

The model for bioconcentration factor (BCF) in fish

Alessandra Roncaglioni

Istituto di Ricerche Farmacologiche “Mario Negri”

CAESAR workshop on QSAR models for REACH
Mario Negri Institute, Milan, Italy
March 10-11, 2009

A vibrant orange and white clownfish is positioned in the top-left corner of the slide, swimming towards the right. The background of the top section is a gradient from orange to yellow.

Outline

- Bioconcentration factor (BCF)
- BCF & REACH
- Data availability & variability
- Modelling BCF in CAESAR
- Comparison with other approaches
- Applet for the BCF model
- Conclusions



Bioconcentration factor (BCF)

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.

$$\text{BCF} = C_f / C_w = k_1 / k_2$$

k_1 = uptake rate constant

k_2 = depuration rate constant

C_f = concentration at steady state conditions

C_w = concentration at at steady state conditions

Experimental test preferred standard: OECD 305
(Bioconcentration flow-through fish test)

- Test duration: **44-116 days**
- Number of likely fish recommended for the test: **132-240 fish**
- Cost for each experiment: **50-100 k€**



BCF in REACH

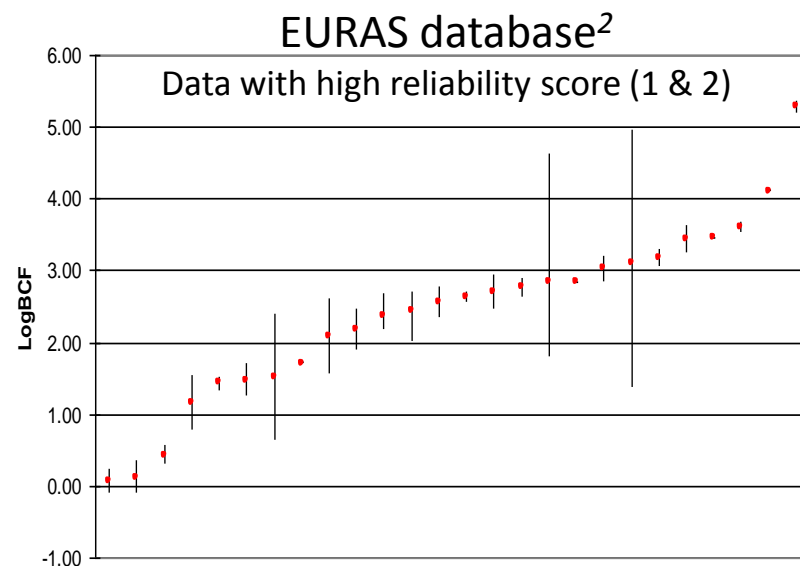
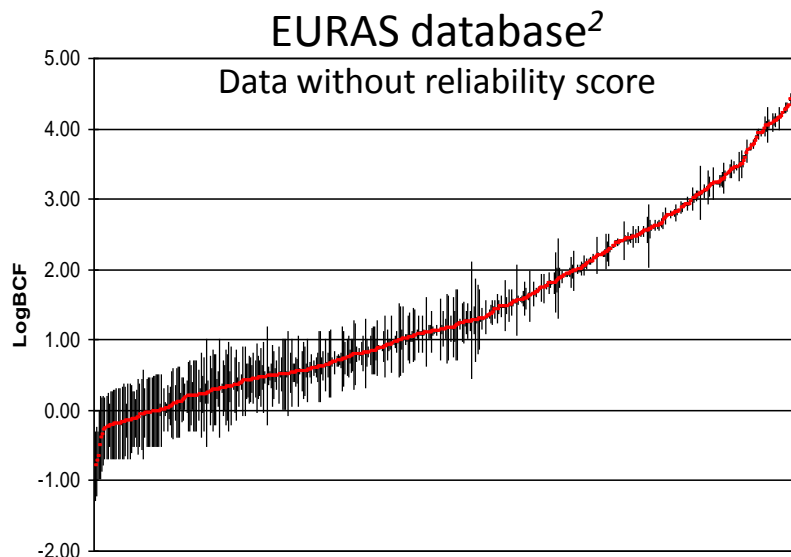
- Potential use of BCF information in REACH is for:
 - C&L
 - Prioritization (PBT, vPvB)
 - Chemical Safety Assessment (CSA)

tonn/year	C&L	B and vB	CSA	BCF value
> 1	X	X		
> 10	X	X	X	
> 100	X	X	X	X

- Quantitative and qualitative (classification) modelling
 - PBT
 - vPvB
 - **B** → $\text{BCF} > 2000 \text{ L/kg} = 3.3 \text{ in Log unit}$
 - **vB** → $\text{BCF} > 5000 \text{ L/kg} = 3.7 \text{ in Log unit}$

