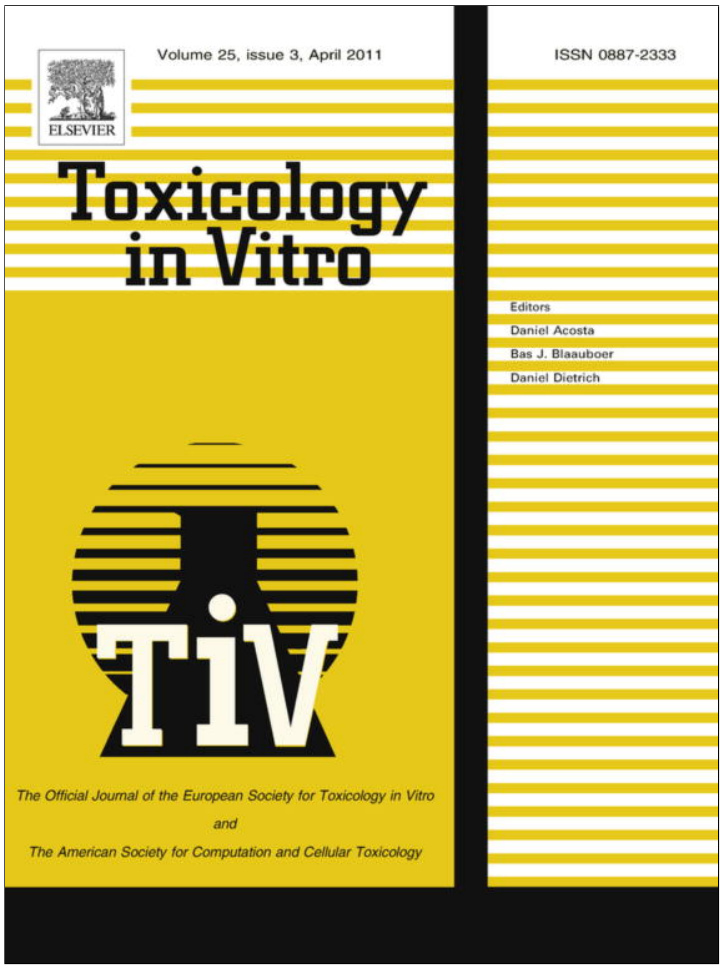

Toxicology in Vitro 25 (2011) 733–744

Contents lists available at ScienceDirect

Toxicology in Vitro

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

The intra- and inter-laboratory reproducibility and predictivity of the KeratinoSens assay to predict skin sensitizers *in vitro*: Results of a ring-study in five laboratories

Andreas Natsch [a,*], Caroline Bauch [b], Leslie Foertsch [c], Frank Gerberick [c], Kimberly Norman [d], Allison Hilberer [d], Heather Inglis [d], Robert Landsiedel [b], Stefan Onken [e], Hendrik Reuter [e], Andreas Schepky [e], Roger Emter [a]

[a] *Givaudan Schweiz AG, Ueberlandstrasse 138, CH-8600 Duebendorf, Switzerland*
[b] *BASF SE, GV/TB – Z570, D-67056 Ludwigshafen, Germany*
[c] *Procter and Gamble Inc., Miami Valley Laboratories, Cincinnati, USA*
[d] *The Institute for In Vitro Sciences, Inc. (IIVS), 30 W. Watkins Mill Rd., Ste. 100, Gaithersburg, MD 20878, USA*
[e] *Beiersdorf AG, Unnastrasse 48, D-20253 Hamburg, Germany*

ARTICLE INFO

Article history:
Received 4 October 2010
Accepted 20 December 2010
Available online 30 December 2010

Keywords:
Skin sensitization
Alternatives to animal tests
Intralaboratory reproducibility
Inter-laboratory reproducibility
Transferability
Reporter gene
KeratinoSens

ABSTRACT

Due to regulatory constraints and ethical considerations, research on alternatives to animal testing to predict the skin sensitization potential of novel chemicals has gained a high priority. Accordingly, different *in vitro, in silico* and *in chemico* approaches have been described in the scientific literature to achieve this goal. To replace regulatory approved animal tests, these alternatives need to be transferable to other labs, their within and between laboratory reproducibility must be assured, and their predictivity should be high. The KeratinoSens assay is a cell-based reporter gene assay to screen substances with a full dose–response assessment. It is based on a stable transgenic keratinocyte cell line. The induction of a luciferase gene under the control of the antioxidant response element (ARE) derived from the human *AKR1C2* gene is determined. Here we report on the results of a ring-study with five laboratories performing the KeratinoSens assay on a set of 28 test substances. The assay was found to be easily transferable to all laboratories. Overall both the qualitative (sensitizer/non-sensitizer categorization) and the quantitative (concentration for significant gene induction) results were reproducible between laboratories. A detailed analysis of the transferability, the within- and between laboratory reproducibility and the predictivity is presented.

1. Introduction

Cosmetic legislation in Europe has imposed a ban on animal testing for the detection of the skin sensitization hazard of cosmetic ingredients by 2013. Therefore, predictive *in vitro* tests are urgently needed to maintain the ability of the industry to invent new products and to guarantee their safety after 2013. At the same time, the REACH regulation (registration, evaluation, authorization and restriction of chemicals) requires the testing of thousands of not previously tested substances for their skin sensitization potential, thus increasing animal testing. Currently, the skin sensitization potential is estimated with the local lymph node assay in mice (LLNA), in which the cellular proliferation in the draining lymph nodes is measured after repeated topical application of the test substance onto the ears. Results are expressed as EC3 values which indicate the concentration which induces a threefold increase in lymph node cell proliferation as measured by ^{3}H-thymidine uptake (Basketter et al., 2002).

Skin sensitization is an immune reaction to small exogenous molecules. In general, skin sensitizing molecules are reactive chemicals (or chemicals metabolically transformed into reactive intermediates) which have the potential to react with skin proteins

Abbreviations: LLNA, local lymph node assay; Nrf2, nuclear factor-erythroid 2-related factor 2; Keap1, Kelch-like ECH-associated protein 1; ARE, antioxidant response element; SOP, standard operating procedure; WLR, within laboratory reproducibility; BLR, between laboratory reproducibility; PC, predictive capacity; ECVAM, European centre for the validation of alternative methods to animal testing; ICCVAM, interagency coordinating committee on the validation of alternative methods; RT-PCR, reverse transcriptase polymerase chain reaction; DMSO, dimethylsulfoxide; DNCB, 2,4-dinitrochlorobenzene; SLS, sodium lauryl sulphate; MCI, (5-Chloro)-methyl-isothiazolinone; IC50, inhibitory concentration for 50% reduction in viability as determined with the MTT assay; EC 1.5, extrapolated concentration for 1.5-fold luciferase induction above threshold; REACH, registration, evaluation, authorization and restriction of chemicals; h-CLAT,human cell line activation test; MUSST, myeloid U937 skin sensitization test.

* Corresponding author. Tel.: +41 44 824 21 05; fax: +41 44 824 29 26.
E-mail address: andreas.natsch@givaudan.com (N. Andreas).

0887-2333/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tiv.2010.12.014

and render them immunogenic (Karlberg et al., 2008). In the sensitization phase, the modified proteins are processed by dendritic cells and presented in the local lymph nodes, where they provoke the expansion of allergen-specific T-cell clones. Upon repeated contact with the skin sensitizer in the elicitation phase, a specific T-cell mediated immune response then leads to the disease status known as contact allergy (Kimber et al., 2002).

Although this complex cascade of events is difficult to model, three non-animal approaches are in advanced development. It is thought that a combination of these three approaches may finally replace the current animal tests (Jowsey et al., 2006; Basketter and Kimber, 2009; Natsch et al., 2009): (i) *In silico* models predict the sensitization potential of chemicals based on a number of rules or structural alerts empirically derived from databases on known skin sensitizers (Roberts et al., 2007). (ii) *In chemico* models determine the sensitization potential based on an assessment of the chemical reactivity of the test chemicals with nucleophiles, normally model peptides (Gerberick et al., 2008). (iii) Lastly, the cell-based *in vitro* models evaluate certain aspects of the cellular response to skin sensitizers. Most of these latter assays address the unspecific innate response of the skin to sensitizing agents rather than the specific T-cell response.

The *in vitro* assays usually use dendritic cells or keratinocytes, the two cell types in first contact with the topically applied sensitizers. The specific markers investigated in these cells may be selected either at the protein level or at the transcription level. At the protein level, the surface markers CD86 and CD54 and the chemokine IL-8 and IL-18 have been investigated in most detail as predictive markers (Aeby et al., 2004; Toebak et al., 2006; Corsini et al., 2009). Based on these studies, the h-CLAT (Ashikaga et al., 2006) and MUSST *in vitro* assays, which measure surface marker expression in THP-1 and U-937 cells, respectively, have entered pre-validation.

