

# International ring trial of the epidermal equivalent sensitizer potency assay: reproducibility and predictive-capacity<sup>1</sup>

Marc A.T. Teunis<sup>1</sup>, Sander W. Spiekstra<sup>2</sup>, Mieke Smits<sup>1</sup>, Els Adriaens<sup>3</sup>, Tobias Eltze<sup>4</sup>, Valentina Galbiati<sup>5</sup>, Cyrille Krul<sup>1,6</sup>, Robert Landsiedel<sup>4</sup>, Raymond Pieters<sup>1,7</sup>, Judith Reinders<sup>2</sup>, Erwin Roggen<sup>8</sup>, Emanuela Corsini<sup>5</sup>, Susan Gibbs<sup>2,9</sup>

<sup>1</sup>University of Applied Sciences, Utrecht, The Netherlands; <sup>2</sup>Dept. Dermatology, VU University Medical Centre, Amsterdam, The Netherlands; <sup>3</sup>Adriaens Consulting, Statistical and Scientific Consultancy, Aalter, Belgium; <sup>4</sup>BASF SE, Ludwigshafen, Germany; <sup>5</sup>Laboratory of Toxicology, DiSFeB, Università degli Studi di Milano, Italy; <sup>6</sup>TNO, Zeist, The Netherlands; <sup>7</sup>Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands; <sup>8</sup>3Rs Management and Consultancy, Denmark; <sup>9</sup>Dept. Oral Cell Biology, Academic Centre for Dentistry Amsterdam (ACTA), The Netherlands

## Summary

*This study describes the international ring trial of the epidermal-equivalent (EE) sensitizer potency assay. This assay does not distinguish a sensitizer from a non-sensitizer, but may classify known skin sensitizers according to their potency. It assesses the chemical concentration resulting in 50% cytotoxicity (EE-EC<sub>50</sub>) or the 2-fold increase in IL-1 $\alpha$  (IL-1 $\alpha$ <sub>2x</sub>). Four laboratories received 13 coded sensitizers. Reproducible results were obtained in each laboratory. A binary prediction model, EC<sub>50</sub>  $\geq$  7 mg/ml = weak to moderate sensitizer and EC<sub>50</sub> < 7 mg/ml = strong to extreme sensitizer had an accuracy of 77%. A superior EE (EC<sub>50</sub> and IL-1 $\alpha$ <sub>2x</sub>) correlation was observed with human in vivo DSA<sub>05</sub> data compared to LLNA-EC<sub>3</sub> data. Human in vivo NOEL and LLNA-EC<sub>3</sub> data correlated to a similar extent to in vitro EE data. Our results indicate that this easily transferable EE potency assay is suitable for testing chemical allergens of unknown potencies and may now be ready for further validation, providing complementary potency information to other assays already undergoing validation for assessing skin sensitization potential.*

*Key words: in vitro, epidermal equivalent, sensitizer potency, NOEL, human DSA<sub>05</sub>*

## 1 Introduction

Repeated exposure to chemical allergens increases the risk of becoming sensitized to that particular chemical. Once an individual has become sensitized, any following exposure to the same chemical may result in allergic contact dermatitis. The contact dermatitis can range from a mild skin rash to extensive skin blistering. Within the North American and Western European populations, the prevalence of skin sensitization to at least one chemical, is approximately 20% (Peiser et al., 2012). The risk to develop allergic contact dermatitis is considered a serious health issue and the identification of potential sensitizing agents within consumer products is therefore crucial.

The preferred animal test is the murine Local Lymph Node Assay (LLNA) as described in the OECD test guideline 429 (OECD TG 429) followed by the Guinea Pig Maximization Test (GPMT) as described in OECD TG 406 (Gerberick et al., 2007a; Rovida et al., 2012). With the enforcement of the 7th Amendment to the EU Cosmetics Directive (76/768/EEC) in March 2013, currently known as: The Cosmetics Regulation (EU 1223/2009), a ban was introduced on the use of animals for identifying repeated dose toxicity endpoints of

---

Received August 2, 2013; accepted in revised form February 10, 2014; Epub February 17, 2014;  
<http://dx.doi.org/10.14573/altex.1308021>

**Abbreviations:** AOO, acetone : olive oil (4:1); BD, broad dose; EE, epidermal equivalent; EC<sub>50</sub>, chemical concentration in mg/ml that results in a decrease in cell viability to 50% compared to vehicle treated epidermal equivalents; DiSFeB, Dipartimento di Scienze Farmacologiche e Biomolecolari, Milan University, Milan; DMSO, 1 % dimethylsulfoxide; DSA<sub>05</sub>, chemical dose per skin area in  $\mu\text{g}/\text{cm}^2$  leading to a sensitization incidence of 5% in the human tested population; FD, fine dose; GPMT; Guinea Pig Maximization Test; HU, University of Applied Sciences, Utrecht; HRIPT, Human Repeat Insult Patch Test; IL-1 $\alpha$ <sub>2x</sub>, Chemical concentration in mg/ml resulting in a 2-fold increase in IL-1 $\alpha$  release into culture supernatant of EE compared to supernatant of vehicle-exposed EE; LLNA, Local Lymph Node Assay; NOEL, human threshold level (no observed effect level) expressed in  $\mu\text{g}/\text{cm}^2$ ; OECD, Organisation for Economic Co-operation and Development; SOP, standard operating procedure; VUMC, VU University Medical Center, Amsterdam

chemicals used in cosmetic ingredients and products, This ban results in an urgent need for the development of suitable non-animal methods for safety testing (Adler et al., 2011). The development of animal alternatives has become even more urgent due to the REACH regulation, which may demand toxicity tests for chemicals produced in quantities of over 1 ton per year, (Grindon, 2007; Grindon et al., 2008; Rovida and Hartung, 2009).

Within the integrated European Framework Program 6 Project *Sens-it-iv*, (LSHB-CT-2005-018681; 2005-2011) a number of potential *in vitro* assays were developed which mimic the key mechanisms of skin sensitization and which therefore may provide alternatives to animal methods (Roggen, 2013). When used in an integrated testing strategy, some of these assays may be able to assess whether or not a chemical is a potential sensitizer (chemical label), and whereas some may also determine the potency of that sensitizer (chemical classification) (Basketter and Kimber, 2009; De Wever et al., 2012). In order to determine whether these assays may actually be suitable to replace the LLNA for risk assessment of potentially sensitizing substances validation according to EURL-ECVAM guidelines of these assays and other assays developed in parallel to *Sens-it-iv* is required. The key mechanisms which the assays are based on are i) chemical penetration to the viable epidermal cell layers to result in cytokine release and cytotoxicity (EE potency assay – the subject of this manuscript) (dos Santos et al., 2011); ii) formation of hapten-protein complexes, the activation of the Keap1/Nrf-2 pathway and triggering keratinocytes to release innate danger signals in the form of cytokines, ATP and reactive oxygen species (e.g. Keratinosens™, IL-18 NCTC assay, Direct Peptide Reactivity Assay (DPRA)) (Natsch et al., 2011; Galbiati et al., 2011; Gerberick et al., 2004; Natsch et al., 2013); iii) dendritic cell maturation and changing biosignatures (e.g. MUTZ-3 GARD assay, hCLAT, MUSST, PBMD) (Maxwell et al., 2011; Lindstedt and Borrebaeck, 2011; Johansson et al., 2013; dos Santos et al., 2009); iv) dendritic cell migration (e.g. MUTZ-DC migration assay) (Gibbs et al., 2013b) and finally T cell priming in the local lymph node (e.g. T cell amplification and differentiation assays) (Martin et al., 2010).

Many assays under development are aimed at distinguishing a sensitizer from a non-sensitizer (YES/ NO answer: chemical label). An assay which addresses sensitizer potency (chemical classification) is of high importance when considering the need to totally replace *in vivo* animal testing for hazard and risk assessment of skin sensitizing chemicals (Mehling et al., 2012). This manuscript describes the international ring trial of an assay that may be able to rank sensitizers according to their potency (dos Santos et al., 2011; Spiekstra et al., 2009). The EE potency assay is a modification of the EURL-ECVAM validated EE assay for assessing the corrosive and irritant properties of a chemical and therefore by definition will not distinguish a sensitizer from a non-sensitizer (Fentem et al., 1998; Spielmann et al., 2007) (for epiCS® see OECD guideline Test Number 439: In vitro skin irritation). The validated skin irritation / corrosion test basically assesses the undiluted test chemical. Our EE potency assay is a modification in the sense that we have expanded the possibility to carry out a dose response of the diluted chemical using the same model (EE) and the same end point (cell viability as assessed by MTT reduction) to address sensitizer potency based on irritant potential. This sensitizer potency classification is based on the clinical observation that there is a clear role for irritancy in contact sensitization due to the irritant properties of many sensitizers (Agner et al., 2002; Basketter et al., 2007; Bonneville et al., 2007; McLelland et al., 1991). The local trauma results in an increase in epidermal cytokine production e.g. IL-1 $\alpha$ . Previously, we have shown that there is a relationship between the strength of the sensitizer and the irritant potential of the chemical (dos Santos et al., 2011; Spiekstra et al., 2009). The primary readout of the EE potency assay is the EC<sub>50</sub> value which is the chemical concentration leading to a 50% decrease in EE viability (MTT assay) compared to vehicle exposed EE. The second readout parameter is the IL-1 $\alpha$ <sub>2x</sub> which is the chemical concentration resulting in a 2-fold increase in the release of the pro-inflammatory cytokine IL-1 $\alpha$  into the culture supernatant. Using a panel of 12 test chemicals, we have shown that the EC<sub>50</sub> value in particular, and IL-1 $\alpha$  release to a lesser extent, correlated well with LLNA-EC<sub>3</sub> data and Human Repeat Insult Patch Test (HRIPT) data with regards to ranking sensitizer potency when using the VUMC-EE model (dos Santos et al., 2011). The EC<sub>3</sub> concentration is the primary parameter used in the murine LLNA and represents the chemical concentration resulting in a three-fold increase of <sup>3</sup>H-thymidine incorporation in the auricular draining lymph node, compared to vehicle control (Gerberick et al., 2007a). The HRIPT is a test which assesses the maximum no observed threshold effect level (NOEL in  $\mu\text{g}/\text{cm}^2$ ) of a chemical in human volunteers (Basketter et al., 2005). Since the EE potency assay does not identify sensitizers it has to be used as a tier 2 assay in combination with one of the above mentioned assays (e.g. NTCT assay, DC maturation or migration assay) in tier 1 which will first identify the sensitizers from the non sensitizers. Since the assay assesses potency it has the potential to identify the maximum safe threshold concentration of a chemical.

Once a potential assay has been developed, the next phase is optimization and testing transferability and reproducibility of the method in different naive laboratories. This is essential for future widespread implementation of the assay. This ring trial set up (also referred to as phase 1 of pre-validation) of the assay involves finalization of a preliminary standard operating procedure, testing the transferability of the assay in different laboratories and finally testing the intra-laboratory and inter-laboratory reproducibility and predictive capacity of the assay with a coded panel of test chemicals. Pre-validation is required before an assay can enter the validation phase with an extended panel of test chemicals in multiple laboratories.

Previously, we have published the transfer phase of the EE potency assay international ring trial (Teunis et al., 2013). The transferability of the standard operating procedure (SOP) was described using two training chemicals (DNCEB and resorcinol) from the lead laboratory (VUMC) to 3 other European laboratories (University of

Applied Sciences Utrecht, The Netherlands (HU), University of Milan, Italy (DiSFeB) and BASF Chemical Company, Ludwigshafen, Germany (BASF)). Furthermore, the transferability of the method from the VUMC-EE to the commercially available epiCS® (previously EST1000™) (CellSystems, Biotechnology GmbH, Troisdorf, Germany) was also described.

In this current study, we report the results obtained from the international ring trial. The intra-laboratory and inter-laboratory reproducibility of the EE potency assay in 4 European laboratories is described along with putative positive and negative acceptance criteria. A test panel of 13 coded chemical sensitizers was used to test the predictive capacity of the assay in ranking sensitizer potency. For this, EC<sub>50</sub> and IL-1 $\alpha$ <sub>2x</sub> values were compared to published mouse LLNA-EC<sub>3</sub> and human NOEL and DSA<sub>05</sub> data. DSA<sub>05</sub> is the chemical dose per skin area in  $\mu\text{g}/\text{cm}^2$  leading to a sensitization incidence of 5% in the human tested population (Schneider and Akkan, 2004). Thereby it was possible to establish a prediction model and to establish a linear correlation graph to rank sensitizers with regards to their weak to extreme sensitizing potencies.

## 2 Materials & Methods

### 2.1 Method outline

For a full description of the technology transfer and standard operating procedure (SOP) for the EE potency assay see supplementary materials in Teunis et al., (2013). Following the SOP, any chemical, which is soluble in DMSO or a mixture (4:1) of acetone:olive oil (AOO) can be tested. The maximum solubility of all test chemicals in this ring trial was determined by an independent laboratory (TNO, Zeist, The Netherlands). For an overview of the EE potency assay method see Fig. 1.

After having assessed the transferability of the EE potency assay (Teunis et al., 2013), the ring trial reported here, involving four European laboratories: VUMC (lead laboratory), HU, DiSFeB and BASF, was started. Thirteen coded well-known sensitizing chemicals were used (Tab. 1). Each chemical was tested in two independent experiments in order to obtain the EC<sub>50</sub> and the IL-1 $\alpha$ <sub>2x</sub> values (Fig. 1). Two independent experiments were defined as two experiments performed on different days and using different EE batches. Two read-out parameters were assessed for each chemical:

*Readout A:* Cell viability measured by MTT assay and expressed as the EC<sub>50</sub> value (effective chemical concentration in mg/ml required to reduce cell viability to 50% compared to vehicle exposed cultures).

*Readout B:* Release of the pro-inflammatory cytokine IL-1 $\alpha$ , measured by ELISA and expressed as IL-1 $\alpha$ <sub>2x</sub> (effective chemical concentration required to result in a 2-fold increase in release of IL-1 $\alpha$  into culture supernatant compared to vehicle-exposed cultures).

