P05-49

External validation of alert profilers for genotoxicity hazard within the OECD QSAR Toolbox

Monika Kemény[2], Colin M. North[1], Felix M. Kluxen[2], Markus Frericks[2], Dragana Vukelic[3], Sunice Cao[4], Roustem Saiakhov[5], Mounika Girireddy[5]

[1]BASF Corporation, Research Triangle Park, USA; [2]BASF SE, Ludwigshafen, Germany; [3]BASF Services Europe GmbH, Berlin, Germany; [4]BASF (China) Company Ltd, Beijing, China; [5]MultiCASE Inc, Mayfield Heights, USA;
Presenting author: Monika Kemény[2], email: monika.kemeny@basf.com

MULTICASE

BASF
We create chemistry

Introduction

Genotoxic hazard identification is a key aspect of regulatory decision-making in many countries. Computational toxicity prediction is a useful first-step approach to hazard assessment of pesticide impurities and metabolites. In addition to validated Quantitative Structure-Activity Relationship (QSAR) models, general mechanistic and endpoint specific profiler information from the OECD QSAR Toolbox can be used for chemical grouping. They contain expert knowledge about structure-activity relationships, however, their **predictive performance** is case specific as described by Yordanova D. et al., 2019. To better understand the general predictive performance and how the profilers contribute to a weight of evidence decision in genotoxic endpoints, genotoxicity-relevant profilers in OECD QSAR Toolbox were compared to experimental results from the CASE Ultra AMES mutagenicity and an in vivo MNT database (Chakravarti S.K. and Saiakhov R.D., 2022).

Material and methods

As external input dataset (see Figure 1, process start) served the almost **30000 compounds** from the CASE Ultra **AMES mutagenicity** database and the commercial **in vivo micronucleus** test (MNT) database trained with further data from EFSA evaluations on pesticides and their metabolites (**1059 compounds**, 330 actives/729 inactive). The compounds were profiled through the relevant OECD QSAR Toolbox profilers for genotoxic profiling as recommended by EFSA in the "Guidance on the establishment of the residue definition for dietary risk assessment" (2016) (Table 1) , with and without considering metabolism via the "Rat liver S9 metabolism simulator".

Table 1: Genotoxicity profiling	AMES	MNT
General mechansitic profilers		
DNA binding by OASIS	X	
DNA binding by OECD	X	
Protein binding by OASIS		X
Protein binding by OECD		X
Endpoint specific profilers		
DNA alerts for AMES, CA and MNT by OASIS	X	X
In vitro mugagenicity (AMES test) alerts by ISS	X	
In vivo Mutagenicity (Micronucleus) alerts by ISS		X
Protein binding alerts for Chromosomal aberration by OASIS		X
Metabolism: Rat liver S9 metabolism simulator	With and without	With and without

First the compounds were classified into four quadrants based on the direct outcome of the profiling versus their experimental results to calculate the **general predictive performance** "Full set-without metabolism" (see Figure 1). Those were 2 contradictory conclusions (False Negative (FN) and False Positive (FP)) and 2 aligned conclusions (True Positive (TP) and True Negative (TN)) (See Table 2).

Table 2: Classification scheme for the four quadrants		
Experimental result: / **Profiler result:**	**Negative**	**Positive**
No alert	True Negative (TN)	False Negative (FN)
Alert	False Positive (FP)	True Positive (TP)

The **FN rate** in the profiling was challenged by checking, if compounds with "no alert" in the profiling would be more accurately accomplished by considering **metabolic breakdown**. FN compounds with alert only after metabolic breakdown were re-categorized as TP ("No alerts" with metabolism and "Full set" with metabolism (see Figure 1)). The **FP rate** in the profiling for AMES mutagenicity was challenged for "**deactivating features**" by manually re-categorizing any FP with relevant deactivating features into TN. After each step, the performance statistics were calculated to get an impression about the impact on the general predictive performance (Given as blue squares in Figure 1 and numerically in Table 3 and 4).