Experimental variability

- According to Dimitrov *et al.*¹: **0.75** Log units
- Assessed in other compilations:
 - EURAS database²
 - Considered the “golden” standard for BCF
 - Reliability scores assigned to judge the quality of the experiments

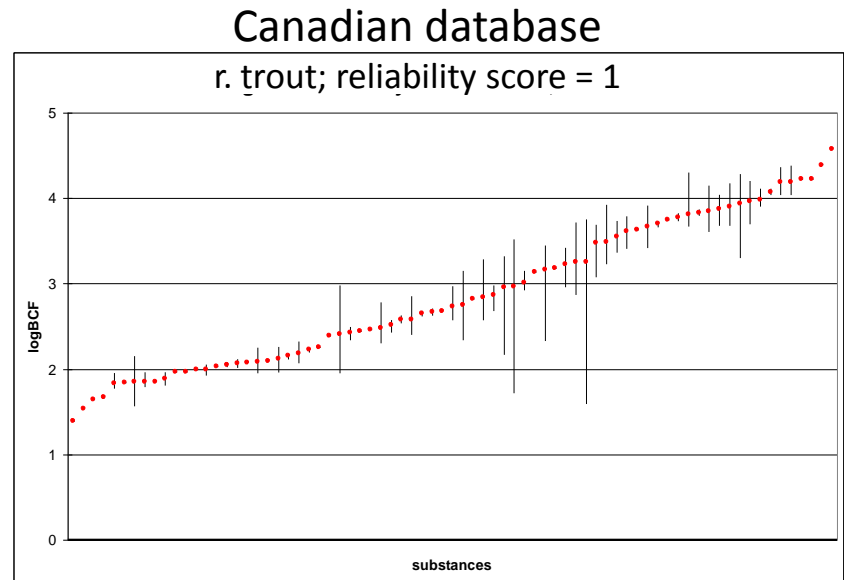
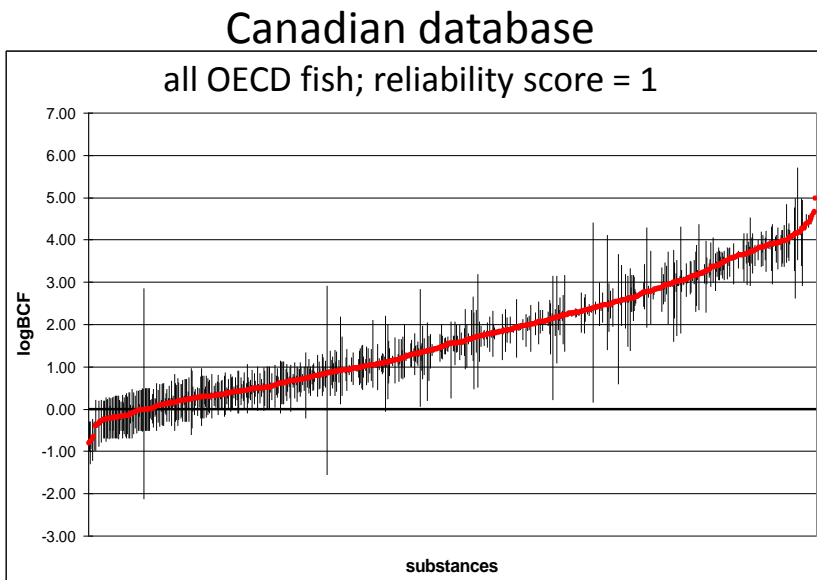


¹SAR QSAR Environ. Res. 16, 2005, 531-554

²<http://www.euras.be/eng/project.asp?ProjectId=92>

Experimental variability

- Assessed in other compilations:
 - Canadian database³
 - Large compilation of bioaccumulation data
 - Reliability scores assigned to judge the quality of the experiments



- **but ...** B – vB range = only **0.4** log units

A vibrant orange and white clownfish is shown swimming in the top-left corner of the slide, set against a dark orange background.

Outline

- Bioconcentration factor (BCF)
- BCF & REACH
- Data availability & variability
- Modelling BCF in CAESAR
- Comparison with other approaches
- Applet for the BCF model
- Conclusions

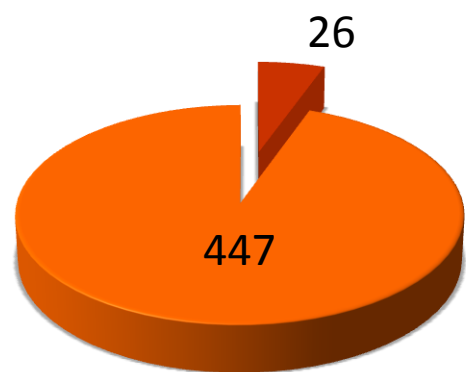


CAESAR modelling for BCF (I)

