Research Article

Integrated Skin Sensitization Assessment Based on OECD Methods (I): Deriving a Point of Departure for Risk Assessment

Andreas Natsch[1] and George Frank Gerberick[2]

[1]Fragrances S&T, Ingredients Research, Givaudan Schweiz AG, Kemptthal, Switzerland; [2]GF3 Consultancy, LLC, Cincinnati, OH, USA

Abstract

Three guidelines covering key events in the skin sensitization adverse outcome pathway are endorsed by the Organisation for Economic Co-operation and Development (OECD). A recent guideline covers defined approaches (DA) to combine data from these tests for regulatory (sub)classification. The guideline methods provide continuous data that could characterize the sensitization potency on a more granular scale beyond (sub)classifications. We assembled a comprehensive database of *in vitro* and *in vivo* results from OECD guideline tests. Building on a previous approach using regression models, we provide quantitative models using input data from the kinetic direct peptide reactivity assay (kDPRA), the KeratinoSens™ (KS) assay, and the human cell line activation test (h-CLAT) to calculate a point of departure (PoD) in the form of a predicted local lymph node assay (LLNA) EC3 value for use in risk assessment. Predictive models include results from either two or all three assays. Detailed analysis versus *in vivo* data estimates redundancy between different tests and helps guide model selection. All models were tested on a set of case studies selected for their availability of multiple LLNA reference data in the OECD database. The predicted PoDs were within or close to the range of the variation of the historical LLNA data for most of these cases studies, and overall, the models predicted the *in vivo* value with a median fold-misprediction factor of around 2.5. The robustness of the models was characterized by comparing a comprehensive historical database versus the curated dataset provided by the OECD working group on DA.

1 Introduction

The progress achieved in developing new approach methodologies (NAM) for assessing the skin sensitization potential of chemicals over the last two decades has been remarkable. The Organisation for Economic Co-operation and Development (OECD) led the development of the adverse outcome pathway for skin sensitization divided into mechanistic key events (OECD, 2014). Three OECD guidelines have been published that cover these mechanistic events (covalent binding to protein, keratinocyte activation and dendritic cell activation). A total of eight non-animal test methods are approved in OECD TG 442C, 442D and 442E (OECD, 2018a,b, 2021d). However, none of the eight test methods are considered standalone replacements for complete hazard identification or potency determination.

The current focus is to find ways to combine *in vitro*, *in chemico* and *in silico* assessments with read-across predictions from similar chemicals to generate integrated approaches to testing and assessment (IATA) or defined approaches (DA). The OECD has published a new guideline (TG 497) that describes two simple DA for assessing skin sensitization (OECD, 2021a). This strategy of combining NAM data will enhance the vigor of skin sensitization hazard identification calls on chemicals and further address the critical need for predicting chemical potency. Potency prediction is needed for the UN Global Harmonized System (GHS) subclassification of sensitizers into 1A (strong sensitizers) and 1B (other sensitizers). For the specific needs of GHS subclassification into subcategory 1A, the kinetic direct peptide reactivity assay (kDPRA) is accepted as a standalone assay (Natsch et al., 2020; OECD, 2021d). However, assessing potency is also required for conducting next generation risk assessments (NGRA) on new chemical entities for which only non-animal information is available (Api et al., 2020; Bernauer et al., 2021; Dent et al., 2018; Gilmour et al., 2020). Potency prediction for risk assessment needs to be more granular than the GHS subclassification, and it is expected that this granularity can only be provided by combining multiple methods. Such

potency information is critical to ensure any new chemical is safe for exposed workers and consumers.

Multiple studies have addressed the application of quantitative *in silico*, *in chemico* and *in vitro* data from the validated assays for potency prediction (Gilmour et al., 2020; Kleinstreuer et al., 2018; Natsch et al., 2018; OECD, 2021a). A few NAMs have been designed to predict potency, including the SENS-IS® (Cottrez et al., 2015, 2016), the genomic allergen rapid detection (GARD) (Zeller et al., 2017), and the kDPRA (Wareing et al., 2017). DA using data from NAMs have been shared for predicting potency, including a Bayesian network approach (Jaworska et al., 2015), regression models with KeratinoSens™ (KS) and peptide reactivity data (Natsch et al., 2015), artificial neural network model (Hirota et al., 2015, 2018), and integrated use of human cell line activation test (h-CLAT), DPRA and DEREK data (Takenouchi et al., 2015). It will be essential for risk assessors and regulators to evaluate these different approaches to identify which DA or individual methods provide an accurate point of departure (PoD) value for conducting sound risk assessments.

One promising approach based on generating linear regression models has been published using KS and kinetic peptide reactivity data to provide a predicted EC3 as a PoD (Natsch et al., 2015, 2018). Predicting an EC3 value offers the advantage of generating continuous potency values compared to predicting a chemical potency class (Cottrez et al., 2015; Jaworska et al., 2015; Zang et al., 2017; Zeller et al., 2017). It also provides the opportunity to manage uncertainty using statistical tools based on knowledge of the accuracy of the prediction. Such uncertainty could be factored in to refine the PoD value for conducting a skin sensitization risk assessment. The determination of potency has been primarily dependent on the use of the LLNA (Loveless et al., 2010; OECD, 2010), which has long been considered the "gold standard" for potency assessment because it yields quantitative data suitable for a dose-response evaluation. An alternative, non-animal approach is urgently needed.

The previous work on regression models (Natsch et al., 2015) used kinetic rate constants generated with the Cor1-C420 assay (Natsch and Gfeller, 2008). In this paper, updated linear regression models based on data from OECD validated methods (kDPRA, KS and/or h-CLAT) are described for predicting a PoD value for risk assessment purposes. A comprehensive database of 322 chemicals was assembled that contains data from previous papers (Natsch et al., 2015, 2020) merged with the database compiled within the framework of the OECD project on DA (OECD, 2021c). The paper addresses (1) contribution of the *in vitro* parameters for predicting LLNA potency, (2) comparison of prediction models based on an inclusive dataset versus the highly curated OECD dataset, (3) guidance in model selection with multiple input data, (4) comparison of kDPRA and Cor1-Cor420 reactivity data, and (5) case studies on chemicals with multiple LLNA EC3 values. This work further advances the 3Rs for skin sensitization testing as a standardized way to derive a PoD from validated methods is still a missing element in the application of NAM for sensitization assessment.

2 Materials and methods

Data sources of existing data
All *in vivo* and *in vitro* data are from our previous publication (Natsch et al., 2015), from data compilations by Urbisch et al. (2015) and Jaworska et al. (2015), and from the database compiled by the OECD working group on DA (OECD, 2021c). The database (ESM1-1[1]) contains data from the DPRA and kDPRA (OECD TG 442C), KS assay (TG 442D), h-CLAT (TG 442E), LLNA (GL 429) and the Cor1-C420 reactivity assay (Natsch and Gfeller, 2008). All individual parameters are described in ESM1-1[1].

