I metodi di non testing per la stima delle proprietà (eco)tossicologiche

Alessandra Roncaglioni

Istituto di Ricerche Farmacologiche “Mario Negri”

Workshop sulle attività di ricerca ISPRA nell’ambito del regolamento REACH

Roma, 13 Dicembre 2010

MARIO NEGRI
ISTITUTO DI RICERCHE FARMACOLOGICHE

Outline

- QSAR requirments for REACH
- Proposal for assessing the AD
- Integrated use of NTM for BCF
- The role of ANTARES project in the validation of NTM for REACH

According to *REACH regulation* (Annex XI) a (Q)SAR is VALID if:

- the model is recognized scientifically valid;
- the substance is included in the applicability domain of the model;
- results are adequate for classification and labelling and for risk assessment;
- adequate documentation of the methods provided.

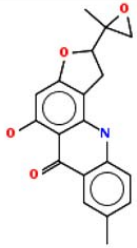
Components for the AD Index in CAESAR

- **Chemiometric check** (descript. space)
- **Similarity index** (chemical; sub-indices)
- **Fragments for outliers** (output space)
- **Prediction Concordance** (tox exploration)
- **Prediction Accuracy** (output space)
- **Uncertainty** (output space)
- **Visualisation of similar substances**

Similarity search

CAESAR QSAR model for Mutagenicity - version 1.0

Prediction for the compound no. 1: Cc1ccc2Nc3c4CC(Oc4cc(O)c3C(=O)c2c1)C1(C)CO1

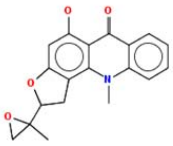


Activity: Mutagen

Remarks for the prediction:

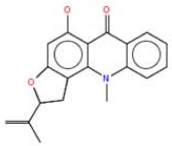
VISUALIZATION OF SIMILAR SUBSTANCES is very useful for the reporting

The following chemicals similar to the query compound have been identified in the CAESAR database:



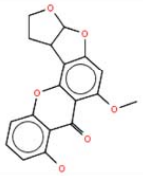
Dataset id: 10
SMILES: O=C5c1ccccc1N(c3c5(c(O)cc2OC(Cc23)C4(OC4)(C)))C
Similarity: 0.99

Experimental class: Mutagen
Predicted class: Mutagen



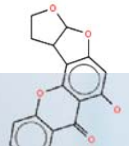
Dataset id: 772
SMILES: O=C2c1ccccc1N(c4c2c(O)cc3OC(C(=C)C)Cc34)C
Similarity: 0.922

Experimental class: Mutagen
Predicted class: Mutagen



Dataset id: 1963
SMILES: O=C1c5c(O)cccc5(Oc3c1c(OC)cc2OC4OCCC4(c23))
Similarity: 0.828

Experimental class: NON-Mutagen
Predicted class: NON-Mutagen



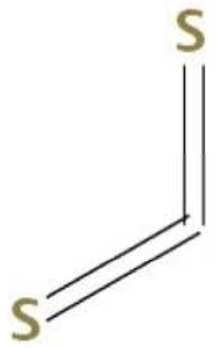
Dataset id: 3769
SMILES: O=C4c5c(O)cccc5(Oc2c4(c(O)cc1OC3OCCC3(c12)))
Similarity: 0.825

Experimental class: NON-Mutagen
Predicted class: NON-Mutagen

Chemicals outside descriptor range

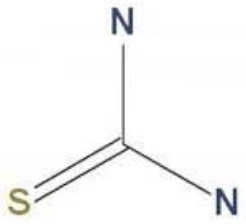
CAESAR QSAR model for bioconcentration factor (BCF) in fish

Prediction for the compound no. 1: S=C=S



BCF value: 2 (L/Kg) whole body weight
Log BCF value: 0.24
Remarks for the prediction:
Descriptors for this compound have values outside the descriptor range for the compounds of the training set.

The following chemicals similar to the query compound have been identified in the CAESAR database:



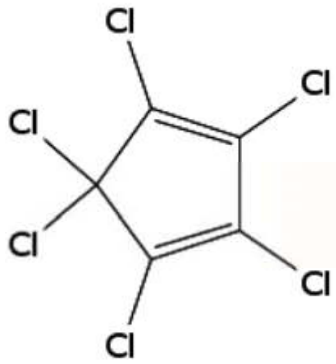
Dataset id: 399
SMILES: NC(=S)N
Similarity: 0.348

Experimental Log BCF: 0.30
Predicted Log BCF: 0.46

Identified fragments linked with low reliability of prediction

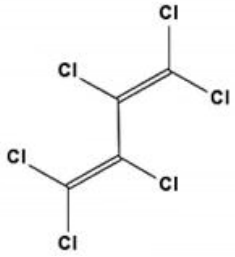
CAESAR QSAR model for bioconcentration factor (BCF) in fish

Prediction for the compound no. 1: ClC=1C(Cl)=C(Cl)C(Cl)(Cl)C=1Cl



BCF value: 25 (L/Kg) whole body weight
Log BCF value: 1.39
Remarks for the prediction:
Presence of chemical features in the compound (6 Cl atoms in the molecule) that might be associated with a lower reliability of the predicted value.