- Dataset
 - Dimitrov *et al*: data according to official guidelines, widest collection available (~ 500 compounds)
 - Structure check and error pruning (removing ~ 50 comp.)
- Descriptors
 - 2D descr. & lipophilicity: DRAGON, CODESSA, ACD, Pallas, MDL
 - Tautomerism issue (example next slide)
- Models
 - Descriptor selection: GA, heuristic method
 - Classification: AFP
 - Quantitative: MLR, NN (SVM, CP, MLP), GMDH

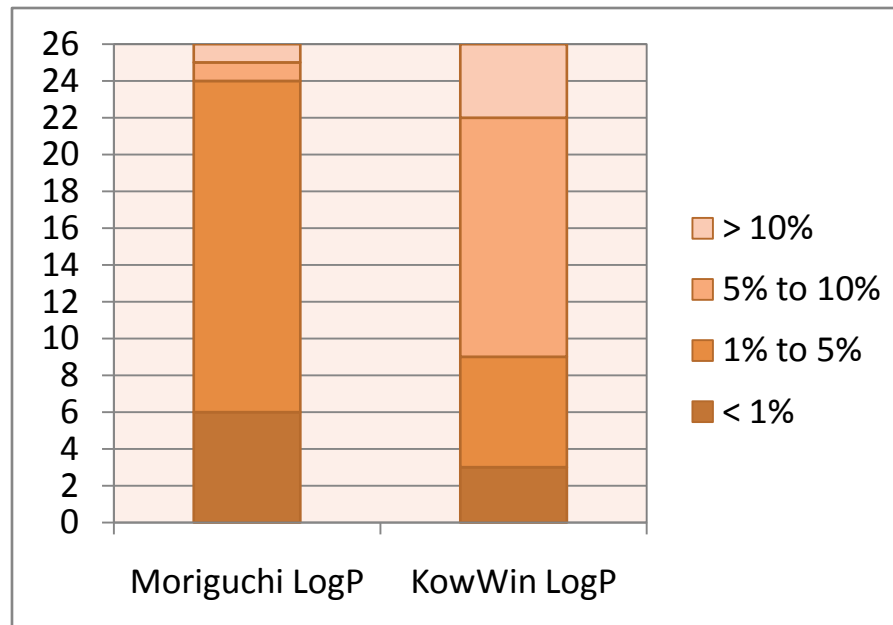
Tautomers behaviour in the BCF dataset

BCF dataset

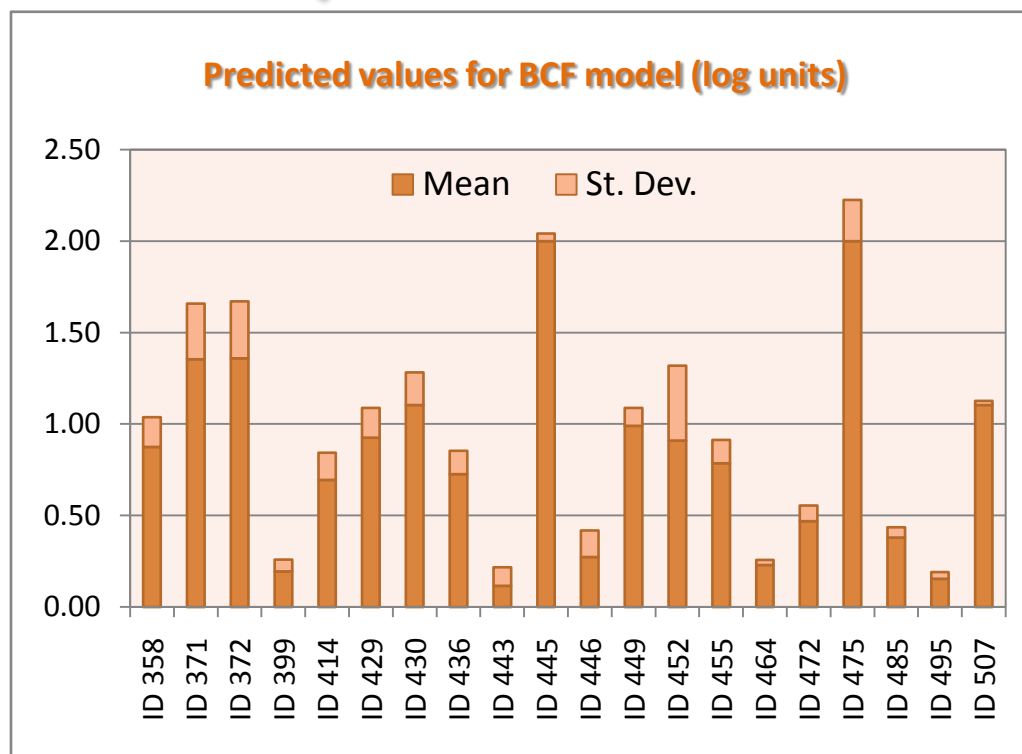
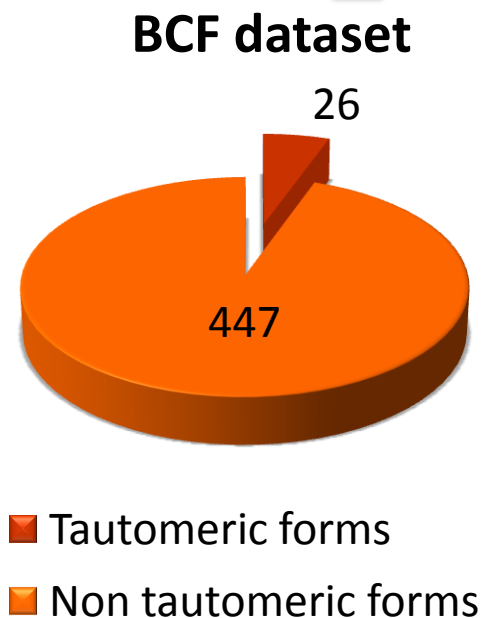