Data transformation and normalization for statistical analysis
All data were log-transformed and normalized as described (Natsch et al., 2015). In case multiple LLNA data were available, we took the geometric mean of the LLNA EC3 values. Data are expressed as pEC3, a logarithmic expression taking molecular weight into account since most *in vitro* data are expressed in molar concentrations and not on a per weight basis as in the LLNA.

$$pEC3 = Log\left(\frac{MW}{EC3}\right)$$

For negative LLNA results, the pEC3 was set to zero, which is the special case for, e.g., molecules with a molecular weight of 100 and an EC3 of 100%. This approach treats all negatives equally, as not all negatives were tested up to 100%. Based on this normalization, the pEC3 spans a range from 0 for non-sensitizers to 4.86 for the potent sensitizer oxazolone.

In the KS assay, the IC50 value (concentration for 50% reduction in cellular viability) and the EC1.5 or EC3 values, i.e., concentration for 1.5- or 3-fold luciferase induction, are determined. Chemicals are tested up to a maximal concentration of 2000 µM. For chemicals with no cytotoxicity at this concentration, the numerical IC50 was set to the arbitrary value of 4000 µM. Similarly, if the luciferase gene was not induced above a given threshold, the EC1.5 or EC3 values were set to 4000 µM. In addition, in the KS prediction model only gene induction observed at non-cytotoxic levels (> 70% viability) is considered relevant. Thus, if 1.5-fold gene induction was only observed at cytotoxic levels,

---

[1] doi:10.14573/altex.2201141s1

Tab. 1: Subsets of chemicals in the dataset used for the different evaluations on the robustness of predictive models

Data subset	n	Data available for all chemicals in a subset	Used for
Full dataset	322	LLNA, KS, Cor1-C420 assay	Overall analysis in ESM4[2]
Full dataset with kDPRA data	203	LLNA, KS, kDPRA, Cor1-C420 assay	Most comprehensive model integrating kDPRA and KS
Full dataset with kDPRA and h-CLAT data	188	LLNA, KS, kDPRA, Cor1-C420 assay, h-CLAT	Models using h-CLAT instead of (or in combination with) KS
OECD LLNA dataset	149	LLNA data from previous database, LLNA data curated by OECD DA group, KS, kDPRA, Cor1-C420 assay, h-CLAT	Comparison of models derived from OECD dataset with models on more comprehensive datasets; evaluation of the impact of data curation
Core set 2015	244	LLNA, KS, Cor1-C420 assay	For direct comparison with the 2015 analysis in ESM4[2]
Core set 2015; subset with kDPRA data	158	LLNA, KS, Cor1-C420 assay, kDPRA	Predictivity with kDPRA on subset of previous publication

then EC1.5 and EC3 were set to 4000 µM. The three parameters were linearized by logarithmic transformation and normalized by multiplication by 1 and the addition of the constant Log(4000).

Thus, as an example, the normalized IC50 is:

$$LogIC50_{norm} = -1 \times Log(IC50_{KS}) + Log(4000)$$

These three normalized KS parameters span a linear range from 0 (no effect at 2000 µM) to 3.61 (effect at lowest tested concentration of 0.98 µM).

In the h-CLAT, four quantitative parameters are determined: The CV75 indicates the concentration for 25% reduction in cellular viability, the EC150 indicates the concentration for 1.5-fold induction of CD86 surface expression, and the EC200 indicates the concentration for 2-fold induction of CD54 surface expression. The MIT reports the smaller of the two values EC150 and EC200. Most chemicals are tested up to 5000 µg/mL, but in some cases only up to 1000 or 2000 µg/mL (Nukada et al., 2012). Data from h-CLAT are reported in µg/mL, and they were thus transformed to µM. Negative values were set to the arbitrary value of 25,000 µM (i.e., 5000 µg/mL for a chemical with a default molecular weight of 200), and normalization was done in the same way. As an example, the normalized MIT value is:

$$LogMIT_{norm} = -1 \times Log(MIT\mu M) + Log(25000)$$

In the kDPRA, the reaction rates are measured at different time points, and the logarithm of the maximal rate in $s^{-1}M^{-1}$ is reported (Log k_{max}). The minimal value for chemicals that are negative in the assay, i.e., reaching less than 13.86% depletion of the cysteine peptide after 24 h is < 3.5. These values were thus normalized according to the following formula:

$$Log\,k_{max\,norm} = Log\,k_{max} + 3.5$$

With these linear transformations, all *in vitro* parameters are set to 0 in case a molecule is non-reactive in kDPRA, non-toxic beyond the selected threshold in the cell-based assays, or it does not induce the cellular markers (luciferase or CD86/CD54 induction), while all positive *in vitro* evidence according to the prediction models of the individual assays has a positive value. These linear transformations of logarithmic data do not distort the data but make reading the regression equations easier.

Vapor pressure was calculated with TIMES SS software (Laboratory of Mathematical Chemistry). Vapor pressure is considered important as chemicals are tested in the LLNA with open application, while evaporation is limited in the *in vitro* assays. The relationship of calculated vapor pressure to evaporation from the LLNA vehicle acetone / olive oil 4:1 (AOO) at the typical skin surface temperature of 32°C was tested experimentally for 37 molecules, and a linear relationship between Log(VP_calc) and log half-life of the chemical in AOO was demonstrated as shown in Figure S2 in Natsch et al. (2015). Chemicals with a Log(VP_calc) < 10 Pa were found not to evaporate significantly from AOO within 60 min. Therefore, vapor pressure data were normalized as follows:

$$LogVP_{norm} = Log(VP) - 1$$

and values below zero were set to 0. With this approach, chemicals predicted to evaporate significantly from mouse ears within 60 min have a positive coefficient, whereas the coefficient for chemicals with low volatility (i.e., not significantly evaporating within 60 min) is set to 0.

Statistical analysis
Multiple linear regression against the pEC3 was performed with the MiniTab 15 software. The R^2-values and F- and *p*-values of the correlations are given throughout the manuscript as measures of correlation strengths. The regression equations were then used

---

to predict the most likely pEC3 and EC3 values of the individual chemicals. The ratio between the larger and the smaller values of the measured and predicted EC3 was calculated in each case to give the fold-misprediction.

3 Results

3.1 A comprehensive database of *in vitro* and *in vivo* data

We compiled a comprehensive database on 322 chemicals (Tab. ESM1-1[1]) with data from the OECD-validated LLNA, KS, h-CLAT, DPRA and kDPRA. This database also contains physicochemical information and reaction rates measured using a method similar to the kDPRA but with the more reactive peptide Cor1-C420, as reported earlier. This database is a merger of our previous database (Natsch et al., 2015) with the database used to validate the kDPRA (Natsch et al., 2020) and the database put together within the frame of the OECD project on developing a DA for skin sensitization in the Series on Testing and Assessment No. 336, Annex 2 (OECD, 2021c). It also contains h-CLAT and DPRA data from Urbisch et al. (2015) and Jaworska et al. (2015). Table 1 indicates the different subsets of this database.