At the transcription level, a number of research groups have reported novel bio-markers for skin sensitizers. Most studies at the transcription level have been based on gene chip analysis of primary cells exposed to sensitizers or irritants (Ryan et al., 2004; Hooyberghs et al., 2008; Python et al., 2009). Several research groups then followed up these initial studies with RT-PCR analysis of the identified markers (Gildea et al., 2006; Python et al., 2009). However, attempts to transfer this genetic data obtained with primary cells into tests based on stable cell lines have proven surprisingly difficult (Lambrechts et al., 2009; Python et al., 2009). To date, no assay based on transcriptional changes has been submitted for (pre)validation, and the reasons for this could be (i) the difficulties in reproducing the gene expression changes in stable cell lines and (ii) the difficulties in fully standardizing RT-PCR based measurements.

We have proposed a pragmatic reporter cell-based approach based on the finding that the majority of the skin sensitizers induce the Nrf2-Keap1-ARE regulatory pathway (Natsch and Emter, 2008; Ade et al., 2009; Natsch, 2010; Vandebriel et al., 2010). The antioxidant response element (ARE) from the human AKR1C2 gene was inserted in front of a SV40 promotor and placed upstream of a luciferase gene. Stable insertion of the resulting construct in HaCaT keratinocytes resulted in the KeratinoSens reporter cell line. Induction of luciferase in this cell line can be used to screen for skin sensitizers. Since the Nrf2-Keap1-ARE regulatory pathway is involved in antioxidant response signalling, the ability to differentiate between sensitisation, irritation and antioxidant potential solely based on the ARE assay may have limitations but so far the assay has shown high predictivity. Indeed, the predictivity has been analyzed on a set of 67 reference test substances and an overall accuracy of 85.1% was determined. Due to this high predictivity and since it is based on the highly reproducible and technically simple luciferase-expression readout, this assay was considered a good candidate to be transferred to other laboratories and to enter (pre)validation. We report here the results of a detailed ring-study in five laboratories on a total of 28 reference substances.

2. Materials and methods

2.1. Test substances and study setup

The substances selected for this study cover (i) all the substances in the publication by Casati et al. (2009), which is largely overlapping with the list used by the Sens-it-iv consortium (Sens-it-iv, 2009) and (ii) all the substances in the draft performance standards for alternative endpoints in the LLNA published by ICCVAM (2008b). Diethyl phthalate was added as additional negative control from the list published by Sens-it-iv to make up the total number to 28 test substances. Supporting information 1 lists these substances, their commercial source and batch, CAS-number, and their sensitization potential as determined by the LLNA as given in the ICCVAM database (ICCVAM, 2008a) and in Basketter et al. (1999)).

This study-set was divided into two groups: A first set of 7 test substances was used to assess the transferability of the assay and for an initial assessment of the reproducibility (Phase I). The remaining 21 substances were sent to the laboratories blind-coded to assess the reproducibility and predictivity in detail (Phase II). The following substances were selected for Phase I: Three negatives (chlorobenzene, methyl salicylate, and sulfanilamide), three clear positives (DNCB, citral, and ethylene glycol dimethacrylate), and a borderline test substance according to the historical data published before (hexyl cinnamic aldehyde). The remaining 21 substances in Supporting information 1 were used for Phase II.

2.2. Cell line

The KeratinoSens cell line is derived from the human keratinocyte cell line HaCaT (Boukamp et al., 1988). It contains a stable insertion of a luciferase gene under the control of the ARE-element of the gene AKR1C2. The optimization of this cell line has been described in detail (Emter et al., 2010).

2.3. Test procedure and standard operating procedure (SOP)

All tests were run according to the previously published SOP (Emter et al., 2010). Briefly, cells were grown for 24 h in 96-well plates. The medium was then replaced with medium containing the test substance and a final level of 1% of the solvent, DMSO. Each test substance was tested at 12 twofold dilutions ranging from 0.98 to 2000 μM. Each 96-well test plate contained 7 serially diluted test substances, 6 wells with the solvent control, 1 well with no cells for background value and 5 wells with the positive control, cinnamic aldehyde, in five different concentrations. In each repetition, three parallel replicate plates were run with this same set-up and a fourth parallel plate was prepared for cytotoxicity determination. Cells were incubated for 48 h with the test substances, and then luciferase activity and cytotoxicity (with the MTT assay) were determined. This full procedure was repeated three times for each chemical, thus generating 9 luciferase induction data points and 3 MTT datapoints for each chemical at each concentration in each lab.

2.4. Controls and acceptance criteria

All the labs performed three repetitions consisting of three replicates on the Phase I substances and sent these data to the lead lab. Data quality was assessed by the lead lab whether they fulfil

the following criteria: (i) Variability in the 18 DMSO-control wells for each triplicate experiment is below 20% in all three repetitions (ii) dose–response curves are reproducible within the laboratory (i.e. increasing gene activation with increasing concentration up to the cytotoxic levels; EC1.5 and EC3 values which are no more then one well up and down in the dilution series from the average), and (iii) the positive control cinnamic aldehyde (contained in each test plate) gives a statistically significant induction above 1.5-fold below 64 μM in all three repetitions. Once this criteria were met by the three consecutive repetitions done by a particular lab, these data were taken as the final data of Phase I, and the individual labs were allowed to move into Phase II evaluation. Each lab then performed three repetitions consisting of three replicates on the Phase II substances and these data were directly used as the final data, and no data were rejected at this stage. This approach was chosen in order to avoid repeated testing and to gain experience on how robust the assay is and how narrowly the acceptance criteria should be defined in the future. More narrow criteria for the EC1.5 value of the positive control cinnamic aldehyde were initially also defined, but experiments were accepted even if these criteria were not met, again to gain experience on how narrow the criteria should be applied in the future. The results for positive and negative controls are all reported in the results section and in the Supporting information.

2.5. Modifications in the different laboratories for luciferase measurements

Each lab used their own quality of fetal calf serum and D-MEM medium but otherwise strictly adhered to the SOP. The only significant difference amongst the laboratories was the use of different approaches for luminescence readings, as not all laboratories were equipped with the same luminometer. The lead lab and lab 3 used the Glomax luminometer from Promega (Duebendorf, Switzerland), lab 1 used the infinity F500 (Tecan, Männedorf, Switzerland), lab 4 used the FLUOstar OPTIMA (BMG Labtech Inc., Cary, NC, USA) and lab 2 used the Orion II/MPL4 microplate luminometer (Berthold Detection Systems GmbH, Pforzheim, Germany). All laboratories used an injector and a flash substrate to inject the substrate to the lysed cells immediately before luciferase readings, with the exception of lab 2 which used the Perkin–Elmer Neolite Assay Kit, which is added simultaneously to all wells of the entire assay plate without the need of an injector.

2.6. Data analysis and statistical evaluation

For each chemical in each repetition and at each concentration, the gene induction compared to DMSO controls and the wells with statistically significant induction over the threshold of 1.5 (i.e. 50% enhanced gene activity) were determined. Furthermore the maximal fold-induction (I_{max}) and the EC1.5 value (concentration in μM for induction above the threshold, based on linear extrapolation as done in the LLNA) were calculated. The following prediction model was applied: A substances is rated as positive if the following three criteria are fulfilled (i) The EC1.5 value is below 1000 μM in all three repetitions or in at least 2 repetitions, (ii) at the lowest concentration with a gene induction above 1.5-fold (i.e. at the EC 1.5 determining value), the cellular viability is above 70% and (iii) there is an apparent overall dose–response for luciferase induction, which is similar between the repetitions.

To quantitatively compare the EC1.5 values and the IC50 values for cytotoxicity from the different laboratories, a logarithmic evaluation of the variability was made. The logarithmic values of the concentration data (EC1.5 and IC50) were calculated with the base 2. This is more intuitive as compared to base 10, since twofold dilutions were tested. Thus the Log_2-transformed values correlate to the number of wells in the plate to reach EC1.5/IC 50 values. Based on these Log_2-transformed values, the logarithmic standard deviations were calculated. These values were then re-transformed calculating the exponential function with base 2 (i.e. $2^{stdev(log\ values)}$), thereby rendering the geometrical standard deviation, which corresponds to a factor. (Numerical example: If the standard deviation of the Log_2 transformed values is 0.5, the geometric standard deviation is 1.414 or the square root of 2. The logarithmic 95.4% confidence interval then becomes ± 1 (i.e. twice the standard deviation) and the geometric (or re-transformed) 95.4% confidence interval is confined by a factor of 2. Thus in this specific case, the 95.4% confidence interval is covered by the concentration range one well in the microtiter plate up and down of the geometric mean.) There are two reasons for this approach: (i) the data approximate a log-normal distribution better than a normal distribution (data not shown) and therefore logarithmic calculations better describe the variability of the data and (ii) the geometric standard deviations are scale-independent and can be compared between substances of differing potency.

