Natsch et al.:

Predictivity of the Kinetic Direct Peptide Reactivity Assay (kDPRA) for Sensitizer Potency Assessment and Subclassification

Supplementary Data

Tab. S1: Chemicals that could not be evaluated based on technical limitations

Name	CAS	Observed problem
2-Nitro-1,4-phenylendiamine	5307-14-2	fluorescence quenching
Vanillin	121-33-5	fluorescence quenching
Tropolone	533-75-5	fluorescence quenching
Thioglycerol	96-27-5	reaction with fluorescent probe
Tetrachlorsalicylanilide	1154-59-2	autofluorescence
Bandrowski's Base (N,N-bis(4-aminophenyl)-2,5-diamino-1,4-quinone-diimine)	20048-27-5	fluorescence quenching

Tab. S2: Full database of results from kDPRA along with *in vivo* data and data from other *in vitro* assays
see doi:10.14573/altex.2004292s2

Tab. S3: Full ROC analysis for different log k_{max} cut-off values to predict GHS Cat 1A versus LLNA and human data

	LLNA			Human		
Cut-off	**Sensitivity**	**Specificity**	**Balanced accuracy**	**Sensitivity**	**Specificity**	**Balanced accuracy**
0.8	8.9%	100.0%	54.4%	0.0%	100.0%	50.0%
0.7	13.3%	99.3%	56.3%	0.0%	96.7%	48.3%
0.6	15.6%	99.3%	57.4%	0.0%	96.7%	48.3%
0.5	22.2%	99.3%	60.7%	3.0%	96.7%	49.8%
0.4	22.2%	99.3%	60.7%	3.0%	96.7%	49.8%
0.3	22.2%	99.3%	60.7%	3.0%	96.7%	49.8%
0.2	22.2%	99.3%	60.7%	3.0%	96.7%	49.8%
0.1	22.2%	99.3%	60.7%	3.0%	96.7%	49.8%
0	24.4%	99.3%	61.9%	3.0%	96.7%	49.8%
-0.1	28.9%	99.3%	64.1%	3.0%	96.7%	49.8%
-0.2	31.1%	97.0%	64.1%	6.1%	95.6%	50.8%
-0.3	37.8%	97.0%	67.4%	12.1%	94.4%	53.3%
-0.4	42.2%	97.0%	69.6%	12.1%	93.3%	52.7%
-0.5	48.9%	96.3%	72.6%	15.2%	93.3%	54.2%
-0.6	53.3%	95.6%	74.4%	21.2%	92.2%	56.7%
-0.7	55.6%	95.6%	75.6%	24.2%	92.2%	58.2%
-0.8	55.6%	95.6%	75.6%	24.2%	92.2%	58.2%
-0.9	55.6%	94.8%	75.2%	24.2%	92.2%	58.2%
-1	60.0%	94.1%	77.0%	30.3%	91.1%	60.7%
-1.1	64.4%	92.6%	78.5%	30.3%	91.1%	60.7%
-1.2	66.7%	91.9%	79.3%	30.3%	90.0%	60.2%
-1.3	71.1%	91.1%	81.1%	39.4%	90.0%	64.7%
-1.4	73.3%	91.1%	82.2%	42.4%	90.0%	66.2%
-1.5	73.3%	91.1%	82.2%	42.4%	90.0%	66.2%
-1.6	75.6%	90.4%	83.0%	45.5%	90.0%	67.7%
-1.7	77.8%	88.1%	83.0%	48.5%	88.9%	68.7%
-1.8	77.8%	86.7%	82.2%	51.5%	88.9%	70.2%
-1.9	77.8%	85.9%	81.9%	54.5%	88.9%	71.7%
-2	**84.4%**	**85.9%**	**85.2%**	**63.6%**	**88.9%**	**76.3%**
-2.1	84.4%	85.2%	84.8%	63.6%	87.8%	75.7%
-2.2	84.4%	83.0%	83.7%	66.7%	87.8%	77.2%
-2.3	84.4%	82.2%	83.3%	66.7%	87.8%	77.2%
-2.4	86.7%	80.0%	83.3%	72.7%	87.8%	80.3%
-2.5	88.9%	77.0%	83.0%	75.8%	84.4%	80.1%

