Natsch and Gerberick:

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

Electronic Supplementary Material (ESM3): Choice of different models

Here more details are provided on four key questions:

1. If two congruent outcomes from two key event-based tests are available (i.e., sufficient to perform hazard assessment), are further data required on the third key event to improve the potency assessment?
2. If incomplete evidence is sufficient: Would the chosen sequence of testing significantly affect the result on potency?
3. If data from all three key events are available: Which of the provided models should be used for the final assessment?
4. Should applicability domain considerations affect the sequence of testing and model selection?

1 Further testing with two concordant results?

Since EQ1 and EQ4 with kDPRA have a stronger correlation to the *in vivo* EC3 as compared to EQ6 and EQ7 without kDPRA, we recommend always performing the DPRA (followed by the kDPRA in case of positive DPRA) along with a cell-based assay first.

The situation is different if

– for *technical* reasons the kDPRA is not applicable or
– for *mechanistic reasons* the kDPRA is considered to have limited predictivity.

This will be further discussed in Section 4 of this document.

If the cell-based assay (either KS or h-CLAT) and the kDPRA are positive, EQ1 or EQ4 can be used. The third test would then be needed only to apply EQ5. However, the additional information would not change the picture significantly: As illustrated in Figure ESM3-1, for chemicals positive in three assays, there is a high correlation between predicted EC3 values of EQ1 (KS and kDPRA) vs EQ5 (all tests) and EQ4 (h-CLAT and kDPRA) vs EQ5.

These strong correlations have a slope close to 1 and an intercept close to 0, indicating that the predicted EC3 not only correlate but are also numerically similar.

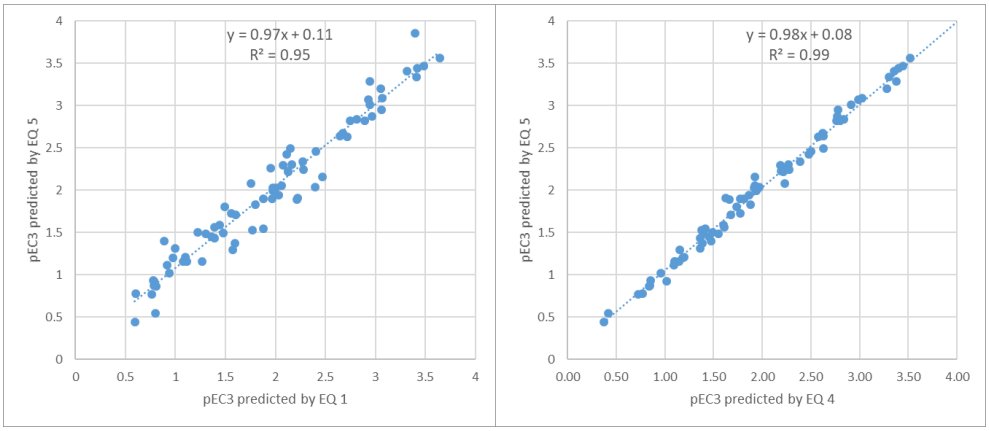

Fig. ESM3-1: Correlation of pEC3 values predicted by two positive tests (KS and kDPRA, left; h-CLAT and kDPRA, right) and pEC3 predicted by all three tests (EQ5)
Data from all chemicals (n = 73) with positive DPRA, KS and h-CLAT tests according to the prediction models of the three test guidelines are shown.

doi:10.14573/altex.2201141s3

Similar to Table 2 in the main manuscript, the predictivity for the three models for this data subset is shown in Table ESM3-1. This analysis further indicates that predictivity vs. *in vivo* data is not significantly increased by moving from either EQ1 or EQ4 to EQ5 for chemicals positive in three assays. Thus, performing the third test will not significantly improve the precision of the prediction for these chemicals.

Tab. ESM3-1: Predictivity on the dataset (n = 73) with positive data from all three tests (KS, h-CLAT and DPRA data)

Model	Input parameters	Fold-mis-prediction[a] (geomean)	Fold-mis-prediction (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	4.3	3.2	16 (22%)	9 (12%)	12 (16%)	5 (7%)
EQ4	kDPRA, h-CLAT	3.9	3.2	16 (22%)	8 (11%)	9 (12%)	5 (7%)
EQ5	kDPRA, KS, h-CLAT	4.0	3.4	16 (22%)	8 (11%)	13 (18%)	5 (7%)

[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 are those for which the measured LLNA EC3 is lower than the predicted EC3; over-predicted chemicals are those with measured LLNA EC3 higher than the predicted value.

2 Effect of the sequence of testing

Also, there is 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) as shown in Figure ESM3-2. This indicates that there would not be a large and systematic difference whether the testing started with 442D (followed by potency assessment with EQ1) or 442E (and using EQ4). Thus, the testing sequence would not significantly affect the result in cases where testing was stopped with partial evidence from two assays.