3. Results

3.1. Phase I – Transferability phase

No face-to-face training was required to transfer the method, and the SOP was found sufficiently detailed to perform the test in all the four external laboratories. There were no significant technical obstacles specific to the method. The key technical issues identified in the transfer phase were due to the different luminometer reading methods which initially yielded variable background readings and/or gradients over the assay plates in some labs. These issues are reviewed in the discussion section. Once they were solved, each lab performed three successive and successful repetitions and no further data needed to be discarded. All dose–response graphs for the 7 test substances tested in Phase I are presented in Supporting information 2. The results for DNCB and sulfanilamide are shown as examples in Figs. 1 and 2. Table 1 shows the numerical analysis of the data from phase I (I_{max}, EC1.5 and IC50 values and number of positive repetitions).

The positive test substances, DNCB, citral and ethylene glycol dimethacrylate, were positive in all 5 laboratories, and in all three repetitions. The three negative substances, chlorobenzene, methyl salicylate and sulfanilamide, were overall negative in all labs, with a few cases of a borderline induction in one of the three repetitions. The dose–response curves clearly confirm the positive rating for the three positive test substances and the negative rating for the three negatives (see Supporting information 2). The borderline test substance, hexyl cinnamic aldehyde, was consistently positive in two labs, negative in one lab, and gave a mixed result in two labs.

The EC1.5 values for DNCB and citral were similar in the four external labs and also close to the historical and new data of the lead lab. Also, for hexyl cinnamic aldehyde, the EC1.5 values in the positive repetitions show little variance (Table 1). However, a somewhat higher variation was observed for ethylene glycol dimethacrylate. The IC50 measure for cytotoxicity also proved to be reliable. For example, for DNCB the historical and the ring-study IC50 values of the lead lab were 8.2 and 10.1 μM, respectively, and the values from the four external labs were all between 6.6 and 12.5 μM.

3.2. Phase II – results for blind-coded substances

Once the laboratories had successfully tested the seven test substances of Phase I, they progressed to testing the 21 blind-coded substances of Phase II. These dose–response curves are all

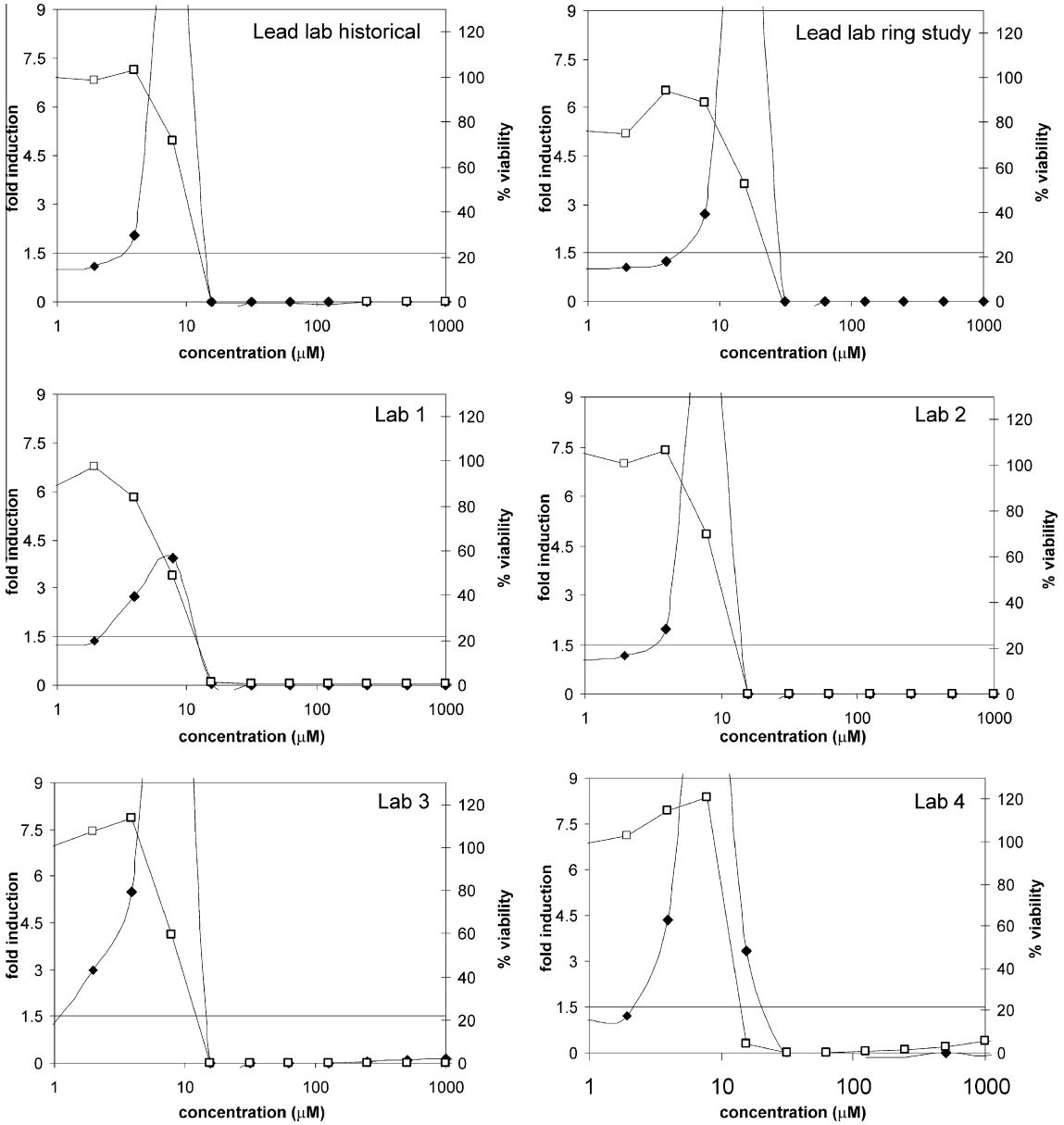

Fig. 1. Induction of luciferase activity (closed diamonds) and cellular viability (open squares) for DNCB in full dose–response analysis according to the SOP. Historical results of the lead lab and new data from the five labs in the transferability phase (Phase I). Graphs represent average results of 9 data points (3 independent trials with 3 replicates per trial).

presented in Supporting information 3, and the I_{max}, EC1.5 and IC50 values are summarized in Table 2. As illustrated by the dose–response graphs, similar overall dose–response results for these blind-coded test substances were obtained by the different labs. The biggest variation was observed in the dynamic range, i.e. the maximal gene induction I_{max} can vary significantly between the laboratories. However, the EC1.5, which had been found the key parameter to quantify the luciferase response (Natsch et al., 2009), appears more reproducible and is further analyzed quantitatively below.

Among the 15 sensitizers tested in Phase II, 11 were rated positive in all three repetitions in all five labs, and for 2 sensitizers there was one single negative repetition in one lab. For these 13 test substances the dose response curves are very clear. Phenyl benzoate is a false-negative in the historical data and this was confirmed with three negative repetitions in four labs, with one single positive repetition in one lab.

The only chemical which gave contradicting results in the between laboratory reproducibility (BLR) assessment of phase II is eugenol. In three labs, this test item was negative in 2 out of 3 repetitions and it was negative in the historical data, but rated positive in all repetitions in two labs with very reproducible dose–response curves. Among the 6 non-sensitizers, 4 were rated negative in all three repetitions in all the five labs. SLS was negative in four labs with 1–3 repetitions giving significant induction at cytotoxic concentrations only. There was one exception: in one lab significant luciferase induction was observed at the same concentration as in the other labs, but the cells were still fully viable in the parallel MTT plates at the inducing concentration. Diethyl phthalate was negative in four labs, but positive in one lab. However, the positive induction was paralleled by an increase of the MTT value to 170% (see Supporting information 3).