Following the SOP, finding the Broad Dose (BD) response range was determined by two consecutive range finding-experiments, named BD-A and BD-B (Fig. 1). From the BD-A, using 10-fold serial dilutions starting from the master starting stock solution, a chemical concentration was identified from the tested range which results in >60%, preferably >80% reduction in EE viability compared to vehicle exposed EE. Then, 3-fold serial dilutions of this starting point (identified from BD-A) were tested in BD-B. From BD-B, again a chemical concentration was identified from the tested range which results in >60%, preferably >80% reduction in EE viability compared to the vehicle. This chemical concentration was then the highest concentration used in the Fine Dose (FD) experiments. Two-fold serial dilutions of this starting point concentration were tested in the FD experiments. If a chemical failed to result in >60% reduction in EE viability in BD-A or BD-B, the chemical was excluded from the assay since it would not be possible to obtain an EC<sub>50</sub> value in the FD.

For each test chemical, BD-A and B are performed in single-fold, whereas the FD experiments were performed in two independent experiments in each laboratory. Only controls (unexposed, vehicle(s) and positive assessment conditions) were tested in duplicate per independent experiment. Statistical analysis and prediction models are described below.

### 2.2 Selection and coding of test chemicals

Chemicals were selected by an independent party (TNO). Initially, over 80 chemicals were short-listed by the project team and from this list, 13 sensitizers were selected and coded by TNO. Each laboratory received a uniquely coded set of test chemicals. The code for the chemicals was communicated directly to the statistician (Adriaens Consulting, Aalter, Belgium), after all data had been received by the statistician. All 13 tested chemicals along with the *in vivo* potency information, vehicles used, maximum solubility and starting concentrations tested in BD-A and BD-B experiments are shown in Tab. 1 and 2. With the exception of 2-mercaptobenzothiazole which was purchased from Fisher-Scientific (ACROS Organics; Loughborough, UK), all chemicals were purchased from Sigma-Aldrich (Sigma, Aldrich, SAFC; St Louis, Missouri, USA). Chemicals were > 95 % pure with the exception of formaldehyde which was 36.5%-38% in H<sub>2</sub>O. Isoeugenol was a 98% mixture of the *cis* and *trans* form and oxazolone was purified by recrystallization.

### 2.3 Epidermal equivalent skin models - maintenance and chemical exposure

epiCS® (Cat.-Nr.: CS-1001) previously known as EST1000™ (Epidermal Skin Test-1000) (CellSystems, Biotechnology GmbH, Troisdorf, Germany) are commercially available reconstructed epidermal skin models, derived from normal human keratinocytes.

#### *Maintenance of EE models*

Upon arrival in the laboratories, the epiCS® skin tissues were handled exactly as recommended by the supplier and as described in detail in the supplementary SOP (Teunis et al., 2013). Maintenance medium (supplied by CellSystems) was used throughout the procedure and was also used for preparing dilutions of the test chemicals. In short, upon receipt, epiCS® cultures were transferred to a 6-well plate containing 1 ml maintenance medium and incubated overnight at 37°C, 5% CO<sub>2</sub>, 95% humidity to allow the cultures to equilibrate. After equilibration, cultures were used for the EE potency assay according to the SOP.

#### *Preparation of chemicals*

DMSO (1% in CellSystems® maintenance medium) or AOO (4:1), the choice depending on which resulted in the highest chemical solubility, were used as vehicles for dissolving the chemicals. Broad dose response A and B and FD response experiments were performed as described in Fig. 1. The positive control chemical was resorcinol (60 mg/ml (545 mM) in 1% DMSO). This concentration was selected from past experience by the VUMC lead laboratory (Teunis et al., 2013; dos Santos et al., 2011). For each experiment, unexposed, vehicle-exposed and positive controls were tested in duplicate and the test chemical concentrations in single fold. In the BD- A and B, 4 concentrations were tested per chemical. In FD, 5 chemical concentrations were tested per chemical.

#### *Exposure to test-chemicals and controls*

Pre-sterilized Finn Chamber filter paper discs of 7.5 mm (Epitest LTD Oy, Finland) were impregnated with 25 µl of the test samples (chemical dilutions, vehicles, positive control). Excess fluid was gently tapped from the filter and the impregnated filters were topically applied to the epiCS® stratum corneum. The epiCS® were then returned to the incubator for 24 hours (37 °C, 5% CO<sub>2</sub>, 95% humidity). After 24 hours of incubation, filter paper disks were gently removed. Culture supernatant was harvested and stored at -20 °C for IL-1α ELISA (FD concentrations only; see below) and epiCS® were harvested for MTT assay in order to assess cell viability (all cultures; see below).

#### *MTT assay and quantification of IL-1α secretion*

The MTT assay and quantification of IL-1α secretion (R&D System Inc. (Minneapolis, Minnesota) by ELISA was performed as described in the supplementary SOP (Teunis et al., 2013).

## **2.4 Acceptance criteria**

*Quality controls of the epiCS® models:* all epiCS® came with a batch control certificate. Models were checked by CellSystems for barrier integrity (defined as within target when viability was >50% after treatment with Triton X-100 for two hours).

*Skin equivalent performance:* in this international ring trial, only putative vehicle and positive control acceptance criteria are defined. Since the acceptance criteria have not been fully tested previously, if an experiment did not fulfill the quality criteria and if an EC<sub>50</sub> value could be obtained, then the EC<sub>50</sub> value was still included in the final analysis.

*Putative acceptance criteria for vehicles:* vehicle exposure alone should not result in more than 30% decrease in cell viability compared to unexposed cultures. If the vehicle results in more than 30% decrease in viability then the EE batch does not fulfill the proposed quality criteria. The percentage difference between the unexposed and the vehicle exposed EE were calculated as follows: ((average viability unexposed – average viability vehicle exposed) / average viability unexposed) x 100.

*Putative acceptance criteria for positive control:* exposure to resorcinol should result in 20-80% (preferably 50%) decrease in cell viability compared to vehicle.

#### *Exclusion of chemicals:*

From the BD-B, a chemical concentration is chosen from the dilution range tested that results in >60%, preferably >80% decrease in EE viability compared to the vehicle. Then, 2x serial dilutions of this starting point chemical concentration were tested in the FD experiments. If a chemical fails to result in >60% decrease in viability in BD-A or BD-B, this chemical was excluded from the assay since no EC<sub>50</sub> value would be determined in FD.

## **2.5 Data management and statistical analysis**

All the data were collected prior to un-coding of the chemicals. For the statistical analyses, a summary template was designed by the statistician, and the results were transferred to this template by each participating laboratory. This summary template contained internal checks that ensured that no mistakes were made in the transfer of the results.

*Reproducibility of the controls:* the viability (MTT assay) of the unexposed, vehicle exposed and positive control were plotted for each batch of EE and the frequency of experiments fulfilling the putative acceptance criteria recorded.

*Reproducibility of the Broad Dose experiments:* the BD experiments provided the dose range for final testing of the chemical in the FD response experiments. The concentrations obtained in the BD experiments were tabulated and compared between the laboratories (exploratory).

*Reproducibility of the Fine Dose experiments:* the FD experiments were performed in duplicate. The agreement in EC<sub>50</sub> concentration between the two independent experiments within each laboratory was assessed with scatter plots. Correlations between the 2 runs were determined by Pearson analysis (2-tailed) in combination with line of equality. Analyses with 95% confidence interval using GraphPad Software, San Diego, CA, USA. Correlations were considered significant for  $p < 0.05$ .

*EE potency assay:* the EC<sub>50</sub> value is the effective chemical concentration required to reduce metabolic activity (corresponding to cell viability) to 50% of the maximum value. The 100% value for cell viability corresponds to the vehicle control (1% DMSO in culture medium or AOO 4:1). EC<sub>50</sub> values were obtained by linear regression analysis based on changes in metabolic activity (MTT). In order to rank the chemicals, correlations between EC<sub>50</sub> and LLNA, NOEL or DSA<sub>05</sub> were determined by nonparametric two-tailed correlation Spearman Analyses using GraphPad Software, San Diego, CA, USA.

*IL-1 $\alpha$  release and potency:* IL-1 $\alpha_{2x}$  values were obtained by linear regression analysis based on the chemical concentration resulting in a 2-fold release in IL-1 $\alpha$ . In order to rank the chemicals, correlations between EC<sub>50</sub> and LLNA, NOEL or DSA<sub>05</sub> were determined by nonparametric two-tailed correlation Spearman Analyses using GraphPad Software, San Diego, CA, USA.

*Prediction model:* In addition to the previously proposed ranking prediction model, in which the lower the EC<sub>50</sub> value, the more cytotoxic (irritant) the chemical and the stronger the sensitizing potency of the chemical is, an additional prediction model was identified in this study where strong and extreme sensitizers had EC<sub>50</sub> values  $< 7$  mg/ml chemical whereas the majority of the moderate and weak sensitizers had EC<sub>50</sub> value  $\geq 7$  mg/ml chemical.

### 3 Results

#### 3.1 Acceptance criteria: reproducibility of data for vehicle and positive control resorcinol

Very little batch variation was observed between the unexposed batches of epiCS® used in each laboratory. In total, 23 different batches were used in the 4 different laboratories, with the same batch often being delivered to multiple laboratories. Average OD<sub>570 nm</sub> values obtained from the MTT assay of unexposed EE for the different batches were as follows - VUMC (n=14): OD<sub>570 nm</sub> 2.866 +/- 0.279; HU (n=12): OD<sub>570 nm</sub> 2.695 +/- 0.266; DiSFeB (n=12): OD<sub>570 nm</sub> 2.881 +/- 0.586; BASF (n=12): OD<sub>570 nm</sub> 3.047 +/- 0.868. Vehicle exposure generally did not result in more than 30% decrease in cell viability compared to unexposed cultures in accordance with the proposed acceptance criteria for this international ring trial. Of the 23 batches used in this study, only 1 batch in the VUMC lab (batch 5) and a different batch in the BASF lab (batch 12) showed slightly more than 30% cytotoxicity after vehicle exposure compared to unexposed cultures (Fig. 2). A putative acceptance criterium was also defined for the positive control resorcinol (545 mM): topical exposure to a single concentration of resorcinol should result in 20 – 80% (preferably 50%) decrease in EE viability. However, variation was observed both within the labs and between the labs (Fig. 2). For VUMC, batch 1 and 4 showed  $< 20\%$  decrease in EE viability and batch 9 showed  $> 80\%$  decrease in EE viability when exposed to resorcinol. For HU, batch 11 showed  $< 20\%$  decrease in EE viability when exposed to resorcinol. For DiSFeB, batches 1, 7 and 11 showed  $> 80\%$  decrease in EE viability when exposed to resorcinol. For BASF, batches 7 and 8 showed  $> 80\%$  decrease in EE viability when exposed to resorcinol. Of note, the batch numbers allotted to each laboratory did not correlate between labs and therefore the deviations observed between laboratories was not due to the same batch of epiCS®. This indicates that variation was due to technical inter-laboratory variation rather than true batch variation. Since the vehicle and positive performance criteria had not been tested before the start of the study, BD and FD data obtained from batches not meeting the putative performance criteria were still included in all further analysis for determining EC<sub>50</sub> values and potency.

#### 3.2 Broad Dose B response

Chemical concentrations were selected from BD-A for further testing in BD-B (Fig. 1; Tab. 2). From BD-B the chemical concentration could be selected by each laboratory for use in FD and identification of the EC<sub>50</sub> (Tab. 2). Of the 13 coded sensitizers selected for the study, in the VUMC, HU and DiSFeB labs, 11 chemicals resulted in  $> 60\%$  decrease in EE viability enabling a chemical concentration to be selected for further testing in FD (Tab. 2). In

the BASF lab only 9 chemicals resulted in > 60% decrease in EE viability. Exposure to p-phenylenediamine and cobalt (II) chloride was reported to give unreliable results or no 60% decrease in EE viability in 3 of the 4 laboratories. For both chemicals this was due to interference with the MTT photometric assay (p-phenylenediamine oxidized spontaneously to a brown compound and cobalt (II) chloride had a strong green colour). When unreliable results were reported, the chemicals were excluded from further analysis in FD in the corresponding laboratories. Furthermore, for unknown reasons, 60% decrease in EE viability was not reached when exposing EE to 2-mercaptobenzothiazole, phenylacetaldehyde or formaldehyde in the BASF laboratory. The 60% decrease in EE viability was also not reached for formaldehyde in the DiSFeB laboratory. The chemical concentration selected to enter the FD was not identical in each laboratory and sometimes differed by up to 10-fold. In conclusion, the start concentrations for the FD experiments for 11 sensitizers were identified in VUMC, HU and DiSFeB laboratories, and for 9 sensitizers in the BASF laboratory (Tab. 2).

### 3.3 Fine Dose response and determination of EC<sub>50</sub> value

*Inter-experiment variability:* the inter-experiment variability within a laboratory and between laboratories is a measure for the robustness of the assay. For each chemical the EC<sub>50</sub> value was determined in two separate runs (FD-1 and FD-2) and the FD-1 and FD-2 results correlated with each other (Fig. 3). In general, many dots (chemicals) were near or touching the line of equality indicating very good reproducibility within a laboratory. The weak sensitizers benzocaine (DiSFeB) and  $\alpha$ -hexylcinnamaldehyde (VUMC, HU) and the strong sensitizer cobalt (II) chloride (BASF) showed poor reproducibility between runs. Of note, cobalt (II) chloride was already identified by VUMC, HU and DiSFeB as giving unreliable results in the MTT assay due to its interference with the spectrophotometric assay readout. Pearson correlations of all chemicals tested yielded a strong correlation between both runs (FD-1 and FD-2) for VUMC, HU and BASF. Pearson r values ranged from 0.965 to 0.989 (p-value: 0.0001) in these three laboratories. DiSFeB showed slightly less but still significant correlation (Pearson r value: 0.688 p-value: 0.019). These results indicate extremely low intra- and inter- laboratory variation with regards to the assay protocol.

### 3.4 EC<sub>50</sub> potency ranking

From the FD experiments, EC<sub>50</sub> values were determined. All individual results for each laboratory and each Fine Dose experiment are shown in Tab. 3. The EC<sub>50</sub> values were used to rank the potency of the chemical. The lower the EC<sub>50</sub> value, the more cytotoxic (irritant) the chemical, and the stronger the sensitizing potency of the chemical. In general there was a good agreement between the EC<sub>50</sub> values obtained in the 4 laboratories. Two laboratories reported unreliable results for EC<sub>50</sub> finding for 2-mercaptobenzothiazole (VUMC, HU) and BASF did not test this chemical as no EC<sub>60</sub> was obtained in the BD experiments. p-phenylenediamine and cobalt (II) chloride were also not tested in 3 of the 4 laboratories since no reliable EC<sub>60</sub> was obtained in the BD experiments.

