. F., Vassallo, J. D., Foertsch, L. M., Price, B. B., Chaney, J. G., and Lepoittevin, J.-P. (2007b). Quantification of chemical peptide reactivity for screening contact allergens: A classification tree model approach. *Toxicol. Sci.* **97,** 414–427.

Gerberick, F., Aleksic, M., Basketter, D., Casati, S., Karlberg, A.-T., Kern, P., Kimber, I., Lepoittevin, J. P., Natsch, A., Ovigne, J. M., *et al*. (2008). Chemical reactivity measurement and the predictive identification of skin

sensitisers: The report and recommendations of ECVAM Workshop 64a. *ATLA Altern. Lab. Anim.* **36,** 215–242.

Gildea, L. A., Ryan, C. A., Foertsch, L. M., Kennedy, J. M., Dearman, R. J., Kimber, I., and Gerberick, G. F. (2006). Identification of gene expression changes induced by chemical allergens in dendritic cells: Opportunities for skin sensitization testing. *J. Invest. Dermatol.* **126,** 1813–1822.

Jowsey, I. R., Basketter, D. A., Westmoreland, C., and Kimber, I. (2006). A future approach to measuring relative skin sensitizing potency: A proposal. *J. Appl. Toxicol* **26,** 341–350.

Kim, H.-J., Barajas, B., Wang, M., and Nel, A. E. Nrf2 activation by sulforaphane restores the age-related decrease of TH1 immunity: Role of dendritic cells. *J. Allergy Clin. Immunol.* **121,** 1255–1261.

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

Mitjans, M., Viviani, B., Lucchi, L., Galli, C. L., Marinovich, M., and Corsini, E. (2008). Role of p38 MAPK in the selective release of IL-8 induced by chemical allergen in naive THP-1 cells. *Toxicol. In Vitro* **22,** 386–395.

Natsch, A., and Emter, R. (2008). Skin sensitizers induce antioxidant response element dependent genes: Application to the *In Vitro* testing of the sensitization potential of chemicals. *Toxicol. Sci.* **102,** 110–119.

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

Roberts, D. W., and Aptula, A. O. (2008). Determinants of skin sensitisation potential. *J. Appl. Toxicol.* **28,** 377–87.

Roberts, D. E., Patlewicz, G., Dimitrov, S. D., Low, L. K., Aptula, A. O., Kern, P. S., Dimitrova, G. D., Comber, M. I. H., Phililips, R. D., Niemela, J., *et al.* (2007a). TIMES-SS—A mechanistic evaluation of an external validation study using reaction chemistry principles. *Chem. Res. Toxicol.* **20,** 1321–1330.

Roberts, D. W., Patlewicz, G., Kern, P. S., Gerberick, F., Kimber, I., Dearman, R. J., Ryan, C. A., Basketter, D. A., and Aptula, A. O. (2007b). Mechanistic applicability domain classification of a local lymph node assay dataset for skin sensitization. *Chem. Res. Toxicol.* **20,** 1019–1030.

Ryan, C. A., Gildea, L. A., Hulette, B. C., Dearman, R. J., Kimber, I., and Gerberick, G. F. (2004). Gene expression changes in peripheral blood-derived dendritic cells following exposure to contact allergens. *Toxicol. Lett.* **150,** 301–316.

Sakaguchi, H., Ashikaga, T., Miyazawa, M., Kosaka, N., Ito, Y., Yoneyama, K., Sono, S., Itagaki, H., Toyoda, H., and Suzuki, H. The relationship between CD86/CD54 expression and THP-1 cell viability in an *in vitro* skin sensitization test—Human cell line activation test (h-CLAT). *Cell Biol. Toxicol.* (in press).

Sakaguchi, H., Ashikaga, T., Miyazawa, M., Yoshida, Y., Ito, Y., Yoneyama, K., Hirota, M., Itakagi, H., Toyoda, H., and Suzuki, H. (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* **20,** 774–784.

Sanderson, D. M., and Earnshaw, C. G. (1991). Computer prediction of possible toxic action from chemical structure; the DEREK system. *Hum. Exp. Toxicol.* **10,** 261–273.

Smith, C. M., and Hotchkiss, S. A. M. (2001). In *Allergic Contact Dermatitis: Chemical and Metabolic Mechanisms.* Taylor and Francis, London.

Wakabayashi, N., Dinkova-Kostova, A. T., Holtzclaw, W. D., Kang, M., Kobayashi, A., Yamamoto, M., Kensler, T., and Talalay, P. (2004). Protection against electrophile and oxidant stress by induction of the phase 2 response: Fate of cysteines of the Keap1 sensor modified by inducers. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 2040–2045.

Wang, X. J., Hayes, J. D., and 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,** 10983–10994.

Zhang, X., Chen, X., Song, H., Chen, H., and Rovin, B. (2005). Activation of the Nrf2/antioxidant response pathway increases IL-8 expression. *Eur. J. Immunol.* **35,** 3258–3267.