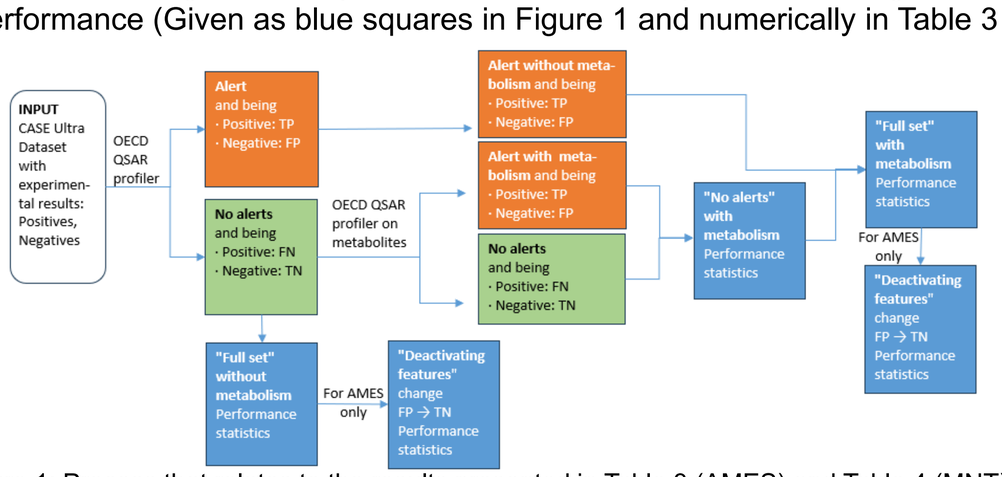

Figure 1: Process that relates to the results presented in Table 3 (AMES) and Table 4 (MNT)

Results and Discussion

Table 3: AMES profilers

General predictive performance	No alerts	Full set	Full set	Deactivating features	Deactivating features
Metabolism	+	-	+	-	+
DNA alerts for AMES, CA and MNT by Oasis					
sensitivity	46,8	56,0	76,6	56,0	76,6
specificity	99,6	93,6	93,2	94,6	94,3
accuracy	91,2	82,9	88,5	83,6	89,3
positive predictivity	95,8	77,8	81,9	80,6	84,4
negative predictivity	90,9	84,2	90,9	84,3	91
Invitro mutagenicity (Ames test) alerts by ISS					
sensitivity	70,6	82,3	94,8	82,3	94,8
specificity	99,0	72,5	71,8	79,9	79,2
accuracy	96,0	75,7	79,2	80,7	84,3
positive predictivity	89,1	59,0	61,7	66,3	68,7
negative predictivity	96,6	89,5	96,6	90,3	96,9
DNA binding by OASIS					
sensitivity	57,8	68,1	86,5	68,1	86,5
specificity	99,3	85,6	84,9	88,8	88,4
accuracy	93,9	80,6	85,4	82,9	87,8
positive predictivity	92,1	65,3	69,6	70,8	74,8
negative predictivity	94,0	87,0	94,0	87,4	94,2
DNA binding by OECD					
sensitivity	69,6	71,1	91,2	71,1	91,2
specificity	98,9	57,9	57,3	63,7	63,2
accuracy	94,1	61,7	67,0	65,8	71,1
positive predictivity	92,8	40,2	46,0	43,8	49,6
negative predictivity	94,2	83,4	94,2	84,7	94,7