In the majority of the runs, strong and extreme sensitizers had EC<sub>50</sub> values < 7 mg/ml chemical whereas the majority of the moderate and weak sensitizers had EC<sub>50</sub> value  $\geq$  7 mg/ml chemical (Tab. 3). Therefore it was next determined whether it was possible to differentiate weak / moderate from moderate/ strong sensitizers using a cut-off of 7 mg/ml (Tab. 4). Since only 2 FD runs were performed in this study, some chemicals scored an ambiguous result. Tab. 4 describes this prediction model excluding ambiguous results and also describing the worst case scenario if ambiguous chemicals would happen to score negative. For correct classification of ambiguous chemicals a 3<sup>rd</sup> FD run would be required in the future. VUMC, HU and BASF showed good sensitivity (60 – 83%), specificity (80 – 100 %) and accuracy (73 – 82%) in the EE potency assay using the 7 mg/ml as cut off. Only the DiSFeB laboratory analysed the strong sensitizer p-phenylenediamine which resulted in an EC<sub>50</sub> value  $\geq$  7 mg/ml and was the main reason for the generally lower sensitivity and accuracy obtained by DiSFeB compared to the other 3 laboratories. For this prediction model, the within laboratory reproducibility of the FD runs had a concordance ranging from 77 – 100 % and the between laboratory concordance was 35% for all laboratories combined and 77% for the 2 best performing laboratories (VUMC and HU) (Tab. 4).

### 3.5 Correlation of EC<sub>50</sub> potency values with *in vivo* LLNA-EC<sub>3</sub>, and human DSA<sub>05</sub>

Next the EC<sub>50</sub> data were correlated to human NOEL, DSA<sub>05</sub> and LLNA-EC<sub>3</sub> (Fig. 4). Clearly very reproducible, good correlating and generally significant results were obtained by each independent laboratory and particularly when the data obtained from all 4 laboratories were averaged (all laboratories combined: EE-EC<sub>50</sub> vs. NOEL spearman r = 0.720; p = 0.034; EE-EC<sub>50</sub> vs. DSA<sub>05</sub> spearman r = 0.845; p = 0.006 compared to EE-EC<sub>50</sub> vs. LLNA-EC<sub>3</sub> spearman r = 0.715; p = 0.016). For the independent laboratories, the *in vitro* EE-EC<sub>50</sub> correlation to the human DSA<sub>05</sub> data was exceptionally high although it should be noted that for the main outlier oxazolone there is only mouse LLNA data available for the correlations.

### 3.6 Correlation of IL-1 $\alpha$ <sub>2x</sub> values with *in vivo* LLNA-EC<sub>3</sub>, and human DSA<sub>05</sub>

Since IL-1 $\alpha$  release is related to cytotoxicity and irritation, and therefore also possibly to sensitizer potency, it was next determined whether a correlation also existed between the IL-1 $\alpha$ <sub>2x</sub> value, and NOEL, DSA<sub>05</sub> or LLNA-EC<sub>3</sub> (Fig. 5; Tab. 5 and 6). IL-1 $\alpha$ <sub>2x</sub> is the chemical concentration which causes a 2-fold release in IL-1 $\alpha$  from the EE into the culture supernatant. Indeed, again reproducible, good correlating and generally significant results were obtained by each independent laboratory and particularly when the data obtained from all 4 laboratories was

averaged. Again, in all cases, the *in vitro* IL-1 $\alpha_{2x}$  correlation to the human DSA<sub>05</sub> data was very high and in the same order of magnitude as that observed for the EC<sub>50</sub> value correlations (all laboratories combined: IL-1 $\alpha_{2x}$  vs. DSA<sub>05</sub> spearman  $r=0.929$ ;  $p=0.002$  compared to IL-1 $\alpha_{2x}$  vs. LLNA-EC<sub>3</sub> spearman  $r=0.770$ ;  $p=0.013$  or IL-1 $\alpha_{2x}$  vs. or NOEL spearman  $r=0.810$ ;  $p=0.022$ ).

#### 4 Discussion

In this international ring trial the intra- and inter- laboratory variation, and the predictive capacity of the EE potency assay were evaluated. Highly reproducible results were obtained in each laboratory. In all laboratories, human EE-EC<sub>50</sub> data showed better correlation to human data than to mouse LLNA-EC<sub>3</sub> data.

Since acceptance criteria had not been previously described, putative acceptance criteria were defined at the start of the study and tested during the study (dos Santos et al., 2011). Batch variation between the unexposed batches of epiCS® was very low indicating that the production procedure and transport of the EE was very standardized. For the vehicle exposure, the acceptance criteria "vehicle exposure alone should not result in more than 30 % decrease in cell viability compared to unexposed EE" was met. This indicates that this putative acceptance criteria can now be accepted as a valid acceptance criteria for vehicle exposure when further implementing this assay. In contrast, the putative acceptance criteria defined for the positive control resorcinol (20-80% cytotoxicity compared to vehicle exposed EE) was found not to be suitable for further studies. The degree of cytotoxicity exhibited often varied from < 20 % to > 80 % cytotoxicity between batches and between laboratories when testing a single concentration of resorcinol. In the future, the problem may possibly be solved by testing at least 2 different resorcinol concentrations thus allowing for slight shifts in the dose response between experiments or by testing a stronger sensitizer, e.g. DNCB, which shows less variation and which is not a prohapten.

The SOP was designed to be able to determine the EC<sub>50</sub> value of any unknown chemical using 3 consecutive dose response experiments. Notably, both the intra-laboratory and inter laboratory variation was low throughout the BD-A, BD-B and FD experiments. The same start chemical concentration for the FD was often identified by all 4 laboratories (e.g. cinnamaldehyde, citral, eugenol). Also, similar results (EC<sub>50</sub> values) were often obtained from the FD between the laboratories despite sometimes up to 10-fold variation in the start chemical concentration being used (e.g. oxazolone). Even though 2 weak sensitizing chemicals (benzocaine,  $\alpha$ -hexylcinnamaldehyde) showed poor reproducibility in the duplicate FD runs these chemicals were still correctly ranked by all laboratories as weak sensitizers. Of note, the EE potency assay appears not only to be reproducible between laboratories but also to a certain extent between different EE cultures (dos Santos et al., 2011; Gibbs et al., 2013a). For example the EC<sub>50</sub> value obtained for DNCB was 0.3 mg/ml in this present study and 1.3 mg/ml in our previous study using in house VUMC EE which when plotted on a log scale represents very little variation. Taken together, these results emphasize the beneficial effect on final reproducibility by starting with a broad dose finding (10 fold dilutions) to identify and fine tune the final fine dose finding (2 fold dilutions). The results created in this study have now been incorporated into our most recent unpublished developments in which we have been able to identify a single extended dose response of 2 fold dilutions starting at 200 mg/ml which will enable all unknown sensitizers from weak to extreme to be tested. Notably, benzocaine has been reported to give highly variable results *in vivo*, in both the LLNA and GPMT (Basketter et al., 1995). Two chemicals proved to be difficult to test. P-phenylenediamine and cobalt (II) chloride both interfered with the spectrophotometric MTT assay. Whereas p-phenylenediamine oxidizes the substrate in the absence of viable EE, cobalt (II) chloride has a green colour. Three of the 4 laboratories excluded these 2 coded chemicals already in BD-B since no EC<sub>50</sub> was obtained. DiSFeB did continue to test p-phenylenediamine, and BASF did continue to test cobalt (II) chloride in the FD. However both laboratories wrongly classified the chemicals as a weak / moderate sensitizer. This suggests a minor modification to the SOP is required specifying in more detail when a chemical should be excluded due to interference with the MTT assay. For example, the SOP should mention prior analysis of the chemical in the MTT assay in the absence of EE in order to determine whether the chemical distorts the spectrophotometric readout.

Until now, no classical prediction model for the EE potency assay has been defined. Using a test panel of chemicals, EC<sub>50</sub> and IL-1 $\alpha_{2x}$  values have been obtained which are correlated to human or LLNA-EC<sub>3</sub> data (dos Santos et al., 2011). By continuously adding values obtained from well-defined chemicals, this graph will provide a golden standard correlation graph for determining the potency of an unknown chemical allergen. The EC<sub>50</sub> and IL-1 $\alpha_{2x}$  values of the unknown chemical can then be correlated to values obtained for the standard test panel and extrapolated to an *in vivo* value. Eventually enough data will be created in order for the EE potency assay to have its own assessment score in a similar manner to LLNA-EC<sub>3</sub> and human NOEL or DSA<sub>05</sub> scores which rank sensitizer potency according to cut off ranges (see Tab. 1). We foresee that such data will eventually enable the maximum safe threshold concentration of a chemical to be identified when sufficient NOEL and *in vitro* data are available. In this study an additional prediction model was identified. It was noticed that the potency of a coded chemical could be determined with high accuracy on the basis of a cut-off value for the EC<sub>50</sub> (EE-EC<sub>50</sub>  $\geq 7$  mg/ml = weak to moderate sensitizer; EE-EC<sub>50</sub> < 7 mg/ml = strong to extreme sensitizer). The average overall accuracy for this approach for the combined results of all 4 laboratories was 77%, meaning that when using the current

assay SOP the chemical was correctly predicted to be either a strong to extreme or weak to moderate sensitizer in 77% of the test situations if 2 similarly scoring FD runs are obtained. A minor modification to the current SOP should allow for ambiguous scoring chemicals: if ambiguous scoring from the 2 FD runs is obtained a 3<sup>rd</sup> deciding FD run should be performed. This prediction model could be very suitable to quickly screen for the most potent sensitizers. Importantly, the discrimination between two classes of sensitizers (weak and strong) coincides with the European Classification, Labeling and Packaging of substances (CLP) regulation which is harmonized with the United Nations Globally Harmonized System (GHS) of Classification and Labeling of Chemicals (UN-GHS) (see review de Groot et al., 2010).

In order to test the EE potency assay prediction model further with regards to correlating the EC<sub>50</sub> value to available human and LLNA-EC<sub>3</sub> data, first a detailed review of the literature was performed to identify human and murine potency data. The majority of the data was found in two extensive reports published by ICCVAM (see ICCVAM, 2013a and b and references listed at footnote of supplementary table). Also a new human classification score ranging from 1 to 6 with 1 being the most potent sensitizer group and 6 being the non sensitizer group has very recently been proposed by Basketter et al., (2014) and has also been incorporated into Tab. 1. When animal and human data were conflicting or limited (e.g. 2-mercaptobenzothiazole), the human data was prioritized in the ranking above animal data (see Tab. 1). It was noticed from the reports that the LLNA-EC<sub>3</sub> potency data was influenced considerably by the vehicle used and the type and duration of the chemical exposure. Therefore, for this study it was decided to use a range of potency data available for LLNA-EC<sub>3</sub>, and all human DSA<sub>05</sub>, and limited NOEL data available as described in the reports and to correlate this, not only to the results obtained from the individual laboratories but also to the average result obtained from the 4 laboratories combined. A very good correlation was observed between each laboratory and *in vivo* data. This was particularly so with regards to DSA<sub>05</sub>. Taken together, the human data showed a notably better correlation to EE-EC<sub>50</sub> data than the murine to EE-EC<sub>50</sub> data although it must be noted that no human data was available for the major outlier oxazolone, thus possibly introducing a minor bias to this result.

Next it was determined whether the release of pro-inflammatory IL-1 $\alpha$  into the culture supernatant could provide an additional potency assessment parameter to the EC<sub>50</sub> value. Therefore, the IL-1 $\alpha_{2x}$  value was correlated with human data and LLNA-EC<sub>3</sub> data. As with the EC<sub>50</sub> values, a significant correlation was found, both generally on the individual laboratory level, as well as in the overall (averaged) correlation. Again the human data showed a notably better correlation than the mouse data.

The major limitation of EE potency assay is that although it can classify chemical allergens according to potency it is not able to determine whether or not the chemical is a potential sensitizer. Previously, we have shown that IL-18 production by epidermal keratinocytes (NCTC2544 cell line) is a biomarker for distinguishing a sensitizer from a non-sensitizer (Corsini et al., 2009; Corsini et al., 2013; Galbiati et al., 2011). Parallel to this study, we found that the EE-EC<sub>50</sub> potency assay could be combined with IL-18 release by EE in a single assay, thus greatly increasing the value of this assay which uses commercially available EE in the future (Gibbs et al., 2013a). Alternatively, the EE potency assay can be combined with any other assay or test battery which can distinguish a sensitizer from a non-sensitizer in a tiered or integrated approach (Corsini et al., 2009; Johansson et al., 2013; Natsch et al., 2013). Another limitation in the assay is that not all chemical exposures result in an EE-EC<sub>50</sub> value being obtained. If there are no solubility issues, e.g. a maximum concentration of 200 mg/ml could be tested, and still no EE-EC<sub>50</sub> value is obtained it is possible that the chemical is a very weak sensitizer. However it cannot be ruled out that the chemical does not penetrate the stratum corneum and therefore cannot be tested properly in the assay. *In vivo*, the penetration route for such a chemical may possibly be via the hair follicle. At the moment such chemicals are considered to fall outside of the applicability domain as they cannot be fully tested according to the SOP.

