Natsch and Gerberick:

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

Electronic Supplementary Material (ESM4): Detailed analysis of regression models with kDPRA data and Cor1-C420 data

The current study replaces the prior reactivity data with the Cor1-C420 assay with the OECD-validated kDPRA. To understand the robustness of the predictions based on different reactivity parameters measured with two different reactive test peptides, different subsets of the data were analyzed using the two different reactivity parameters, and a weight of evidence from both parameters.

Overall, the correlation between the kDPRA and the Cor1-C420 assay is high, especially for those chemicals with significant reactivity (Figure ESM4-1). Thus, a predicted Log k_{max} in kDPRA can be calculated from measured Cor1-C420 data according to the formula:

EQ14: Log k_{max} (kDPRA) = 0.9 × log k_{max} (Cor1-C420 assay) - 0.59

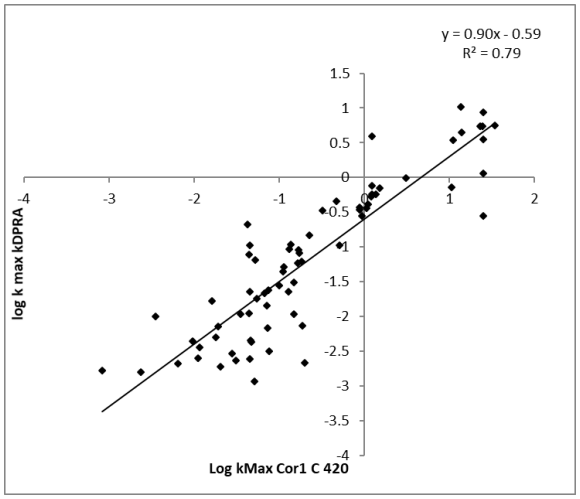

Fig. ESM4-1: Correlation of measured reactivity rates with the Cor1-C420 peptide and the kDPRA
Note: In the kDPRA a higher number of time points are measured giving a higher overall resolution.

The following columns with reactivity parameters were thus added to the database in Table ESM1-1, with the parameters shaded in grey used for statistics:

Tab. ESM4-1: Reactivity parameters added to the database based on measured kDPRA and Cor1-C420 reactivity data

LOG rate kDPRA	Logarithmic rate constant by SOP from kDPRA (in $s^{-1}M^{-1}$)
LOG_Norm rate kDPRA	= Log rate kDPRA + 3.5; non-reactive chemicals have a score of 0; reactive chemicals have positive scores
k max Cor1	Indicates the maximal rate constant (**in $min^{-1}mM^{-1}$**) for Cor1-C420 peptide depletion, calculated based on kinetic measurements taken at earlier time points for chemicals with > 50% depletion at 24 h, k is calculated from 24h depletion value in case < 50% depletion is observed at 24 h
LOG_norm K max Cor1	= Log K max Cor1 + 5.16 (parameter used in previous study, based on data in **$min^{-1}mM^{-1}$**)
LOG rate kDPRA_est_by_Cor1	Predicted Log k_{max} kDPRA estimated from measured Cor1C420 value, **in $s^{-1}M^{-1}$**; calculated by EQ13: Log kmax (kDPRA) predicted = 0.9 × log kmax (Cor1-C420 assay, $s^{-1}M^{-1}$) – 0.59
LOG_Norm rate kDPRA_est_by_Cor1	= LOG rate kDPRA_est_by_Cor 1 + 3.5; values < 0 set to 0
LOG_Norm rate kDPRA combined	= LOG_Norm rate kDPRA; if no kDPRA data available = Log_Norm rate kDPRA_est_by_Cor 1
AVG LOG_Norm rate kDPRA_ Cor1	= average (LOG_Norm rate kDPRALOG; Log_Norm rate kDPRA_est_by_Cor 1)

Note: "**LOG_Norm rate kDPRA_est_by_Cor1**" is the probably most important parameter from this analysis, as these data can be directly used to fill data-gaps with the kDPRA to have a good estimate of reactivity on the same scale.

Analysis of the different datasets and reactivity parameters on models with KS and reactivity data (analogous to EQ1)

a) Analysis on the 203 chemicals used to train EQ1

As shown in Table ESM4-2 and Figure ESM4-2, **EQ15** with the Cor1-C420 reactivity rate transformed to the kDPRA scale (LOG_Norm rate kDPRA_est_by_Cor1) gives overall very similar predicted pEC3 values as EQ1. Using the mean of the two reactivity measurements in **EQ16** (AVG LOG_Norm rate kDPRA_ Cor1) taking the evidence from both tests into account gives only a marginal improvement of overall predictivity for *in vivo* data, but reactivity has a slightly higher weight as compared to EQ1. This shows that the two reactivity measurements overall have similar information content.