- Tautomeric forms
- Non tautomeric forms

Lipophilicity descriptor variation



Tautomers behaviour in the BCF dataset



CAESAR modelling for BCF (II)

- Hybrid model

- If/then rules in different area of the relation (increase the slope and reduce Y intercept)
- GMDH – self organization

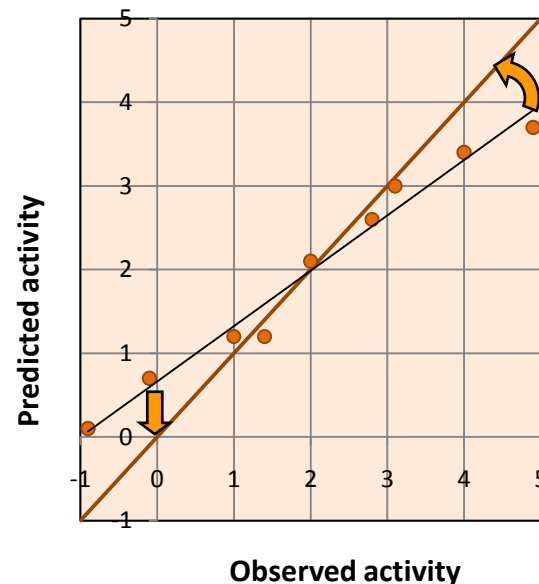
- Validation

- Training / test splitting based on the chemical composition (atomic fragments)

Training set → n = 378

Test set → n = 95

- Cross-validation & test set prediction



Results of CAESAR modelling

Modelling method	Nr. of variables	Descriptors used (SW)	Acc./R ² training set	Acc./R ² _{cv} (loo) training set	Acc./R ² test set
SAR (AFP)	3	Dragon ACD	0.86	0.74	0.81
QSAR (RBF)	5	Dragon MDL	0.81	0.78	0.77
QSAR (CP-NN)	8	Dragon MDL ACD	0.95	0.70	0.76
QSAR (MLP)	5	Dragon	0.80	0.80	0.79
QSAR (GMDH)	4	Dragon MDL	0.76	0.76	0.77
HM (2 models)	8	Dragon MDL	0.83	0.82	0.80
HM (5 models)	36	Dragon MDL KowWin	0.85	0.85	0.80



