CASE STUDY: using and integrating two QSAR models for assessing bioconcentration factor

31/05/2023, LIFE CONCERT REACH Web-Seminars - (Q)SAR Models under REACH: Practical Examples

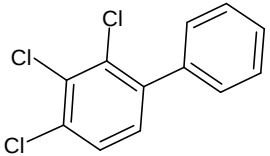
Claudia Ileana Cappelli
knoell Germany GmbH
qsar@knoell.com

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

TABLE OF CONTENTS

Case study definition

- **Aim:** prediction of **bioconcentration factor (BCF) in fish** of **2,3,4-Trichlorobiphenyl** using two QSAR models available in the LIFE CONCERT REACH Gateway, **VEGA/CAESAR** and **OCHEM/Gramatica & Papa (2005)**

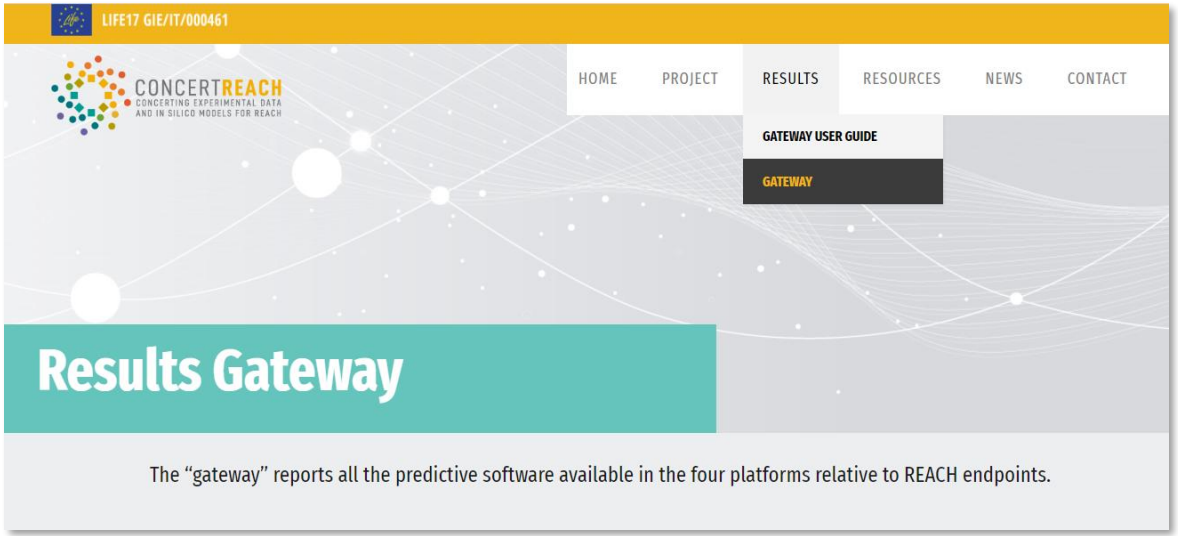


Input: Clc1ccc(c2ccccc2)c(Cl)c1Cl

- **Out of scope:** documentation of results in IUCLID
  - **Case studies:** mutagenicity (R. Gonella-Diaza, 17/05); logKow (A. Szymoszek, 31/05)
  - **Appendix**

The Gateway in LIFE CONCERT REACH

- Web-based system to guide users to **integrate data** from four platforms, i.e., **VEGAHUB, AMBIT, OCHEM and Danish (Q)SAR Database**, for evaluation of substance(s) under REACH
- Four main categories of endpoints: **physicochemical**, **toxicological**, **ecotoxicological** and **endocrine disruption properties**
- **Freely accessible** at https://www.life-concertreach.eu/results/results-gateway/



Model(s) selection

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

9. ECOTOXICOLOGICAL INFORMATION

- 9.1.1. Short-term toxicity testing on invertebrates (preferred species Daphnia)
- 9.1.2. Growth inhibition study aquatic plants (algae preferred)
- 9.1.3. Short-term toxicity testing on fish: the registrant may consider long-term toxicity testing instead of short-term.
- 9.1.4. Activated sludge respiration inhibition testing
- 9.1.5. Long-term toxicity testing on invertebrates (preferred species Daphnia)
- 9.1.6. Long-term toxicity testing on fish
- 9.2.1.1. Ready biodegradability
- 9.2.2.1. Hydrolysis as a function of pH.
- 9.3.1. Adsorption/desorption screening
- 9.3.2. Bioaccumulation in aquatic species, preferably fish
- 9.3.4. Further information on the environmental fate and behaviour of the substance and/or degradation products