doi:10.14573/altex.2004292s1

-2.6	91.1%	75.6%	83.3%	75.8%	83.3%	79.5%
-2.7	91.1%	72.6%	81.9%	75.8%	80.0%	77.9%
-2.8	91.1%	70.4%	80.7%	75.8%	76.7%	76.2%
-2.9	93.3%	69.6%	81.5%	78.8%	75.6%	77.2%
-3	93.3%	68.1%	80.7%	78.8%	74.4%	76.6%
-3.1	93.3%	67.4%	80.4%	78.8%	74.4%	76.6%
-3.2	93.3%	65.9%	79.6%	81.8%	73.3%	77.6%
-3.3	93.3%	65.2%	79.3%	84.8%	73.3%	79.1%
-3.4	93.3%	63.7%	78.5%	87.9%	73.3%	80.6%
-3.5	93.3%	60.7%	77.0%	87.9%	70.0%	78.9%

Tab. S4: Human GHS Cat 1A sensitizers underpredicted by applying a cut-off of log k_{max} = -2.0 [a]

Name	CAS	Log k_{max}	LLNA GHS Cat
4-Phenylenediamine	106-50-3	-2.81	1A
δ-Damascone	57378-68-4	-2.16	1B
Diethylenetriamine	111-40-0	-3.50	1B
2-Hexylidene cyclopentanone	17373-89-6	-2.36	1B
Methylanisylidene acetone	104-27-8	-3.10	1B
Phenylacetaldehyde [b]	122-78-1	-2.36	1B
Glutaric aldehyde	111-30-8	-3.50	1A
2-Aminophenol	95-55-6	-2.46	1A
3-Dimethylaminopropylamine	109-55-7	-3.50	1B
Lyral	31906-04-4	-3.31	1B
6-Methyl-3,5-heptadien-2-one	1604-28-0	-3.29	NC
2-Methoxy-4-methylphenol	93-51-6	-3.50	1B

[a] An additional Human GHS 1A chemical based on OECD review and underpredicted by kDPRA is butyl glycidyl ether, CAS 2426-08-6. [b] Human GHS 1B based on OECD review

Discussion of the false-negative human 1A chemicals in Table S4 according to the Basketter classification

For creosol (2-methoxy-4-methylphenol) Cat 1A classification is based on a low NOEL value only (i.e., from a study showing no sensitization reactions). Thus this class attribution is due to the arbitrarily chosen low test concentration, and it is highly likely that this chemical would only sensitize at much higher doses as similar molecules like eugenol or dihydroeugenol are weak to moderate sensitizers falling into Cat 1B. Thus, it is probably an incorrect assignment.

Lyral was classified as Cat 1A based on clinical observations, while predictive human tests had not found sensitization reactions. Thus, neither LLNA nor human predictive testing would have led to 1A. The false-positives include two clear pro-haptens (diethylenetriamine and 3-dimethylaminopropylamine) and two pre-haptens (4-phenylenediamine, which reacts more slowly in the kinetic assay as it requires abiotic oxidation, and 2-aminophenol).

They also include three Michael acceptor chemicals with lowest observed effect level (LOEL) values close to the human cut-off for Cat 1A chemicals of 500 µg/cm^2, and an extrapolated DSA_{05} (extrapolated value leading to induction of sensitization in 5% of the panelists) therefore closely below the cut-off. (δ-damascone (human LOEL = 500 µg/cm^2), 2-hexylidene cyclopentanone (human LOEL= 500 µg/cm^2), methylanisylidene acetone (human LOEL= 550 µg/cm^2), which are also 1B in LLNA, so these are clearly borderline chemicals. Phenylacetaldehyde was rated 1B by the Basketter et al. compilation and by the OECD data review, but it is 1A based on the ICCVAM evaluation of the RIFM data and here included in 1A.