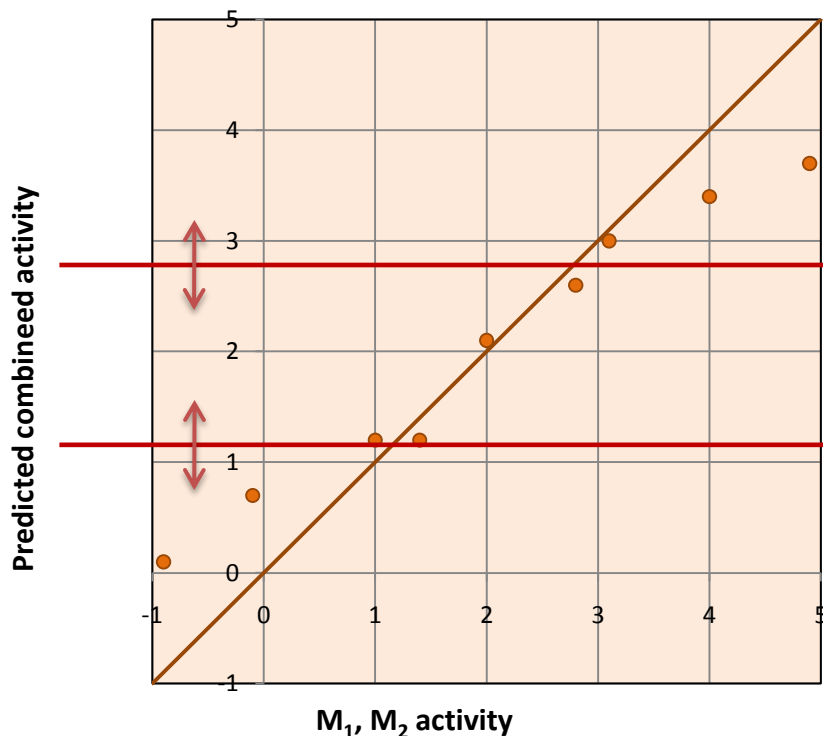
Description of selected model

- Combination of 2 RBF-NN models with 5 descriptors each

MlogP	Moriguchi log of the octanol–water partition coefficient (logP)	} In common in M ₁ and M ₂
BEHp2	Highest eigenvalue n. 2 of Burden matrix/weighted by atomic polarizabilities	
AEige	Absolute eigenvalue sum from electronegativity weighted distance matrix	
GATS5v	Geary autocorrelation – lag 5/weighted by atomic van der Waals volumes	
Cl-089	Cl attached to C1(sp ²)	
X0sol	Solvation connectivity index chi-0	
MATS5v	Moran autocorrelation – lag 5/weighted by atomic van der Waals Volumes	
SsCl	Sum of all (–Cl) E-State values in molecule	