Filter the models by the endpoints list

9.3.2. Bioaccumulation in aquatic species, preferably fish

All | VEGA AND ToxRead | DANISH QSAR DATABASE | AMBIT | OCHEM

End Point	Model	Type	Dataset size	Training set size	Test set size	Cross-validation procedure	Platform	Remarks
BCF	BCF (L/kg wet-wt) (EPI)	continuous					DanishQSARDatabase	

- For each endpoint, the gateway reports a list of models with basic information: **model name**, **type of prediction** (categorical; continuous), **size of the dataset** used to develop the model, **split in training and test set** (if available), **link to the model/platform and documentation** (QSAR Model Reporting Format, QMRF; papers)
- The user can look at the **whole list or** select models belonging to a **specific platform**
- Currently **12 BCF-related models** from VEGA, Danish (Q)SAR Database and OCHEM

Model(s) selection and prediction(s)

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

9.3.2. Bioaccumulation in aquatic species, preferably fish

All | VEGA AND ToxRead | DANISH QSAR DATABASE | AMBIT | OCHEM

End Point	Model	Type	Dataset size	Training set size	Test set size	Cross-validation procedure	Platform	Remarks
2.Environmental fate parameters 4. Bioconcentration 2.4.a.BCF fish	BCF model (CAESAR)	continuous	473	378	95		VEGA	
2.Environmental fate parameters 2.4.a.Bioconcentration . BCF fish	BCF model (Meylan)	continuous	662	516	146		VEGA	
Bioconcentration: OECD 305	BCF model (Arnot-Gobas)	continuous	692	692	0		VEGA	
ENV FATE 5.3.1. Bioaccumulation: aquatic	BCF model (KNN-Read-Across)	continuous	860	860	0		VEGA	

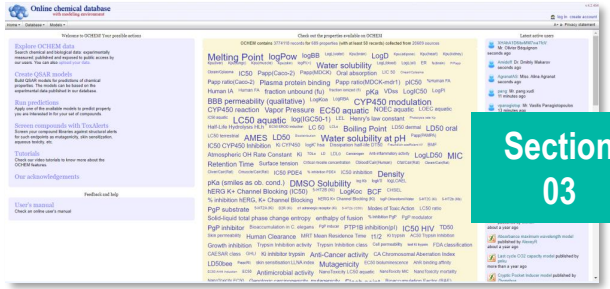
Section 02

After **model(s) selection**, click on the link in the **"Model" column(s)**: you will be redirected to the access page of the platform(s)

9.3.2. Bioaccumulation in aquatic species, preferably fish

All | VEGA AND ToxRead | DANISH QSAR DATABASE | AMBIT | OCHEM

End Point	Model	Type	Dataset size	Training set size	Test set size	Cross-validation procedure	Platform	Remarks
Bioaccumulation in aquatic species, preferably fish	Linear	Continuous		179	59		OCHEM	



Section 03

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

TABLE OF CONTENTS

Access VEGA

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

9.3.2. Bioaccumulation in aquatic species, preferably fish

All | VEGA AND ToxRead | DANISH QSAR DATABASE | AMBIT | OCHEM

End Point	Model	Type	Dataset size	Training set size	Test set size	Cross-validation procedure	Platform	Remarks
2.Environmental fate parameters 4. Bioconcentration 2.4.a.BCF fish	BCF model (CAESAR)	continuous	473	378	95		VEGA	

1

https://www.vegahub.eu/portfolio-item/vega-qsar/

Introduction
Screenshots
Interpretation
How to quote VEGA QSAR
VEGA QSAR for KNIME
2 Download
Previous Versions

3 DOWNLOAD Ver. 1.2.3

Previous releases of VEGA:

Please visit the Previous Version page in case you need older releases of the VEGA QSAR software