Finally, 9 of 12 of these under-predicted chemicals are rated as Cat 1B by the LLNA, too. Thus, overall, only a limited number of important and clear-cut human 1A sensitizers are missed by this refined cut-off (4-phenylenediamine, diethylenetriamine, glutaric aldehyde, 2-aminophenol, 3-dimethylaminopropylamine, 6-methyl-3,5-heptadien-2-one).

Based on all these evaluations, a refined cut-off of log k_{max} = -2.0 appears to be an optimal prediction model to balance accuracy for LLNA and human data.

Alternative calculations for identification of chemicals in the GHS 1A potency class

Since LLNA values are in weight % and k_{max} values are based on molar concentration, we performed two additional ways of calculating which chemicals have a predicted EC3 value < 2% (i.e., fall into the GHS 1A category) to test the impact of this simplification.

a) We transformed the measured k-values to a percentage value (by multiplying k_{max} with 10 and dividing it by the molecular weight), and then performed the ROC-analysis versus the LLNA based on k_{max} values calculated in %.
b) We used the predictive formula derived by regression analysis (Eq. S1: pEC3 = 2.652 + 0.3491 × log k_{max}) to derive a predicted pEC3, which was then transformed to the EC3 and used for classification according to the 2% threshold.

Approach a): Table S5 indicates the ROC analysis for different thresholds of the transformed k_{max} values.

Tab. S5: Predictivity of different thresholds of log k_{max} calculated in $\%^{-1}s^{-1}$ instead of $M^{-1}s^{-1}$

Threshold (k in $\%^{-1}s^{-1}$)	Sensitivity	Specificity	Balanced accuracy
-3	80.0%	86.7%	83.3%
-3.1	80.0%	85.2%	82.6%
-3.2	80.0%	84.4%	82.2%
-3.3	82.2%	84.4%	83.3%
-3.4	84.4%	83.7%	84.1%
-3.5	**86.7%**	**82.2%**	**84.4%**
-3.6	86.7%	80.7%	83.7%
-3.7	88.9%	79.3%	84.1%
-3.8	88.9%	75.6%	82.2%
-3.9	91.1%	73.3%	82.2%
-4	91.1%	70.4%	80.7%

In this analysis, a threshold of -3.5 in $\%^{-1}s^{-1}$ has the best predictivity. With the approach b), no threshold needs to be determined, but rather chemicals are classified by EC3 values calculated from the predicted pEC3 value according Eq. S1.

Table S6 shows the predictivity of the original approach using the threshold of log k_{max} $M^{-1}s^{-1}$ compared to the predictivity with the two alternative calculations. In each case, the same result is obtained for 174 of the 180 chemicals. However, the balanced accuracy is slightly reduced (from 85.2% to 84.4% for approach a) and to 83% for approach b)). We thus propose to remain with the threshold of log k_{max} = -2 for regulatory classification, not least for its simplicity and most importantly for its predictivity, acknowledging that calculating everything in molar terms may be a scientifically preferred approach.

As the predictivity is optimal with the approach using log k_{max} [based on $M^{-1}s^{-1}$] and the threshold -2, this approach is proposed to be taken forward for regulatory use of the kDPRA.

Tab. S6: Predictivity for different approaches to transform the measured log k_{max} values into GHA 1A subclassifications

	Threshold log k_{max} [based on $M^{-1}s^{-1}$] = -2	(a) Threshold logk_{max} [based on $\%^{-1}s^{-1}$] = -3.5	(b) EC3 calculated based on Equation S1
Sensitivity	84.4%	86.7%	82.2%
Specificity	85.9%	82.2%	83.7%
Balanced accuracy	85.2%	84.4%	83.0%

Table S7 shows the chemicals for which the alternative calculations lead to a different outcome. As expected, this is the case for chemicals with a log k_{max} close to the threshold of -2 or those with a relatively low or high molecular weight.