In the analysis below, we present regression analysis on these different datasets, and we provide a prediction spreadsheet (ESM2[3]) integrating the key equations, which can be used to estimate a predicted LLNA pEC3 and an EC3 value as a PoD based on *in vitro* input parameters using the different regression equations.

3.2 A quantitative global regression model integrating kDPRA data with KS data

The database contains kDPRA and KS data for a set of 203 chemicals. For this dataset, a regression model based on the normalized EC1.5 and IC50 from KS, Log k_{max} from the kDPRA and normalized vapor pressure was derived, similar to the global model in the 2015 study (EQ1), by replacing reaction rates with the Cor1-C420 peptide with data from the validated kDPRA.

EQ1 $\quad$ pEC3 = 0.42 + 0.40 × Log $k_{max\ norm}$ + 0.15 × Log EC1.5$_{norm}$ + 0.36 × Log IC50$_{norm}$ − 0.21 × Log VP$_{norm}$

Constant	T = 4.61, $p < 0.0005$
Log $k_{max\ norm}$	T = 9.1, $p < 0.0005$
Log EC1.5$_{norm}$	T = 2.5, $p = 0.013$
Log IC50$_{norm}$	T = 4.47, $p < 0.0005$
Log VP$_{norm}$	T = -3.44, $p = 0.001$
N = 203; S = 0.73; R^2 = 62.0%; R^2 (adj) = 61.3%	

In this equation, k_{max} from kDPRA and IC50 from KS have the highest statistical weight. This equation is derived from the most comprehensive dataset with available kDPRA data. EQ1 is therefore proposed as the key model in the prediction spreadsheet (ESM2[3]) for chemicals with available kDPRA and KS data only.

For comparison, the same regression analysis was performed on the subset of data from the core set used in the 2015 study with kDPRA data (n = 158; EQ2). Similarly, the same model was also calculated on the subset of data with available h-CLAT data. The latter can then be directly (i.e., based on the same chemicals) compared to the models below integrating h-CLAT data (n = 188; EQ3).

EQ2 $\quad$ pEC3 = 0.34 + 0.38 × Log $k_{max\ norm}$ + 0.16 × Log EC1.5$_{norm}$ + 0.42 × Log IC50$_{norm}$ − 0.18 × Log VP$_{norm}$
$\quad$ n = 158; S = 0.72; R^2 = 63.0%; R^2 (adj) = 61.2%

EQ3 $\quad$ pEC3 = 0.40 + 0.40 × Log $k_{max\ norm}$ + 0.14 × Log EC1.5$_{norm}$ + 0.38 × Log IC50$_{norm}$ − 0.21 × Log VP$_{norm}$
$\quad$ n = 188; S = 0.75; R^2 = 58.4%; R^2 (adj) = 57.5%

The resulting equations are very similar to EQ1, indicating that the chemicals added to the core set used previously have no major impact on the overall predictive model and that the model is relatively stable when used with different subsets of the input data.

3.3 A quantitative global regression model integrating kDPRA data with h-CLAT data

Next, a similar model was calculated using only data from the kDPRA and h-CLAT to predict the potency of chemicals with these two sources of data available, thus normalized EC1.5 and IC50 were replaced with the normalized MIT and the CV75 from h-CLAT, both calculated in μM (n = 188; EQ4).

EQ4 $\quad$ pEC3 = 0.18 + 0.36 × Log $k_{max\ norm}$ + 0.21 × Log MIT$_{norm}$ + 0.35 × Log CV75$_{norm}$ − 0.19 × Log VP$_{norm}$

Constant	T = 1.65, $p = 0.102$
Log $k_{max\ norm}$	T = 7.59, $p < 0.0005$
Log MIT$_{norm}$	T = 3.06, $p = 0.004$
Log CV75$_{norm}$	T = 3.69, $p < 0.0005$
Log VP$_{norm}$	T = -3.0, $p = 0.003$
n = 188; S = 0.72; R^2 = 61.1%; R^2 (adj) = 60.3%	

Compared to EQ3, on the same dataset, k_{max} has a similar weight in this equation, and MIT and CV75 can replace the KS data with a very similar overall statistical weight. EQ4 is thus added to the prediction spreadsheet (ESM2[3]) to be used for those chemicals for which only kDPRA and h-CLAT data are available.

3.4 A quantitative global regression model integrating kDPRA data with KS and h-CLAT data

For some chemicals, comprehensive data from all three OECD TG will be available. Thus, the most comprehensive model integrates kDPRA, KS and h-CLAT data (n = 188; EQ5).

Tab. 2: Predictivity on the dataset (n = 188) with KS, h-CLAT and kDPRA data

Model	Input parameters	Fold-misprediction[a] (geomean)	Fold-misprediction (median)	Chemicals > 5-fold under-predicted[b] n, (%)	Chemicals > 10-fold under-predicted n, (%)	Chemicals > 5-fold over-predicted[b] n, (%)	Chemicals > 10-fold over-predicted n, (%)
EQ1	kDPRA, KS	3.3	2.5	33 (18%)	20 (11%)	16 (9%)	7 (4%)
EQ4	kDPRA, h-CLAT	3.2	2.4	30 (16%)	17 (9%)	16 (9%)	7 (4%)
EQ5	kDPRA, KS, h-CLAT	3.1	2.3	35 (19%)	17 (9%)	18 (10%)	6 (3%)
EQ6	KS, h-CLAT	3.5	2.6	33 (18%)	19 (10%)	19 (11%)	8 (4%)
EQ7	KS, h-CLAT	3.4	2.7	31 (16%)	19 (10%)	18 (10%)	6 (3%)
Global model 2015	Cor1-C420 assay, KS	3.5	2.7	39 (21%)	24 (13%)	19 (10%)	4 (2%)

[a] The ratio between the higher and the lower values of the measured and predicted EC3 value. Predicted EC3 > 100% were set to 100%; [b] Under-predicted chemicals: those for which the measured LLNA EC3 is less than the predicted EC3; over-predicted chemicals: those with measured LLNA EC3 greater than the predicted value.