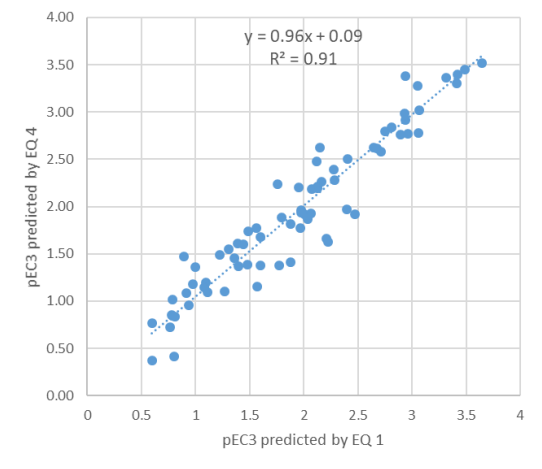

Fig. ESM3-2: Correlation of pEC3 values predicted by two positive tests (KS and kDPRA, EQ1 vs. h-CLAT and kDPRA, EQ4)
Data from all chemicals (n = 73) with positive DPRA, KS and h-CLAT tests according to the prediction models of the three test guidelines are shown.

3 Model selection if data from all three key events are available

There are two possible cases:

a) Data have already been collected from all three tests, and all tests are positive. The question then is which predicted EC3 value should be used as all four models (EQ1,4,5 and 6) can be calculated.
b) In a "2 out of 3" situation, two tests are positive, while the third is negative. Should the negative evidence be included?

a) Three positive tests

If all three tests are positive and data is available, using EQ5 and taking all evidence into account appears most appropriate, although EQ1 and EQ4 would give a similar outcome as illustrated above in Figure ESM3-1. However, a conservative alternative would be, e.g., to use the lowest EC3 value from EQ1, 4 and 5 to perform a more stringent risk assessment. However, as illustrated in Figure ESM3-3, this would lead to a higher pEC3 (lower EC3) for only a minority of chemicals, and for those the predicted values would still be close to that predicted by EQ5. Thus, using EQ5 appears most appropriate.

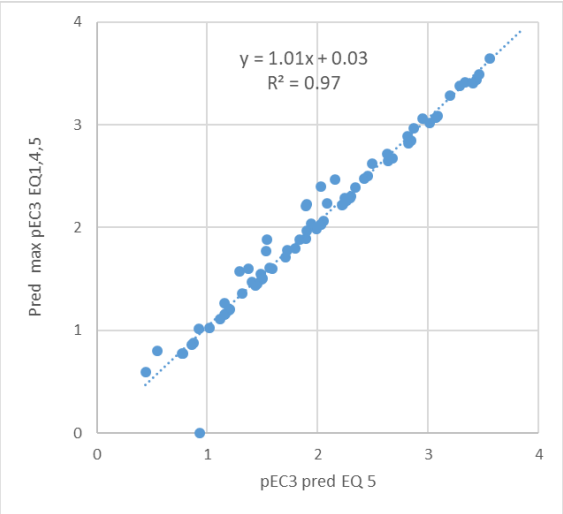

Fig. ESM3-3: Correlation of pEC3 values predicted by EQ5 with the maximal pEC3 (= minimal predicted EC3) from the three models EQ1, EQ4 and EQ5
Data from all chemicals (n = 73) with positive DPRA, KS and h-CLAT tests according to the prediction models of the three test guidelines are shown.

b) Two positive, one negative test

In this case, the choice is to (i) use the model based on the input data from the two positive tests only (EQ1, EQ4 or EQ6, depending on which test is negative) or (ii) use EQ5 (all data from KS, h-CLAT and kDPRA). Theoretically, the former choice is the more conservative one as negative evidence is not factored in, while the latter choice uses all available data and takes into account the information that the chemical was not able to trigger activation of the third KE.

Interestingly, there is a high correlation between the predicted pEC3 for both choices (R^2 value of 0.85 with a slope close to 1) (n = 43 chemicals with 2 positive tests; Fig. ESM3-4). The correlation with the *in vivo* data is slightly better for the latter option (Tab. ESM3-2), though. Thus, the number of chemicals overpredicted (Lower EC3 value predicted as compared to the *in vivo* outcome) is higher if only the evidence from the two positive tests is used, and the geometric mean of the misprediction is enhanced from 2.9-fold to 3.6-fold. This is also mirrored by a positive intercept of the equation in Figure ESM3-4, indicating a more conservative assessment if only the positive tests are factored in. Based on this outcome, it may make sense, in the presence of all three results, to give higher weight to the result from EQ5. However, this also leads back to the question of a “2 out of 2” situation discussed above: If the testing was stopped after two positive tests, we do not know the outcome of the third test: Thus, we have the possibility of an overestimation of the sensitization potential based on using only two tests if the third would turn out negative. However, the results in Figure ESM3-4 and Table ESM3-2 indicate that this effect is minor and would lead to a conservative conclusion in risk assessment (higher pEC3 / lower EC3 predicted).