The data from both study phases were then further analyzed quantitatively to evaluate different aspects of the reproducibility,

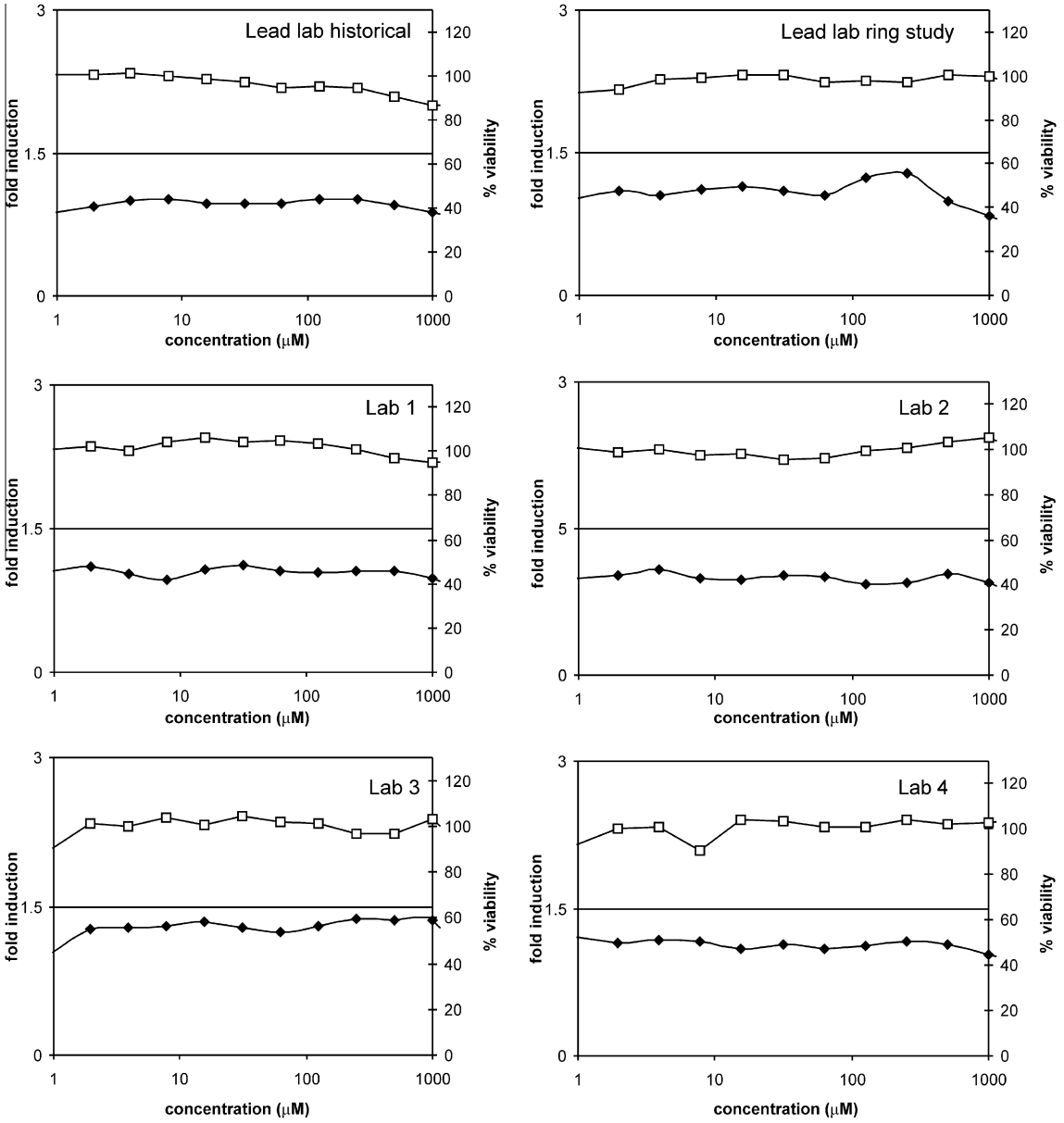

Fig. 2. Induction of luciferase activity (closed diamonds) and cellular viability (open squares) for sulfanilamide in full dose–response analysis according to the SOP. Historical results of the lead lab and new data from the five labs in the transferability phase.

and the data calculated below refer to results from both study phases.

3.3. Assessment of within laboratory reproducibility (WLR) of EC1.5 values

The lead lab had previously screened the set of test substances (Emter et al., 2010) and repeated this analysis within the ring-study. These data were first analyzed to determine the WLR of the EC1.5 values in the lead lab. For each chemical 5–7 repetitions were run in these two studies and these individual results are summarized in Supporting information 4. Based on these data, the geometric mean and the geometric standard deviation of the EC1.5 values were calculated as described in the methods section. The results are presented in Supporting information 5. The geometric standard deviation on the average for all positive chemicals of Phase I and Phase II is at 1.61, with one significant outlier (MCI). Without this outlier it is at 1.38. This is slightly below the square root of 2, indicating that, on the average, the 95.4% confidence interval for within laboratory variation of the EC1.5 value lies within one well up and down of the geometric mean in the dilution series (for details see the methods section).

A separate analysis of the WLR was then made in each of the five laboratories based on the three repetitions of the ring study. For each test substance and each laboratory the geometric standard deviation of the EC1.5 values was calculated (Table 3). The average of these geometric standard deviations for WLR was then calculated for each test substance (column WLR in Table 3). This value ranged between 1.1 and 1.8 for all of the test substances, with an average for all positive substances of 1.44. This is again close to the square root of 2. The average for all test substances within a lab is in the range between 1.27 and 1.69 for the different labs and generally close to 1.38 found above for the WLR in the

Table 1
Luciferase induction and cytotoxicity in the KeratinoSens assay for the seven chemicals of the transferability phase (Phase I).

		I_{max} (fold induction)[a]	EC 1.5 (μM)[b]	Positive repetitions[c]	IC 50 (μM)[d]
2,4-Dinitro-chlorobenzene	Lead_lab_historical	14.8	2.5	2 of 2	8.2
	Lead_lab	12.9	3.3	3 of 3	10.1
	Lab1	4.3	2.1	3 of 3	6.6
	Lab2	12.2	3.0	3 of 3	9.6
	Lab3	19.5	1.4	3 of 3	8.5
	Lab4	15.6	2.1	3 of 3	12.5
Citral	Lead_lab_historical	96.4	23.1	2 of 2	182.8
	Lead_lab	60.2	17.2	3 of 3	103.5
	Lab1	22.3	12.6	3 of 3	159.3
	Lab2	40.0	20.4	3 of 3	171.8
	Lab3	104.4	16.1	3 of 3	166.9
	Lab4	50.2	16.3	3 of 3	238.8
Ethylene glycol dimethacrylate	Lead_lab_historical	188.4	56.5	2 of 2	1655.8
	Lead_lab	176.9	81.5	3 of 3	909.1
	Lab1	25.8	112.2	3 of 3	879.9
	Lab2	102.1	54.6	3 of 3	871.5
	Lab3	363.0	10.6	3 of 3	963.2
	Lab4	204.9	42.0	3 of 3	1770.2
Hexyl cinnamic aldehyde	Lead_lab_historical	2.7	17.2	2 of 2	26.3
	Lead_lab	1.8	n.i.	1 of 3	30.9
	Lab1	1.4	n.i.	0 of 3	39.3
	Lab2	1.5	n.i.	1 of 3	31.1
	Lab3	4.2	23.5	3 of 3	62.7
	Lab4	5.3	17.2	3 of 3	90.9
Methyl salicylate	Lead_lab_historical	1.2	n.i.	0 of 2	>2000
	Lead_lab	1.4	n.i.	0 of 3	>2000
	Lab1	2.7	n.i.	1 of 3	>2000
	Lab2	1.5	n.i.	0 of 3	>2000
	Lab3	1.5	n.i.	1 of 3	>2000
	Lab4	1.5	n.i.	1 of 3	>2000
Chlorobenzene	Lead_lab_historical	1.2	n.i.	0 of 2	>2000
	Lead_lab	1.3	n.i.	0 of 3	>2000
	Lab1	1.5	n.i.	0 of 3	>2000
	Lab2	1.3	n.i.	0 of 3	>2000
	Lab3	1.3	n.i.	0 of 3	>2000
	Lab4	1.8	n.i.	1 of 3	>2000
Sulfanilamide	Lead_lab_historical	1.4	n.i.	0 of 2	>2000
	Lead_lab	1.1	n.i.	0 of 3	>2000
	Lab1	1.2	n.i.	0 of 3	>2000
	Lab2	1.1	n.i.	0 of 3	>2000
	Lab3	1.5	n.i.	1 of 3	>2000
	Lab4	1.3	n.i.	0 of 3	>2000

[a] Maximal fold-induction of luciferase activity in any of the 12 test concentrations, given are the averages of the three repetitions within each lab.
[b] Concentration in μM to reach 1.5-fold induction of gene activity, n.i. indicates no statistically significant induction above the threshold. Given are the geometric means of the three repetitions within each lab.
[c] Number of repetitions rating a chemical positive according to the prediction model (Significant induction > 1.5-fold below 1000 μM and at non-cytotoxic concentrations).
[d] Concentration in μM which reduces cellular viability by 50%. Given are the geometric means of the three repetitions within each lab.