EQ5 $pEC3 = 0.20 + 0.34 \times Log\ k_{max\ norm} + 0.20 \times Log\ MIT_{norm} + 0.09 \times Log\ EC1.5_{norm} + 0.21 \times Log\ CV75_{norm} + 0.11 \times Log\ IC50_{norm} - 0.19 \times Log\ VP_{norm}$

Constant	$T = 1.98, p = 0.080$
Log $k_{max\ norm}$	$T = 6.88, p < 0.0005$
Log $EC1.5_{norm}$	$T = 1.41, p = 0.159$
Log $IC50_{norm}$	$T = 0.94, p = 0.347$
Log MIT_{norm}	$T = 2.74, p = 0.004$
Log $CV75_{norm}$	$T = 1.63, p = 0.119$
Log VP_{norm}	$T = -2.84, p = 0.005$

$n = 188; S = 0.72; R^2 = 61.6\%; R^2\ (adj) = 60.6\%$

The predictive power of EQ5 is only very marginally improved vs EQ3 and EQ4, indicating that with both KS and h-CLAT data, the overall predictivity improves only slightly as compared to the situation where only one cellular assay is available. This has already been observed in our previous comprehensive analysis (Natsch et al., 2020), indicating significant data redundancy between these two cell-based assays. This is especially the case for the cytotoxicity readout. When both IC50 and CV75 are used in EQ5, they contribute additively, but the coefficient and statistical parameters indicate a weak contribution of each parameter. Alternative models using either CV75 or IC50 alone have the same R^2 value (61%), but in both cases, the contribution of the cytotoxicity parameter has more statistical weight ($p = 0.014$ for IC50 and $p = 0.005$ for CV75; regression equations not shown), confirming data redundancy especially for the two cytotoxicity indicators, which is not a surprising finding.

3.5 A quantitative global regression model integrating KS and h-CLAT data in the absence of kDPRA data

Finally, in some cases, chemicals are rated as positive in a DA based on a positive result in the cell-based assays only despite a negative result in the kDPRA and the DPRA. For these and complex extracts or multi-constituent substances with data gaps in the kDPRA due to incompatibility with the assay, a model based on KS and h-CLAT data only may be of interest. This model is presented in Equation 6.

EQ6 $pEC3 = 0.09 + 0.276 \times Log\ MIT_{norm} + 0.22 \times Log\ EC1.5_{norm} + 0.34 \times Log\ CV75_{norm} + 0.06 \times Log\ IC50_{norm} - 0.12 \times Log\ VP_{norm}$

Constant	$T = 0.67, p = 0.50$
Log $EC1.5_{norm}$	$T = 3.22, p = 0.002$
Log $IC50_{norm}$	$T = 0.43, p = 0.671$
Log MIT_{norm}	$T = 3.75, p < 0.0005$
Log $CV75_{norm}$	$T = 2.29, p = 0.025$
Log VP_{norm}	$T = -1.81, p = 0.072$

$n = 188; S = 0.81; R^2 = 51.8\%; R^2\ (adj) = 50.5\%$

All the above models (EQ1, 4 and 5) were also calculated with KS EC3 derived from KS, i.e., the concentration for threefold stimulation of the luciferase signal in KS, instead of EC 1.5, and models with very similar R^2 were obtained (data not shown). However, in the situation without reactivity data, EQ7 using KS EC3 has improved statistical power, with KS EC3 having the highest statistical weight.

EQ7 $pEC3 = 0.202 + 0.222 \times Log\ MIT_{norm} + 0.40 \times Log\ EC3_{norm} + 0.313 \times Log\ CV75_{norm} + 0.023 \times Log\ IC50_{norm} - 0.151 \times Log\ VP_{norm}$

Constant	$T = 1.67, p = 0.098$
Log $EC3_{norm}$	$T = 5.29, p < 0.0005$
Log $IC50_{norm}$	$T = 0.18, p = 0.855$
Log MIT_{norm}	$T = 3.02, p = 0.003$
Log $CV75_{norm}$	$T = 2.28, p = 0.024$
Log VP_{norm}	$T = -2.28, p = 0.023$

$S = 0.77; R^2 = 55.9\%; R^2\ (adj) = 54.7\%$

Tab. 3: Regression coefficients and statistics for predictive models implemented in the Prediction Spreadsheet compared to models derived from the OECD dataset only

Model	Constant	k_{max}	EC1.5	IC50	MIT	CV75	VP_{norm}	R^2	N
EQ1 (KS+kDPRA)	0.42	0.40	0.15	0.36			-0.21	62%	203
KS+kDPRA, OECD set, published LLNA values (EQ 8)	0.45	0.32	0.26	0.25			-0.20	55%	149
KS+kDPRA, OECD set, curated LLNA value (EQ 9)	0.50	0.33	0.21	0.29			-0.25	54%	149
EQ4 (h-CLAT+kDPRA)	0.18	0.36			0.21	0.35	-0.19	61%	188
h-CLAT+kDPRA, OECD set, published LLNA values (EQ10)	0.24	0.31			0.22	0.34	-0.16	57%	149
h-CLAT+kDPRA, OECD set, curated LLNA values (EQ11)	0.28	0.30			0.19	0.37	-0.22	57%	149
EQ5 (KS+h-CLAT+kDPRA)	0.20	0.34	0.09	0.11	0.20	0.21	-0.19	62%	188
KS+h-CLAT+kDPRA, OECD set, published LLNA values (EQ12)	0.23	0.26	0.21	−0.03	0.22	0.19	-0.17	59%	149
KS+h-CLAT+kDPRA, OECD set, curated LLNA values (EQ13)	0.27	0.26	0.15	0.00	0.19	0.25	-0.22	58%	149

These models without kDPRA data have less statistical power, but they would primarily be used either in the case of data gaps for the kDPRA, for chemicals that are negative in the kDPRA, or for chemicals outside of the applicability domain of the kDPRA (see ESM3[4]). These latter chemicals, in most cases, are weak or moderate sensitizers, as the majority of strong sensitizers have a direct peptide reactivity in the kDPRA and DPRA (Natsch et al., 2020).

3.6 Overall predictive capacity of the different models

Equations 1, 4, 5 and 6 (with 7 as an alternative) are implemented in the prediction spreadsheet (ESM2[3]) to predict the potency of new chemicals depending on the set of available *in vitro* data. The overall predictive capacity of these models is summarized in Table 2. For all models, the median of the fold-misprediction is around 2.5-fold, while the geometric mean is around 3.3-fold.

3.7 Models trained on the subset of data in the OECD database

Very recently, the data curated by the OECD expert group on DA was made publicly available (OECD, 2021c), and this database will probably be used as a gold standard in the future. It contains data on a total of 196 chemicals, but a discrete LLNA EC3 value or an unambiguous rating as non-sensitizer by the LLNA is available only for 154 chemicals, as for some chemicals no LLNA data are included or the LLNA data were not considered sufficient to conclude on an EC3 value by the data curation team. For 149 of these 154 chemicals, data are available in our database on kDPRA, KS and h-CLAT. This dataset is useful to investigate two questions:

a) Does the smaller data subset collected by the OECD group lead to significantly different equations compared to those presented above on a more comprehensive set?
b) Do the LLNA EC3 values after the data curation lead to significantly different predictive equations compared to the LLNA data in the previously published databases?