Whereas many assays are being developed or are under pre-validation for determining whether or not a chemical is a potential sensitizer, relatively few assays address sensitizer potency (Mehling et al., 2012). A complicating factor in comparing the potency data from other studies and our study is that correlation of the data to LLNA data is performed using different statistical means in the different studies, and few as well as different chemicals have been investigated (Kolle et al., 2012). One very promising assay is the Genomic Allergen Rapid Detection (GARD) assay (Johansson et al., 2011). This dendritic cell based assay uses a genomic biomarker signature of chemical exposed MUTZ-3 cells to determine whether a chemical is a potential sensitizer and also the potency of the sensitizer. Similar to our EE potency assay, oxazolone was a major outlier and  $\alpha$ -hexylcinnamaldehyde scored as a weak sensitizer rather than as sometimes reported in LLNA as a moderate sensitizer. Another assay is the KeratinoSens<sup>TM</sup> assay (ARE-regulated luciferase activity assay using the cell line HaCaT containing a stable insert of the luciferase gene under control of the ARE element of the gene AKR1C2) which has recently undergone international pre-validation (Natsch et al., 2011). Also, a very different type of *in vitro* assay is the non-cell based peptide reactivity assay based on the ability of a chemical to react with 2 synthetic peptides containing either a single cysteine or lysine (Gerberick et al., 2007b; Gerberick et al., 2009). Interestingly, whereas these last 2 assays could correctly classify oxazolone as an extreme/ strong sensitizer they both have difficulty identifying and therefore also assessing potency of pro-haptens (Emter et al., 2010). For example resorcinol and eugenol are false negatives whereas in our EE potency assay these 2 pro-haptens can be accurately assessed (dos Santos et al., 2011). Taken together, these results indicate that if limitations are taken into account such as chemical solubility, instability and metabolism, then *in vitro* assays may have the

potential to assess sensitizer potency. At the moment it is too early to say whether one assay performs better than another. However, it would be interesting and important to determine whether the different assays are able to complement each other with regards to chemicals which perform poorly in one particular assay and well in another assay (Bauch et al., 2011 and 2012). Such an approach to assess skin sensitizer potency may involve the inclusion of multiple toxicological parameters and a weight of evidence approach, using the data from multiple assays as suggested by Natsch and co-workers who have started by analysing the assay-end points for 145 chemicals, tested in the U937, the DRPA and in the KeratinoSens™ assay (Natsch et al., 2013). The EE potency assay could not only be able to provide additional *in vitro* information, but may also increase the relevance of the information for humans, since it involves penetration (bioavailability) of a chemical through the stratum corneum in order for the chemical to exert a cytotoxic / irritant effect on the viable epidermal layers below.

In conclusion, our results indicate that the EE potency assay is a robust assay for testing chemical sensitizers of unknown potencies and that only minor modifications are required before entering further validation. Little intra- and inter-laboratory variation was observed and a good correlation was observed between our *in vitro* EC<sub>50</sub> potency data and that derived from human and animal studies. At present since only a few chemicals have been tested, it is too early to say whether a combined readout of EE-EC<sub>50</sub> and IL-1 $\alpha$ <sub>2x</sub> will further improve the prediction model. Our results suggest that this assay may now be suitable for validation as it will provide additional and complementary information to other assays already undergoing such developments.

## References

- Adler, S., Basketter, D., Creton, S., et al. (2011). Alternative (non-animal) methods for cosmetics testing: current status and future prospects-2010. *Arch Toxicol* 85, 367-485. <http://dx.doi.org/10.1007/s00204-011-0693-2>
- Agner, T., Johansen, J. D., Overgaard, L., et al. (2002). Combined effects of irritants and allergens. Synergistic effects of nickel and sodium lauryl sulfate in nickel-sensitized individuals. *Contact Dermatitis* 47, 21-26. <http://dx.doi.org/10.1034/j.1600-0536.2002.470105.x>
- Basketter, D.A., Alépée, N., Ashikaga, T., et al. (2014). Categorization of Chemicals According to Their Relative Human Skin Sensitizing Potency. *Dermatitis* 25,11-21.
- Basketter, D. A., Clapp, C., Jefferies, D. et al. (2005). Predictive identification of human skin sensitization thresholds. *Contact Dermatitis* 53, 260-267. <http://dx.doi.org/10.1111/j.0105-1873.2005.00707.x>
- Basketter, D. A., Kan-King-Yu, D., Dierkes, P., et al. (2007). Does irritation potency contribute to the skin sensitization potency of contact allergens? *Cutan Ocul Toxicol* 26, 279-286. <http://dx.doi.org/10.1080/15569520701555359>
- Basketter, D. A. and Kimber, I. (2009). Updating the skin sensitization *in vitro* data assessment paradigm in 2009. *J Appl Toxicol* 29, 545-550. <http://dx.doi.org/10.1002/jat.1443>
- Basketter, D. A., Lea, L. J., Dickens, A., et al. (1999). A comparison of statistical approaches to the derivation of EC3 values from local lymph node assay dose responses. *J Appl Toxicol* 19, 261-266. [http://dx.doi.org/10.1002/\(SICI\)1099-1263\(199907/08\)19:4<261::AID-JAT572>3.0.CO;2-5](http://dx.doi.org/10.1002/(SICI)1099-1263(199907/08)19:4<261::AID-JAT572>3.0.CO;2-5)
- Basketter, D. A., Scholes, E. W., Wahlkvist, H., et al. (1995). An evaluation of the suitability of benzocaine as a positive control skin sensitizer. *Contact Dermatitis* 33, 28-32. <http://dx.doi.org/10.1111/j.1600-0536.1995.tb00443.x>
- Bauch, C., Kolle, S. N., Ramirez, T., et al. (2012). Putting the parts together: combining *in vitro* methods to test for skin sensitizing potentials. *Regul Toxicol Pharmacol* 63, 489-504. <http://dx.doi.org/10.1016/j.yrtph.2012.05.013>
- Bauch, C., Kolle, S. N., Fabian, E., et al. (2011). Intralaboratory validation of four *in vitro* assays for the prediction of the skin sensitizing potential of chemicals. *Toxicology in Vitro* 25, 1162-1168. <http://dx.doi.org/10.1016/j.tiv.2011.05.030>
- Bonneville, M., Chavagnac, C., Vocanson, M., et al. (2007). Skin contact irritation conditions the development and severity of allergic contact dermatitis. *J Invest Dermatol* 127, 1430-1435. <http://dx.doi.org/10.1038/sj.jid.5700726>
- Corsini, E., Galbiati, V., Mitjans, M., et al. (2013). NCTC 2544 and IL-18 production: A tool for the identification of contact allergens. *Toxicol In Vitro* 27, 1127-1134. <http://dx.doi.org/10.1016/j.tiv.2012.05.018>
- Corsini, E., Mitjans, M., Galbiati, V., et al. (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. <http://dx.doi.org/10.1016/j.tiv.2009.04.005>
- De Wever, B., Fuchs, H. W., Gaca, M., et al. (2012). Implementation challenges for designing integrated *in vitro* testing strategies (ITS) aiming at reducing and replacing animal experimentation. *Toxicol In Vitro* 26, 526-534. <http://dx.doi.org/10.1016/j.tiv.2012.01.009>
- de Groot R., Brekelmans, P., Herremans, J., et al. (2010). The changes in hazard classification and product notification procedures of the new European CLP and Cosmetics Regulations. *Clin Toxicol (Phila)*. 48, 28-33. <http://dx.doi.org/10.3109/15563650903376097>

- dos Santos, G. G., Reinders, J., Ouweland, K., et al. (2009). Progress on the development of human in vitro dendritic cell based assays for assessment of the sensitizing potential of a compound. *Toxicol Appl Pharmacol* 236, 372-382. <http://dx.doi.org/10.1016/j.taap.2009.02.004>
- dos Santos, G. G., Spiekstra, S. W., Sampat-Sardjoepersad, S. C., et al. (2011). A potential in vitro epidermal equivalent assay to determine sensitizer potency. *Toxicol In Vitro* 25, 347-357. <http://dx.doi.org/10.1016/j.tiv.2010.10.008>
- 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. <http://dx.doi.org/10.1016/j.taap.2010.03.009>
- Fentem, J. H., Archer, G. E., Balls, M., et al. (1998). The ECVAM International Validation Study on In Vitro Tests for Skin Corrosivity. 2. Results and Evaluation by the Management Team. *Toxicol In Vitro* 12, 483-524. [http://dx.doi.org/10.1016/S0887-2333\(98\)00019-8](http://dx.doi.org/10.1016/S0887-2333(98)00019-8)
- Galbiati, V., Mitjans, M., Lucchi, L., et al. (2011). Further development of the NCTC 2544 IL-18 assay to identify in vitro contact allergens. *Toxicol In Vitro* 25, 724-732. <http://dx.doi.org/10.1016/j.tiv.2010.12.011>
- Gerberick, G. F., Ryan, C. A., Dearman, R. J., et al. (2007a). Local lymph node assay (LLNA) for detection of sensitization capacity of chemicals. *Methods* 41, 54-60. <http://dx.doi.org/10.1016/j.ymeth.2006.07.006>
- Gerberick, G. F., Troutman, J. A., Foertsch, L. M., et al. (2009). Investigation of peptide reactivity of pro-hapten skin sensitizers using a peroxidase-peroxide oxidation system. *Toxicol Sci* 112, 164-174. <http://dx.doi.org/10.1093/toxsci/kfp192>
- Gerberick, G. F., Vassallo, J. D., Bailey, R. E., et al. (2004). Development of a peptide reactivity assay for screening contact allergens. *Toxicol Sci* 81, 332-343. <http://dx.doi.org/10.1093/toxsci/kfh213>
- Gerberick, G. F., Vassallo, J. D., Foertsch, L. M., et al. (2007b). Quantification of chemical peptide reactivity for screening contact allergens: a classification tree model approach. *Toxicol Sci* 97, 417-427. <http://dx.doi.org/10.1093/toxsci/kfm064>
- Gibbs, S., Corsini, E., Spiekstra, S. W., et al. (2013a). An epidermal equivalent assay for identification and ranking potency of contact sensitizers. *Toxicol Appl Pharmacol* 272, 529-541. <http://dx.doi.org/10.1016/j.taap.2013.07.003>
- Gibbs, S., Spiekstra, S., Corsini, E., et al. (2013b). Dendritic cell migration assay: a potential prediction model for identification of contact allergens. *Toxicol In Vitro* 27, 1170-1179. <http://dx.doi.org/10.1016/j.tiv.2012.05.016>
- Grindon, C. (2007). The new EU REACH regulation has finally been adopted: is this the end of the campaign trail... or just the beginning? *Altern Lab Anim* 35, 239-242.
- Grindon, C., Combes, R., Cronin, M. T., et al. (2008). An integrated decision-tree testing strategy for skin sensitisation with respect to the requirements of the EU REACH legislation. *Altern Lab Anim* 36 Suppl 1, 75-89.
- ICCVAM (2013a). Comparative LLNA, Guinea Pig, and Human Data Used in the Performance Evaluation. Annex II-1. LLNA Data for 196 Substances Used for the Evaluation of Skin Sensitization Potency [http://iccvam.niehs.nih.gov/docs/immunotox\\_docs/LLNA-pot/3b-AppC-BRD-AnnexII-1.pdf](http://iccvam.niehs.nih.gov/docs/immunotox_docs/LLNA-pot/3b-AppC-BRD-AnnexII-1.pdf). Last assessed 11 December 2013.
- ICCVAM (2013b). Human Data for LLNA Potency Evaluation. Annex II-2. [http://iccvam.niehs.nih.gov/docs/immunotox\\_docs/LLNA-pot/3b-AppC-BRD-AnnexII-2.pdf](http://iccvam.niehs.nih.gov/docs/immunotox_docs/LLNA-pot/3b-AppC-BRD-AnnexII-2.pdf). Last assessed 11 December 2013.
- Johansson, H., Albrekt, A. S., Borrebaeck, C. A., et al. (2013). The GARD assay for assessment of chemical skin sensitizers. *Toxicol In Vitro* 27, 1163-1169. <http://dx.doi.org/10.1016/j.tiv.2012.05.019>
- Johansson, H., Lindstedt, M., Albrekt, A. S., et al. (2011). A genomic biomarker signature can predict skin sensitizers using a cell-based in vitro alternative to animal tests. *BMC Genomics* 12, 399. <http://dx.doi.org/10.1186/1471-2164-12-399>
- Kolle, S. N., Basketter, D. A., Casati, S., et al. (2013). Performance standards and alternative assays: Practical insights from skin sensitization. *Regulatory Toxicology and Pharmacology* 65, 278-85. <http://dx.doi.org/10.1016/j.yrtph.2012.12.006>
- Lindstedt, M. and Borrebaeck, C. (2011). Pattern rules: biomarker signatures for sensitization as an alternative to animal testing. *Biomark Med* 5, 809-811. <http://dx.doi.org/10.2217/bmm.11.82>
- Martin, S. F., Esser, P. R., Schmucker, S., et al. (2010). T-cell recognition of chemicals, protein allergens and drugs: towards the development of in vitro assays. *Cell Mol Life Sci* 67, 4171-4184. <http://dx.doi.org/10.1007/s00018-010-0495-3>
- Maxwell, G., Aeby, P., Ashikaga, T., et al. (2011). Skin sensitisation: the Colipa strategy for developing and evaluating non-animal test methods for risk assessment. *ALTEX* 28, 50-55. <http://www.altex.ch/All-issues/Issue.50.html?iid=122&aid=5>
- McLelland, J., Shuster, S., Matthews, J. N. (1991). 'Irritants' increase the response to an allergen in allergic contact dermatitis. *Arch Dermatol* 127, 1016-1019. <http://dx.doi.org/10.1001/archderm.1991.01680060090010>

- Mehling, A., Eriksson, T., Eltze, T., et al. (2012). Non-animal test methods for predicting skin sensitization potentials. *Arch Toxicol* 86, 1273-1295. <http://dx.doi.org/10.1007/s00204-012-0867-6>
- Natsch, A., Bauch C., Foertsch, L., et al. (2011). The intra- and inter-laboratory reproducibility and predictivity of the KeratinoSens assay to predict skin sensitizers in vitro: results of a ring-study in five laboratories. *Toxicol In Vitro* 25, 733-744. <http://dx.doi.org/10.1016/j.tiv.2010.12.014>
- Natsch, A., Ryan, C. A., Foertsch, L., et al. (2013). A dataset on 145 chemicals tested in alternative assays for skin sensitization undergoing prevalidation. *J Appl Toxicol*. <http://dx.doi.org/10.1002/jat.2868> (Epub ahead of print)
- Peiser, M., Tralau, T., Heidler, J., et al. (2012). Allergic contact dermatitis: epidemiology, molecular mechanisms, in vitro methods and regulatory aspects. Current knowledge assembled at an international workshop at BfR, Germany. *Cell Mol Life Sci* 69, 763-781. <http://dx.doi.org/10.1007/s00018-011-0846-8>
- Roggen, E. L. (2013). Application of the acquired knowledge and implementation of the Sens-it-iv toolbox for identification and classification of skin and respiratory sensitizers. *Toxicol In Vitro* 27, 1122-1126. <http://dx.doi.org/10.1016/j.tiv.2012.09.019>
- Rovida, C. and Hartung, T. (2009). Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals - a report by the transatlantic think tank for toxicology (t(4)). *ALTEX* 26, 187-208. <http://www.altex.ch/All-issues/Issue.50.html?iid=107&aid=4>
- Rovida, C., Ryan, C., Cinelli, S., et al. (2012). The local lymph node assay (LLNA). *Curr Protoc Toxicol*. Chapter 20, Unit.20.7. <http://dx.doi.org/10.1002/0471140856.tx2007s51>
- Spiekstra, S. W., dos Santos, G. G., Scheper, R. J., et al. (2009). Potential method to determine irritant potency in vitro - Comparison of two reconstructed epidermal culture models with different barrier competency. *Toxicol In Vitro* 23, 349-355. <http://dx.doi.org/10.1016/j.tiv.2008.12.010>
- Spielmann, H., Hoffmann, S., Liebsch, M., et al. (2007). The ECVAM international validation study on in vitro tests for acute skin irritation: report on the validity of the EPISKIN and EpiDerm assays and on the Skin Integrity Function Test. *Altern Lab Anim* 35, 559-601.
- Teunis, M., Corsini, E., Smits, M., et al. (2013). Transfer of a two-tiered keratinocyte assay: IL-18 production by NCTC2544 to determine the skin sensitizing capacity and epidermal equivalent assay to determine sensitizer potency. *Toxicol In Vitro* 27, 1135-1150. <http://dx.doi.org/10.1016/j.tiv.2012.06.004>

#### **Acknowledgements**

Dr. A. Penninks, Dr. F. Kuper and M. van den Wijngaarden (TNO, Zeist The Netherlands) selected, coded and shipped the chemicals in this pre-validation study. Dr. S. Casati of EURL-ECVAM is acknowledged for her help in designing the study. Dr D.A. Basketter is acknowledged for discussions on *in vivo* human data for NOEL and the Human Category Scale. C. Bernhard Madsen (National Food Institute, Technical University of Denmark) and J. Ezendam (RIVM, National Institute for Public Health and the Environment, Bilthoven, The Netherlands) are acknowledged for useful discussions throughout this study.