Lead Lab. Therefore, on average a similar WLR for EC 1.5 values was found in all laboratories.

3.4. Assessment of between laboratory reproducibility (BLR) of EC 1.5 values

For each chemical the logarithmic average of the EC1.5 values per lab was calculated, and the standard deviation between the laboratories over these logarithmic averages was determined for each test substance. This value was then re-transformed to yield the geometric standard deviation of the EC1.5 between the laboratories (column BLR in Table 3). This parameter varied between 1.2 and 2.6 with an average of 1.64 (1.56 without the outlier MCI), indicating that the between laboratory variability is slightly higher as compared to the within laboratory variability. Still, for the majority of the substances it is below 1.41. Therefore, also for the BLR the 95.4% confidence interval of the EC1.5 value is, for most substances, within one well up and down from the geometric mean.

3.5. Assessment of WLR and BLR of IC50 values

The geometric means of the IC50 values for cytotoxicity of all test substances are listed in Tables 1 and 2, and in general congruent results within and between the labs were obtained. To quantify this, the within and between lab geometric standard deviations of the IC50 values were calculated (as done for the EC1.5 values in Table 3). These values could be calculated for those 18 test substances with a reduction of viability >50% at any of the test concentrations and they are listed in Table 4. The average of these geometric standard deviations within individual labs ranged between 1.09 and 1.81 for all of the test substances, with an average for all cytotoxic substances of 1.35, indicating that also for the WLR of the cytotoxicity values the 95.4% confidence interval is confined by a factor of less than 2. The BLR geometric standard deviations were between 1.08 and 1.94 for all test substances, with an average of 1.42. Thus, the variability of the IC50 values between the laboratories is even lower as compared to the variability of the EC1.5 values.

Table 2
Luciferase induction and cytotoxicity in the KeratinoSens assay for the 21 blind-coded chemicals of the reproducibility phase (Phase II).

		I_{max} (fold induction)[a]	EC 1.5 (μM)[b]	Positive repetitions[c]	IC 50 (μM)[d]
Non-sensitizers					
Lactic acid	Lead Lab	1.7	n.i.	0 of 3	>2000
	Lab1	1.3	n.i.	0 of 3	>2000
	Lab2	1.3	n.i.	0 of 3	>2000
	Lab3	1.2	n.i.	0 of 3	>2000
	Lab4	1.1	n.i.	0 of 3	>2000
Glycerol	Lead Lab	1.4	n.i.	0 of 3	>2000
	Lab1	1.2	n.i.	0 of 3	>2000
	Lab2	1.3	n.i.	0 of 3	>2000
	Lab3	1.2	n.i.	0 of 3	>2000
	Lab4	1.1	n.i.	0 of 3	>2000
Diethyl phthalate	Lead Lab	1.2	n.i.	0 of 3	>2000
	Lab1	1.2	n.i.	0 of 3	>2000
	Lab2	1.4	n.i.	1 of 3	>2000
	Lab3	1.7	n.i.	0 of 3	>2000
	Lab4	1.9	582.7	2 of 3	>2000
Isopropanol	Lead Lab	1.3	n.i.	0 of 3	>2000
	Lab1	1.3	n.i.	0 of 3	>2000
	Lab2	1.4	n.i.	0 of 3	>2000
	Lab3	1.5	n.i.	0 of 3	>2000
	Lab4	1.1	n.i.	0 of 3	>2000
Salicylic acid	Lead Lab	1.1	n.i.	0 of 3	>2000
	Lab1	1.4	n.i.	0 of 3	>2000
	Lab2	1.3	n.i.	0 of 3	>2000
	Lab3	1.3	n.i.	0 of 3	>2000
	Lab4	1.3	n.i.	0 of 3	>2000
Sodium lauryl sulphate	Lead Lab	5.9	35.3	3 at cytotox	57.0
	Lab1	1.4	n.i.	1 at cytotox	74.5
	Lab2	5.1	n.i.	1 at cytotox	54.5
	Lab3	2.7	n.i.	1 at cytotox	46.2
	Lab4	8.1	33.9	3 of 3	91.1
Sensitizers					
4-Phenylenediamine	Lead Lab	45.2	9.9	3 of 3	400.7
	Lab1	23.2	10.1	3 of 3	396.4
	Lab2	19.4	8.4	3 of 3	577.2
	Lab3	39.0	6.0	3 of 3	648.9
	Lab4	20.0	10.1	3 of 3	427.0
Cinnamic aldehyde	Lead Lab	22.4	14.3	3 of 3	164.1
	Lab1	9.8	14.8	3 of 3	172.6
	Lab2	32.4	6.0	3 of 3	132.6
	Lab3	44.7	7.5	3 of 3	348.9
	Lab4	22.2	6.5	3 of 3	113.1
4-Nitrobenzylbromide	Lead Lab	7.0	1.4	3 of 3	7.2
	Lab1	4.7	<0.98	3 of 3	5.7
	Lab2	13.6	1.4	3 of 3	13.7
	Lab3	14.0	<0.98	3 of 3	10.8
	Lab4	9.7	<0.98	3 of 3	8.3
Metol	Lead Lab	5.9	8.7	3 of 3	52.7
	Lab1	4.5	3.2	3 of 3	31.8
	Lab2	9.4	5.7	3 of 3	53.6
	Lab3	12.2	6.6	3 of 3	137.0
	Lab4	6.4	3.5	3 of 3	24.0
Isoeugenol	Lead Lab	13.2	18.4	3 of 3	550.9
	Lab1	9.5	23.3	3 of 3	560.6
	Lab2	23.2	10.4	3 of 3	997.5
	Lab3	56.8	4.5	3 of 3	791.6
	Lab4	14.4	22.9	3 of 3	515.7
Eugenol	Lead Lab	1.4	n.i.	1 of 3	1363.2
	Lab1	1.4	n.i.	1 of 3	1281.4
	Lab2	2.5	309.1	3 of 3	1496.6
	Lab3	2.5	231.2	3 of 3	1487.1
	Lab4	1.7	n.i.	1 of 3	1538.0
Oxazolone	Lead Lab	24.3	191.2	3 of 3	1667.8
	Lab1	6.5	185.2	3 of 3	1761.0
	Lab2	19.3	240.8	3 of 3	>2000
	Lab3	46.3	152.9	3 of 3	>2000
	Lab4	6.4	84.3	3 of 3	777.9

(*continued on next page*)

Table 2 (*continued*)

		I_{max} (fold induction)[a]	EC 1.5 (μM)[b]	Positive repetitions[c]	IC 50 (μM)[d]
2-Mercapto-benzothiazole	Lead Lab	24.5	44.3	3 of 3	983.4
	Lab1	4.9	108.0	3 of 3	1099.6
	Lab2	54.2	54.4	3 of 3	1576.6
	Lab3	64.1	65.5	3 of 3	1656.1
	Lab4	21.5	226.9	3 of 3	1429.1
(5-Chloro)-methyl-isothiazolinone	Lead Lab	7.6	7.6	3 of 3	32.6
	Lab1	4.9	1.0	3 of 3	8.9
	Lab2	4.8	4.2	3 of 3	25.6
	Lab3	13.5	0.8	3 of 3	14.5
	Lab4	7.8	2.8	3 of 3	12.8
Imidazolidinyl urea	Lead Lab	1.8	50.0	2 of 3	92.7
	Lab1	3.6	35.9	3 of 3	80.5
	Lab2	5.1	29.9	3 of 3	125.2
	Lab3	11.5	31.3	3 of 3	103.0
	Lab4	9.2	32.2	3 of 3	94.7
Methyldibromo glutaronitrile	Lead Lab	5.2	12.4	3 of 3	34.4
	Lab1	2.1	7.2	3 of 3	24.7
	Lab2	6.1	10.9	3 of 3	53.5
	Lab3	7.8	7.4	3 of 3	43.5
	Lab4	3.3	6.7	3 of 3	31.8
Glyoxal	Lead Lab	67.1	134.1	3 of 3	610.5
	Lab1	14.5	77.0	3 of 3	523.2
	Lab2	58.3	120.0	3 of 3	721.9
	Lab3	195.0	95.6	3 of 3	>800
	Lab4	69.7	169.1	3 of 3	700.4
Cinnamyl alcohol	Lead Lab	10.8	104.1	3 of 3	1705.3
	Lab1	1.7	106.6	2 of 3	1726.3
	Lab2	17.1	86.3	3 of 3	1827.8
	Lab3	9.1	130.9	3 of 3	>2000
	Lab4	8.6	118.2	3 of 3	>2000
Tetramethyl-thiuramdisulfide	Lead Lab	22.3	<0.98	3 of 3	23.8
	Lab1	8.1	<0.98	3 of 3	26.8
	Lab2	17.7	4.1	3 of 3	53.7
	Lab3	67.9	1.9	3 of 3	36.6
	Lab4	9.3	<0.98	3 of 3	11.8
Phenyl benzoate	Lead Lab	1.0	n.i.	0 of 3	185.2
	Lab1	1.2	n.i.	0 of 3	263.3
	Lab2	1.2	n.i.	0 of 3	309.6
	Lab3	1.2	n.i.	0 of 3	239.9
	Lab4	2.4	n.i.	1 of 3	808.6