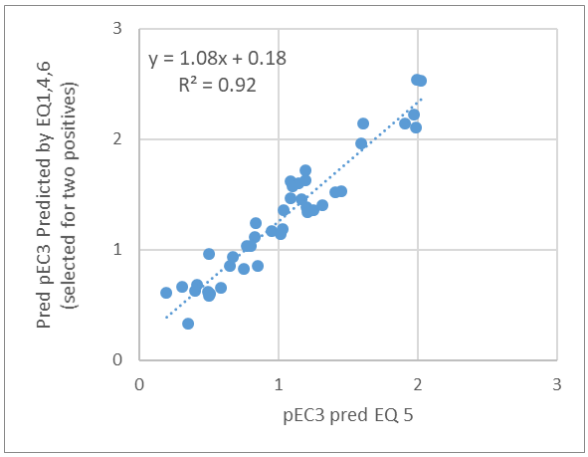

Fig. ESM3-4: Correlation of pEC3 values predicted by EQ5 with the pEC3 from either EQ1, EQ4 or EQ5, selected based on two positive results for the given chemical
Data from all chemicals (n = 43) with two positive tests from DPRA, KS or h-CLAT.

Tab. ESM3-2: Predictivity on the dataset (n = 43) with two positive and one negative test (from KS, h-CLAT and DPRA)

Model	Fold-mis-prediction[a] (geomean)	Fold-mis-prediction (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, 4 or 6 (selected according to the two positive tests)	3.6	3.6	3 (7%)	0 (0%)	10 (24%)	4 (10%)
EQ5	2.9	2.6	5 (12%)	2 (5%)	3 (7%)	0 (0%)

[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 are those for which the measured LLNA EC3 is lower than the predicted EC3; over-predicted chemicals are those with measured LLNA EC3 higher than the predicted value.

4 Should applicability domain considerations affect the sequence of testing and model selection?

All the analyses above are on the global predictivity and correlation between different testing sequences and different levels of evidence. Focusing on *specific chemicals*, there may still be a reason for a given testing sequence or preference for a model based on a specific domain. Thus, Table ESM3-3 lists the chemicals for which EQ6 results in a more than 2-fold lower EC3 as compared to EQ5. Thus for these chemicals, the result from only the cell-based assays is more conservative and closer to the *in vivo* result as compared to the models using kDPRA.

These chemicals include three putative prohaptens (ethylenediamine, 3-dimethyl-amino-1-propylamine, 1-naphtol) and a pre-hapten (1,4-phenylenediamine). In addition, glutaraldehyde, which is known to be a specifically amine-reactive compound mispredicted in the Cys-peptide-based kDPRA, is in this group. Thus, although the evidence is based on very few chemicals, an expert judgment may be applied and EQ6 or EQ7 given preference over EQ5 for chemicals with alerts to act as pro/pre-hapten or which have a specific amine-reactivity (as shown by predominant depletion of the Lys-peptide in the DPRA). This is in line with the known limitations of the kDPRA summarized in APPENDIX III, ANNEX 1 of OECD TG 442D.

Tab. ESM3-3: Chemicals with predicted EC3 from EQ6 < 2-fold lower than predicted EC3 from EQ5

Name	CAS	LLNA EC3	EC3 EQ5	EC3 EQ6
Chemicals positive in 3 assays				
1-Naphtol	90-15-3	1.3	6.97	2.60
1,4-Phenylenediamine	106-50-3	0.16	1.72	0.58
Glutaraldehyde	111-30-8	0.10	3.98	0.66
Tetrachlorsalicylanilide	1154-59-2	0.04	1.80	0.34
Chemicals positive in KS and h-CLAT only				
Ethylenediamine	107-15-3	2.2 [a]	38.2	15.3
3-Dimethyl-amino-1-propylamine	109-55-7	2.2 [b]	32.04	11.3

[a] OECD DB reports NC for ethylenediamine, as only one of three LLNA studies was positive, indicating that for this chemical there is also uncertainty in the LLNA. [b] OECD DB reports 3.5%.

Conclusions on model choice and sequence of testing

Due to significant data redundancy between KeratinoSens and h-CLAT, the analysis of the data provided here indicates:

- The *sequence of testing* for potency prediction can start with either kDPRA and KS or kDPRA and h-CLAT with a similar final outcome.
- *Testing can stop after 2 positive tests* – if the third test were positive, the result would be very similar and if the third test were negative, the outcome with stopping the testing is slightly more conservative.
- In the case of *all three tests being positive, using EQ5 appears most appropriate*, both for a “2 out of 3” and a “3 out of 3” situation.
- *Using a lower value predicted by EQ6 may be appropriate* in case the chemical is a putative pro-hapten and/or outside of the AD of kDPRA.