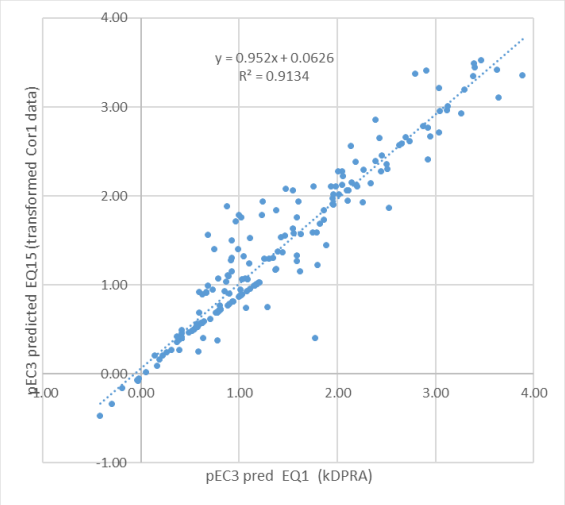

Fig. ESM4-2: pEC3 predicted by KS and either kDPRA (EQ1) or Cor1-C420 data transformed to the kDPRA scale (EQ15)

b) Analysis on the 244 chemicals used to train models in the previous analysis

On the dataset with 244 chemicals used for the global model in the 2015 analysis (Natsch et al., 2015), **EQ18** corresponds to the original global model presented in 2015, and it has a similar predictivity as **EQ17**, in which the rate from the Cor1-C420 assay is transformed to a predicted k_{max} in kDPRA. Thus, as to be expected, this linear data transformation has no significant effect on predictivity.

In **EQ19**, the transformed Cor1-C420 values are replaced by kDPRA values if available, and in **EQ20**, the average values are taken if both are available. These two equations again have a very similar predictivity to EQ17. Thus, overall, the transformed Cor1-C420 rates and the kDPRA rates can be used interchangeably based on the analysis of this data sub-set used previously. Furthermore, the weight-of-evidence from both reactivity assays does not significantly improve predictivity, indicating largely redundant information as also indicated in Figure ESM4-2.

c) Analysis on the complete dataset on 317 chemicals

Focusing on the complete dataset on 317 chemicals, the picture again is very similar: **EQ21** using the transformed values from the Cor1-C420 assay on 317 chemicals is very similar to EQ15 using the same input parameters on 203 chemicals, with a slightly higher weight for KS data. If the Cor1-C420 values are replaced by kDPRA data if available (**EQ22**) or if average values are taken (**EQ23**), the predictivity and the regression equation does only change slightly.

EQ22 is comparable to EQ1, since the same input data are used for 203 chemicals in EQ1 and EQ22, and for the additional 114 chemicals, the data from the Cor1-C420 assay transformed to the kDPRA scale are used for data-gap filling. Thus, this comparison indicates how robust the key EQ1 used in the prediction Spreadsheet is to the addition of more data. KS EC1.5 receives a slightly higher weight in EQ22, while the weight of the IC50 parameter is slightly reduced, but else the predictive equation does not significantly change by the addition of 114 more data points using the reactivity data obtained by read-across from the Cor1-C420 assay, indicating that this global model is quite stable. We have no h-CLAT data to test whether this is also true for EQ4, EQ5, EQ6 and EQ7, but since they are all based already on a large set of 188 chemicals and based on the stability shown here of models with KS and different reactivity parameters we would expect them also to be robust.

This analysis shows that the reactivity parameter is probably well covered by either assay. The estimated kDPRA values derived from the Cor1-C420 testing can be used for data-gap filling for modeling and read-across if no kDPRA values are available.

Tab. ESM4-2: Regression models on different data subsets integrating the reactivity parameters from the kDPRA and the Cor1-C420 assay

Model	Dataset	Reactivity parameter	Constant	k_{max}	EC1.5	IC50	VP_{norm}	R^2	N
EQ1	**Full dataset with kDPRA data**	**LOG_Norm rate kDPRA**	**0.42**	**0.40**	**0.15**	**0.36**	**-0.21**	**62.0**	**203**
EQ15	Full dataset with kDPRA data	LOG_Norm rate kDPRA_est_by_Cor1	0.40	0.49	0.12	0.28	-0.22	61.5	203
EQ16	Full dataset with kDPRA data	AVG LOG_Norm rate kDPRA_ Cor1	0.41	0.48	0.12	0.31	-0.22	63.3	203
EQ17	Key 244	LOG_Norm rate kDPRA_est_by_Cor1	0.15	0.488	0.208	0.323	-0.181	61.8	244
EQ18	Key 244	LOG_norm K max Cor1	0.07	0.373	0.22	0.289	-0.194	60.9	244
EQ19	Key 244	LOG_Norm rate kDPRA combined (extr if no kDPRA data)	0.15	0.408	0.252	0.371	-0.168	61.8	244
EQ20	Key 244	AVG LOG_Norm rate kDPRA_ Cor1	0.15	0.482	0.213	0.338	-0.179	62.7	244
EQ21	Full dataset	LOG_Norm rate kDPRA_est_by_Cor1	0.33	0.455	0.199	0.260	-0.207	54.1	317
EQ22	**Full dataset**	**LOG_Norm rate kDPRA combined (data gap filled by Cor1_C420 if no kDPRA data)**	**0.33**	**0.399**	**0.227**	**0.301**	**-0.195**	**54.6**	**316**
EQ23	Full dataset	AVG LOG_Norm rate kDPRA_ Cor1	0.33	0.456	0.199	0.271	-0.206	55.1	317

Reference

Natsch, A., Emter, R., Gfeller, H., Haupt, T. 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**,** 319-32. doi: 10.1093/toxsci/kfu229