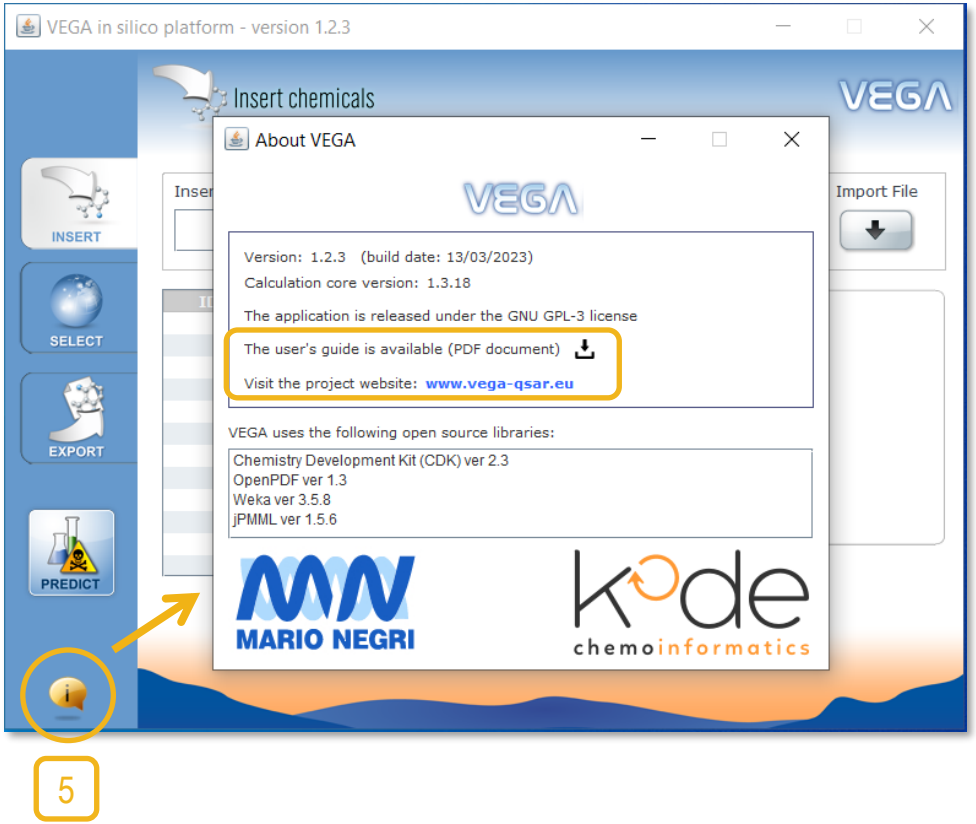
… And unzip the folder

VEGA-GUI-1.2.3 > VEGA GUI 1.2.3

Name

4
jdk-11.0.6
README.TXT
Vega-GUI-1.2.3.jar
Vega-launcher-LINUX.sh
Vega-launcher-WIN.bat

Access VEGA

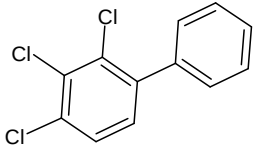


5

VEGA: Virtual models for Evaluating the properties of chemicals within a Global Architecture

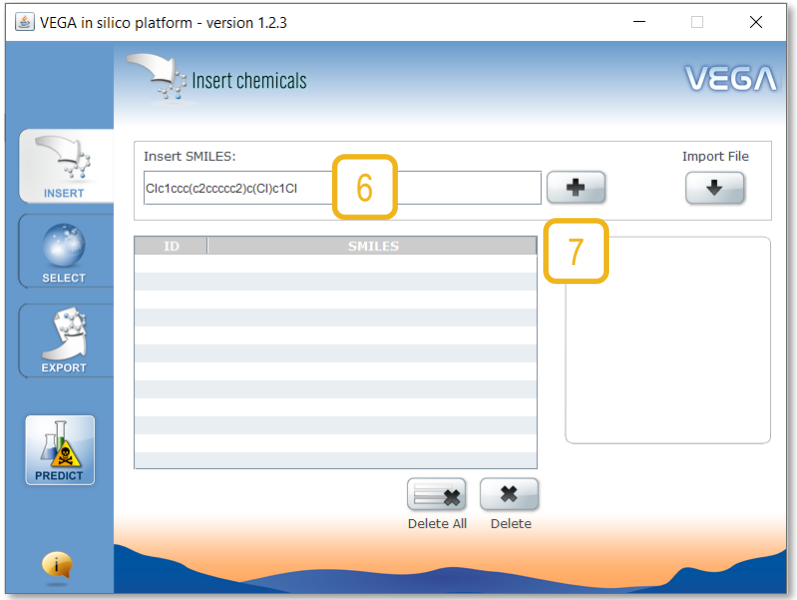
- Developed mainly by Mario Negri Institute (Milan) and Kode s.r.l. (Pisa)
- **Free platform** developed based on contributions from EU projects
- Includes **more than 100 statistical and knowledge-based (Q)SAR models** for the prediction of (eco)toxicity, environmental fate and physicochemical properties of chemicals.