Description of selected model

- Combination of 2 RBF-NN models with 5 descriptors each



$$\text{Pred}_{\text{Comb.}} = 1.05 \text{ mean}(M_1, M_2) - 0.065$$

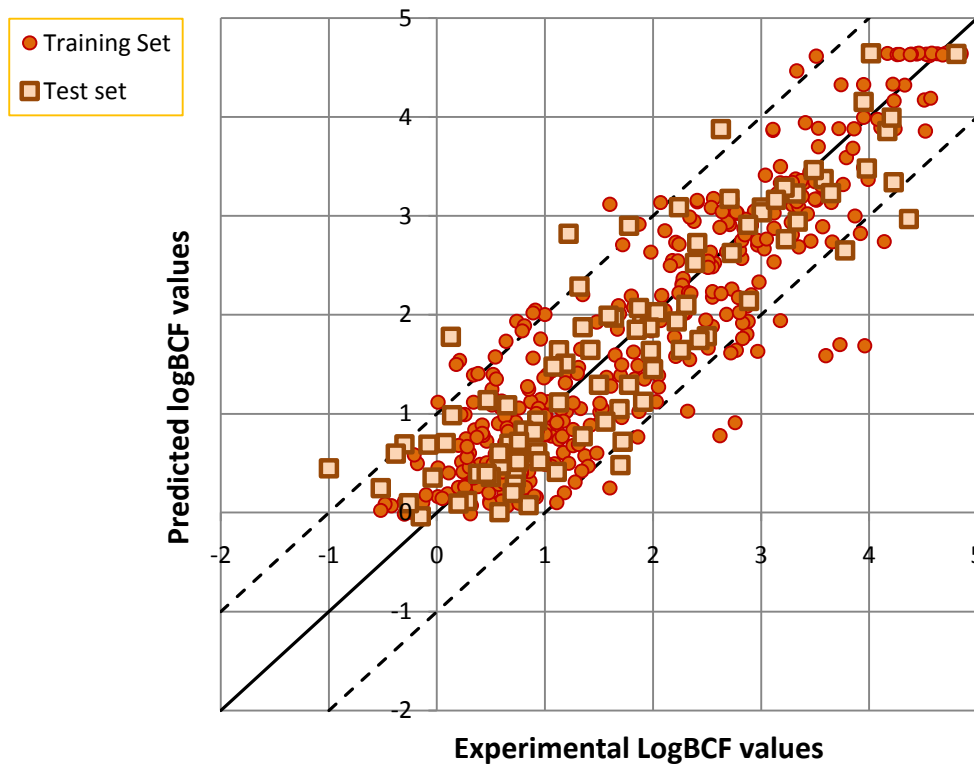
2.41

$$\text{Pred}_{\text{Comb.}} = 0.996 \text{ min}(M_1, M_2) + 0.042$$

1.355

$$\text{Pred}_{\text{Comb.}} = 0.936 \text{ mean}(M_1, M_2) - 0.123$$

Description of selected model





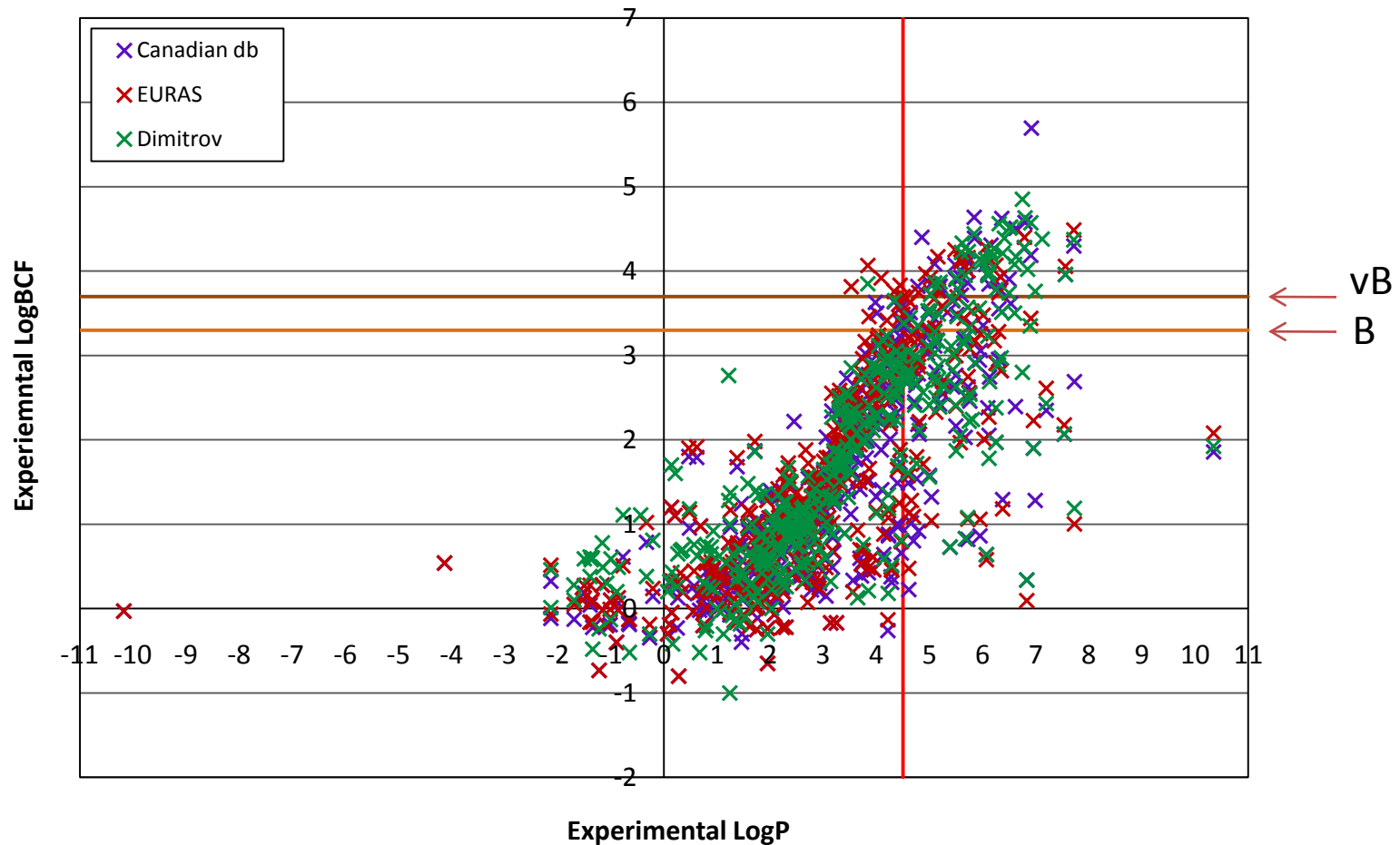
Outline

- Bioconcentration factor (BCF)
- BCF & REACH
- Data availability & variability
- Modelling BCF in CAESAR
- Comparison with other approaches
- Applet for the BCF model
- Conclusions