The following chemicals similar to the query compound have been identified in the CAESAR database:



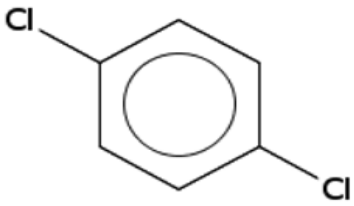
Dataset id: 233
SMILES: ClC(=C(Cl)Cl)C(=C(Cl)Cl)Cl
Similarity: 0.483

Experimental Log BCF: 3.35
Predicted Log BCF: 2.69

Compounds with large similarity index

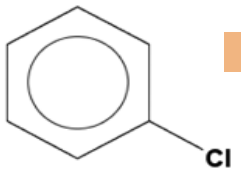
CAESAR QSAR model for bioconcentration factor (BCF) in fish

Prediction for the compound no. 1: Clc1ccc(Cl)cc1



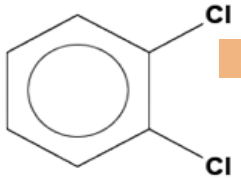
BCF value: 174 (L/Kg) whole body weight
Log BCF value: 2.24
Remarks for the prediction:

The following chemicals similar to the query compound have been identified in the CAESAR database:



Dataset id: 205
SMILES: Clc1ccccc1
Similarity: 0.961

Experimental Log BCF: 1.13
Predicted Log BCF: 1.64



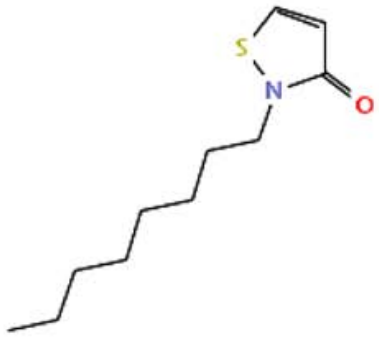
Dataset id: 209
SMILES: Clc1c(cccc1)Cl
Similarity: 0.959

Experimental Log BCF: 2.24
Predicted Log BCF: 2.22

Compounds with a low similarity index

CAESAR QSAR model for bioconcentration factor (BCF) in fish

Prediction for the compound no. 6: O=C1C=CSN1CCCCCCCC

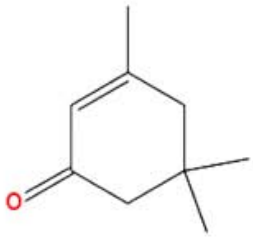


BCF value: 103 (L/Kg) whole body weight

Log BCF value: 2.01

Remarks for the prediction:

The following chemicals similar to the query compound have been identified in the CAESAR database:



Dataset id: 477

SMILES: CC1(C)CC(=CC(=O)C1)C

Similarity: 0.467

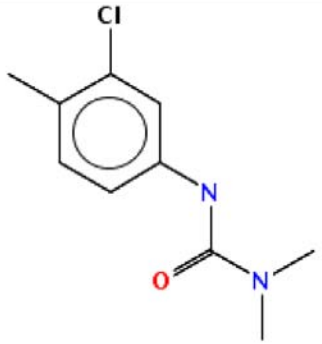
Experimental Log BCF: 0.14

Predicted Log BCF: 0.99

Concordance of experimental value for similar molecules

CAESAR QSAR model for Carcinogenicity - version 1.0

Prediction for the compound no. 1: CN(C)C(=O)NC1=CC(Cl)=C(C)C=C1

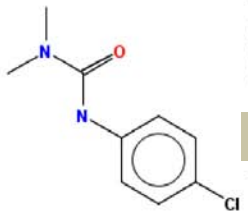


Carcinogenic: Non-Positive

Class indices: Positive=0.079, Non-Positive=0.921

Remarks for the prediction:

The following chemicals similar to the query compound have been identified in the CAESAR database:

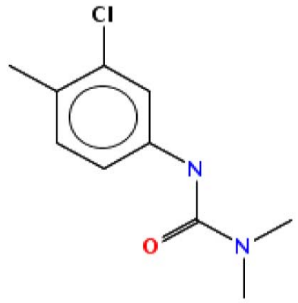


Dataset id: 162
SMILES: Clc1ccc(cc1)NC(=O)N(C)C
Similarity: 0.884

Experimental class: Positive
Predicted class: Non-Positive

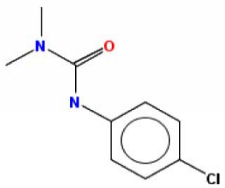
Accuracy of prediction for similar molecules

Prediction for the compound no. 1: CN(C)C(=O)NC1=CC(Cl)=C(C)C=C1

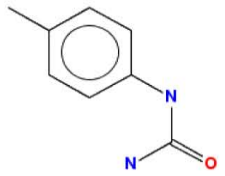


Carcinogenic: Non-Positive
Class indices: Positive=0.079, Non-Positive=0.921
Remarks for the prediction:

The following chemicals similar to the query compound have been identified in the CAESAR database:



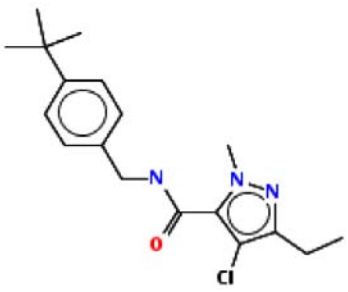
Dataset id: 162
SMILES: Clc1ccc(cc1)NC(=O)N(C)C
Similarity: 0.884
Experimental class: Positive
Predicted class: Non-Positive



Dataset id: 754
SMILES: O=C(Nc1ccc(cc1)C)N
Similarity: 0.857
Experimental class: Non-Positive
Predicted class: Positive

Uncertainty

Prediction for the compound no. 1: CCC1=NN(C)C(C(=O)NCC2=CC=C(C=C2)C(C)(C)C)=C1Cl



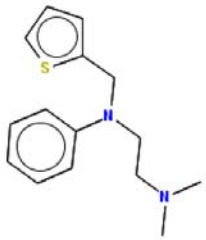
Carcinogenic: Non-Positive

Class indices: Positive=0.491, Non-Positive=0.509

Remarks for the prediction:

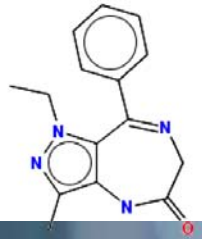
Borderline output, in the this example the carcinogenicity probability values predicted: positive=0.491 and negative=0.509 show a **VERY HIGH UNCERTAINTY IN THE PREDICTION**

The following chemicals similar to the query compound have been identified in the CAESAR database:



Dataset id: 433
SMILES: CN(CCN(c1ccccc1)Cc1cccs1)C
Similarity: 0.693

Experimental class: Non-Positive
Predicted class: Non-Positive



Dataset id: 696
SMILES: O=C1Nc2c(n(nc2C)CC)C(=NC1)c1ccccc1
Similarity: 0.69

Experimental class: Non-Positive
Predicted class: Positive

Developmental Toxicity Prediction

Developmental Toxicity

STEP 1 2 3 4

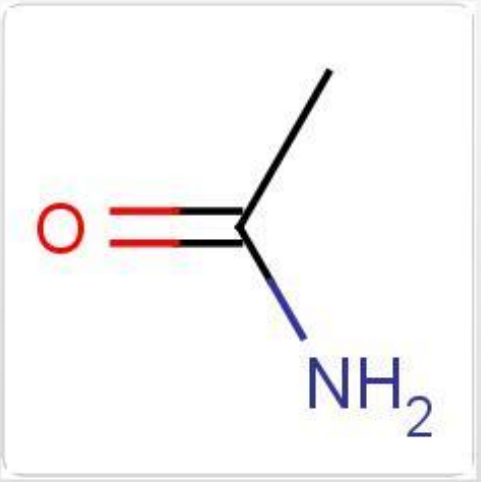
LOAD

Get Started

IMPORT STRUCTURE(S) FROM FILE

DRAW STRUCTURE

ID	IUPAC Name	SMILES
1	5-(butan-2-yl)-5-ethyl-1,3-diazinane-2,...	CCC(C)C1(CC)C(=O)NC(=O)NC1=O
2	acetamide	CC(N)=O
3	(E)-N-ethylidenehydroxylamine	C\C=N\O
4	acetonitrile	CC#N
5	acetaldehyde	CC=O



CLEAR

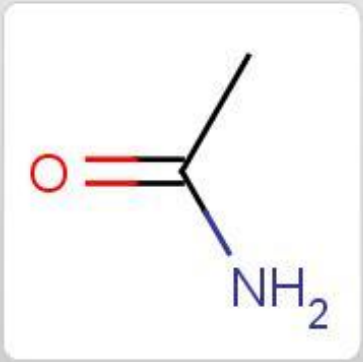
CLEAR ALL

Developmental Toxicity Prediction

Developmental Toxicity

STEP 1 2 3 4

ID	IUPAC Name	SMILES	Assessment	
1	5-(butan-2-yl)-5-ethyl-1,3-...	CCC(C)C1(CC)C(=O)NC...	Developmental toxicant (compound into AD)	
2	acetamide	CC(N)=O	Developmental toxicant (compound possibly out of AD)	
3	(E)-N-ethylidenehydroxyl...	C\C=N\O	n.a. (compound out of AD)	
4	acetonitrile	CC#N	Developmental NON-toxicant (compound possibly out of AD)	
5	acetaldehyde	CC=O	Developmental NON-toxicant (compound into AD)	



MOLECULE ID	2
SMILES	CC(N)=O
IUPAC NAME	acetamide
PREDICTED VALUE	Developmental toxicant
APPLICABILITY DOMAIN	Predicted substance could be out of the Applicability Domain of the model.
ASSESSMENT	Developmental toxicant (compound possibly out of AD)