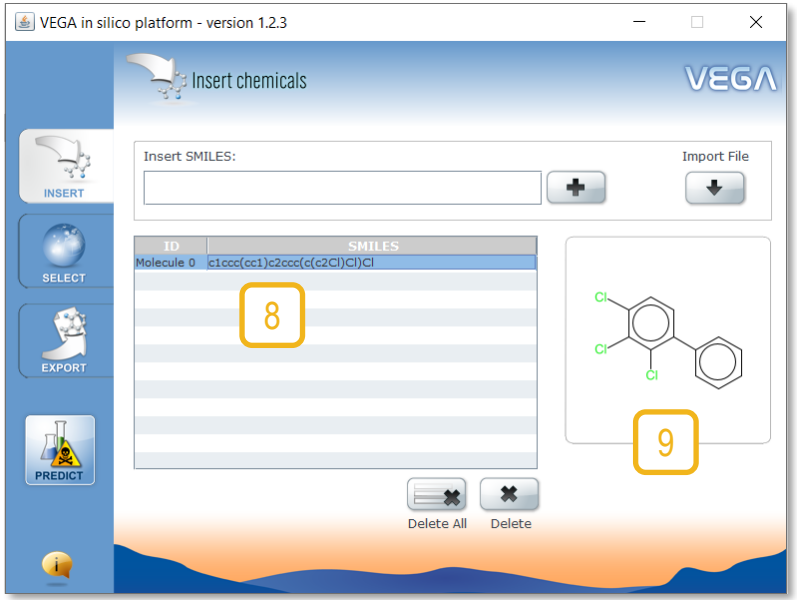
Insert the target chemical



Name: 2,3,4-Trichlorobiphenyl

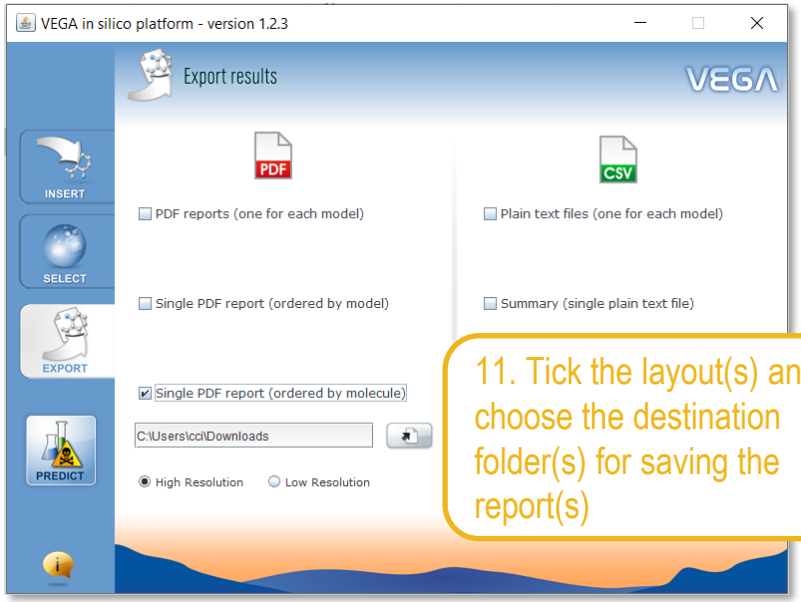
SMILES: Clc1ccc(c2ccccc2)c(Cl)c1Cl





Select the model(s) of interest and preferred output(s) settings

Model information in QMRF, available at https://www.vegahub.eu/portfolio-item/vega-qsar-models-qrmf/



11. Tick the layout(s) and choose the destination folder(s) for saving the report(s)

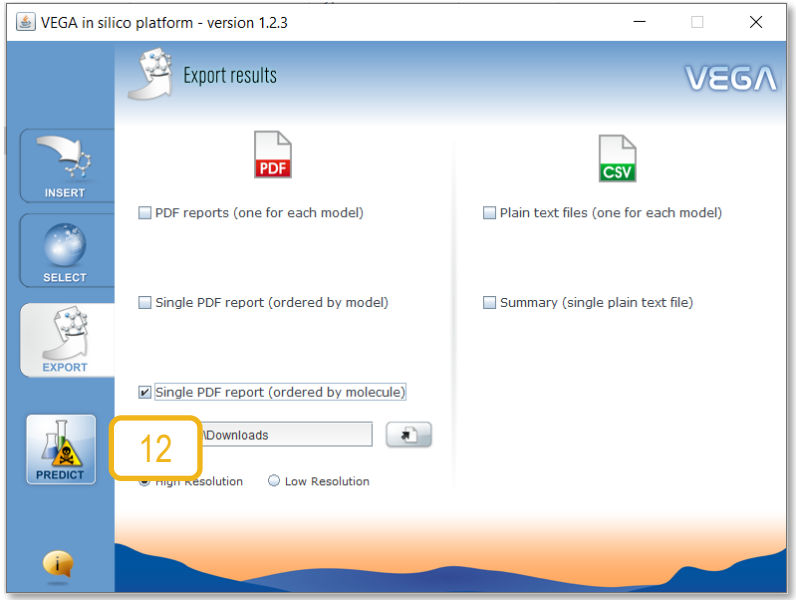
Full PDF reports:

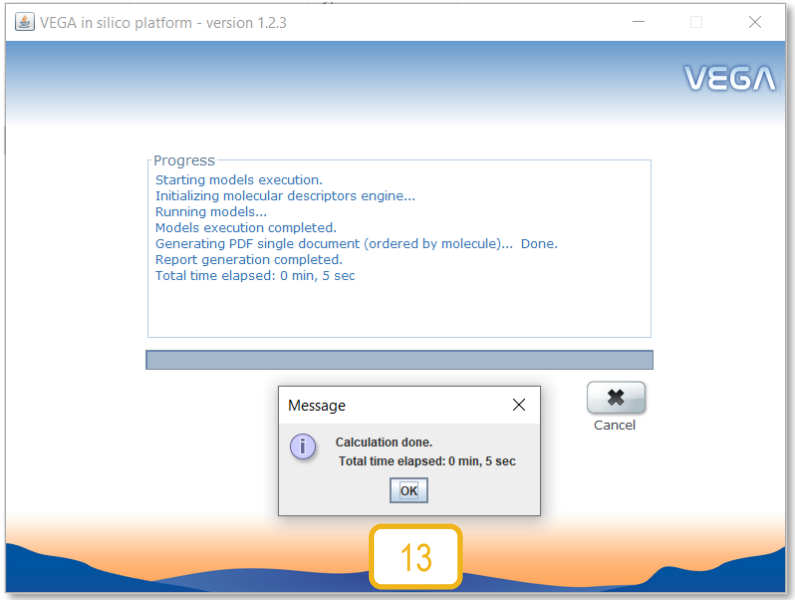
- prediction(s) results
- applicability domain
- experimental data of the target (if any)
- most similar substances
- other supporting info (if any)

Simplified text reports:

- essential information
- useful for excel import

Predict the target chemical





Assessment: scientific validity of the QSAR model (CAESAR)

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

9.3.2. Bioaccumulation in aquatic species, preferably fish

All | VEGA AND ToxRead | DANISH QSAR DATABASE | AMBIT | OCHEM

End Point	Model	Type	Dataset size	Training set size	Test set size	Cross-validation procedure	Platform	Remarks
2.Environmental fate parameters 4. Bioconcentration 2.4.a.BCF fish	BCF model (CAESAR)	continuous	473	378	95		VEGA	

QMRF

QMRF	***QMRF identifier (JRC Inventory):*** **To be entered by JRC**
	QMRF Title: **BCF model (CAESAR) (version 2.1.15)**
	Printing Date: **Apr 15, 2022**

1.QSAR identifier

1.1. QSAR identifier (title):

Model to predict bioconcentration factors (BCF) v 2.1.15.

1.2. Other related models:

Two models, model A and model B, have been used to build hybrid model, model C.

In the proposed approach, the outputs of the individual models (model A and B) were used as inputs of the hybrid model.

Model A was developed by Radial Basis Function Neural Networks (RBFNN) using a heuristic method to select the optimal descriptors; model B was developed by Radial Basis Function Neural Networks (RBFNN) using genetic algorithm for the descriptors selection.

https://www.vegahub.eu/portfolio-item/vega-qsar-models-qrmf/

- Quantitative prediction of **BCF in fish (log of L/kg)**
- **Hybrid model,** implementing the detection of **structural alerts** for reasoning (outliers, chemical classes with particular BCF behavior, polar groups)
- **Applicability Domain Index (ADI)** [0 - not reliable; 1 - fully reliable]
- **Goodness-of-fit**
  - n training set = 378; R2 = 0.81; RMSE = 0.58
- **Robustness - leave many out (20%) cross validation**
  - R2cv= 0.79; SDEP = 0.66
- **Predictivity - external validation**
  - n test set = 95; R2 = 0.78; RMSE = 0.62

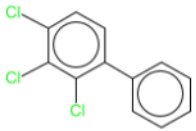
Assessment: QSAR model applicability to the query chemical

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

VEGA — BCF model (CAESAR) 2.1.15 — page 1

1. Prediction Summary

Prediction for compound Molecule 0 -



Prediction: Reliability:

Prediction is 4.33 log(L/kg), the result appears reliable. Anyhow, you should check it through the evaluation of the information given in the following sections.