[a] Maximal fold-induction of luciferase activity in any of the 12 test concentrations, given are the averages of the three repetitions within each lab.

[b] Concentration in μM to reach 1.5-fold induction of gene activity, n.i. indicates no statistically significant induction above the threshold. Given are the geometric means of the three repetitions within each lab.

[c] Number of repetitions rating a chemical positive according to the prediction model (Significant induction > 1.5-fold below 1000 μM and at non-cytotoxic concentrations).

[d] Concentration in μM which reduces cellular viability by 50%. Given are the geometric means of the three repetitions within each lab.

3.6. Assessment of BLR of the predictive capacity (PC)

In Fig. 3, the positive (red and orange) and negative (dark and faint green) ratings for all the 28 substances are summarized, and the Cooper statistics calculated. The accuracy was between 85.7 and 96.4% in the different laboratories. The main reason for the difference in accuracy between the laboratories is the different rating of the borderline chemical hexyl cinnamic aldehyde, and the different results obtained for eugenol. The outliers for diethyl phthalate and SLS in lab 4 affected the Cooper statistics somewhat for this lab.

3.7. Quality control values

Finally, the controls included in all assay plates (DMSO controls and positive control cinnamic aldehyde) were compared across laboratories (see Supporting information 6). Once the test was set up and running in the different laboratories, the performance criteria for a variability of <20% in the DMSO-control wells was fulfilled in 54 of the 60 runs, and it was between 20% and 26% in the remaining 6 runs. The average variability for all runs was 13.3%. The performance criteria that cinnamic aldehyde was significantly positive (EC1.5 < 64 μM) was fulfilled in all of the 60 runs. The quantitative performance criteria for the induction by cinnamic aldehyde (a) Induction at 64 μM between 2 and 8-fold was fulfilled in 51 of the 60 runs, and (b) EC1.5 between 7 μM and 30 μM was fulfilled in 51 of the 60 runs. Generally these results indicate that cinnamic aldehyde is a very robust positive control to verify that the test is working and it is reliably positive in all the runs in all the labs. However, the quantitative variability for the EC1.5 value for cinnamic aldehyde when tested as positive reference was clearly higher as compared to some other substances studied here. Interestingly, this variability was lower when cinnamic aldehyde was tested as a blind coded chemical (see Table 2). The reason for this is currently unknown.

4. Discussion

The results of this study indicate that the KeratinoSens assay, based on a reporter gene read-out, is easily transferable between

Table 3
Geometric standard deviations of the EC1.5 values for the consistently positive[a] chemicals.

	Geometric standard deviations						
	Lead lab[b]	Lab 1	Lab 2	Lab 3	Lab 4	WLR[c]	BLR[d]
Metol	1.62	1.71	2.23	1.17	2.24	1.74	1.53
(5-Chloro)-methylisothiazolinone	1.98	2.33	1.62	1.75	1.44	1.80	2.56
Imidazolidinyl urea	1.00	1.08	1.52	1.07	1.03	1.13	1.23
Oxazolone	1.03	1.91	1.76	1.20	1.58	1.46	1.49
4-Phenylenediamine	1.13	1.29	1.77	1.07	2.88	1.52	1.25
Cinnamic aldehyde	1.10	1.06	1.32	1.32	2.01	1.33	1.55
Isoeugenol	1.57	1.82	1.68	2.07	1.50	1.71	2.02
2-Mercaptobenzothiazole	1.25	1.62	3.09	1.05	1.38	1.55	1.92
Cinnamyl alcohol	1.17	2.28	1.26	1.27	1.09	1.36	1.17
Glyoxal	1.04	2.70	1.05	1.17	1.34	1.36	1.35
Methyldibromo glutaronitrile	1.36	1.51	1.41	1.13	1.16	1.30	1.32
Citral	1.12	1.73	1.61	1.90	1.09	1.45	1.19
Ethylene glycol dimethacrylate	1.19	1.22	1.10	1.42	1.16	1.21	2.49
2,4-Dinitrochlorobenzene	1.29	1.34	1.27	1.43	1.02	1.26	1.40
Average	1.27	1.69	1.62	1.36	1.49	1.44	1.60

[a] Hexyl cinnamic aldehyde and eugenol, showed induction in some labs only, thus not evaluated; tetramethylthiuramdisulfide and 4-nitrobenzyl-bromide, data contain several EC1.5 values <0.98, thus could not be used for this statistical analysis. However also for these four test substances the observed EC1.5 values were reproducible as shown in Tables 1 and 2.
[b] For each lab the geometric standard deviation of the three repetitions was calculated.
[c] Indicates the average geometric standard deviation within the different labs.
[d] Indicates the geometric standard deviation of the logarithmic averages of each lab.

Table 4
Geometric standard deviations of the IC50 values for the chemicals which were cytotoxic at the tested doses in the majority of the labs.

	Geometric standard deviations						
	Lead lab[a]	Lab 1	Lab 2	Lab 3	Lab 4	WLR[b]	BLR[c]
2,4-Dinitrochlorobenzene	1.10	1.73	1.24	1.30	1.02	1.28	1.27
4-Phenylenediamine	1.05	1.01	1.40	1.40	2.14	1.40	1.26
Cinnamic aldehyde	1.20	1.56	1.18	1.49	1.06	1.30	1.54
4-nitrobenzylbromide	1.41	2.83	1.73	1.39	1.70	1.81	1.41
4-Methylaminophenol sulphate (METOL)	1.68	1.96	1.02	1.27	2.34	1.65	1.94
Isoeugenol	1.17	2.03	1.44	1.13	1.37	1.43	1.33
Eugenol	1.04	1.33	1.01	1.03	1.03	1.09	1.08
2-Mercaptobenzothiazole	1.08	1.36	1.06	1.10	1.27	1.17	1.26
(5-chloro)-Methylisothiazolinone	2.38	2.07	1.01	1.51	1.57	1.71	1.70
Sodium lauryl sulphate	1.10	1.30	1.46	1.04	1.05	1.19	1.31
Imidazolidinyl urea	1.02	1.28	1.48	1.05	1.01	1.17	1.18
Methyldibromo glutaronitrile	1.35	1.28	1.08	1.06	1.28	1.21	1.34
Glyoxal	1.03	1.38	1.08	n.c. [d]	1.01	1.12	1.16
tetramethylthiuramdisulfide	1.07	1.66	1.02	1.32	1.19	1.25	1.75
Phenyl benzoate	1.02	1.56	2.25	1.45	2.06	1.67	1.76
Citral	1.04	1.27	1.43	1.07	1.05	1.17	1.35
Ethylene glycol dimethacrylate	1.06	1.68	1.04	1.37	n.c.	1.29	1.35
Hexyl cinnamic aldehyde	1.19	1.60	1.45	1.44	1.08	1.35	1.60
Averages	1.22	1.60	1.30	1.26	1.36	1.35	1.42

[a] For each lab the geometric standard deviation of the three repetitions was calculated.
[b] Indicates the average geometric standard deviation within the different labs.
[c] Indicates the geometric standard deviation of the logarithmic averages of each lab.
[d] n.c. not cytotoxic at maximal test concentration.

labs. In general, the predictive capacity is similar between labs, and more importantly, also the quantitative dose–response data were reproduced in the different laboratories. The fact that the between laboratory variability for EC1.5 values was only slightly above the within laboratory variability indicates that transfer of the assay to other labs does not affect the results significantly.