Table 3 provides comparisons of the model parameters as calculated for the chemicals in the OECD database only (n = 149) vs (i) published LLNA data and (ii) curated LLNA data, and these models are compared to the models established on the more comprehensive sets of chemicals (EQ1, EQ4 and EQ5).

Overall, the additional models (EQ8-EQ13) derived from the subset in the OECD database are similar to EQ1, EQ4 and EQ5, integrating the larger datasets of all available data for a given data combination. Especially, only minor differences are observed between the models on the historical and the curated LLNA data (EQ8 vs 9 / EQ10 vs 11 / EQ12 vs 13).

3.8 Practical application: An open spreadsheet for calculating a PoD and case studies

To further facilitate application, here we provide a prediction spreadsheet as ESM2[3]. Users can simply enter the experimental data from the different *in vitro* assays along with the molecular weight and a calculated vapor pressure to calculate a predicted EC3, which can be used as a PoD and, if applying adequate safety factors and taking into account all available information and applicability domain limitations, for a quantitative risk assessment. To illustrate its use, the prediction spreadsheet also contains the input data for two examples, DNCB and cinnamic aldehyde.

---

[4] doi:10.14573/altex.2201141s3

Tab. 4: Case studies: Predictions by the different models for chemicals with multiple LLNA EC3 (n ≥ 5) values in the OECD curated database

	OECD MLLP LLNA EC3[a]	LLNA studies (n)	LLNA EC3 measured range[b]	LLNA EC3 statistical range[c]	Pred. EC3 EQ1[d]	Pred. EC3 EQ4[e]	Pred. EC3 EQ5[f]	Pred. EC3 EQ13[g]	Pred. global model 2015[i]
Aniline	**NC**	14	13.25 - (> 100)	NC	60	52	57	52	> 100
Penicillin G	**31.3**	8	11.2 - 46.5	15.2 - 47.2	> 100	> 100	> 100	> 100	> 100
Hydroxycitronellal	**21.1**	8	18.8 - 33	18.5 - 27.2	18.7	11.3	10.9	8.0	3.5
Geraniol	16.1	6	5.6 - 57	7.6 - 37.5	18.3	14.3	14.2	10.3	5.1
Eugenol	**11.6**	16	3.8 - 16.6	5.8 - 15.4	19.9	6.8	10.4	10.2	14.1
Alpha-hexyl cinnamic aldehyde	**10.8**	29	1.2 - 33.8	4.9 - 17.4	5.9	(25)[h]	17.4	15.0	1.8
Lilial	**8.6**	5	3 - 18.6	3.6 - 16.8	20.5	9.3	12.5	13.9	6.2
Citral	**5.8**	16	1.5 - 26.8	3 - 12.7	9.4	5.0	4.8	3.5	12.6
Formaldehyde	**3.8**	15	0.35 - 14.5	0.6 - 9.6	1.5	0.8	1.0	2.0	4
3-dimethylaminopropylamine	**3.5**	7	1.8 - (> 10)	1.9 - 5.7	40	37	32	27	53
Isoeugenol	**1.3**	31	0.5 - 6.4	0.8 - 3	1.8	(4.6)[h]	4.2	4.1	1.6
Cinnamic aldehyde	**1**	12	0.5 - 3.1	0.7 - 1.8	1.0	0.8	0.8	0.9	1.1
Hydroquinone	**0.19**	20	0.07 - 1.67	0.1 - 0.5	0.9	0.4	0.4	0.4	0.8
PPD	**0.11**	10	0.06 - 0.2	0.07 - 0.17	3.5	1.9	1.7	1.2	0.72
DNCB	**0.054**	20	0.012 - 0.096	0.02 - 0.08	0.18	0.19	0.17	0.24	0.21
Kathon CG	**0.008**	10	0.0049 - 0.063	0.004 - 0.035	0.05	0.05	0.05	0.08	n.d.
Oxazolone	**0.002**	7	0.0011 - 0.0026	0.001 - 0.003	1.5	0.5	0.7	0.9	0.9

[a] LLNA EC3, median-like location parameter derived by OECD group from multiple EC3 values; [b] range between minimal and maximal measured LLNA EC3 in OECD DB; [c] geometric mean and geometric standard deviation were calculated, range around the geometric mean defined by the geometric standard deviation is given; [d] based on KS and kDPRA; [e] based on h-CLAT and kDPRA; [f] based on KS, h-CLAT and kDPRA; trained on comprehensive database (n = 188) and historical LLNA database; [g] based on KS, h-CLAT and kDPRA; trained on curated OECD LLNA data only (n = 149); EQ13: $pEC3 = 0.26 + 0.25 \times Log\ k_{max\ norm} + 0.19 \times Log\ MIT_{norm} + 0.17 \times EC1.5_{norm} + 0.27 \times Log\ CV75_{norm} - 0.02 \times Log\ IC50_{norm} - 0.19 \times Log\ VP_{norm}$; [h] values in brackets: chemicals negative in h-CLAT, EQ4 is not recommended for chemicals negative in h-CLAT; [i] predictive model using the Cor1-C420 assay published previously

For all chemicals, additional spreadsheets in ESM1 (Tab. ESM1-2[1] – ESM2-4[3]) give the individual predictions by the different equations on the different data subsets shown in Table 1. Here we show individual data on key case studies (Tab. 4). We selected all chemicals for which the OECD LLNA database provides n ≥ 5 individual EC3 values. Selecting these chemicals as case studies has two advantages: (i) The overall weight of evidence for the LLNA EC3 is strong as it comes from multiple independent tests and (ii) for these chemicals, we have a good understanding of the variability of the target, i.e., the range of LLNA EC3 measured, and thus the predictions can also be compared to this intrinsic variability of the prediction target.

For most of the 16 case studies, the predicted EC3 values fall within the experimental variability of the LLNA or are close to this range. In general, the predictions based on kDPRA and either cell-based assay are quite close for these chemicals.

3.9 Model selection with multiple input data

Here we show how different combinations of source data from OECD TG 442C, 442D and 442E can be used in different regression equations. Depending on the available data, a user is faced with two critical questions:

a) Once data from two positive tests are available, is it worth collecting data from the third test for a more accurate potency assessment?

b) If the available data allows calculating a predicted PoD with different equations, which result should be taken forward?