3

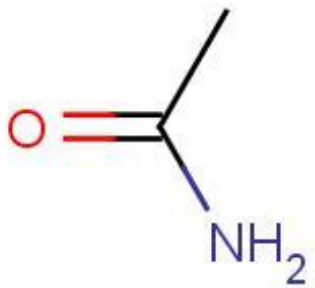
Developmental Toxicity Prediction

Developmental Toxicity

STEP 1 2 3 4

ID	IUPAC Name	SMILES	Assessment	
1	5-(butan-2-yl)-5-ethyl-1,3-...	CCC(C)C1(CC)C(=O)NC...	Developmental toxicant (compound into AD)	
2	acetamide	CC(N)=O	Developmental toxicant (compound possibly out of AD)	
3	(E)-N-ethylidenehydroxyl...	C\C=N\O	n.a. (compound out of AD)	
4	acetonitrile	CC#N	Developmental NON-toxicant (compound possibly out of AD)	
5	acetaldehyde	CC=O	Developmental NON-toxicant (compound into AD)	

SIMILAR MOLECULES



Global Applicability Domain Index	0.718
Similar molecules with known experimental value index	0.749
Concordance for similar molecules index	0.474
Accuracy of prediction for similar molecules Index	1
Model descriptors range check	✓

4

Results with the AD Index

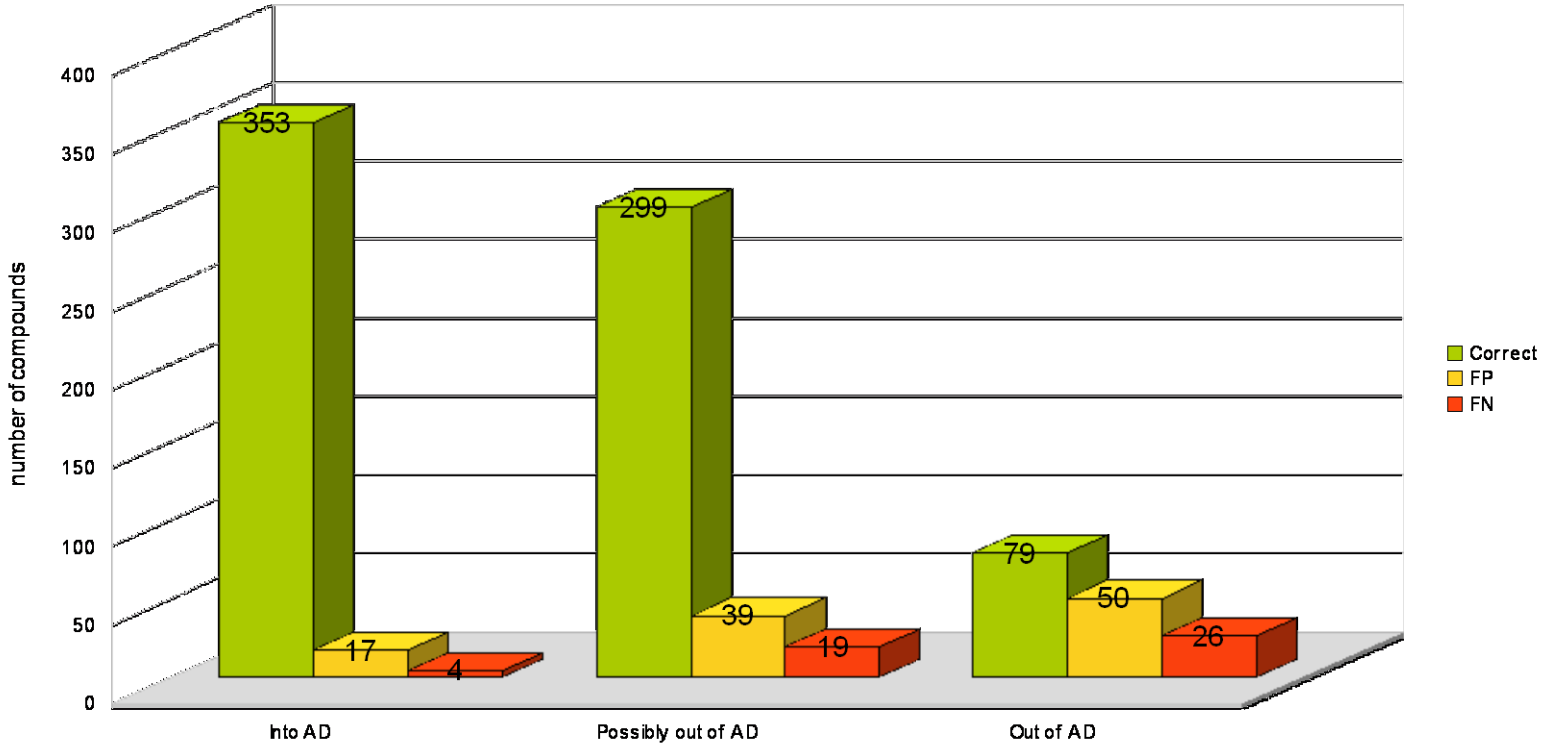
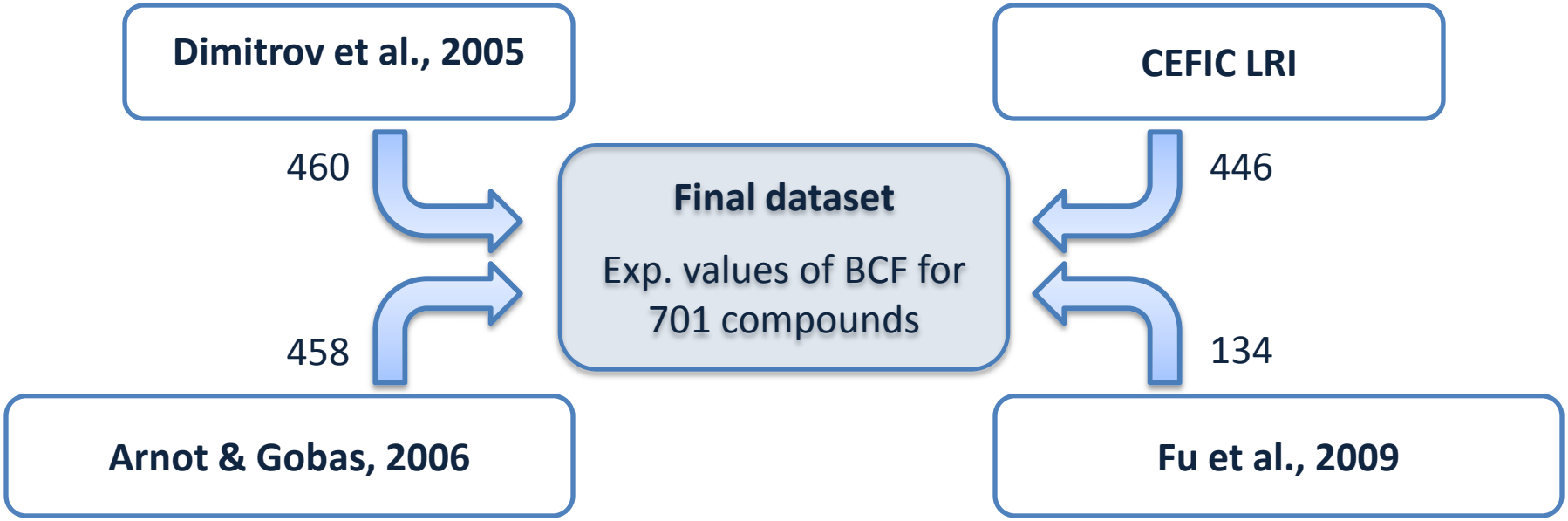