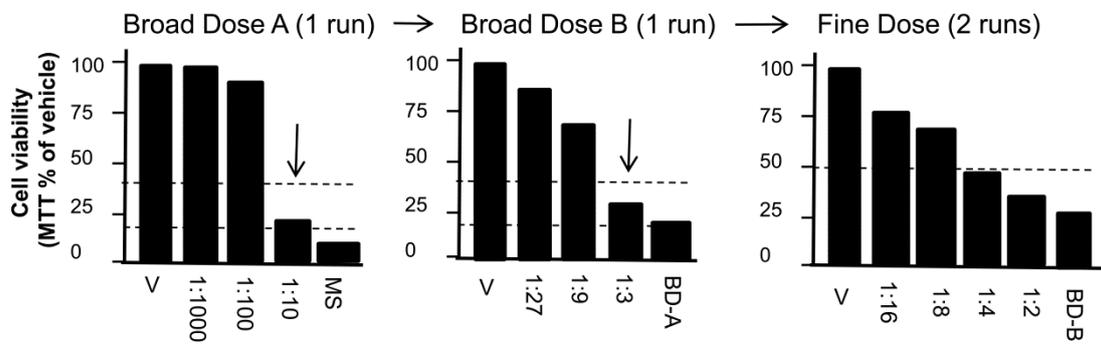
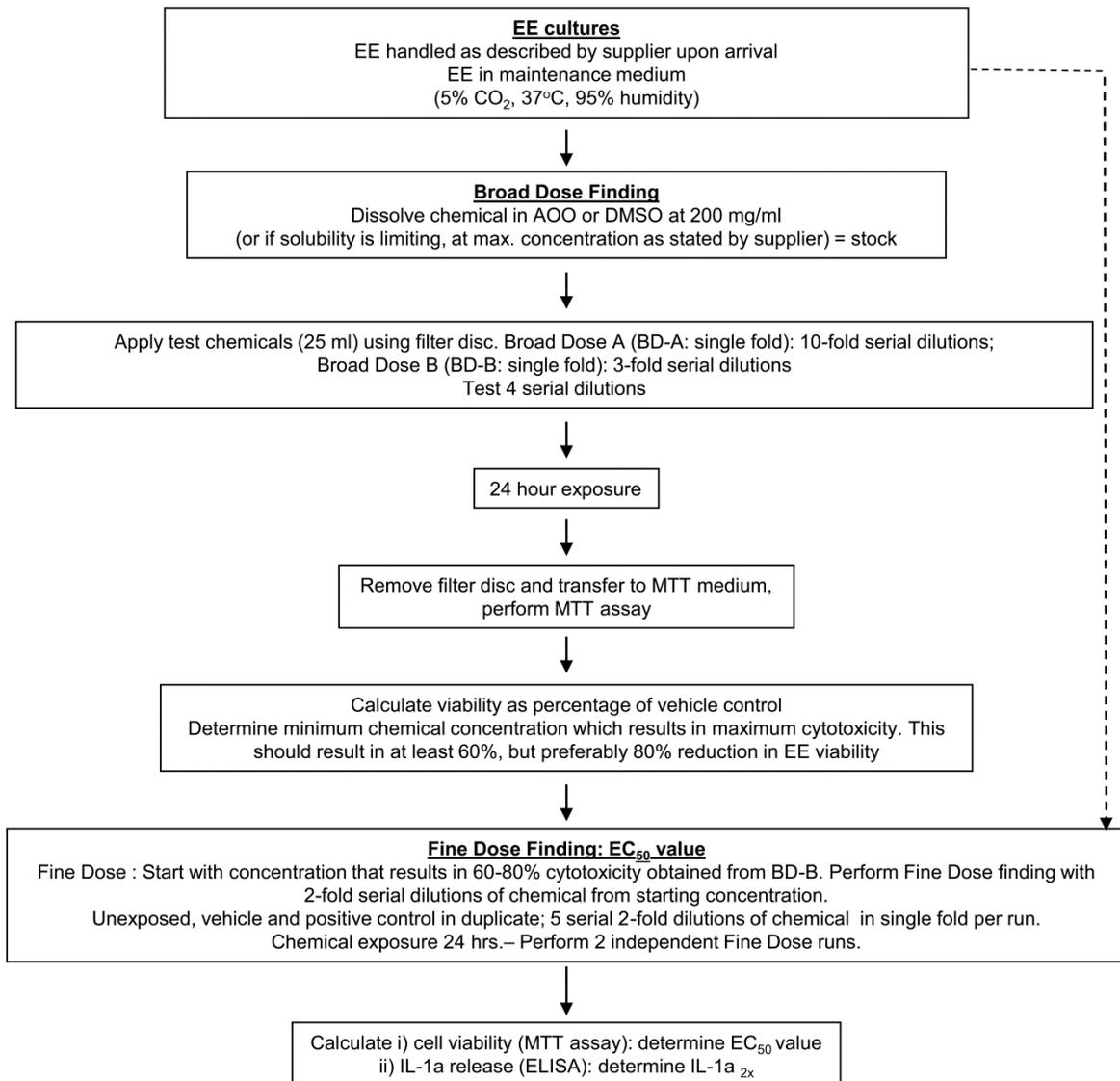
The study was financed by the Dutch Government (ZonMw programme DBIII -114011015), the FP6 project *Sens-it-iv* (LSHB-CT-2005-018681) and the participating laboratories: VUMC, HU, DiSFeB and BASF.

#### **Correspondence to**

Susan Gibbs, PhD  
Department of Dermatology  
Room 2BR-028  
VU University Medical Center  
De Boelelaan 1117  
1081 HV Amsterdam  
Phone: +31 20 444 2815  
Fax: +31 20 444 2816  
e-mail: S.Gibbs@vumc.nl

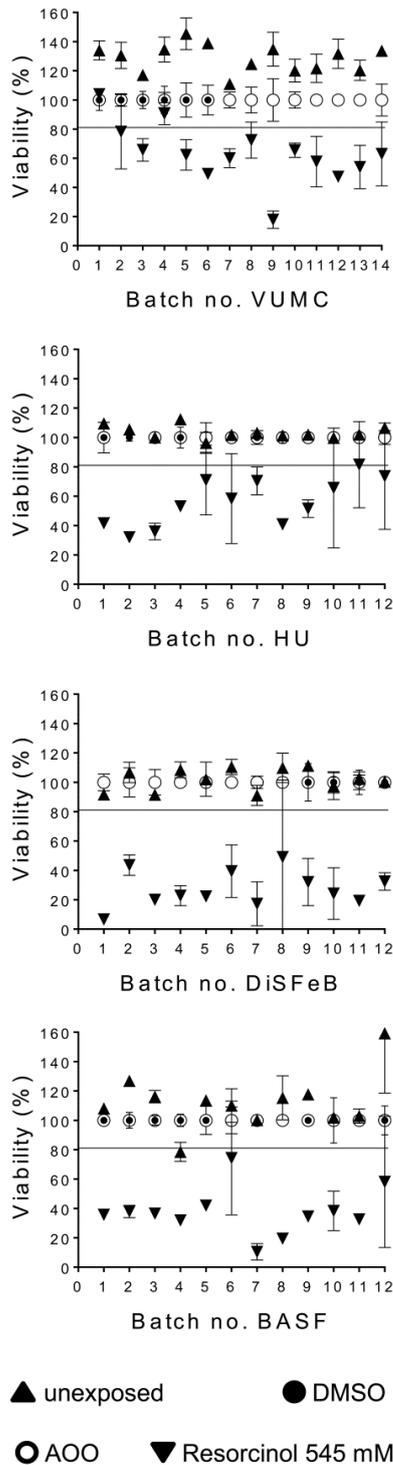
**Figures**

**Fig. 1: Flow diagram for the pre-validation study illustrating the method used for chemical exposure, Broad Dose A and B finding and Fine Dose finding**



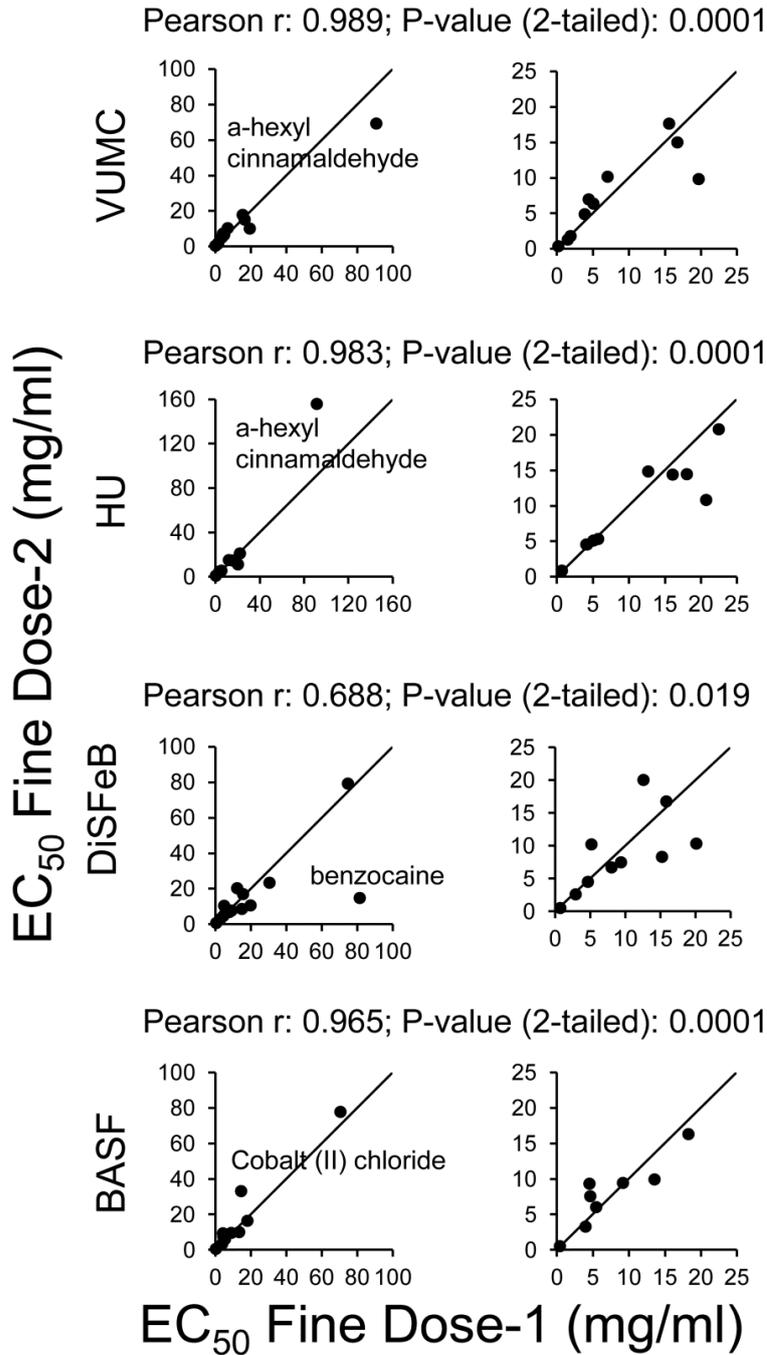
**Fig. 2: Individual viability values for unexposed, vehicle exposed (1 % DMSO or AOO 4:1) and resorcinol exposed (positive control) epiCS®**

The batch numbers allotted to each laboratory did not correlate between laboratories and therefore the deviations observed between laboratories was not due to the same batch of epiCS®. The black line corresponds to the upper limit of 80 % viability of the positive control exposed EE.