Tab. S7: Chemicals with differing classifications using the different calculations

Name	CAS	log k_{max}	MW [g/mol]	LLNA classification	Threshold log k_{max} [based on $M^{-1}s^{-1}$] = -2	(a) Threshold log k_{max} [based on $\%^{-1}s^{-1}$] = -3.5	(b) EC3 calculated based on Eq. S1
Imidazolidinyl urea	39236-46-9	-1.106	388.29	1B/2	FP	FP	TN
δ-Damascone	57378-68-4	-2.162	192.3	1B/2	TN	FP	TN
Phenylacetaldehyde	122-78-1	-2.363	120.15	1B/2	TN	FP	FP
Methylmethacrylate	80-62-6	-2.001	100.12	1B/2	TN	FP	FP
Propyl gallate	121-79-9	-1.960	212.2	1A	TP	TP	FN
2-Aminophenol	95-55-6	-2.460	109.13	1A	FN	TP	TP
Toluene diamine sulphate	615-50-9	-1.964	220.25	1A	TP	TP	FN
2,3-Butanedione	431-03-8	-2.618	86.09	1B/2	TN	TN	FP
2-Ethylhexyl acrylate	103-11-7	-2.133	184.28	1B/2	TN	FP	TN
Methyl methanesulphonate	66-27-3	-2.145	110.13	1B/2	TN	FP	FP

Tab. S8: F-values for the linear regression of logarithmic *in vitro* parameters versus pEC3

		Set I (n = 173)	Set II (n = 154)	Set I EC3 < 30% (n = 121)	Set II EC3 < 30% (n = 107)
kDPRA	k_{max}	191.14	126.05	84.34	50.55
KeratinoSens	EC1.5	77.84	57.41	18.58	13.28
	EC3	95.91	80.24	25.78	20.47
	IC50	95.48	78.72	20.45	17.24
h-CLAT	EC150		59.00		20.99
	EC200		29.71		4.35[1)]
	MIT		85.23		27.04
	CV75		115.24		28.66
DPRA	kCys		71.9		23.57
	kLys		28.12		20.87

All correlations are statistically highly significant at $p \leq 0.0005$ (with the exception of EC200 / Set II / EC < 30%, where $p = 0.039$).

Tab. S9: *P*-values of multiple linear regression of logarithmic *in vitro* parameters versus pEC3: Dataset stratified for LLNA positives between 10-100% and LLNA positives at < 10%

***p*-values of the single parameters in the multiple regression**	**LLNA 10-100%**	**LLNA < 10%**
a) Combining kDPRA with hClat		
Log_CV75	**0.000**	0.966
Log_MIT	0.659	0.186
Log k_{max}	0.523	**0.000**
Overall r^2	45.1%	30.6%
b) Combining kDPRA with KeratinoSens		
Log_IC50	**0.000**	0.022
Log_EC3	0.099	0.379
Log k_{max}	0.945	**0.000**
Overall r^2	40.1%	34.8%

Multiple regression analysis

This analysis is the basis for Table 8 and (partly) Table 7 in the main manuscript.

Set I, EC3 < 30%, regression pEC3 versus log k_{max} only

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.702302	40.48%	40.00%	38.38%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.652	0.103	25.85	0.000	
Log k_{max}	0.3491	0.0380	9.18	0.000	1.00

Regression Equation

Eq. S2: pEC3 = 2.652 + 0.3491 log k_{max}

Set I, EC3 < 30%, regression pEC3 versus log k_{max} and KeratinoSens

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.677795	45.68%	44.34%	41.91%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.131	0.194	16.17	0.000	
Log kl	0.3206	0.0431	7.44	0.000	1.37
Log EC3	-0.0142	0.0741	-0.19	0.849	1.68
Log IC50	-0.2227	0.0882	-2.53	0.013	1.34