Chart reports statistics for **Mutagenicity** model Test-Set (**836** compounds) divided by three Applicability Domain classes:

* Prediction into AD
* Prediction possibly out of AD
* Prediction out of AD

Experimental data sources for BCF



The salts were lead up to the acid form and the following data were eliminated:

- All substances with no sufficient data to generate correct structure,
- Mixtures,
- Inorganic compounds,
- Duplicates
- All data reported as “not reliable” in the original dataset.

Experimental variability

Dimitrov et al., 2005	Variability of exp. data = 0.75 l.u.
Fu et al., 2009	Variability of exp. data not indicated
Arnot & Gobas, 2006	OECD 305 fish, all score: exp. variability = 0.61 l.u. OECD 305 fish, score 1: exp. variability = 0.48 l.u.
CEFIC LRI	Variability of exp. data = 0.45 l.u.
Final dataset	Variability of exp. data = 0.63 l.u.

Difference between B (3.3 l.u.) and vB (3.7 l.u.) threshold = 0.4 l.u.

QSAR models used (I)

CAESAR
- Built on 378 compounds of Dimitrov et al., 2005
- Based on 8 descriptors
- 2 models combined together
- Applicability domain clearly indicated (chemical range, fragments, tox results)

T.E.S.T. v3.3
- Built on 610 compounds of Dimitrov et al., 2005, CEFIC LRI and Arnot & Gobas, 2006
- Consensus model based on 5 models
- Applicability domain not clearly indicated

BCFBAF v3.00 (in EPISuite v4.0)
- Built on compounds of Arnot & Gobas, 2006
- Meylan et al., 1999 model (logP-based): based on fragments
- Arnot & Gobas, 2006 (based on biotransformation rate)
- Applicability domain clearly indicated (chemical range)

QSAR models used (II)