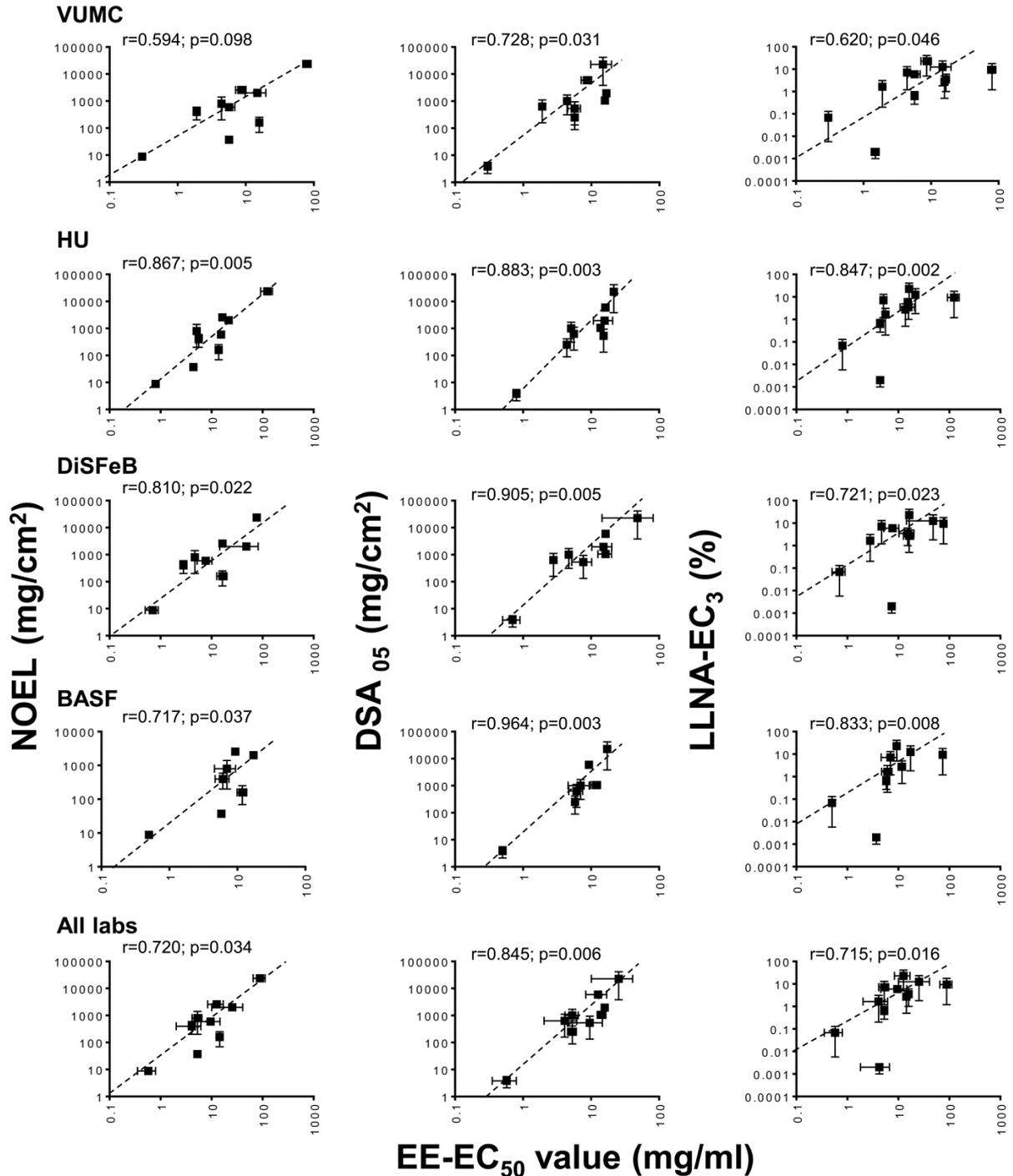
**Fig. 3: Agreement in EC<sub>50</sub> values between the Fine Dose run 1 and run 2**

Dots refer to the values obtained for the different test chemicals in Tab. 3. Only chemicals are included which could be tested in FD. The line corresponds to the equality line. Note: dots falling on the line or near the line indicate a good reproducibility within laboratories. Left side plots show the full range and right side plots show the range to 25 mg/ml.



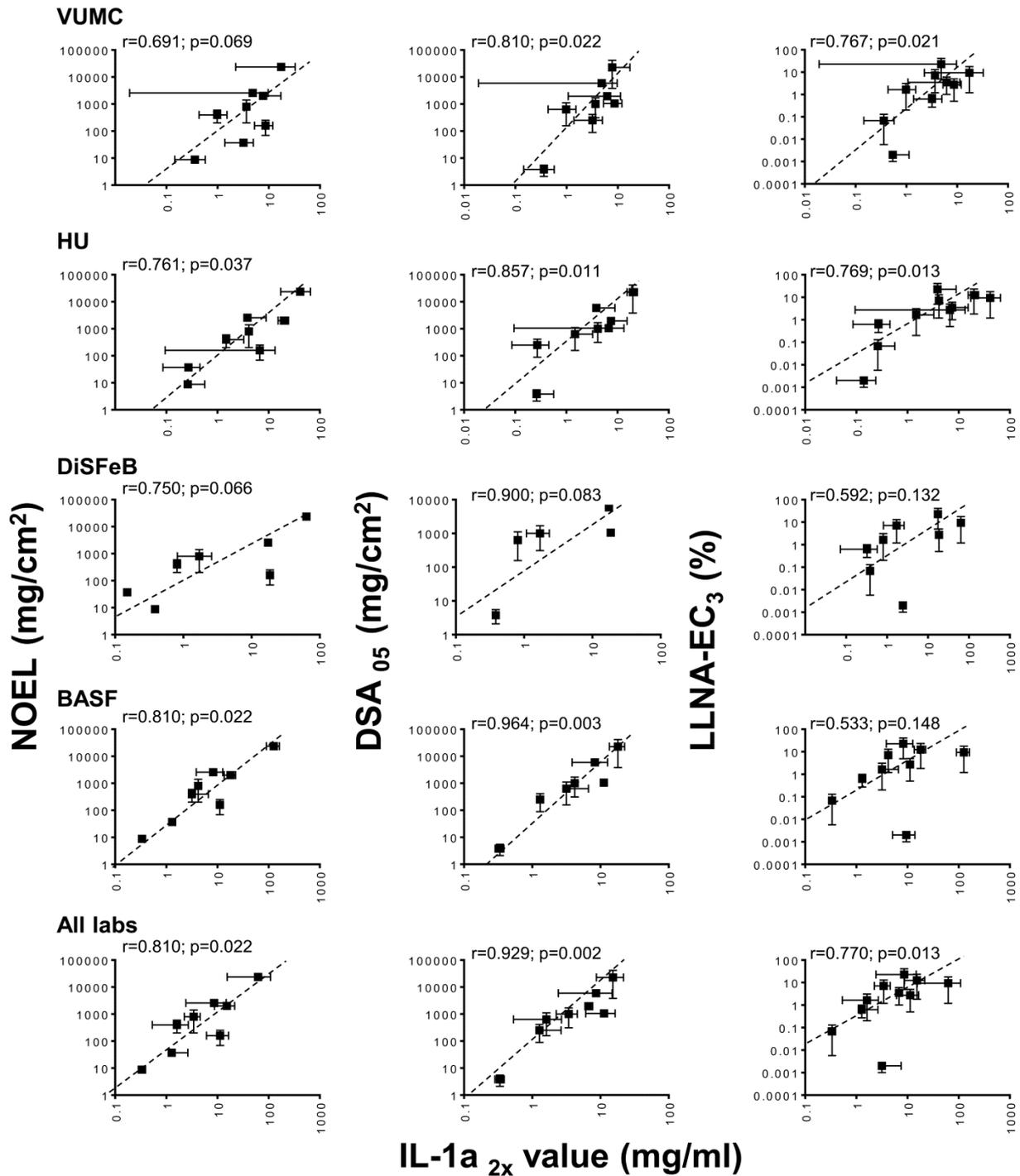
**Fig. 4: Correlation of EE-EC<sub>50</sub> values with human NOEL and DSA<sub>05</sub> data and murine LLNA-EC<sub>3</sub> data**

In vivo data are derived from Tab. 6 and represent the average +/- range of values described in ICCVAM reports (see refs: ICCVAMa and ICCVAMb). *In vitro* data are derived from Tab. 3 and 6; EE-EC<sub>50</sub> values are obtained by linear regression analysis based on viability changes (MTT assay). For individual laboratories, EE-EC<sub>50</sub> data represents the average obtained from the 2 FD runs +/- range of the 2 values. For all laboratories combined, data represents the average of the 4 laboratories +/- SD. Since the data are used to rank chemical potency, Spearman correlation (*r*) and *p* value (two tailed) using all data are shown. Line represents visual line of best fit in which the major deviating chemical oxazolone is excluded from the line for LLNA-EC<sub>3</sub>.



**Fig. 5: Correlation of IL-1 $\alpha_{2x}$  values with human NOEL and DSA<sub>05</sub> data and murine LLNA-EC<sub>3</sub> data**

In vivo data are derived from Tab. 6 and represent average  $\pm$  range of values described in ICCVAM reports (see refs: ICCVAMa and ICCVAMb). *In vitro* data are derived from Tab. 5 and 6 and IL-1 $\alpha_{2x}$  values were obtained by linear regression analysis based on 2-fold increase in IL-1 $\alpha$  release into culture supernatants. For individual laboratories, IL-1 $\alpha_{2x}$  data represents the average obtained from the 2 FD runs  $\pm$  range of the 2 values. For all laboratories combined, data represents the average of the 4 laboratories  $\pm$  SD. Since the data are used to rank chemical potency, Spearman correlation ( $r$ ) and  $p$  value (two tailed) using all data are shown. Line represents visual line of best fit in which the major deviating chemical oxazolone is excluded from the line for LLNA-EC<sub>3</sub>.



**Table 1: Chemical information – human category, NOEL ( $\mu\text{g}/\text{cm}^2$ ), DSA<sub>05</sub> ( $\mu\text{g}/\text{cm}^2$ ), LLNA-EC<sub>3</sub> (%).**

Chemical (CAS N°)	Human Category	Human NOEL ( $\mu\text{g}/\text{cm}^2$ )	Human DSA <sub>05</sub> ( $\mu\text{g}/\text{cm}^2$ )	LLNA-EC <sub>3</sub> (%)
<b>Extreme</b>				
Oxazolone (15646-46-5)	ND	ND	ND	0.001 - 0.003
1-chloro-2,4-dinitrobenzene (97-00-7)	1	8.8	2.1 - 5.5	0.0058 - 0.131
p-Phenylenediamine (106-50-3)	1	10	6.9 - 345	0.001 - 2.2
<b>Strong</b>				
Cobalt (II) chloride (7646-79-9)	2	ND	172 - 453	0.4 – 0.8
Formaldehyde (50-00-0)	2	37	89 - 411	0.27 - 0.99
Cinnamaldehyde (104-55-2)	2	200, 400, 591	157 - 1111	0.2 - 3.1
Phenyl acetaldehyde (122-78-1)	ND	591	133 - 938	3 – 8.8
Isoeugenol (97-54-1)	2	69, 250	775 - 1333	0.5 – 5.0
<b>Moderate</b>				
Citral (5392-40-5)	3	200, 779, 1400	310 - 1691	1.2 – 13.0
Eugenol (97-53-0)	3	1938, 3200, 1938	5926	4.9 – 40.9
2-mercaptobenzothiazole (149-30-4)	3	ND	1642 - 2269	1.0 - 6
<b>Weak</b>				
Benzocaine (94-09-7)	4	2000	3831 - 41667	1.8 - 37
$\alpha$ -hexylcinnamaldehyde (101-86-0)	5	23622	ND	1.2 – 17.6

The chemicals are listed according to their potency values obtained from a combined assessment of all data available from the human category scale, human NOEL, human DSA<sub>05</sub> and murine LLNA-EC<sub>3</sub> experiments. When human and murine data were conflicting or limited, the human data were prioritized in the ranking above murine data.

Human Category Scale: 1 = Extensive evidence of contact allergy in relation to degree of exposure and size of exposed population; 2 = A frequent cause of contact allergy, but of less significance compared with induction of skin sensitization in a HRIPT category 1; 3 = A common cause of contact allergy, perhaps requiring higher exposure compared with category 2; 4 = Infrequent cause of contact allergy in relation to level of exposure; 5 = A rare cause of contact allergy except perhaps in special circumstances (Basketter et al., 2014 In Press).

Human NOEL ( $\mu\text{g}/\text{cm}^2$ ) = no observed effect level; all available data for NOEL is shown.

Human DSA<sub>05</sub> ( $\mu\text{g}/\text{cm}^2$ ) = induction dose per skin area (DSA) that produces a positive response in 5% of the tested population.

The LLNA-EC<sub>3</sub> values are expressed as % according to Basketter et al. (1999): potency classification is based on the mathematical estimation of the concentration of chemical necessary to obtain a threshold positive response (SI=3); this is termed the EC<sub>3</sub> value. Chemicals with an EC<sub>3</sub> value (%)  $\geq 10$  to  $\leq 100$  are classified as weak,  $\geq 1$  to  $< 10$  moderate,  $\geq 0.1$  to  $< 1$  strong,  $< 0.1$  extreme.

In vivo data represents Cobalt (II) Sulphate whereas in the EE potency assay Cobalt (II) Chloride was tested.

ND: indicates no data available;

References: Due to the extreme amount of LLNA data available a range of values is shown which was obtained from ICCVAM report Annex II-1 (see ref. ICCVAMa).

For NOEL and DSA<sub>05</sub> references, see Basketter et al., 2014 In Press and ICCVAM report Annex II-2 (see ref. ICCVAMb).

A full list of references for this table is available in the supplementary table information.

**Table 2: Chemical information: vehicle, maximum solubility, starting concentrations used in Broad Dose and Fine Dose finding**

Chemical	Vehicle <sup>a</sup>	Max. sol. (mg/ml)	Start conc. BD-A <sup>b</sup> (mg/ml)	Start conc. BD-B <sup>b</sup> (mg/ml)	Start conc. FD (mg/ml)			
					VUMC	HU	DISFEB	BASF
<b>Extreme</b>								
Oxazolone	AOO	200	200	20-200	2	20	20	22
1-chloro-2,4-dinitrobenzene	AOO	200	200	2-200	0.67	2	2	0.6
p-Phenylenediamine	AOO	50	50	NT-50	NT	Not reliable	50	NT
<b>Strong</b>								
Cobalt (II) chloride	DMSO	125	125	NT-125	NT	Not reliable	Not reliable	42
Formaldehyde	DMSO	200	200	0.2-20	6.67	6.7	Not reached	6.7
Cinnamaldehyde	AOO	200	200	20	7	6.7	6.67	6.67
Phenyl acetaldehyde	AOO	200	200	NT-20	7	20 <sup>rep</sup>	6.67	NT
Isoeugenol	AOO	200	200	20-200	20	66.7	20	22
<b>Moderate</b>								
Citral	AOO	200	200	20	6.67	6.7	6.67	6.67 <sup>rep</sup>
Eugenol	AOO	200	200	20	20	20	20	20
2-mercaptobenzothiazole	AOO	100	100	NT-100	11	100	100	NT
<b>Weak</b>								
Benzocaine	AOO	200	200	200	22	200	200	22
α-hexylcinnamaldehyde	AOO	200	200	200	67	200	200	200

Start concentrations for BD-B and FD are the chemical concentrations which result in > 60 %, preferably > 80 % reduction in EE viability compared to vehicle exposed EE in the prior run (BD-A and BD-B respectively).

