ASSESSMENT OF QSAR MODELS FOR FISH TOXICITY BASED ON MODE OF ACTION

E. Benfenati, A. Roncaglioni, A. Lombardo
Istituto di Ricerche Farmacologiche "Mario Negri", via La Masa 19 Milan, Italy

Introduction

Typically QSAR models to predict aquatic toxicity give acceptable results for non-reactive modes of action identifying baseline toxicity. Indeed, compounds may be classified on the basis of their Mode of Action (MOA). Some classification schemes, based on the chemical structure of the compound, exist. One of them was originally developed by Verhaar *et al*. and it is nowadays implemented in a software developed by ECB called Toxtree. For the compounds acting through a narcosis-like MOA, with a baseline toxicity, many logP-based QSAR models give good predictions, but for the other compounds it is more questionable to obtain reliable QSAR predictions.
In this work we compared the predictions for the fish toxicity of some free (DEMETRA [1] and ECOSAR) and commercial (TOPKAT) models with four MOA-specific equations for narcosis MOA on a pool of industrial chemicals.

Material & Methods

DATASET:
LC50 96h data for *Oncorhynchus mykiss were* extracted from the OECD-HPV through the OECD QSAR Toolbox beta version from the HPVC inventory. The data were pruned eliminating mixtures, inorganics, data on compounds with chemical purity < 80% or without purity indications, or formulations. The salts were neutralized. After this pruning 174 compounds with an average LC50 96h were retained for our study.

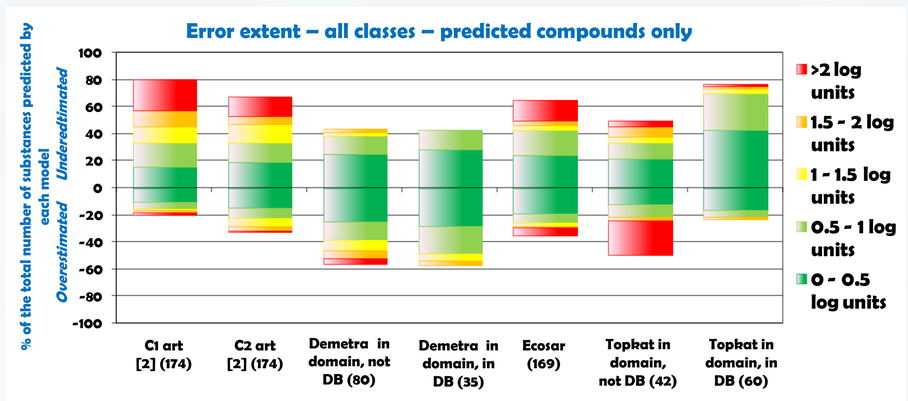
SOFTWARE USED:

- DEMETRA → used to perform the fish toxicity (the predictions are referred to *Oncorhynchus mykiss*)
- TOPKAT v6.1 → used to perform the fish toxicity (the predictions are referred to *Pimephales promelas*)
- ECOSAR v0.99h for EPI Suite v3 → used to predict the fish toxicity
- TOXTREE v1.51 → used to apply the Verhaar classification
- DRAGON v5.5 → used to predict the logP values (MlogP)

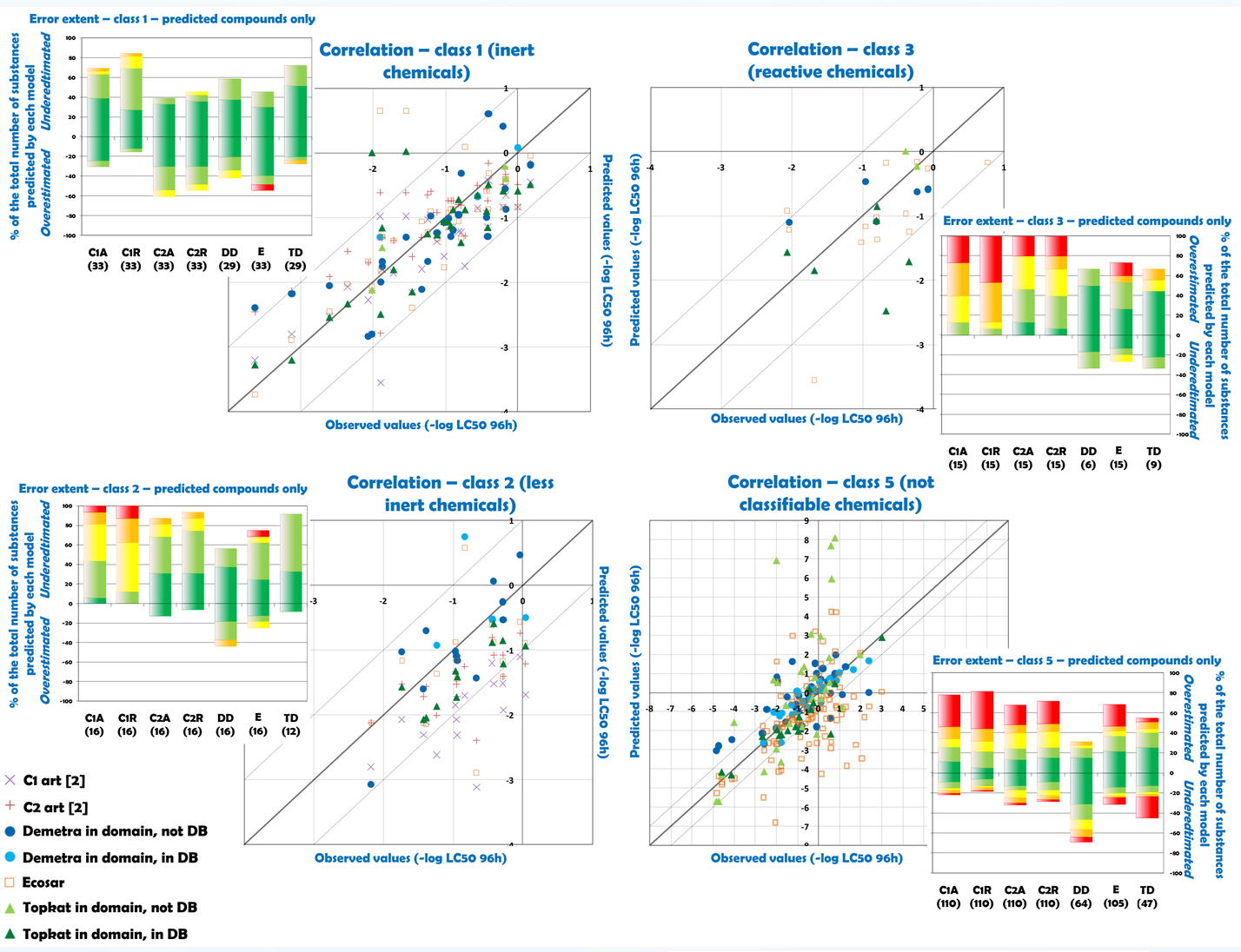
CHEMICAL DOMAIN:
One important characteristic to ensure reliable predictions is to analyze the applicability domain of the QSAR models. This was done by applying the proper domain concept to each model. ECOSAR has not specific chemical domain identification because specific equations are formulated on the basis of the chemical class of the compounds. TOPKAT predictions are considered reliable when included in the Optimum Prediction Space (OPS) and with all fragments covered by the set of data used to build the model. In DEMETRA some classes of compounds are identified with a greater uncertainty in the predictions. The chemical domain of the logP-based models was identified by means of the Verhaar classification scheme considering compounds assigned to class 1 and class 2.