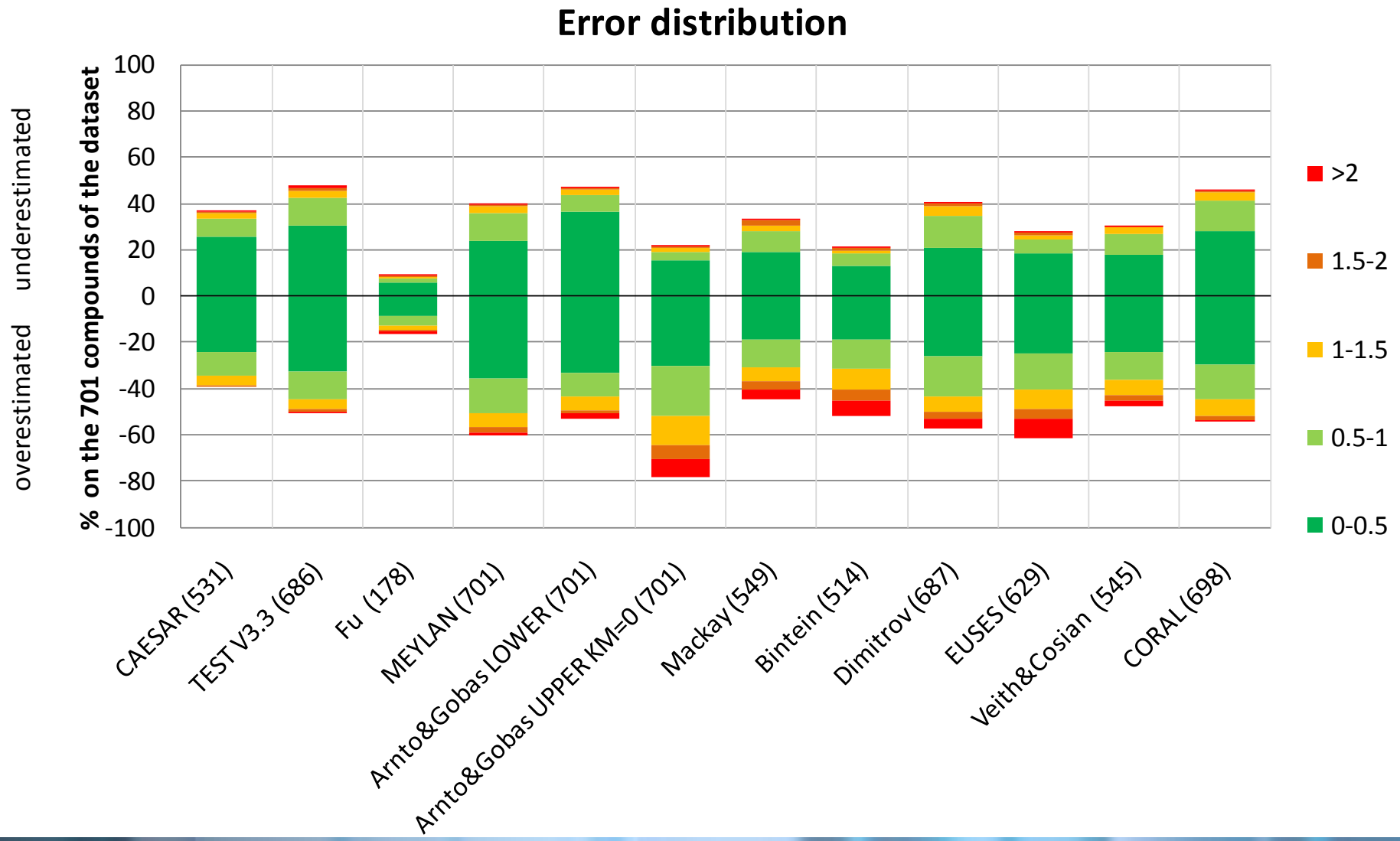
Fu et al., 2009
- Built on 73 monovalent acids and 65 monovalent bases (Fu et al., 2009 DB)
- Based on logP and pKa

ChemProp v5.1.5
- LogP-based equations:
  - Mackay, 1982
  - Bintein et al., 1993
  - Dimitrov et al., 2002
  - EUSES, 1996

LogP-BASED EQUATIONS
- Veith & Kosian, 1983

CORAL
- Built on 1037 compounds of Dimitrov et al., 2005, CEFIC LRI, Arnot & Gobas, 2006 and Footprint
- Based on indices of presence of atoms calculated using SMILES code

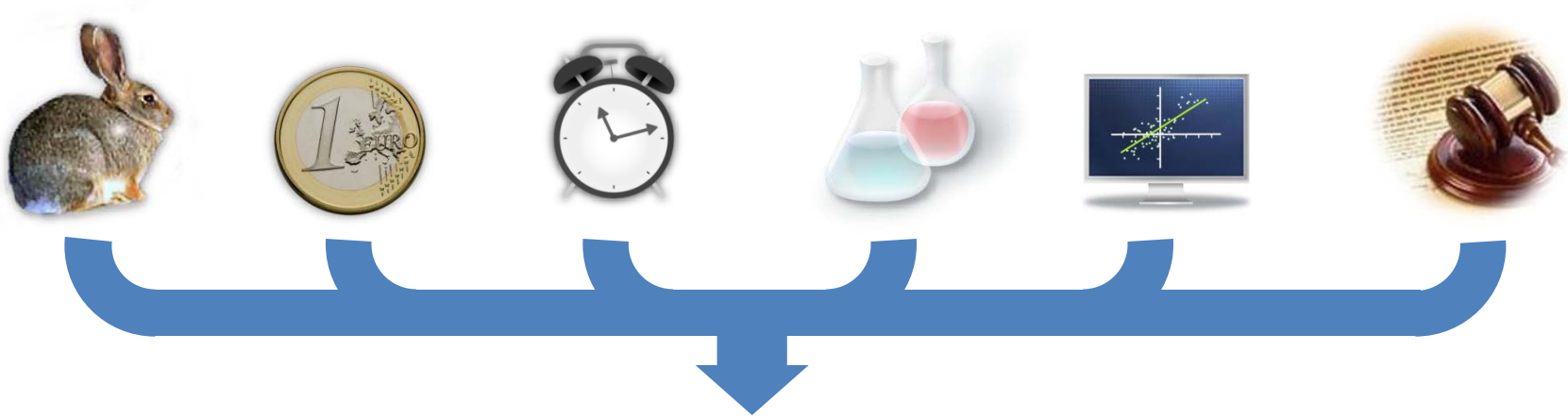
Regression models - errors



Consensus evaluation (11 models)