4.1. Hurdles for transferability

There were some minor hurdles in the transferability phase, but these were not specific to the assay but rather specific to the use of any luciferase based assay. The key issues identified in the transfer phase were due to three reasons (a) The use of differing luminometers (b) the use of a glow-substrate instead of a flash-substrate for the luciferase measurements. Glow-substrates emit low levels of light for prolonged periods, which contributes to lower sensitivity and thus higher variability. Longer reading times for plates with the glow-substrate then caused a gradient over the assay plate in one lab. This lab then switched to a flash-substrate. Nevertheless, one laboratory did successfully use the glow-approach. (c) The use of assay plates which do not properly fit the height of the luminometer, or allow light to scatter into adjacent wells. Especially when using the glow-approach, the initial plates used allowed light scattering into the adjacent rows, and thus initial false-positive results in two labs. To ensure accurate determinations of EC1.5 values (i) a high sensitivity, (ii) a low variability (and thus a stable background) and (iii) no gradient over the plate were found to be absolutely critical. A training experiment was designed based on these experiences and this may be used in future transfers to ensure the above three parameters are met. Once these hurdles (in

	Study phase	Positive with EC 1.5 up to 1000 µM					
		Lead Lab hist.	Lead lab	Lab 1	Lab 2	Lab 3	Lab 4
Sensitizers							
Hexyl cinnamic aldehyde	MT	2 of 2	1 of 3	0 of 3	1 of 3	3 of 3	3 of 3
Citral	MT	2 of 2	3 of 3	3 of 3	3 of 3	3 of 3	3 of 3
Ethylene glycol dimethacrylate	MT	2 of 2	3 of 3	3 of 3	3 of 3	3 of 3	3 of 3
2,4-Dinitrochlorobenzene	MT	2 of 2	3 of 3	3 of 3	3 of 3	3 of 3	3 of 3
4-Methylaminophenol sulphate	BC	2 of 2	3 of 3	3 of 3	3 of 3	3 of 3	3 of 3
(5-Chloro)-methylisothiazolinone	BC	2 of 2	3 of 3	3 of 3	3 of 3	3 of 3	3 of 3
Phenyl benzoate	BC	1 of 4	0 of 3	0 of 3	0 of 3	0 of 3	1 of 3
Imidazolidinyl urea	BC	3 of 4	2 of 3	3 of 3	3 of 3	3 of 3	3 of 3
Oxazolone	BC	4 of 4	3 of 3	3 of 3	3 of 3	3 of 3	3 of 3
4-Phenylenediamine	BC	2 of 2	3 of 3	3 of 3	3 of 3	3 of 3	3 of 3
Cinnamic aldehyde	BC	4 of 4	3 of 3	3 of 3	3 of 3	3 of 3	3 of 3
Isoeugenol	BC	4 of 4	3 of 3	3 of 3	3 of 3	3 of 3	3 of 3
Tetramethylthiuramdisulfide	BC	2 of 2	3 of 3	3 of 3	3 of 3	3 of 3	3 of 3
2-Mercaptobenzothiazole	BC	4 of 4	3 of 3	3 of 3	3 of 3	3 of 3	3 of 3
Eugenol	BC	0 of 4	1 of 3	1 of 3	3 of 3	3 of 3	1 of 3
Cinnamyl alcohol	BC	4 of 4	3 of 3	2 of 3	3 of 3	3 of 3	3 of 3
Glyoxal	BC	4 of 4	3 of 3	3 of 3	3 of 3	3 of 3	3 of 3
4-Nitrobenzylbromide	BC	2 of 2	3 of 3	3 of 3	3 of 3	3 of 3	3 of 3
Methyldibromo glutaronitrile	BC	2 of 2	3 of 3	3 of 3	3 of 3	3 of 3	3 of 3
Non-sensitizers							
Isopropanol	BC	0 of 2	0 of 3	0 of 3	0 of 3	0 of 3	0 of 3
Methyl salicylate	MT	0 of 2	0 of 3	1 of 3	0 of 3	1 of 3	1 of 3
Chlorobenzene	MT	0 of 2	0 of 3	0 of 3	0 of 3	0 of 3	1 of 3
Sulfanilamide	MT	0 of 2	0 of 3	0 of 3	0 of 3	1 of 3	0 of 3
Salicylic acid	BC	0 of 2	0 of 3	0 of 3	0 of 3	0 of 3	0 of 3
Sodium lauryl sulfate	BC	0 of 2	3 at cyto-tox. conc.	1 at cyto-tox. conc.	1 at cyto-tox. conc.	1 at cyto-tox. conc.	3 of 3
Lactic acid	BC	1 of 4	0 of 3	0 of 3	0 of 3	0 of 3	0 of 3
Glycerol	BC	0 of 4	0 of 3	0 of 3	0 of 3	0 of 3	0 of 3
Diethyl phthalate	BC	0 of 2	0 of 3	0 of 3	1 of 3	0 of 3	2 of 3
Cooper statistics							
correct positive		17	16	16	17	18	17
correct negative		9	9	9	9	9	7
false positive		0	0	0	0	0	2
false negative		2	3	3	2	1	2
n		28	28	28	28	28	28
Sensitivity (%)		89.5	84.2	84.2	89.5	94.7	89.5
Specificity (%)		100.0	100.0	100.0	100.0	100.0	77.8
Accuracy (%)		92.9	89.3	89.3	92.9	96.4	85.7

Fig. 3. The predictive capacity in the different labs. If 2 of 3 reps are positive and overall dose response is given in all reps, compound is considered positive (red and orange). If only one rep is positive and dose response is not evident compound is considered negative (dark and faint green). The induction at cytotoxic concentrations for SDS was not considered positive. MT: method transfer/phase I data (7 test substances), BC: blind study/phase II data (21 test substances). Detailed results are presented in the dose–response curves in Supporting information 2 and 3 and in Tables 1 and 2. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

the physical, not the biological part of the experiment) were overcome, the transferability was found to be very good, and the biological system did not pose any challenges in the lab-to-lab transfer.

4.2. Outliers

Not too surprisingly in an extended study with 5 × 28 individual test substance assessments, a few outliers were recorded. The borderline chemical hexyl cinnamic aldehyde gave mixed results among labs, but was a known borderline test item in the historical data and it varied between historical and new data in the lead lab also. Thus in this case, the variability between the labs was reproducing a variability already observed within the lead lab. Hexyl cinnamic aldehyde becomes positive very close to the cytotoxic level, and the IC50 value for this test item was higher in the labs reporting a consistently positive result. The result for eugenol, a reported weak sensitizer, was more surprising. This chemical was positive in the AREc32 assay (Natsch and Emter, 2008) and positive in two testing labs, but negative in three labs including the lead lab. We had observed that HaCaT keratinocytes and EpiDerm™ skin models convert eugenol into a metabolite of high molecular weight (our unpublished observation), and efficient detoxification could be the reason for a lack of gene induction and the low cytotoxicity for eugenol. However, this does not explain the differing results in the participating labs. For SLS, our observation that the luciferase expression is induced at a very narrow range of cytotoxic concentrations was reproduced in several labs. This phenomenon had been previously illustrated by a detailed dose–response analysis (Emter et al., 2010). Still, there was an outlier in one lab with a positive induction at a concentration which was rated as non-cytotoxic in the parallel MTT-plate. Comparing the data of the different labs, the outlier appears rather at the level of the IC50 value, and not at the level of the EC1.5 value. This discrepancy could be due to the fact that cytotoxicity is not measured directly within the assay plates, and a possible improvement would be to include a cytotoxicity measure directly in the assay plates prior to luciferase readings as demonstrated by Uibel et al. (2010). Finally, one lab found a positive induction for diethyl phthalate which was paralleled by a strong increase in the MTT values. This might indicate that cell numbers increased, and that the luciferase production per cell remained constant. The reason for this effect is unknown.

4.3. Prediction model

Two changes were made to the prediction model, which had already been anticipated in our previous publication (Emter et al., 2010): (i) Test substances are rated positive if EC1.5 is below 1000 μM (previously 2000 μM), since occasional false-positive results were recorded at 2000 μM and (ii) only if the viability at the EC1.5-determining value is >70%, the result is considered positive. The reasons for these modifications had been discussed (Emter et al., 2010).