Compound: Molecule 0
Compound SMILES: c1ccc(cc1)c2ccc(c(c2Cl)Cl)Cl
Experimental value: -
Predicted BCF [log(L/kg)]: 4.33
Predicted BCF [L/kg]: 21600
Predicted BCF from sub-model 1 (HM) [log(L/kg)]: 4.25
Predicted BCF from sub-model 2 (GA) [log(L/kg)]: 4.12
Predicted LogP (MLogP): 5.47
Structural Alerts: -
Reliability: The predicted compound is into the Applicability Domain of the model
Remarks:
none

Compound predicted as bioaccumulative (LogBCF > 3.3)

In applicability domain (AD)

Individual models give consistent predictions

LogP as support (≤ 6)

No alerts identified

Global AD Index — ADI > 0.85
AD index = 1
Explanation: the predicted compound is into the Applicability Domain of the model.

Similar molecules with known experimental value
Similarity index = 0.992
Explanation: strongly similar compounds with known experimental value in the training set have been found.

Accuracy of prediction for similar molecules
Accuracy index = 0.244
Explanation: accuracy of prediction for similar molecules found in the training set is good.

Concordance for similar molecules
Concordance index = 0.249
Explanation: similar molecules found in the training set have experimental values that agree with the predicted value.

Maximum error of prediction among similar molecules
Max error index = 0.376
Explanation: the maximum error in prediction of similar molecules found in the training set has a low value, considering the experimental variability.

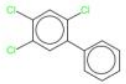
Model's descriptors range check
Descriptors range check = True
Explanation: descriptors for this compound have values inside the descriptor range of the compounds of the training set.

Atom Centered Fragments similarity check
ACF index = 1
Explanation: all atom centered fragment of the compound have been found in the compounds of the training set.

Assessment: QSAR model applicability to the query chemical

VEGA — BCF model (CAESAR) 2.1.15

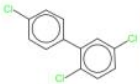
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values



Compound #1

CAS: 15862-07-4
Dataset id:185 (Training Set)
SMILES: c1ccc(cc1)c2cc(c(cc2Cl)Cl)Cl
Similarity: 0.998
Experimental value : 4.22
Predicted value : 4.332

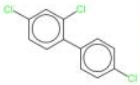
Alerts (not found also in the target): Moiety (SMILES: Clc1cc(c2ccccc2)c(Cl)cc1) (S



Compound #2

CAS: 16606-02-3
Dataset id:184 (Training Set)
SMILES: c1cc(ccc1c2cc(ccc2Cl)Cl)Cl
Similarity: 0.986
Experimental value : 3.95
Predicted value : 4.326

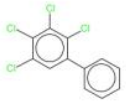
Alerts (not found also in the target): Moiety (SMILES: Clc1ccccc1c1ccc(Cl)cc1) (SF
Moiety (SMILES: Clc1cc(c2ccccc2)c(Cl)cc1) (SR 09)



Compound #3

CAS: 7012-37-5
Dataset id:183 (Training Set)
SMILES: c1cc(ccc1c2ccc(cc2Cl)Cl)Cl
Similarity: 0.979
Experimental value : 4.33
Predicted value : 4.322

Alerts (not found also in the target): Moiety (SMILES: Clc1ccccc1c1ccc(Cl)cc1) (SF



Compound #4

CAS: 33284-53-6
Dataset id:194 (Training Set)
SMILES: c1ccc(cc1)c2cc(c(c(c2Cl)Cl)Cl)Cl
Similarity: 0.963
Experimental value : 4.39
Predicted value : 4.636

Alerts (not found also in the target): Moiety (SMILES: Clc1cc(c2ccccc2)c(Cl)cc1) (SR 09)

A similarity index (SI) of the target with respect to similar molecules with known experimental value is calculated. It takes into account how similar are the first two most similar compounds. Values near 1 mean that the predicted compound is well represented in the dataset used to build the model, otherwise the prediction could be an extrapolation.

The 2 mostly similar compounds from the training set:

- exhibit high similarity to the target

Similar molecules with known experimental value SI > 0.9
Similarity index = 0.992
Explanation: strongly similar compounds with known experimental value in the training set have been found.

- have experimental values that agree with target prediction…

Concordance for similar molecules < 0.5
Concordance index = 0.249
Explanation: similar molecules found in the training set have experimental values that agree with the predicted value.

- …and their prediction accuracy is good.

Accuracy of prediction for similar molecules < 0.5
Accuracy index = 0.244
Explanation: accuracy of prediction for similar molecules found in the training set is good.

In the literature, experimental BCF values varies between 0.45 log units (EURAS; gold standard) and ±0.75 log units (Dimitrov et al., 2005) (Lombardo et al., 2010).