LogP based BCF estimations

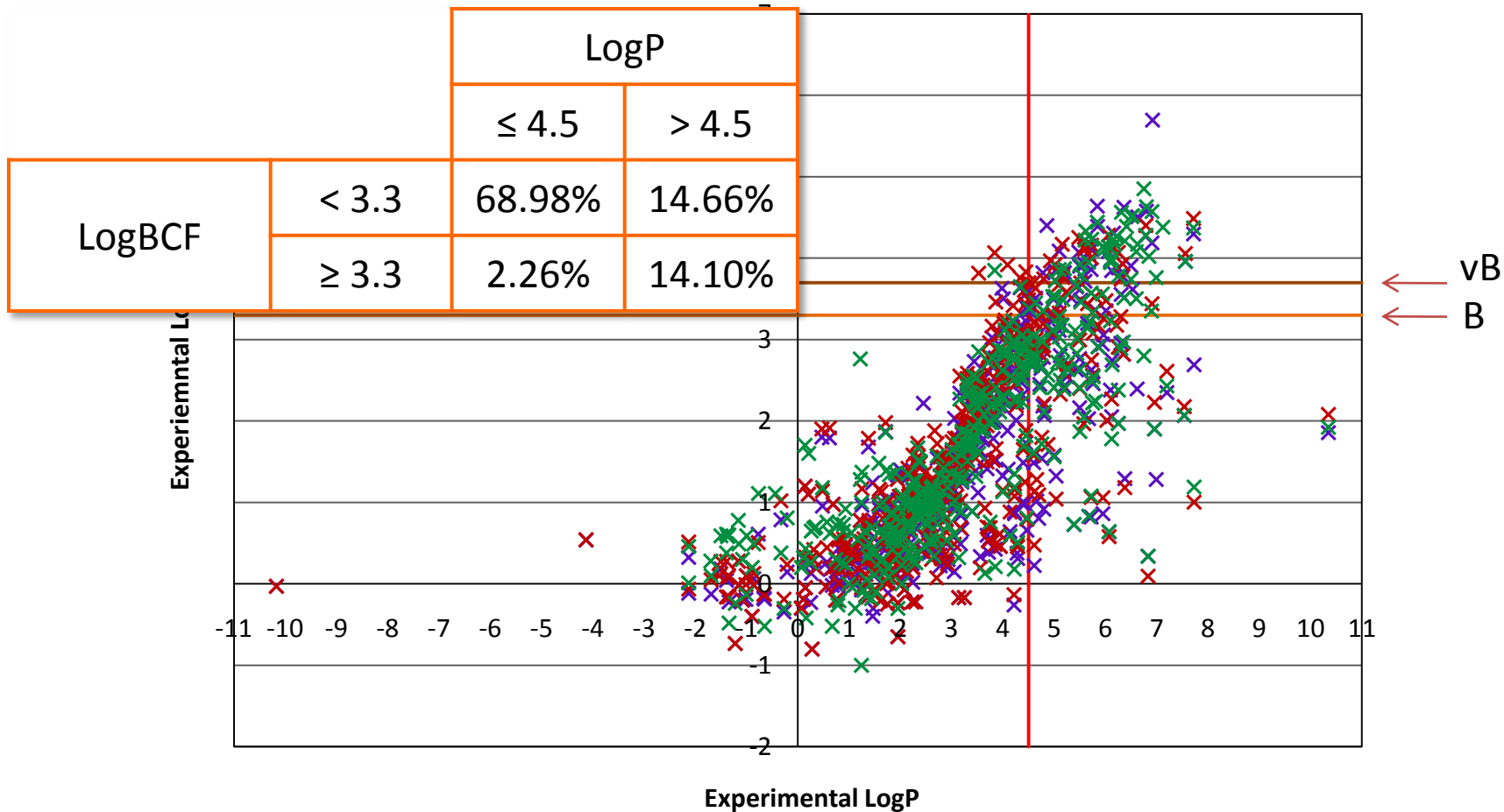
Correlation between LogP and LogBCF





LogP based BCF estimations

Correlation between LogP and LogBCF





External validation & EPI Suite comparison

- Experimental BCF values (median) coming from:
 - Dimitrov *et al.*
 - Canadian
 - EURAS