The detailed data analysis to answer these questions is presented in ESM3[4], and a summary is given here. If two positive results are available, including kDPRA data, studying the third key event, in general, is not needed because (i) if the third outcome were negative, the output of the two positive tests could be used directly with the corresponding equation EQ1 or EQ4, and if (ii) the

third outcome were positive, either equation EQ1, EQ4 or EQ5 could then be used, but these would lead to very similar predictions. Thus, for the 73 chemicals with three positive tests, the correlation between the predicted pEC3 from EQ1 (KS + kDPRA) and EQ5 (KS, h-CLAT and kDPRA) has an R^2 value of 0.95 with a slope close to 1 and y-intercept close to zero. Similarly, the correlation between EQ4 (h-CLAT and kDPRA) and EQ5 has an R^2 value of 0.99. Thus, adding the third test and moving from EQ1 or EQ4 to EQ5 has little effect on the final prediction. There would also not be a significant difference, whether the testing started with either 442D or 442E, as there is also a high correlation (R^2 value of 0.91 with a slope close to 1 and y-intercept close to zero) between the predicted pEC3 from EQ1 (KS and kDPRA) and EQ4 (h-CLAT and kDPRA) for chemicals positive in all assays. This relatively close alignment of EQ1, EQ4 and EQ5 for chemicals positive in all three tests is also illustrated by the case studies in Table 4. If all three tests are positive and data available, then using EQ5 appears most appropriate. A conservative alternative would be using the lowest EC3 value from EQ1, EQ4 and EQ5 to perform a more stringent risk assessment (see ESM3[4]).

A different question is what to do when all three tests are available and one of them is negative. Then the choice is to (i) use the model with the data from the two positive tests only or (ii) use EQ5 (KS, h-CLAT and kDPRA). Theoretically, the former choice is the more conservative one, as negative evidence is not factored in, while the latter uses all available data. Interestingly, there is also a high correlation between the predicted pEC3 for both choices (R^2 value of 0.92 with a slope close to 1) (n = 43 chemicals with 2 positive tests) and the correlation with the *in vivo* data is better for the latter option (see ESM3[4]). Thus, based on this analysis, it appears that using EQ5 is optimal in all cases when all data are available (2 or 3 positive tests). However, further expert-guided analysis can lead to different choices in selected cases, e.g., in cases of putative pro- or pre-haptens, which may be better predicted by the cellular assays and best predicted by EQ6.

3.10 Comparison of the reaction rate measured with the kDPRA and the Cor1-C420 assay used previously

Our previous study (Natsch et al., 2015) was entirely based on Cor1-C420 peptide reactivity data. In a final analysis, we compared data from the two peptide reactivity assays and regression models developed with the two reactivity assays. This helps to understand assay redundancy and similarity between the reactivity assays and between the current and the previous model, which provides information on the robustness of the approach proposed here and an indication of how the two peptide assays can be used interchangeably. This analysis is detailed in ESM4[2], and here we give a brief summary of the results.

A linear correlation is observed when comparing the reaction constants from kDPRA and the Cor1-C420 assay. Due to the higher reactivity of the Cor1-C420 peptide as compared to the Cys-peptide (Natsch et al., 2007), the correlation has a negative y-intercept:

EQ14 $\text{Log } k_{max} \text{ (kDPRA)} = 0.9 \times \text{Log } k_{max}$
      (Cor1-C420 assay) $- 0.59$

Thus, based on the measured Cor1-C420 data, we calculated predicted kDPRA values based on EQ14 and used these data to train the same regression model on the same chemicals as in EQ1 (EQ15).

EQ15 $\text{pEC3} = 0.40 + 0.49 \times \text{Log } k_{max \, norm}$ (calculated from
      Cor1-420 result) $+ 0.12 \times \text{Log } EC1.5_{norm} + 0.28 \times$
      $\text{Log } IC50_{norm} - 0.22 \times \text{Log } VP_{norm}$
      $S = 0.73; R^2 = 61.5\%; R^2 \text{ (adj)} = 60.7\%$

This equation has a similar predictive capacity as EQ1 (calculated on the same dataset) and overall a similar contribution of the individual parameters, although contribution of the reactivity parameter is slightly higher and the contribution of KS slightly weaker in EQ15 vs EQ1. The reactivity rates for the Cor1-C420 assay presented in the full database are all presented as measured values and also transformed to predicted kDPRA k_{max} values according to EQ14. In another data column, data for measured kDPRA k_{max} are shown and data gaps were filled with the predicted kDPRA k_{max} values according to EQ14. Due to data redundancy between the two tests, these data can be used in the absence of a true kDPRA experimental value for any further modeling on the larger database, e.g., to perform domain-based assessments or quantitative read-across assessment based on data from the target and the read-across substance using all the chemicals in the full database (For details see ESM4[2]).

4 Discussion

This paper presents global regression models generated from different NAM datasets for producing a PoD value for risk assessment purposes. The models are based on the use of OECD-validated NAMs that include the recently accepted kDPRA (OECD, 2021d) as well as the KS and h-CLAT (OECD, 2018a,b). A comprehensive database of 322 chemicals was assembled for this analysis (Natsch et al., 2015, 2020; OECD, 2021c) that may also be useful for further refinements or the development of alternative models.

4.1 Robustness of the models when compared to previous models

Previous work describing the development of regression models using NAM data used a more reactive peptide Cor1-C420 (Natsch and Gfeller, 2008) for assessing reactivity (Natsch et al., 2015, 2018). With the update of OECD TG 442C (OECD, 2021d) that now includes the kDPRA, it made sense to evaluate its use for input into these regression models. A linear correlation is observed when comparing reaction constants from the kDPRA and Cor1-C420 assays (see ESM4[2]). Thus, the two reactivity assays can be used interchangeably. To illustrate this, using EQ14, measured Cor1-C420 rate constants were used to calculate predicted kDPRA values for use in generating the same regression model as in EQ1. Not surprisingly, the equation generated (EQ15) has the same predictive capacity as EQ1 and, overall, a similar contribution of the individual parameters, indicating

similar information content of the two reactivity assays and a robustness of the overall approach.

The regression model using a comprehensive dataset with available kDPRA and KS data (n = 203) was based on the normalized EC1.5 and IC50 from KS, Log k_{max} from the kDPRA, and normalized vapor pressure. This model (EQ1) is based on the highest number of chemicals compared to other models and is very similar to the global model published previously (Natsch et al., 2015), but replaces Cor1-C420 reactivity data with kDPRA data.

4.2 Complementary and redundancy of the cell-based assays

A vital feature described in the paper is the flexibility in which input NAM data is used to make a PoD prediction. A model integrating kDPRA data with h-CLAT instead of KS showed a similar statistical weight using normalized MIT and CV75 data (n = 188; EQ4). This shows that either KS or h-CLAT data can be used to obtain comparable predictions. It is well known that having data from all three assays is not always possible due to assay compatibility factors (Kolle et al., 2019). However, in cases where data from all three OECD TG data is available, a comprehensive model can be obtained that integrates kDPRA, KS and h-CLAT data (n = 188; EQ5).