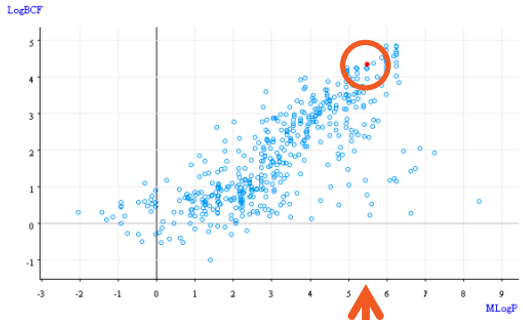
Assessment: QSAR model applicability to the query chemical

VEGA BCF model (CAESAR) 2.1.15 page 6

4.2 Reasoning:
Analysis of Molecular Descriptors

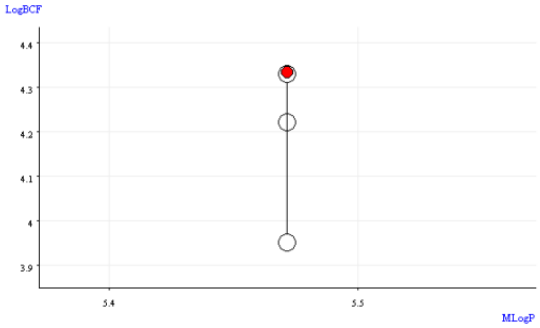
Descriptor name: MLogP
Descriptor name: LogP is directly correlated to the logBCF value.

Following, a scatterplot of MLogP against response values; experimental values are reported for the training set, predicted value for the studied compound. Light blue dots represent values of compounds from training set, red dot is the value of the studied compound.



(5.47 ≤ 6)
LogP as support

Following, a scatterplot of MLogP against response values only for 3 most similar compounds in the training set. Red dot is the value of the studied compound, black outlined circles represents experimental values of compounds from training set, black dots represents predicted value of the same compound; the size of the circle is proportional to the similarity to the studied compound.



The most similar chemicals to the target also share the same logP

Assessment: relevance of the QSAR model to regulatory purpose

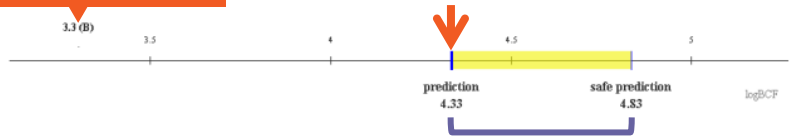
VEGA BCF model (CAESAR) 2.1.15 page 2

2. Possible Use and Uncertainty

Threshold 3.3 (bioaccumulative)

Following, a chart showing the predicted value together with its conservative confidence interval for safe classification.
For the threshold logBCF = 3.3, the current compound can be associated (due to its Applicability Domain index value) to a conservative interval of 0.5 log units.
On this basis, the compound can be classified as bioaccumulative.

LogBCF = 3.3 (B) < Predicted LogBCF = 4.33



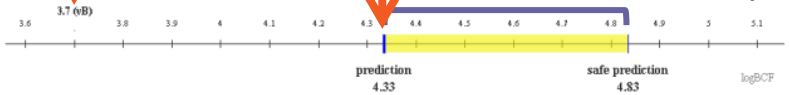
Uncertainty interval = 0.5 log units

Threshold 3.7 (very bioaccumulative)

Following, a chart showing the predicted value together with its conservative confidence interval for safe classification.
For the threshold logBCF = 3.7, the current compound can be associated (due to its Applicability Domain index value) to a conservative interval of 0.5 log units.
On this basis, the compound can be classified as very bioaccumulative.

LogBCF = 3.7 (vB) < Predicted LogBCF = 4.33

Uncertainty interval = 0.5 log units



The target **fulfils the criteria** for **B*** and **vB****

- **Screening criterion** ($LogK_{ow}$ = 5.47 > 4.5) **also fulfilled**.

** B: bioaccumulative*
*** vB: very bioaccumulative*

Assessment: adequacy of the QSAR result

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

Highly reliable → based on model validity, ADI and information to derive it

- Target well represented in the training set (i.e., strongly similar compounds, target descriptors within the range of the training set, all structural fragments found in the training set)
- Training set analogues experimental values: consistent with target prediction
- Training set analogues prediction accuracy: good
- The maximum error in prediction of training set analogues has a low value, considering the experimental variability.

Relevant → based on the purpose

- The target fulfils B and vB criteria.

Adequate

Klimisch 2 - results derived from a valid QSAR model and falling into its applicability domain, with adequate and reliable documentation/justification.

