EVALUATION OF A NEW QSAR MODEL FOR BIOCONCENTRATION FACTOR IN FISH (BCF)

E. Benfenati[1], E. Boriani[1], A. Roncaglioni[1], A. Lombardo[1]

[1] Istituto di Ricerche Farmacologiche "Mario Negri", Milan, Italy

Introduction

Bioconcentration is the uptake of the test substance in an organism relative to the concentration of test substance in the surrounding medium leading to an increase in concentration.

BCF is used in REACH for:

- C & L
- Prioritization (PBT, vPvB)
- Chemical Safety Assessment (CSA)

→ B is BCF > 2000 L/kg = 3.3 in Log unit

→ vB is BCF > 5000 L/kg = 3.7 in Log unit

Both quantitative and qualitative (classification) evaluation may be requested to define if a chemical is PBT (persistent, bioaccumulative, toxic) and vPvB (very persistent, very bioaccumulative). We checked the performances of the QSAR model for BCF developed within the project CAESAR to get: 1) a continuous value (regression method) or 2) a category (classifier): not B, B, or vB.

Materials & Methods

We checked a new QSAR model [1] based on:

- BCF data: 473 compounds produced only according to official guidelines [2].
- Chemical structures & descriptors: structures checked by two persons (10% of the structures pruned); bidimensional chemical descriptors (from Dragon and MDL) have been used.
- CAESAR QSAR model: Combined model including 8 descriptors (combination of 2 individual models comprising 5 descriptors each). The larger role is due to LogP.
- Validation: to further check the QSAR model we used the European database developed within EURAS [3]. Further data on LogP variability has been obtained using the database developed by Arnot *et al.* [4].

Results & Discussions

Regression model

CAESAR model for BCF already produced very good results in the prediction of a test set of compounds (R^2 = 0.80) [1] - see Fig. 1. Here we further checked its predictivity with a new data set. Results are shown in Fig. 2. Also in this case CAESAR showed good results in the prediction.

Fig. 1 Predictions for training and test set for Dimitrov DB

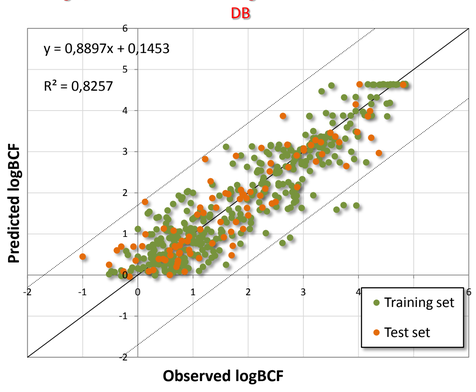


Fig. 2 Predictions of Hybrid Model for compounds in Euras DB

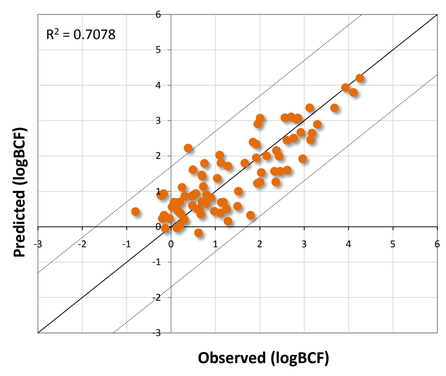
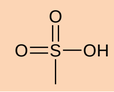
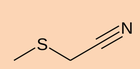
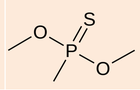
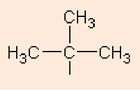
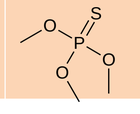
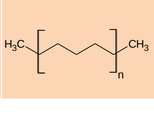


Table 1. Exclusion rules.

COD	RULES	STRUCTURES	COD	RULES	STRUCTURES	COD	RULES	STRUCTURES
1	Sulfonic acid		4	Thiocyanate		7	Br	
2	Phosphothioate		5	T-butyl		8	Cyanate	
3	Phosphorothioate		6	Peroxide		9	Long aliphatic chain	

The CAESAR model is of general applicability, but some restrictions apply. We identified a series of chemical classes which are associated to a higher uncertainty. Their list is given in Table 1. Such a behavior is not unusual. For instance, fluorinated compounds are known to be difficult ones for predictions [5]. In addition we eliminated pesticides because out of REACH.

Classifier

We evaluated the use of CAESAR model as a classifier. While in regression models a confidence interval can be used, in this case we added an uncertainty value of 0.4 log unit, to keep into account experimental and model uncertainty. Table 2 shows the predicted and experimental classes.