No. of compounds with all models with an error < 0.5	91 (13.0%)
No. of compounds with all models with an error > 1	4 (0.6%)
No. of compounds with uncertain prediction	152 (21.7%)
No. of compounds with most of models with an error < 0.5	383 (54.6%)
No. of compounds with most of models with an error between 0.5 and 1	73 (10.4%)
No. of compounds with most of models with an error > 1	93 (13.3%)

Aims & scope



Promotion of non testing methods (NTM) for their use in the REACH context linking *scientists*, *regulators* and *industries*

to evaluate and validate existing NTM for their application according to REACH needs

Actions

Action 1: **Survey of current methods** for the compliance to REACH.

Action 2: Identification of the **criteria for the non-testing methods**.

Action 3: **Identification of** suitable **experimental databases/data sets** for the (eco)toxicological and environmental endpoints for REACH.

Action 4: **List of (Q)SAR models** for the ecotoxicological, toxicological and environmental endpoints for REACH, and their review.

Action 5: **Validation** of non-testing methods (incl. read-across).

Action 6: Identification of **boundaries for best use** of models (applicability domain) **and** of the **assessment factors**.

Action 7: **Architecture for integration** of different non-testing methods for best performances and coverage of applicability.

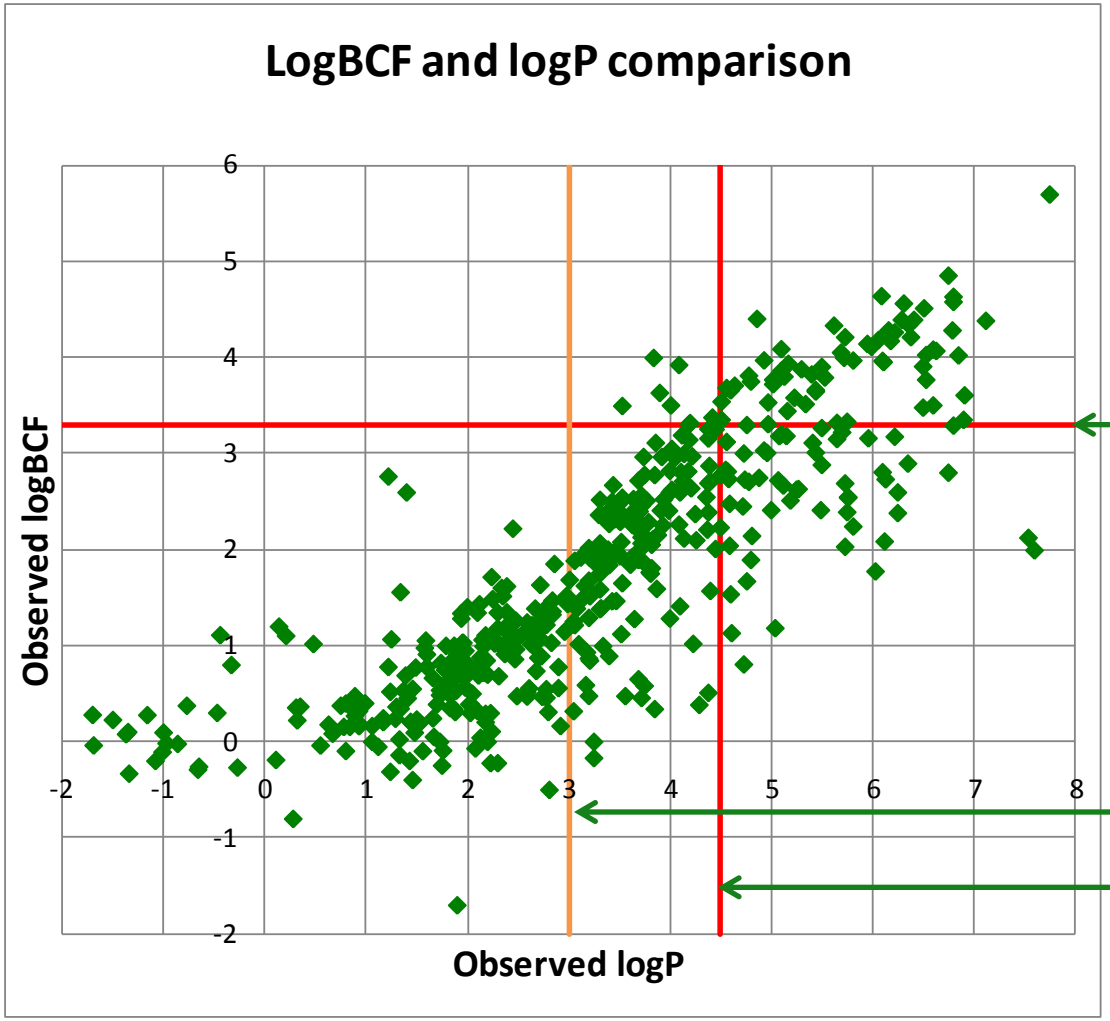
Actions for management & dissemination

Grazie!