Interestingly, having data from both KS and h-CLAT along with kDPRA data only slightly improves the predictive power compared to one cellular assay. This redundancy has been observed previously (Natsch et al., 2020) and indicates that there is significant redundancy between the two cell-based assays. For situations where kDPRA data is absent, possibly due to an incompatibility issue using complex mixtures or technical issues (e.g., thiols), a regression model integrating KS and h-CLAT data is available (n = 188; EQ6), too. In the absence of reactivity data, the parameters for luciferase induction in KS or surface marker induction in h-CLAT receive more statistical weight. The best statistical power is obtained using KS EC3 instead of EC1.5, suggesting that this parameter for strong Nrf2-dependent luciferase induction can partially compensate for the lack of reactivity data (EQ7).

For each model based on a dataset of 188 chemicals with data from KS, h-CLAT or kDPRA, their predictive capacity is similar. For all models, the median of the fold-misprediction is around 2.5-fold, while the geometric mean is around 3.3-fold (Tab. 2). This finding demonstrates the flexibility in the utility of using these models based on what test data is available. A description of how to select an appropriate model is provided in ESM3[2], and a parallel paper describes how the models can be combined with the "2 out of 3" DA (Natsch and Gerberick, 2022). The analysis shows that for chemicals with three positive tests, adding the third test has little effect on the final prediction. Thus, a third key event study is not needed if two positive tests, including kDPRA data, are available. There is also no significant difference whether one starts with the kDPRA or either of the two cell-based assays. When multiple PoD values are available, the most predictive choice would be to use EQ5, integrating all three assays, although a different choice might be made when the chemical is a putative pro- or pre-hapten, in which case EQ6 might be optimal.

4.3 Assessment of using the comprehensive database vs the curated OECD DA database

Very similar models to EQ1 were obtained using the core set of data from Natsch et al. (2015) for which kDPRA data are available (n = 158; EQ2) and the set of chemicals with available h-CLAT data (n = 188; EQ3), indicating the model is robust and not highly dependent on the training set used. A curated database developed by an OECD expert group has been made publicly available (OECD, 2021c), which contains 196 chemicals, of which 154 have LLNA data and 149 are available with NAM data in our database. The regression coefficients for the models described (EQ1, EQ4, EQ5) are very similar to equivalent models derived from the OECD curated subset (Tab. 3). However, the contribution of reactivity from kDPRA tends to be lower. This may be attributed to the fact that the OECD group had stringent criteria to allow for extrapolation of LLNA results (OECD, 2021b). This led to the exclusion of LLNA data for several strong and extreme sensitizers, which are highly reactive chemicals and for which the reaction rate has a high weight in the assessment. On the other hand, the KS EC1.5 has a higher weight on this selected subset. When, for the same chemical set, the LLNA EC3 from previously published databases is used as the target as compared to the curated EC3 values (e.g., EQ8 vs EQ9), there is only a very small difference in the resulting regression equation, indicating that the data curation had no major impact on the overall picture. In addition, models established on a large number of chemicals are not inferior to the models built upon the OECD curated data (see results for EQ13 and EQ5 in Tab. 4), although the data curation certainly makes a significant difference for some individual chemicals. Going forward, it therefore appears appropriate to use models incorporating the maximal weight of evidence from the largest databases publicly available rather than to base models only on the curated OECD data.

4.4 Transparency of the regression models

As compared to complex tools like neural networks or Bayesian nets, linear regression models may appear to be relatively simple statistical tools. However, they have the advantage of high transparency. The prediction spreadsheet (ESM2[3]) is designed to allow one to enter experimental data from the different NAMs along with molecular weight and a calculated vapor pressure to calculate a PoD value. This should facilitate practical application.

4.5 Key misprediction

For most of the 16 case studies, it is evident that the PoD values, based on using kDPRA and one or both cellular assays, are quite similar to the LLNA values based on ≥ 5 individual EC3 values. Clear outliers under-predicted by the models are 3-dimethylaminopropylamine and p-phenylendiamine (see discussion on applicability domain in ESM3[4]). The extreme sensitizer oxazolone is predicted as a strong sensitizer by the models, but the predicted PoD value is clearly underestimated. The unique reactivity of oxazolone with lysine residues, explaining its unsurpassed sensitization potential, has been investigated before and is not captured by the kDPRA (Natsch et al., 2010). A more detailed discussion

on mispredicted chemicals is provided in a parallel paper (Natsch and Gerberick, 2022).

4.6 Outlook

For risk assessors responsible for assessing the skin sensitization risk of new chemical entities or chemicals lacking sufficient data, it is critical to have tools available that are predominantly dependent on using NAM data. The PoD value obtained from these regression models can thus be used to assist in the conduct of skin sensitization risk assessments. The predictive regression models (EQ1, 4, 5, 6 and 7 as an alternate) have been built from a comprehensive database of 188-203 chemicals along with a comparative analysis using the curated OECD database (OECD, 2021c). They show similar statistical strength and prediction accuracy. Performance is similar when using different data input parameters or when comparing models generated from different datasets. The PoD derived from these models may, using appropriate assessment factors to account for uncertainty and taking all information into account, be used as a starting point to determine safe use levels in products. The application and guidance on how to use these regression models when using the "2 out 3" DA is covered in a separate paper (Natsch and Gerberick, 2022).

Electronic supplementary material

ESM1[1] contains the full database in Sheet 1 and the predictions by the different models for the individual chemicals for different data subsets and vs different LLNA datasets in Sheets 2-4.

ESM2[3] is the prediction spreadsheet to calculate the PoD from *in vitro* data.

ESM3[4] discusses choice of the different regression models based on test availability.

ESM4[2] compares predictivity of the different reactivity assays.

References

Api, A. M., Basketter, D., Bridges, J. et al. (2020). Updating exposure assessment for skin sensitization quantitative risk assessment for fragrance materials. *Regul Toxicol Pharmacol 118*, 104805. doi:10.1016/j.yrtph.2020.104805

Bernauer, U., Bodin, L., Chaudhry, Q. et al. (2021). The SCCS notes of guidance for the testing of cosmetic ingredients and their safety evaluation, 11th revision, 30-31 March 2021, SCCS/1628/21. *Regul Toxicol Pharmacol 127*, 105052. doi:10.1016/j.yrtph.2021.105052

Cottrez, F., Boitel, E., Auriault, C. et al. (2015). Genes specifically modulated in sensitized skins allow the detection of sensitizers in a reconstructed human skin model. Development of the SENS-IS assay. *Toxicol In Vitro 29*, 787-802. doi:10.1016/j.tiv.2015.02.012

Cottrez, F., Boitel, E., Ourlin, J. C. et al. (2016). SENS-IS, a 3D reconstituted epidermis based model for quantifying chemical sensitization potency: Reproducibility and predictivity results from an inter-laboratory study. *Toxicol In Vitro 32*, 248-260. doi:10.1016/j.tiv.2016.01.007