Regression Equation

Eq. S3: pEC3 = 3.131 + 0.3206 log k_{max} - 0.0142 log EC3 - 0.2227 log IC50

Set II, EC3 < 30%, regression pEC3 versus log k_{max} only

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.728177	32.50%	31.85%	29.48%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.554	0.122	20.90	0.000	

Log k_{max}	0.3143	0.0442	7.11	0.000	1.00

Regression Equation

Eq. S4: pEC3 = 2.554 + 0.3143 log k_{max}

Set II, EC3 < 30%, regression pEC3 versus log k_{max} and KeratinoSens

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.703640	38.39%	36.58%	32.95%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.085	0.230	13.40	0.000	
Log k_{max}	0.2794	0.0512	5.46	0.000	1.42
Log EC3	-0.0281	0.0859	-0.33	0.744	1.68
Log IC50	-0.2307	0.0996	-2.32	0.023	1.30

Regression Equation

Eq. S5: pEC3 = 3.085 + 0.2794 log k_{max} - 0.0281 log EC3 - 0.2307 log IC50

Set II, EC3 < 30%, regression pEC3 versus log k_{max} and h-CLAT

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.693353	39.96%	38.22%	33.70%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.998	0.184	16.27	0.000	
Log k_{max}	0.2382	0.0472	5.05	0.000	1.26
Log MIT	-0.1313	0.0825	-1.59	0.114	1.88
Log CV75	-0.195	0.123	-1.59	0.114	1.92

Regression Equation

Eq. S6: pEC3 = 2.998 + 0.2382 log k_{max} - 0.1313 log MIT - 0.195 log CV75

Set II, EC3 < 30%, regression pEC3 versus log k_{max} and KeratinoSens and h-CLAT

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.695337	41.02%	38.07%	31.96%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.085	0.228	13.51	0.000	
Log k_{max}	0.2545	0.0524	4.86	0.000	1.53
Log MIT	-0.1330	0.0848	-1.57	0.120	1.95
Log CV75	-0.079	0.157	-0.50	0.616	3.07
Log EC3	0.0192	0.0878	0.22	0.827	1.80
Log IC50	-0.135	0.127	-1.06	0.290	2.16

Regression Equation

Eq. S7: pEC3 =3.085 + 0.2545 log kmax - 0.1330 log MIT - 0.079 log CV75 + 0.0192 log EC3 - 0.135 log IC50

Set II, EC3 < 30%, regression pEC3 versus h-CLAT and KeratinoSens

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.769119	27.12%	24.23%	18.26%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.071	0.253	12.16	0.000	
Log EC3	-0.1489	0.0893	-1.67	0.098	1.52
Log IC50	-0.063	0.139	-0.45	0.650	2.13
Log MIT	-0.1653	0.0935	-1.77	0.080	1.94
Log CV75	-0.218	0.171	-1.28	0.204	2.97

Regression Equation

Eq. S8: pEC3 = 3.071 - 0.1489 log EC3 - 0.063 log IC50 - 0.1653 log MIT - 0.218 log CV75

Set II, EC3 < 30%, regression pEC3 versus DPRA and h-CLAT and KeratinoSens

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.752573	27.28%	23.49%	15.84%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	3.274	0.299	10.95	0.000	
Cys log k	0.240	0.106	2.26	0.026	1.84
Log EC3	-0.014	0.105	-0.13	0.898	1.98
Log IC50	-0.033	0.140	-0.24	0.814	2.05
Log MIT	-0.1235	0.0929	-1.33	0.187	1.89
Log CV75	-0.202	0.170	-1.19	0.236	2.90

Regression Equation

Eq. S9: pEC3 = 3.274 + 0.240 Cys log k - 0.014 log EC3 - 0.033 log IC50 - 0.1235 log MIT - 0.202 logCV75