<sup>a</sup> The vehicles used in this study for dissolving chemicals before applying topically to EE. DMSO = 1 % DMSO in culture medium; AOO = acetone: olive oil (4:1)

<sup>b</sup> Start chemical concentration used in the different laboratories

NT: not tested in BD-B in some laboratories as no starting concentration was obtained from BD-A.

<sup>Rep</sup> BD-B run repeated because all concentrations in FD-1 resulted in > 50 % viability (HU, BASF).

Not reliable: result not reliable due to colour interference with the MTT assay

**Table 3: Chemical concentrations (mg/ml) which resulted in 50 % reduction in EE viability (EC<sub>50</sub>) compared to vehicle treated EE**

Chemical	Run	VUMC	HU	DiSFeB	BASF
<b>Extreme</b>					
Oxazolone	1	1.6	4.5	6.7	4.1
	2	1.3	4.2	8.1	3.3
1-chloro-2,4-dinitrobenzene	1	0.3	0.8	0.8	0.5
	2	0.3	0.8	0.5	0.5
p-Phenylenediamine	1	NT	NT	30.7	NT
	2	NT	NT	23.1	NT
<b>Strong</b>					
Cobalt (II) chloride	1	NT	NT	NT	14.7
	2	NT	NT	NT	32.9
Formaldehyde	1	5.2	4.3	NT	5.6
	2	6.3	4.5	NT	6.0
Cinnamaldehyde	1	2.0	5.8	3.0	4.7
	2	1.8	5.3	2.5	7.5
Phenylacetaldehyde	1	4.5	16.1	5.2	NT
	2	6.9	14.4	10.2	NT
Isoeugenol	1	14.9	12.8	12.6	13.6
	2	16.8	14.8	20	9.9
<b>Moderate</b>					
Citral	1	4.0	5.1	4.8	4.6
	2	4.8	5.0	4.5	9.3
Eugenol	1	10.1	18.1	16.0	9.2
	2	7.1	14.4	16.7	9.4
2-mercaptobenzothiazole	1	NR	All >50%	20.2	NT
	2	NR	20.8	10.7	NT
	1R	15.6	All >50%		NT
	2R	17.6	10.8		NT
<b>Weak</b>					
Benzocaine	1	9.8	22.5	81.7	18.3
	2	19.7	20.7	14.5	16.3
α-hexylcinnamaldehyde	1	69.0	91.8	74.9	70.7
	2	91.0	155.6	78.8	77.7

Results are shown from two independent fine dose experiments (1, 2) with the exception of 2-mercaptobenzothiazole where the 2 runs were repeated (1R, 2R) due to inconclusive data in VUMC and HU laboratories. Areas with dark grey background represent chemicals from which EC<sub>50</sub> <7 mg/ml and areas with light grey background represent chemicals from which ≥ 7 mg/ml.

NT: chemical not tested in FD as no EC<sub>60</sub> concentration was obtained in BD-B; NR = EC<sub>50</sub> value not reached in FD; All >50%: all concentrations in FD resulted in more than 50% reduction in viability so no EC<sub>50</sub> could be obtained. For DiSFeB isoeugenol FD2: an EC<sub>50</sub> value was not obtained and therefore the maximum tested FD concentration (20 mg/ml) identified from BD-B (>EC<sub>60</sub>) is used as the run was not repeated within the study.

**Table 4. Predictive capacity of EE potency for each laboratory based on an EC<sub>50</sub> cut-off value of <7 mg/ml for strong/ extreme sensitizers and ≥ 7 mg/ml for weak/ moderate sensitizers.**

Reference result for chemical	VUMC		HU		DiSFeB		BASF		Average
	E/S < 7	M/W ≥ 7	E/S < 7	M/W ≥ 7	E/S < 7	M/W ≥ 7	E/S < 7	M/W ≥ 7	
strong /extreme chemicals	<u>5</u>	1	<u>4</u>	2	<u>2</u>	2 <sup>(+2)</sup>	<u>3</u>	2 <sup>(+1)</sup>	
weak/ moderate chemicals	1	<u>4</u>	1	<u>4</u>	1	<u>4</u>	0 <sup>(+1)</sup>	<u>3</u>	
Number of chemicals tested	6	5	5	6	3	6 <sup>(+2)</sup>	3 <sup>(+1)</sup>	5 <sup>(+1)</sup>	
<b>Sensitivity (%)</b>	<b>83</b>		<b>66</b>		<b>50 (33)</b>		<b>60 (50)</b>		<b>69 (58)</b>
<b>Specificity (%)</b>	<b>80</b>		<b>80</b>		<b>80</b>		<b>100 (75)</b>		<b>84 (74)</b>
<b>Accuracy (%)</b>	<b>82</b>		<b>73</b>		<b>65 (56)</b>		<b>80 (63)</b>		<b>77 (66)</b>
<b>Within lab reproducibility</b>									<b>All runs</b>
Same predictions	13/13		12/13		10/13		11/13		
% concordance	100		92		77		85		
<b>Between lab reproducibility</b>									
Same predictions									5/13 (10/13)
% concordance									35 (77)

Data are based on results obtained for the total number of chemicals tested per laboratory for all chemicals from which an EC<sub>50</sub> value could be obtained (see Table 3). Where ambiguous results from the 2 independent runs were obtained for a single chemical, the result was neither correct nor incorrect (non-conclusive) and are indicated as worst case scenario as follows: <sup>+1</sup> = +1 ambiguous chemical; <sup>+2</sup> = +2 ambiguous chemicals. Bold underlined numbers indicate the number of chemicals showing correct potency classification according to in vivo data shown in Tab. 1. E/S <7 = extreme / strong sensitizer with EC<sub>50</sub> cut-off value <7 mg/ml. M/W ≥7 = moderate / weak sensitizer with EC<sub>50</sub> cut-off value of ≥7 mg/ml. Sensitivity = percentage of correctly identified strong/ extreme sensitizers; specificity = percentage of correctly identified weak/ moderate sensitizers; accuracy = average of sensitivity and specificity. In the determination of sensitivity, specificity, and accuracy: no brackets = both ambiguous results and chemicals not tested (no EC<sub>50</sub>) are excluded; with brackets = worst case scenario is shown with incorporation of ambiguous results possibly happening to score negative and exclusion of chemicals not tested (no EC<sub>50</sub>).

Within laboratory reproducibility: number of chemicals having same prediction in FD1 and FD2 is shown; Between laboratory reproducibility: number of chemicals having same FD prediction in all laboratories (without brackets) and between only VUMC and HU (with brackets) is shown.

**Table 5: Chemical concentrations (mg/ml) which result in 2-fold increase in IL-1 $\alpha$  release (IL-1 $\alpha_{2x}$ ) compared to vehicle treated EE.**

Chemical	run	VUMC	HU	DiSFeB	BASF
<b>Extreme</b>					
Oxazolone	1	< 0.42	< 1.25	NR	12.60
	2	0.96	< 1.25	2.44	6.39
1-chloro-2,4-dinitrobenzene	1	0.21	0.48	NR	0.30
	2	0.51	< 1.25	0.38	0.37
p-Phenylenediamine	1	NT	NT	NR	NT
	2	NT	NT	NR	NT
<b>Strong</b>					
Cobalt (II) chloride	1	NT	NT	NT	NR
	2	NT	NT	NT	NR
Formaldehyde	1	4.48	< 0.42	NT	< 1.25
	2	1.92	< 0.42	NT	1.41
Cinnamaldehyde	1	< 1.25	< 0.42	NR	0.70
	2	1.37	2.74	0.81	6.64
Phenyl acetaldehyde	1	NR	NR	NR	NT
	2	NR	NR	NR	NT
Isoeugenol	1	11.03	< 4.17	18.57	NR
	2	6.25	11.43	NR	11.18
<b>Moderate</b>					
Citral	1	3.98	4.33	1.08	4.20
	2	3.36	3.89	2.34	NR
Eugenol	1	8.29	< 1.25	NR	5.14
	2	1.44	7.42	17.47	11.40
2-mercaptobenzothiazole	1	4.25	< 6.25	NR	NT
	2	4.47	16.28	NR	NT
	1R	< 6.25	< 6.25		
	2R	13.72	< 6.25		
<b>Weak</b>					
Benzocaine	1	< 1.39	16.68	NR	14.86
	2	14.44	23.06	NR	21.25
$\alpha$ -hexylcinnamaldehyde	1	< 12.5	58.58	NR	151.61
	2	28.21	24.31	63.70	101.03

Values shown indicate the chemical concentration (mg/ml) obtained from the dose response experiments where a 2-fold increase in IL-1 $\alpha$  release (IL-1 $\alpha_{2x}$ ) was observed. When the IL-1 $\alpha_{2x}$  correlated to a lower chemical concentration to the lowest concentration tested this is shown by the sign < .

NT: chemical not tested in FD as no EE-EC<sub>60</sub> concentration was obtained in BD-B;

NR = IL-1 $\alpha_{2x}$  value not reached in FD.

Results are from two independent FD experiments (1, 2) with the exception of 2-mercaptobenzothiazol where the 2 runs were repeated (1R, 2R) due to inconclusive data in VUMC and HU laboratories.

**Table 6: Comparison of average EE-EC<sub>50</sub> values, average IL-1 $\alpha$ <sub>2x</sub> values, human category, NOEL and DSA<sub>05</sub> and murine LLNA-EC<sub>3</sub> for all sensitizers tested in EE potency assay**

Substance	Human Category	Human NOEL ( $\mu\text{g}/\text{cm}^2$ )	Human DSA <sub>05</sub> ( $\mu\text{g}/\text{cm}^2$ )	LLNA-EC <sub>3</sub> (%)	EE-EC <sub>50</sub> (mg/ml)	IL-1 $\alpha$ <sub>2x</sub> (mg/ml)
<b>Extreme</b>						
Oxazolone	ND	ND	ND	0.001 - 0.003	4.3 +/- 2.4	3.2 +/- 4.4
1-chloro-2,4-dinitrobenzene	1	8.8	2.1 - 5.5	0.0058 - 0.131	0.6 +/- 0.2	0.3 +/- 0.1
p-Phenylenediamine	1	10	6.9 - 345	0.001 - 2.2	-	-
<b>Strong</b>						
Cobalt (II) chloride	2	ND	172 - 453	0.4 - 0.8	-	-
Formaldehyde	2	37	89 - 411	0.27 - 0.99	5.3 +/- 0.8	1.3 +/- 1.4
Cinnamaldehyde	2	200, 400, 591	157 - 1111	0.2 - 3.1	4.1 +/- 2.1	1.6 +/- 1.1
Phenyl acetaldehyde	ND	591	133 - 938	3 - 8.8	9.6 +/- 5.1	3.6 +/- 1.9
Isoeugenol	2	69, 250	775 - 1333	0.5 - 5.0	14.4 +/- 2.1	11.3 +/- 5.2
<b>Moderate</b>						
Citral	3	200, 779, 1400	310 - 1691	1.2 - 13.0	5.3 +/- 1.2	3.4 +/- 1.1
Eugenol	3	1938, 3200, 1938	5926	4.9 - 40.9	12.7 +/- 4.3	8.6 +/- 6.2
2-mercatopbenzothiazole	3	ND	1642 - 2269	1.0 - 6	15.8 +/- 0.7	6.8 +/- 0.9
<b>Weak</b>						
Benzocaine	4	2000	3831 41667	1.8 - 37	25.5 +/- 15.4	15.2 +/- 6.5
$\alpha$ -hexylcinnamaldehyde	5	23622	ND	1.2 - 17.6	88.7 +/- 23.5	62.2 +/- 46.7

Sensitizers values are obtained and ranked as described in Table 1.

EE-EC<sub>50</sub> and IL-1 $\alpha$ <sub>2x</sub> represent values obtained by linear regression analysis, mean +/- SD of combined data obtained from the 4 laboratories. For each laboratory, the average value obtained from the independent runs is used in the analysis.

**Supplementary Table 1 with full references:**

**Chemical information: Human Category, NOEL ( $\mu\text{g}/\text{cm}^2$ ), DSA<sub>05</sub> ( $\mu\text{g}/\text{cm}^2$ ), LLNA-EC<sub>3</sub> (%).**

Chemical + CAS	Human Category	Human NOEL ( $\mu\text{g}/\text{cm}^2$ )	Human DSA <sub>05</sub> ( $\mu\text{g}/\text{cm}^2$ )	LLNA-EC <sub>3</sub> (%)	Reference <sup>a</sup>
<b>Extreme</b>					
Oxazolone (15646-46-5)	ND	ND	ND	0.001 - 0.003	1-3
1-chloro-2,4-dinitrobenzene (97-00-7)	1	8.8	2.1 - 5.5	0.0058 – 0.131	1-15
p-Phenylenediamine (106-50-3)	1	10	6.9 - 345	0.001 - 2.2	2, 6,9, 11, 16-20
<b>Strong</b>					
Cobalt (II) chloride (7646-79-9)	2	ND	172 - 453	0.4 – 0.8	18, 20-22
Formaldehyde (50-00-0)	2	37	89 - 411	0.27 - 0.99	9, 11, 18-20, 23-26
Cinnamaldehyde (104-55-2)	2	200, 400, 591	157 - 1111	0.2 - 3.1	6, 11, 18-20, 24, 27-31
Phenyl acetaldehyde (122-78-1)	ND	591	133 - 938	3 – 8.8	2, 11, 24, 25, 31, 32
Isoeugenol (97-54-1)	2	69, 250	775 - 1333	0.5 – 5.0	6, 9, 11, 18, 19, 29, 33,
<b>Moderate</b>					
Citral (5392-40-5)	3	200, 779, 1400	310 - 1691	1.2 – 13.0	2, 11, 18, 27, 31, 34, 35,
Eugenol (97-53-0)	3	1938, 3200, 1938	5926	4.9 – 40.9	1, 3, 6, 9, 18, 19, 36, 37,
2-mercaptobenzothiazole (149-30-4)	3	ND	1642 - 2269	1.0 - 6	2,11, 18, 21, 38, 39
<b>Weak</b>					
Benzocaine (94-09-7)	4	2000	3831 - 41667	1.8 - 37	8, 11, 19, 20, 40, 41, 42
$\alpha$ -hexylcinnamaldehyde (101-86-0)	5	23622	ND	1.2- 17.6	11, 23, 24, 43-46

The chemicals are listed according to their potency values obtained from a combined assessment of all data available from the human category scale, human NOEL, human DSA<sub>05</sub> and murine LLNA-EC<sub>3</sub> experiments. When human and murine data were conflicting or limited, the human data were prioritized in the ranking above murine data.