Table 4: MNT profilers

General predictive performance	No alert	Full set	Full set
Metabolism	+	-	+
DNA alerts for AMES, CA and MNT by Oasis			
sensitivity	29,8	31,7	52,1
specificity	99,2	87,7	86,9
accuracy	80,6	69,8	75,8
positive predictivity	92,8	54,6	65,1
negative predictivity	79,5	73,3	79,5
Protein binding alerts for Chrom Abs by Oasis			
sensitivity	45,6	31,4	62,7
specificity	98,6	86,9	85,7
accuracy	84,3	69,2	78,3
positive predictivity	92,5	52,9	67,3
negative predictivity	83,0	73,0	83,0
Invivo mutagenicity Micronucleus alerts by ISS			
sensitivity	82,8	81,6	96,8
specificity	92,7	22,3	20,7
accuracy	89,9	41,3	45,0
positive predictivity	81,4	33,1	36,5
negative predictivity	93,3	72,1	93,3
Protein binding by OASIS			
sensitivity	83,6	40,3	90,2
specificity	97,5	65,3	63,6
accuracy	93,3	57,3	72,1
positive predictivity	93,5	35,3	53,9
negative predictivity	93,2	70,0	93,2
Protein binding by OECD			
sensitivity	77,6	35,2	85,4
specificity	98,8	59,9	59,1
accuracy	91,6	52,0	67,5
positive predictivity	97,0	29,1	49,5
negative predictivity	89,6	66,4	89,6

Sensitivity %= TP/(TP+FN);
Sensitivity %= TN/(TN+FP);
Accuracy %= (TP + TN)/(TP+FP+TN+TP)
Positive predictivity %= TP/(TP+FP);
Negative predictivity %= TN/(TN+FN)

Positive predictive performance as measure to identify the percentage of true positives in the pool of all profiled positives varied from 40-78% for AMES and from 29-55% for MNT, indicating substantial differences in the relative performance of profilers for predicting genotoxicity. Consideration of **metabolic simulation** for chemicals without an alert improved positive predictive performance slightly by 2-6% points for AMES and by 3-20% points for MNT, resulting in a positive predictive performance of 46-82% for AMES and 37-67% for MNT profilers. Considering expert-derived **deactivating rules** implemented with CASE Ultra GT_EXPERT model (Hasselgren et al., 2020) provided an additional improvement of 3-7% points for bacterial mutagenicity profiling with regard to positive predictivity. In general, it became obvious that:

- OASIS profilers have overall better accuracy than OECD profilers
- Endpoint specific profilers by OASIS are consistently better in accuracy than their general mechanistic profilers.
- Invivo mutagenicity Micronucleus alerts by ISS have the lowest accuracy and show a high grade of overconservatism, with sensitivity of >80% on costs of low positive predictivity (33%).
- Performance of some individual alerts within a profiler is sometimes so low that it is not sufficient to simply consider the existence or the absence of alerts in a weight of evidence approach.

Evaluators who fail to recognize that OECD QSAR Toolbox profilers are not considered to be directly used for prediction purposes and build up their evidence by combining QSAR assessment with the OECD QSAR Toolbox profiling may be unintentionally over-weighting expert knowledge models in the overall assessment.

Conclusion

Understanding the predictive performance of OECD QSAR Toolbox profilers is important to calibrate the scientific confidence one should apply in a read-across assessment for genotoxicity. The validation against the CASE Ultra datasets clearly confirms that the OECD QSAR Toolbox profilers should not directly be used for genotoxicity prediction or in a weight of evidence approach in combination with QSAR predictions but are considered for building chemical categories for subsequent grouping and Read-across. An in-depth analysis and rework of the profilers using a larger compound space is advisable. The new functionality of alert performance is necessary to select only those alerts with high performance in a Read-across and to disregard those of low performance.

References

Yordanova, D. et al., 2019; Alert performance: A new functionality in the OECD QSAR Toolbox; Computational Toxicology 10, 26-37.
Chakravarti, S.K., Saiakhov, R.D. (2022). MultiCASE Platform for In Silico Toxicology. In: Benfenati, E. (eds) In Silico Methods for Predicting Drug Toxicity. Methods in Molecular Biology, vol 2425. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-1960-5_19.
EFSA Guidance on the establishment of the residue definition for dietary risk assessment, 2016; EFSA Journal 2016; 14(12):4549.
Hasselgren, C. et al., 2020. "Management of pharmaceutical ICH M7 (Q)SAR predictions – The impact of model updates." Regulatory Toxicology and Pharmacology 118: 104807.