Dent, M., Amaral, R. T., Da Silva, P. A. et al. (2018). Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients. *Comput Toxicol 7*, 20-26. doi:10.1016/j.comtox.2018.06.001

Gilmour, N., Kern, P. S., Alépée, N. et al. (2020). Development of a next generation risk assessment framework for the evaluation of skin sensitisation of cosmetic ingredients. *Regul Toxicol Pharmacol 116*, 104721. doi:10.1016/j.yrtph.2020.104721

Hirota, M., Fukui, S., Okamoto, K. et al. (2015). Evaluation of combinations of in vitro sensitization test descriptors for the artificial neural network-based risk assessment model of skin sensitization. *J Appl Toxicol 35*, 1333-1347. doi:10.1002/jat.3105

Hirota, M., Ashikaga, T. and Kouzuki, H. (2018). Development of an artificial neural network model for risk assessment of skin sensitization using human cell line activation test, direct peptide reactivity assay, KeratinoSens™ and in silico structure alert parameter. *J Appl Toxicol 38*, 514-526. doi:10.1002/jat.3558

Jaworska, J. S., Natsch, A., Ryan, C. et al. (2015). Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: A decision support system for quantitative weight of evidence and adaptive testing strategy. *Arch Toxicol 89*, 2355-2383. doi:10.1007/s00204-015-1634-2

Kleinstreuer, N. C., Hoffmann, S., Alépée, N. et al. (2018). Non-animal methods to predict skin sensitization (II): An assessment of defined approaches*. *Crit Rev Toxicol 48*, 359-374. doi:10.1080/10408444.2018.1429386

Kolle, S. N., Natsch, A., Gerberick, G. F. et al. (2019). A review of substances found positive in 1 of 3 in vitro tests for skin sensitization. *Regul Toxicol Pharmacol 106*, 352-368. doi:10.1016/j.yrtph.2019.05.016

Loveless, S. E., Api, A. M., Crevel, R. W. et al. (2010). Potency values from the local lymph node assay: Application to classification, labelling and risk assessment. *Regul Toxicol Pharmacol 56*, 54-66. doi:10.1016/j.yrtph.2009.08.016

Natsch, A., Gfeller, H., Rothaupt, M. et al. (2007). Utility and limitations of a peptide reactivity assay to predict fragrance allergens in vitro. *Toxicol In Vitro 21*, 1220-1226. doi:10.1016/j.tiv.2007.03.016

Natsch, A. and Gfeller, H. (2008). LC-MS-based characterization of the peptide reactivity of chemicals to improve the in vitro prediction of the skin sensitization potential. *Toxicol Sci 106*, 464-478. doi:10.1093/toxsci/kfn194

Natsch, A., Gfeller, H., Kuhn, F. et al. (2010). Chemical basis for the extreme skin sensitization potency of (E)-4-(Ethoxymethylene)-2-phenyloxazol-5(4H)-one. *Chem Res Toxicol 23*, 1913-1920. https://pubs.acs.org/doi/10.1021/tx1002707

Natsch, A., Emter, R., Gfeller, H. et al. (2015). Predicting skin sensitizer potency based on in vitro data from KeratinoSens and kinetic peptide binding: Global versus domain-based assessment. *Toxicol Sci 143*, 319-332. doi:10.1093/toxsci/kfu229

Natsch, A., Emter, R., Haupt, T. et al. (2018). Deriving a no expected sensitization induction level for fragrance ingredients without animal testing: An integrated approach applied to specific case studies. *Toxicol Sci 165*, 170-185. doi:10.1093/toxsci/kfy135

Natsch, A., Haupt, T., Wareing, B. et al. (2020). Predictivity of the kinetic direct peptide reactivity assay (kDPRA) for sensitizer potency assessment and GHS subclassification. *ALTEX 37*, 652-664. doi:10.14573/altex.2004292

Natsch, A. and Gerberick, G. F. (2022). Integrated skin sensitization assessment based on OECD methods (II): Hazard and potency by combining kinetic peptide reactivity and the "2 out of 3" defined approach. *ALTEX 39*, 647-655. doi:10.14573/altex.2201142

Nukada, Y., Ashikaga, T., Miyazawa, M. et al. (2012). Prediction of skin sensitization potency of chemicals by human cell line activation test (h-CLAT) and an attempt at classifying skin sensitization potency. *Toxicol In Vitro 26*, 1150-1160. doi:10.1016/j.tiv.2012.07.001

OECD (2010). Test No. 429: Skin Sensitisation: Local Lymph Node Assay. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/9789264071100-en

OECD (2014). The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. *OECD Series on Testing and Assessment, No. 168*. OECD Publishing, Paris. doi:10.1787/9789264221444-en

OECD (2018a). Test No. 442D: In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/9789264229822-en

OECD (2018b). Test No. 442E: In Vitro Skin Sensitisation: In Vitro Skin Sensitisation assays addressing the Key Event on activation of dendritic cells on the Adverse Outcome Pathway for Skin Sensitisation. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/9789264264359-en

OECD (2021a). Guideline No. 497: Defined Approaches on Skin Sensitisation. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/b92879a4-en

OECD (2021b). Series on Testing and Assessment No. 336: Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation – Annex 3. OECD Publishing, Paris. https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2021c). Series on Testing and Assessment No. 336: Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation – Annex 2. OECD Publishing, Paris. https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2021d). Test No. 442C: In Chemico Skin Sensitisation: Assays addressing the Adverse Outcome Pathway key event on covalent binding to proteins. *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing, Paris. doi:10.1787/9789264229709-en

Takenouchi, O., Fukui, S., Okamoto, K. et al. (2015). Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin sensitizing potential and potency of chemicals. *J Appl Toxicol 35*, 1318-1332. doi:10.1002/jat.3127

Urbisch, D., Mehling, A., Guth, K. et al. (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods. *Regul Toxicol Pharmacol 71*, 337-351. doi:10.1016/j.yrtph.2014.12.008

Wareing, B., Urbisch, D., Kolle, S. N. et al. (2017). Prediction of skin sensitization potency sub-categories using peptide reactivity data. *Toxicol In Vitro 45*, 134-145. doi:10.1016/j.tiv.2017.08.015

Zang, Q., Paris, M., Lehmann, D. M. et al. (2017). Prediction of skin sensitization potency using machine learning approaches. *J Appl Toxicol 37*, 792-805. doi:10.1002/jat.3424

Zeller, K. S., Forreryd, A., Lindberg, T. et al. (2017). The GARD platform for potency assessment of skin sensitizing chemicals. *ALTEX 34*, 539-559. doi:10.14573/altex.1701101

Conflict of interest

The authors declare no competing interests.

Data availability

All data of this publication are made publicly available and all models and tests used are freely available.