Human Category Scale: 1 = Extensive evidence of contact allergy in relation to degree of exposure and size of exposed population; 2 = A frequent cause of contact allergy, but of less significance compared with induction of skin sensitization in a HRIPT category 1; 3 = A common cause of contact allergy, perhaps requiring higher exposure compared with category 2; 4 = Infrequent cause of contact allergy in relation to level of exposure; 5 = A rare cause of contact allergy except perhaps in special circumstances (Basketter et al., 2014 In Press).

Human NOEL ( $\mu\text{g}/\text{cm}^2$ ) = no observed effect level; all available data for NOEL is shown.

Human DSA<sub>05</sub> ( $\mu\text{g}/\text{cm}^2$ ) = induction dose per skin area (DSA) that produces a positive response in 5% of the tested population.

The LLNA-EC<sub>3</sub> values are expressed as % according to Basketter et al. (1999): potency classification is based on the mathematical estimation of the concentration of chemical necessary to obtain a threshold positive response (SI=3); this is termed the EC<sub>3</sub> value. Chemicals with an EC<sub>3</sub> value (%)  $\geq 10$  to  $\leq 100$  are classified as weak,  $\geq 1$  to  $< 10$  moderate,  $\geq 0.1$  to  $< 1$  strong,  $< 0.1$  extreme.

In vivo data represents Cobalt (II) Sulphate whereas in the EE potency assay Cobalt (II) Chloride was tested.

ND: indicates no data available;

<sup>a</sup>**References:** For LLNA data, a range of values is shown which was obtained from ICCVAM report Annex II-1 (see ref. ICCVAMa). For NOEL and DSA<sub>05</sub> data, references were obtained from Basketter et al., 2014 In Press and ICCVAM report Annex II-2 (see ref. ICCVAMb).

The full list of references for this table are:

1) Gerberick et al., 1992; 2) Gerberick et al., 2005; 3) Loveless et al., 1996; 4) Betts et al., 2006; 5) Kimber et al., 1995; 6) Basketter et al., 2007; 7) Montelius et al., 1994; 8) van Och et al., 2000; 9) Kimber et al., 1991; 10) Basketter et al., 1997; 11) Basketter et al., 1014 In Press; 12) Friedmann et al., 1983; 13) Rees et al., 1989, 14) Basketter and Scholes, 1992; 15) NTP. 1997; 16) White et al., 2006; 17) Warbrick et al., 1999; 18) Schneider and Akkan. (2004); 19) Marzulli and Maibach, 1974; 20) Kligman, 1966a; 21) Kligman, 1966b; 22) Ikarashi et al 1992; 23) Basketter et al., 2005; 24) Basketter and Kimber, 2001; 25) Ryan et al., 2000; 26) Ryan et al., 2002; 27) RIFM 2007; 28) Kimber et al., 1989; 29) Wright et al., 2001; 30) Danneman et al., 1983; 31) Opdyke, 1979; 32) Basketter et al., 2002; 33) Basketter and Cadby, 2004; 34) Ashby et al., 1995; 35) Steltenkamp et al., 1980; 36) Lalko et al., 2004; 37) Kimber and Weisenberger, 1991; 38) Basketter et al., 1993; 39) Scholes et al., 1992; 40) Basketter et al., 1995; 41) Kimber et al., 1989; 42) Griem et al., 2003; 43) Basketter et al., 1999; 44) Dearman et al., 2001; 45) Gamer et al 2008; 46) ECPA 2007.

### Supplementary Reference List

- Ashby, J., Basketter, D. A., Paton, D. et al. (1995). Structure activity relationships in skin sensitization using the murine local lymph node assay. *Toxicology*. 103, 177-194.
- Basketter, D.A., Alépée, N., Ashikaga, T. et al., (2014). Categorization of Chemicals According to Their Relative Human Skin Sensitizing Potency. *Dermatitis* 25(1):11-21
- Basketter, D. A. and Cadby, P. (2004). Reproducible prediction of contact allergenic potency using the local lymph node assay. *Contact Dermatitis*. 50, 15-17.
- Basketter, D. A., Clapp, C., Jefferies, D. et al. (2005). Predictive identification of human skin sensitization thresholds. *Contact Dermatitis*. 53, 260-267.
- Basketter, D. A., Dearman, R. J., Hilton, J. et al. (1997). Dinitrohalobenzenes: evaluation of relative skin sensitization potential using the local lymph node assay. *Contact Dermatitis*. 36, 97-100.
- Basketter, D. A., Evans, P., Gerberick, G. F. et al. (2002). Factors affecting thresholds in allergic contact dermatitis: safety and regulatory considerations. *Contact Dermatitis*. 47, 1-6.
- Basketter, D. A., Gerberick, F., Kimber, I. (2007). The local lymph node assay and the assessment of relative potency: status of validation. *Contact Dermatitis*. 57, 70-75.
- Basketter, D. A. and Kimber, I. (2001). Predictive testing in contact allergy: facts and future. *Allergy*. 56, 937-943.
- Basketter, D. A., Lea, L. J., Dickens, A. et al. (1999). A comparison of statistical approaches to the derivation of EC3 values from local lymph node assay dose responses. *J Appl Toxicol*. 19, 261-266.
- Basketter, D. A. and Scholes, E. W. (1992). Comparison of the local lymph node assay with the guinea-pig maximization test for the detection of a range of contact allergens. *Food Chem Toxicol*. 30, 65-69.
- Basketter, D. A., Scholes, E. W., Wahlkvist, H. et al. (1995). An evaluation of the suitability of benzocaine as a positive control skin sensitizer. *Contact Dermatitis*. 33, 28-32.
- Basketter, D. A., Selbie, E., Scholes, E. W. et al. (1993). Results with OECD recommended positive control sensitizers in the maximization, Buehler and local lymph node assays. *Food Chem Toxicol*. 31, 63-67.
- Betts, C. J., Dearman, R. J., Heylings, J. R. et al. (2006). Skin sensitization potency of methyl methacrylate in the local lymph node assay: comparisons with guinea-pig data and human experience. *Contact Dermatitis*. 55, 140-147.
- Danneman, P. J., Booman, K. A., Dorsky, J. et al. (1983). Cinnamic aldehyde: a survey of consumer patch-test sensitization. *Food Chem Toxicol*. 21, 721-725.
- Dearman, R. J., Wright, Z. M., Basketter, D. A. et al. (2001). The suitability of hexyl cinnamic aldehyde as a calibrant for the murine local lymph node assay. *Contact Dermatitis*. 44, 357-361.
- Friedmann, P. S., Moss, C., Shuster, S. et al. (1983). Quantitative relationships between sensitizing dose of DNCB and reactivity in normal subjects. *Clin Exp Immunol*. 53, 709-715.
- ECPA. (2007). Hexyl Cinnamaldehyde (HCA), Potassium Dichromate (PDC) and Formaldehyde (FRM) Evaluation of Potential Dermal Sensitization in the Local Lymph Node Assay in the Mouse. Study SA 06231. Report by Bayer.
- Gamer AO, Nies E, Vohr HW. (2008). Local lymph node assay (LLNA): comparison of different protocols by testing skin-sensitizing epoxy resin system components. *Regul Toxicol Pharmacol* 52(3): 290-298.
- Gerberick, G. F., House, R. V., Fletcher, E. R. et al. (1992). Examination of the local lymph node assay for use in contact sensitization risk assessment. *Fundam Appl Toxicol*. 19, 438-445.
- Gerberick, G. F., Ryan, C. A., Kern, P. S. et al. (2005). Compilation of historical local lymph node data for evaluation of skin sensitization alternative methods. *Dermatitis*. 16, 157-202.
- Griem P, Goebel C, Scheffler H. (2003). Proposal for a risk assessment methodology for skin sensitization based on sensitization potency data. *Regul Toxicol Pharmacol* 38(3): 269-290.
- ICCVAM (2013a): Comparative LLNA, Guinea Pig, and Human Data Used in the Performance Evaluation. Annex II-1. LLNA Data for 196 Substances Used for the Evaluation of Skin Sensitization Potency [http://iccvam.niehs.nih.gov/docs/immunotox\\_docs/LLNA-pot/3b-AppC-BRD-AnnexII-1.pdf](http://iccvam.niehs.nih.gov/docs/immunotox_docs/LLNA-pot/3b-AppC-BRD-AnnexII-1.pdf). Last assessed 11 december 2013.

- ICCVAM (2013b): Human Data for LLNA Potency Evaluation. Annex II-2. [http://iccvam.niehs.nih.gov/docs/immunotox\\_docs/LLNA-pot/3b-AppC-BRD-AnnexII-2.pdf](http://iccvam.niehs.nih.gov/docs/immunotox_docs/LLNA-pot/3b-AppC-BRD-AnnexII-2.pdf). Last assessed 11 december 2013.
- Ikarashi Y, Ohno K, Tsuchiya T, Nakamura A. (1992). Differences of draining lymph node cell proliferation among mice, rats and guinea pigs following exposure to metal allergens. *Toxicology* 76(3): 283-292.
- Kimber, I., Hilton, J., Botham, P. A. et al. (1991). The murine local lymph node assay: results of an inter-laboratory trial. *Toxicol Lett.* 55, 203-213.
- Kimber, I., Hilton, J., Dearman, R. J. et al. (1995). An international evaluation of the murine local lymph node assay and comparison of modified procedures. *Toxicology.* 103, 63-73.
- Kimber, I., Hilton, J., Weisenberger, C. (1989). The murine local lymph node assay for identification of contact allergens: a preliminary evaluation of in situ measurement of lymphocyte proliferation. *Contact Dermatitis.* 21, 215-220.
- Kimber I, Weisenberger C. (1991). Anamnestic responses to contact allergens: application in the murine local lymph node assay. *J Appl Toxicol* 11(2): 129-133.
- Kligman, A. M. (1966a). The identification of contact allergens by human assay. 3. The maximization test: a procedure for screening and rating contact sensitizers. *J Invest Dermatol.* 47, 393-409.
- Kligman, A. M. (1966b). The identification of contact allergens by human assay. II. Factors influencing the induction and measurement of allergic contact dermatitis. *J Invest Dermatol.* 47, 375-392.
- Lalko, J., Isola, D., Api, A. M. (2004). Ethanol and diethyl phthalate: vehicle effects in the local lymph node assay. *Int J Toxicol.* 23, 171-177.
- Loveless, S. E., Ladics, G. S., Gerberick, G. F. et al. (1996). Further evaluation of the local lymph node assay in the final phase of an international collaborative trial. *Toxicology.* 108, 141-152.
- Marzulli, F. N. and Maibach, H. I. (1974). The use of graded concentrations in studying skin sensitizers: experimental contact sensitization in man. *Food Cosmet Toxicol.* 12, 219-227.
- Montelius, J., Wahlkvist, H., Boman, A. et al. (1994). Experience with the murine local lymph node assay: inability to discriminate between allergens and irritants. *Acta Derm Venereol.* 74, 22-27.
- NTP. (1997). Assessment of Contact Hypersensitivity to Dinitrochlorobenzene, Potassium Dichromate and Methyl Salicylate in BALB16 Female Mice. Research Triangle Park, NC: National Institute of Environmental Health Sciences
- Opdyke, D. L. (1979). Monographs on fragrance raw materials. *Food Cosmet Toxicol.* 17, 241-275.
- Rees, J. L., Friedmann, P. S., Matthews, J. N. (1989). Sex differences in susceptibility to development of contact hypersensitivity to dinitrochlorobenzene (DNCB). *Br J Dermatol.* 120, 371-374.
- RIFM. (2007). Data submission to NICEATM. (NICEATM, ed). Woodcliff Lake, NJ.
- Ryan CA, Cruse LW, Skinner RA, Dearman RJ, Kimber I, Gerberick GF. (2002). Examination of a vehicle for use with water soluble materials in the murine local lymph node assay. *Food Chem Toxicol* 40(11): 1719-1725.
- Ryan, C. A., Gerberick, G. F., Cruse, L. W. et al. (2000). Activity of human contact allergens in the murine local lymph node assay. *Contact Dermatitis.* 43, 95-102.
- Scholes, E. W., Basketter, D. A., Sarll, A. E. et al. (1992). The local lymph node assay: results of a final inter-laboratory validation under field conditions. *J Appl Toxicol.* 12, 217-222.
- Schneider K, Akkan Z. (2004). Quantitative relationship between the local lymph node assay and human skin sensitization assays. *Regul Toxicol Pharmacol* 39(3): 245-255
- Steltenkamp, R. J., Booman, K. A., Dorsky, J. et al. (1980). Cinnamic alcohol: A survey of consumer patch-test sensitization. *Food Cosmet Toxicol.* 18, 419-424.
- van Och, F. M., Slob, W., de Jong, W. H. et al. (2000). A quantitative method for assessing the sensitizing potency of low molecular weight chemicals using a local lymph node assay: employment of a regression method that includes determination of the uncertainty margins. *Toxicology.* 146, 49-59.
- Warbrick, E. V., Dearman, R. J., Lea, L. J. et al. (1999). Local lymph node assay responses to paraphenylenediamine: intra- and inter-laboratory evaluations. *J Appl Toxicol.* 19, 255-260.
- White, J. M., Kullavanijaya, P., Duangdeeden, I. et al. (2006). p-Phenylenediamine allergy: the role of Bandrowski's base. *Clin Exp Allergy.* 36, 1289-1293.
- Wright, Z. M., Basketter, P. A., Blaikie, L. et al. (2001). Vehicle effects on skin sensitizing potency of four chemicals: assessment using the local lymph node assay. *Int J Cosmet Sci.* 23, 75-83.