Classification based on logP

LogP thresholds

REACH ⇒ **logP ≤ 3 ⇒ nB** **ECHA guidelines** ⇒ **logP ≤ 4.5 ⇒ nB**



logBCF 3.3

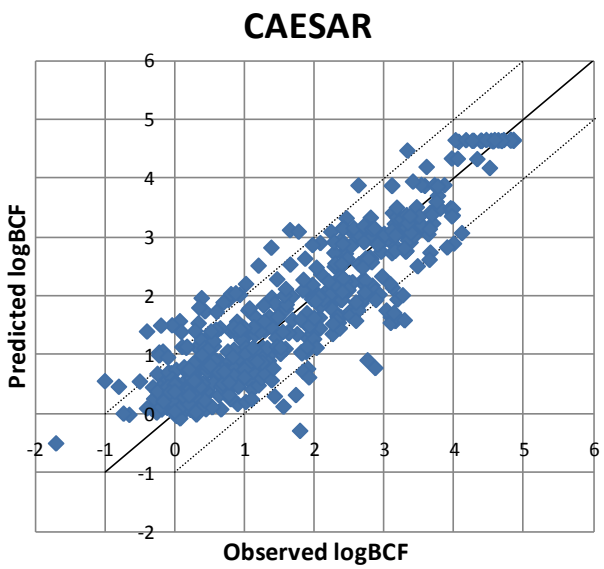
logP 3

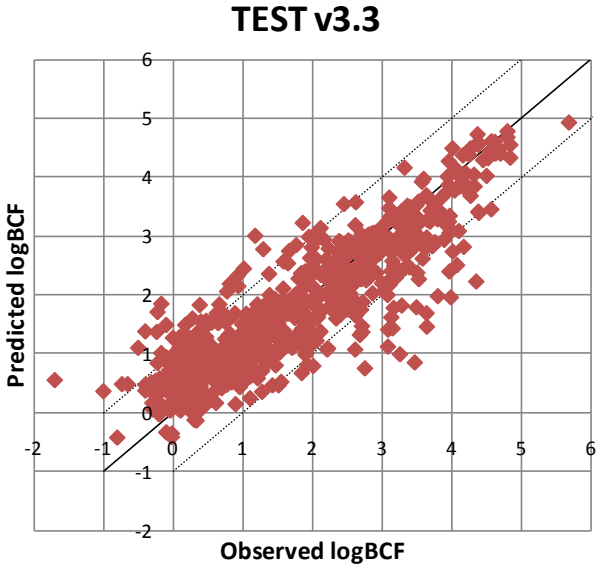
logP 4.5

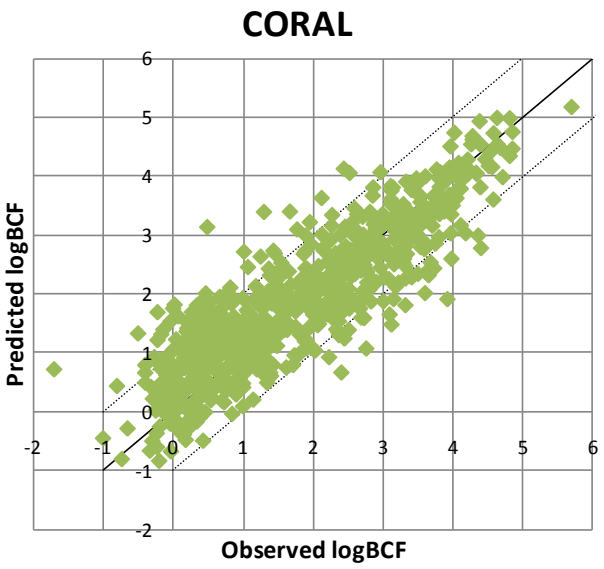
0 B/vB compounds with logP ≤ 3

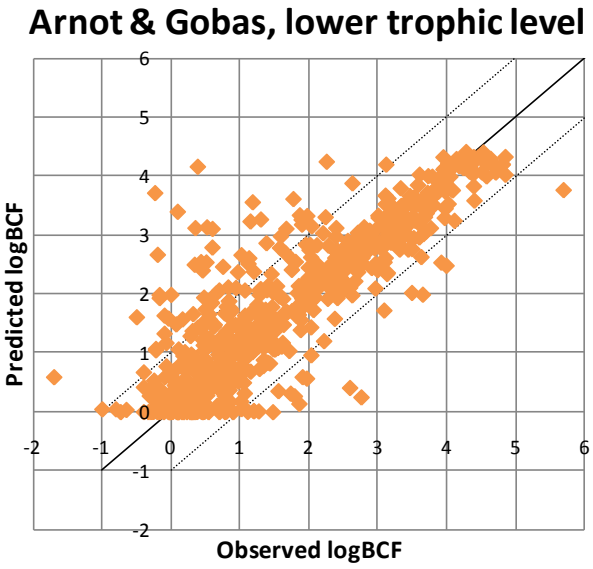
9 B/vB compounds with logP ≤ 4.5

Regression models









Regression models - classification