- Tested models

- CAESAR BCF model
- EPI Suite v4.0*

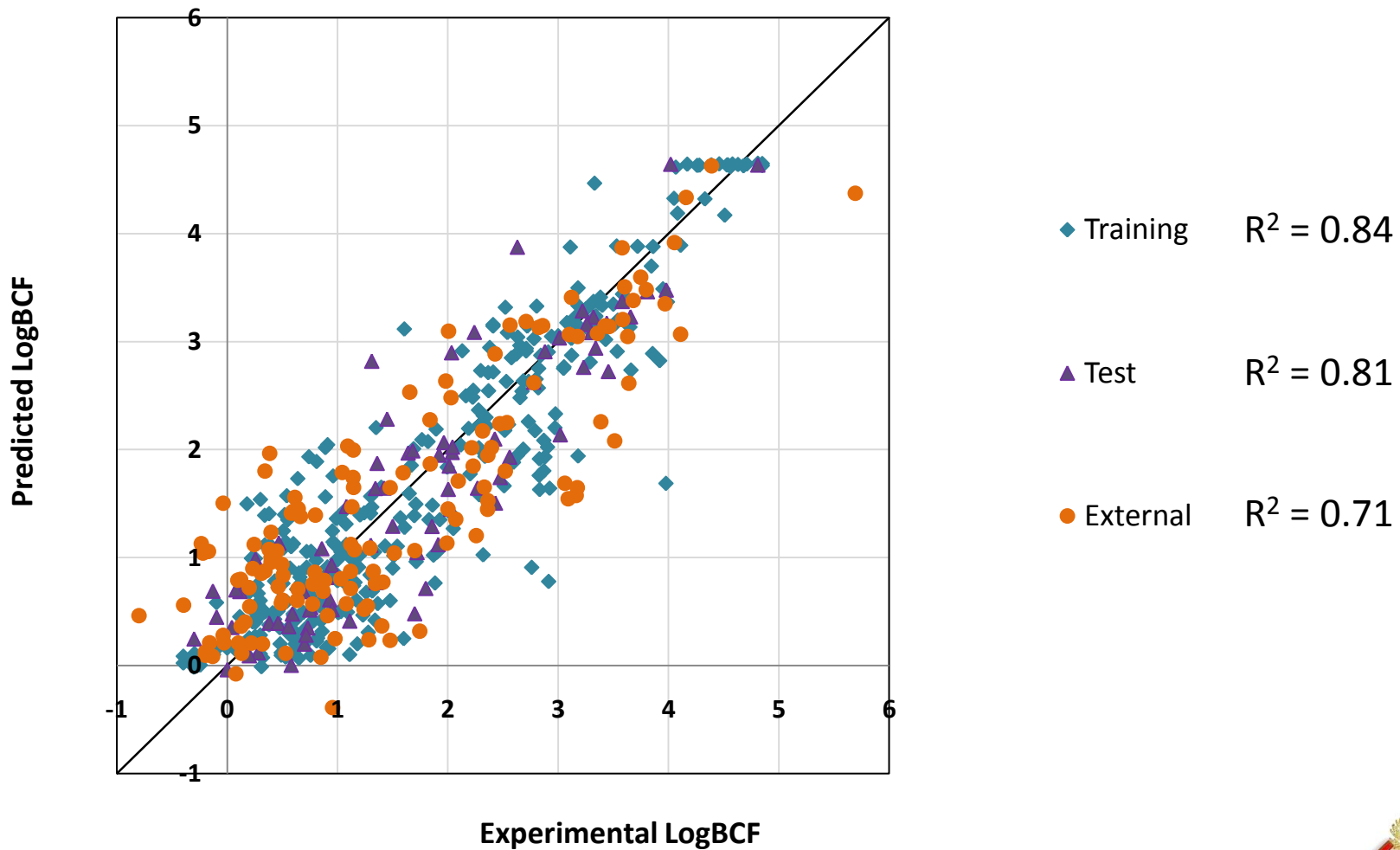


- Three series analyzed

- Training set
- Test set
- Compounds not contained in the databases used respectively to develop the model

	EPI Suite v4.0	CAESAR
Training set	$R^2 = 0.72$ (451)	$R^2 = 0.83$ (368)
Test set	$R^2 = 0.67$ (112)	$R^2 = 0.82$ (91)
External compounds	$R^2 = 0.56$ (71)	$R^2 = 0.61$ (184)

CAESAR model refinement



Classification

CAESAR BCF model (entire dataset)		Estimated LogBCF		
		nB	B	vB
Experimental LogBCF	nB	79.63%	1.09%	0.47%
	B	5.29%	1.40%	0.93%
	vB	2.80%	2.02%	6.38%



643 compounds

634 compounds



EPI Suite v4.0		Estimated LogBCF		
		nB	B	vB
Experimental LogBCF	nB	76.03%	2.68%	3.00%
	B	4.89%	1.10%	1.26%
	vB	2.05%	2.37%	6.62%

CAESAR BCF model (pruned dataset)		Estimated LogBCF		
		nB	B	vB
Experimental LogBCF	nB	83.39%	1.29%	0.37%
	B	4.80%	1.66%	0.55%
	vB	0.74%	1.48%	5.72%



542 compounds

A vibrant orange and white clownfish swimming in an aquarium setting, positioned in the top-left corner of the slide.

Outline

- Bioconcentration factor (BCF)
- BCF & REACH
- Data availability & variability
- Modelling BCF in CAESAR
- Comparison with other approaches
- Applet for the BCF model
- Conclusions



CAESAR BCF model applet

<http://www.caesar-project.eu/>



Conclusions

- New integrated models for BCF with better performance than available methods
- Final model fully implemented and appropriate documentation (QMRF) ensures transparency and reproducibility
- Appreciation of similarity and confidence in prediction
- Feasible to use output as a definitive value or in classification
- Experimental data & quality check
- Use within ITS in collaboration with OSIRIS project



Acknowledgments

- Lab. team & sw developers
 - E. Benfenati
 - C. Zhao
 - E. Boriani
 - A. Lombardo
 - A. Chana
 - O. Schifanella
 - C. Milan
 - R. Gonella Diaza
 - A. Manganaro
 - A. Gomez Delgado
 - D. Bigoni
 - A. Cassano
- CAESAR partners
 - P2 CSL DEFRA
 - P3 BCX
 - P4 POLIMI
 - P5 KM
 - P6 LJMU
 - P7 UFZ
 - P8 NIC-LJ
 - P9 TNO
- OSIRIS project

Thank you!