The results shows that CAESAR can be used for classification with good performances. We notice that the few wrong assignments are due to relatively small errors in absolute values. We also compared the results obtained applying simple thresholds based on logP, as indicated in some regulations, including REACH [6]. Table 3 shows the results using LogP = 3 and LogP = 4.5 as thresholds. LogP = 3 avoids false negatives, which are somehow present with the threshold of LogP = 4.5. However, the number of false positives is huge, and this reduces the accuracy of the approach.

DIMITROV DB

(a)

Observed \ Predicted	nB	B	vB
nB	288	2	3
B	11	5	4
vB	1	4	33

(b)

Observed \ Predicted	nB	B	vB
nB	257	31	5
B	2	9	9
vB	1	0	37

Table 2. Prediction of CAESAR for Dimitrov DB [2] without (a) and with (b) the 0.4 uncertainty value. For EURAS [3] data results are same without and with the uncertanty (c).

EURAS DB

(c)

Observed \ Predicted	nB	B	vB
nB	80	1	0
B	0	1	0
vB	0	0	3

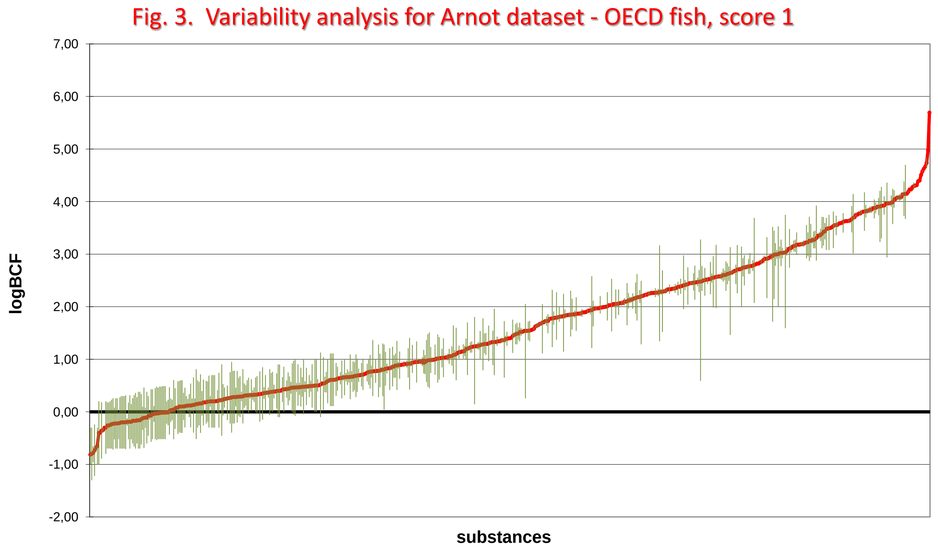
Table 3. LoP and LogBCF thresholds for reliable data as in [4] for the fishes listed by OECD.

ARNOT DB

LogBCF \ LogP	< 3	> 3
nB	201	160
B	0	37

LogBCF \ LogP	< 4.5	> 4.5
nB	319	42
B	4	33

Fig. 3. Variability analysis for Arnot dataset - OECD fish, score 1



Furthermore, we notice that experimental data do have a relevant variability/uncertainty, which is about 0,75 according to Dimitrov [2], 0.41 for Arnot dataset [4] (only reliable data for the fishes suggested by OECD, see Fig. 3) or 0.3-0.4 log units according to EURAS. The uncertainty of the CAESAR model is within this range. The variability/uncertainty of experimental data may limit the distinction between B and vB.

Conclusions

CAESAR model for BCF is a powerful predictive tool, both as a classifier and to get reliable continuous values. Further check with a second test set, using high quality data, confirmed previous results of regression models. As a classifier CAESAR is much more accurate than the simple use of logP as a threshold.

[1] Zhao C., Boriani E., Chana A., Roncaglioni A., Benfenati E., A new hybrid QSAR model for predictioning of bioconcentration factors (BCF), Chemosphere, in press.

[2] Dimitrov S., Dimitrova N., Parkerton T., Comber M., Bonnell M., Mekenyan O., Base-line model for identifying the bioaccumulation potential of chemicals. SAR QSAR Environ. Res. 16, 531–554, 2005.

[3] http://www.euras.be/eng/project.asp?ProjectId=92.

[4] Arnot J.A., Gobas F.A.P.C., A review of bioconcentration factor (BCF) and bioaccumulation factor (BAF) assessments for organic chemicals in aquatic organisms. Environ. Rev . 14, 257–297, 2006.

[5] http://www.pbtprofiler.net.

[6] Guidance on information requirements and chemical safety assessment, Chapter R.11:PBT Assessment, ECHA, May 2008 (http://reach.jrc.it/docs/guidance_document/information_requirements_r11_en.pdf).