4.4. Cytotoxicity of the test molecules

The data in Tables 1 and 2 indicate that with the exception of SLS all tested non-sensitizers have very low cytotoxicity. One may suspect, based on this dataset, that more cytotoxic substances do activate the Nrf2 pathway, and that the difference between sensitizers and non-sensitizers is largely based on their cytotoxic properties. However, this is only due to this particular test set of non-sensitizers which was selected based on published lists and which consists of molecules of low cytotoxicity. We had screened large numbers of novel molecules in our discovery process, and many non-reactive but cytotoxic substances were identified, having IC50 values far below 2000 μM with no significant ARE-dependent luciferase activation. Thus, luciferase induction does not correlate to cytotoxicity as might be suspected based on the present data (our unpublished data).

4.5. Quality controls

In Phase II, all the experiments were accepted, even if slightly outside of the targeted performance criteria. With this approach, analysis of the data now indicates how sensitive the assay is relative to these performance criteria. This will allow an eventual redefining of the performance criteria for full validation studies, whereas if the performance criteria would have been strictly applied in the ring-study, it would not have been possible to determine how narrow the criteria need to be in order to obtain useful and reproducible results. The results indicated that the dose–response curves in a run with control variability of 8.4% are clearly smoother as compared to the data in a parallel run with 26.8% control variability (i.e. the maximal variability observed), which indicates that one should in the future strictly adhere to the 20% variability criterion. On the other hand, the data from runs with an EC1.5 value for the positive control cinnamic aldehyde outside of the target range were still of very good quality and this performance criterion may by defined less stringent based on the current data.

5. Conclusions

Reporter gene assays have widely been used to screen for hormone-active substances, and the corresponding tests had been assessed for transferability (van der Burg et al., 2010). This study is the first inter-laboratory study on gene expression changes induced by skin sensitizers in a stable cell line. The assay was reproducible between the laboratories for 26 out of the 28 chemicals and an overall accuracy for these 28 chemicals between 85.4 and 96.7% was reported from the five labs. More importantly, also the dose–response curves and the quantitative parameters (concentration for significant gene induction and IC50 values for cytotoxicity) were reproducible, and the between laboratory variability was only slightly higher as compared to the within laboratory variability. This high reproducibility of the results from the KeratinoSens assay in this extensive study with many substances and five participating labs encourages us to progress to official (pre)validation.

Supporting information

Supporting information 1 gives details on the test chemicals.

Supporting information 2 gives all the dose response curves in all the labs for Phase I.

Supporting information 3 gives all the dose response curves in all the labs for Phase II.

Supporting information 4 gives the detailed results on the WLR in the lead lab.

Supporting information 5 gives the statistical analysis of the WLR in the lead lab.

Supporting information 6 gives the results for the control values in all the labs and all the repetitions.

Funding

No external funding was received for this study. The organization of the study and evaluation of the data was funded by Givaudan Schweiz AG. Each participating lab covered its own costs. The Institute for *In Vitro* Sciences would like to thank the Colgate-Palmolive Company and The Clorox Company for their support of the ring study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tiv.2010.12.014.

References

Ade, N., Leon, F., Pallardy, M., Peiffer, J.L., Kerdine-Romer, S., Tissier, M.H., Bonnet, P.A., Fabre, I., Ourlin, J.C., 2009. HMOX1 and NQO1 genes are upregulated in response to contact sensitizers in dendritic cells and THP-1 cell line: role of the Keap1/Nrf2 pathway. Toxicol. Sci. 107, 451–460.

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

Ashikaga, T., Yoshida, Y., Hirota, M., Yoneyama, K., Itagaki, H., Sakaguchi, H., Miyazawa, M., Ito, Y., Suzuki, H., Toyoda, H., 2006. Development of an in vitro skin sensitization test using human cell lines: the human cell line activation test (h-CLAT) I. Optimization of the h-CLAT protocol. Toxicol. In Vitro 20, 767–773.

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

Basketter, D.A., Kimber, I., 2009. Updating the skin sensitization in vitro data assessment paradigm in 2009. J. Appl. Toxicol. 29, 545–550.

Basketter, D.A., Lea, L.J., Cooper, K., Stocks, J., Dickens, A., Pate, I., Dearman, R.J., 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.

Boukamp, P., Petrussevska, R.T., Breitkreutz, D., Hornung, J., Markham, A., Fusenig, N.E., 1988. Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. J. Cell Biol. 106, 761–771.

Casati, S., Aeby, P., Kimber, I., Maxwell, G., Ovigne, J.M., Roggen, E., Rovida, C., Tosti, L., Basketter, D., 2009. Selection of chemicals for the development and evaluation of in vitro methods for skin sensitisation testing. Altern. Lab. Anim. 37, 305–312.

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 sensitizers from irritants and low molecular weight respiratory allergens. Toxicol. In Vitro 23, 789–796.

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

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

Gildea, L.A., Ryan, C.A., Foertsch, L.M., Kennedy, J.M., Dearman, R.J., Kimber, I., 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.

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

ICCVAM, 2008a. ICCVAM database, downloaded in December 2008 from http://iccvam.niehs.nih.gov/methods/immunotox/LLNA-pot/appx/LLNApotencyAppB18Jan08FD.xls.

ICCVAM, 2008b. Revised Draft ICCVAM LLNA Performance Standards. Downloaded in December 2008 from http://iccvam.niehs.nih.gov/methods/immunotox/PerfStds/LLNAPerfStd07Jan08FD.pdf.

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

Karlberg, A.T., Bergstrom, M.A., Borje, A., Luthman, K., Nilsson, J.L., 2008. Allergic contact dermatitis–formation, structural requirements, and reactivity of skin sensitizers. Chem. Res. Toxicol. 21, 53–69.

Kimber, I., Basketter, D.A., Gerberick, G.F., Dearman, R.J., 2002. Allergic contact dermatitis. Int. Immunopharmacol. 2, 201–211.

Lambrechts, N., Verstraelen, S., Lodewyckx, H., Felicio, A., Hooyberghs, J., Witters, H., Van Tendeloo, V., Van Cauwenberge, P., Nelissen, I., Van Den Heuvel, R., Schoeters, G., 2009. THP-1 monocytes but not macrophages as a potential alternative for CD34+ dendritic cells to identify chemical skin sensitizers. Toxicol. Appl. Pharmacol. 236, 221–230.

Natsch, A., 2010. The Nrf2-Keap1-ARE toxicity pathway as a cellular sensor for skin sensitizers–functional relevance and a hypothesis on innate reactions to skin sensitizers. Toxicol Sci 113, 284–292.

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

Natsch, A., Emter, R., Ellis, G., 2009. 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. Toxicol. Sci. 107, 106–121.

Python, F., Goebel, C., Aeby, P., 2009. Comparative DNA microarray analysis of human monocyte derived dendritic cells and MUTZ-3 cells exposed to the moderate skin sensitizer cinnamaldehyde. Toxicol. Appl. Pharmacol.

Roberts, D.W., Patlewicz, G., Dimitrov, S.D., Low, L.K., Aptula, A.O., Kern, P.S., Dimitrova, G.D., Comber, M.I.H., Phillips, R.D., Niemelä, J., Madsen, C., Wedebye, E.B., Bailey, P.T., Mekenyan, O.G., 2007. TIMES-SS – A mechanistic evaluation of an external validation study using reaction chemistry principles. Chem. Res. Toxicol. 20, 1321–1330.

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

Sens-it-iv, 2009. Sens-it-iv Newsletter 32, October, 12, 2009. Downloaded June 2010 from http://www.sens-it-iv.eu/files/newsletter/Sens-it-iv_Newsletter_32.html.

Toebak, M.J., Pohlmann, P.R., Sampat-Sardjoepersad, S.C., von Blomberg, B.M., Bruynzeel, D.P., Scheper, R.J., Rustemeyer, T., Gibbs, S., 2006. CXCL8 secretion by dendritic cells predicts contact allergens from irritants. Toxicol. In Vitro 20, 117–124.

Uibel, F., Muhleisen, A., Kohle, C., Weimer, M., Stummann, T.C., Bremer, S., Schwarz, M., 2010. ReProGlo: a new stem cell-based reporter assay aimed to predict embryotoxic potential of drugs and chemicals. Reprod. Toxicol. 30, 103–112.

van der Burg, B., Winter, R., Man, H.Y., Vangenechten, C., Berckmans, P., Weimer, M., Witters, H., van der Linden, S., 2010. Optimization and prevalidation of the in vitro AR CALUX method to test androgenic and antiandrogenic activity of compounds. Reprod. Toxicol. 30, 18–24.

Vandebriel, R.J., Pennings, J.L., Baken, K.A., Pronk, T.E., Boorsma, A., Gottschalk, R., Van Loveren, H., 2010. Keratinocyte gene expression profiles discriminate sensitizing and irritating compounds. Toxicol. Sci. 117, 81–89.