Results and Discussion

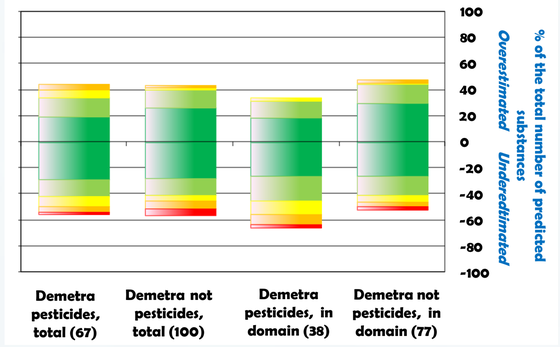
The figure below shows the error extent for the predicted compounds (the number of compounds for each series is between brackets). Overall Demetra has the lowest percentage of errors, with a lower extent and fewer false negatives compared with the other models.



Plots on the right show the correlation of the predictions of the different models once the compounds are sorted by Verhaar class scheme. The histograms report the error trend for the different models analyzed (QSAR for class 1 from Ref [2]: C1A, QSAR for class 1 from Ref [3]: C1R, QSAR for class 2 from Ref [2]: C2A, QSAR for class 2 from Ref [3]: C2R, Demetra compounds in the domain: DD, Ecosar: E and Topkat compounds in the domain: TD). The number of compounds used to calculate the percentages are reported between brackets.
As expected LogP-based models are performing well only for compounds in class 1 and 2, and surprisingly models for class 2 are as good as those for class 1 for non polar narcotics with a lower tendency to underestimate the toxicity. A possible explanation of this odd behavior is that the LogP values used in this study and calculated with the approach by Moriguchi *et al*. do not represent a good approximation of experimental LogP for these compounds. Moreover, also the class assignment done by Toxtree might be doubted for some compounds.
Among the three generalist models (Demetra, Ecosar and Topkat) the best model seems to be Demetra. It produced a lower percentage of errors and a lower amount of compounds with an underestimation of their toxicity, also for the reactive chemicals or for those unable to be classified. Unfortunately, this investigation does not cover class 4 (specifically acting chemicals) since no compounds are assigned to this class.



Demetra model was developed to specifically address ecotoxicity for pesticides. To understand the relevance of this model for REACH compounds pesticide and non pesticide-like compounds have been analyzed separately as reported in the figure on the right. It is clear that the performances of this model, although developed on the basis of pesticide compounds, are equally good for standard organic compounds.



Conclusions

For non-reactive chemicals it is possible to obtain reasonable QSAR approximation of fish acute toxicity using Log P based models but overall other more generalists QSARs seem performing better and in particular Demetra model seems to be the most conservative one, with a lower extent of errors demonstrating its utility not only in addressing pesticide ecotoxicity but also in the study of industrial chemical compounds relevant for REACH legislation.

References

1. Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes, Benfenati, E. (Ed.), Elsevier, Amsterdam, The Netherlands (2007).
2. Overview of structure-activity relationship for environmental endpoints – part 1: general outline and procedure. Report of the EU-DG-XII Project QSAR for Predicting Fate and Effects of Chemicals in the Environment (1995).
3. D.W. Roberts, J.F. Costello, 2003. Mechanisms of action for general and polar narcosis: a difference in dimension. QSAR Comb. Sci. 22 (2003).

We acknowledge financial support from the EC for OSIRIS project.