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

TABLE OF CONTENTS

Access OCHEM

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

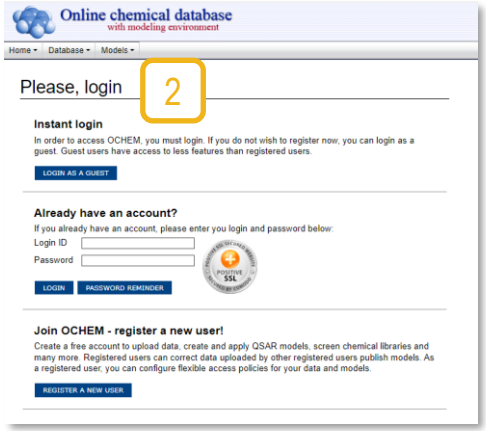
9.3.2. Bioaccumulation in aquatic species, preferably fish

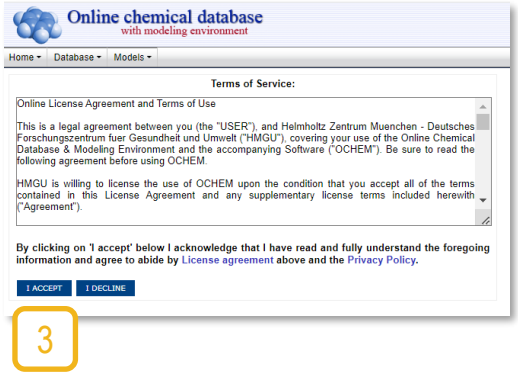
All | VEGA AND ToxRead | DANISH QSAR DATABASE | AMBIT | OCHEM

End Point	Model	Type	Dataset size	Training set size	Test set size	Cross-validation procedure	Platform	Remarks
Bioaccumulation in aquatic species, preferably fish	Linear	Continuous		179	59		OCHEM	

1 (option A) Access to the portal

1 (option B) Direct access to the model: Gramatica & Papa (2005)





https://ochem.eu/login/show.do?render-mode=full

Access OCHEM

OCHEM: Online CHEmical database and Modeling environment

- **Free** web-based platform that provides tools for automation of steps to **create** a predictive **QSAR/QSPR model**
- It consists of a **database of experimental measurements** (> 1M chemical structures and 3M data points) **integrated with a modeling framework**, which supports all the steps to create a predictive model
- **> 150 models** are published on the web site, which can be used to predict new molecules.

Model page layout – “Overview” tab

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

Online chemical database
with modeling environment

v.4.2.454

Welcome, Guest! Logout

Home ▾ Database ▾ Models ▾ A+ a- Privacy statement

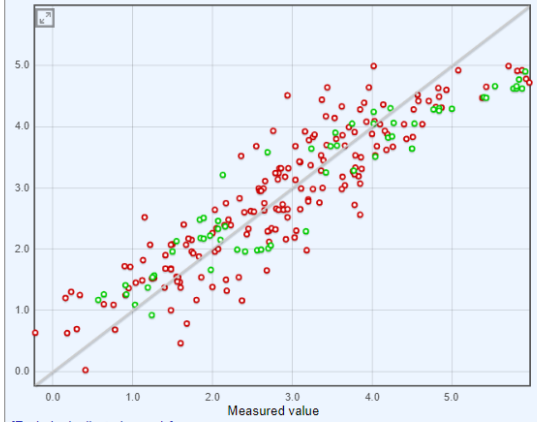
Model profile
Statistical parameters, tables, charts - all the information related to the model.

Overview | Applicability domain

Model name: BCF tutorial , published in An Update of the BCF QSAR Model Based on Theoretical Molecular Descriptors
Public ID is 8

Predicted property: BCF
Training method: MLRA

Data Set	#	R2	q2	RMSE	MAE
Training set: BCF Tutorial (training)	179 records	0.8 ± 0.02	0.8 ± 0.02	0.58 ± 0.03	0.48 ± 0.02
Test set: BCF Tutorial (test) [x]	59 records	0.9 ± 0.02	0.85 ± 0.02	0.61 ± 0.04	0.52 ± 0.04



[Exclude duplicated records]

Export this model | View configuration XML | Export configuration XML | MMP-based analysis (experimental)

APPLY THE MODEL TO NEW COMPOUNDS

[Dragon6 (blocks: 1-29)]
Correl. limit: 0.0 Unique values: 0, Variance threshold: 0.0, Maximum value: 2147483647, ,
[IDDM, HIC, nHAcc, GATS1e, MATS1p]

No validation
-
5 pre-filtered descriptors

Y = -1.14 + 2.43*IDDM - 0.88*HIC - 0.481*nHAcc - 1.17*GATS1e - 1.95*MATS1p

Normalised Y = 2.85 + 1.41*IDDM - 0.629*nHAcc - 0.59*HIC - 0.319*GATS1e - 0.27*MATS1p
(5 variables in equation)

Calculated in 30 seconds
Size: 11 Kb

Reference

Predicted property, model type, access to training and test sets, statistics

Equation and descriptors.

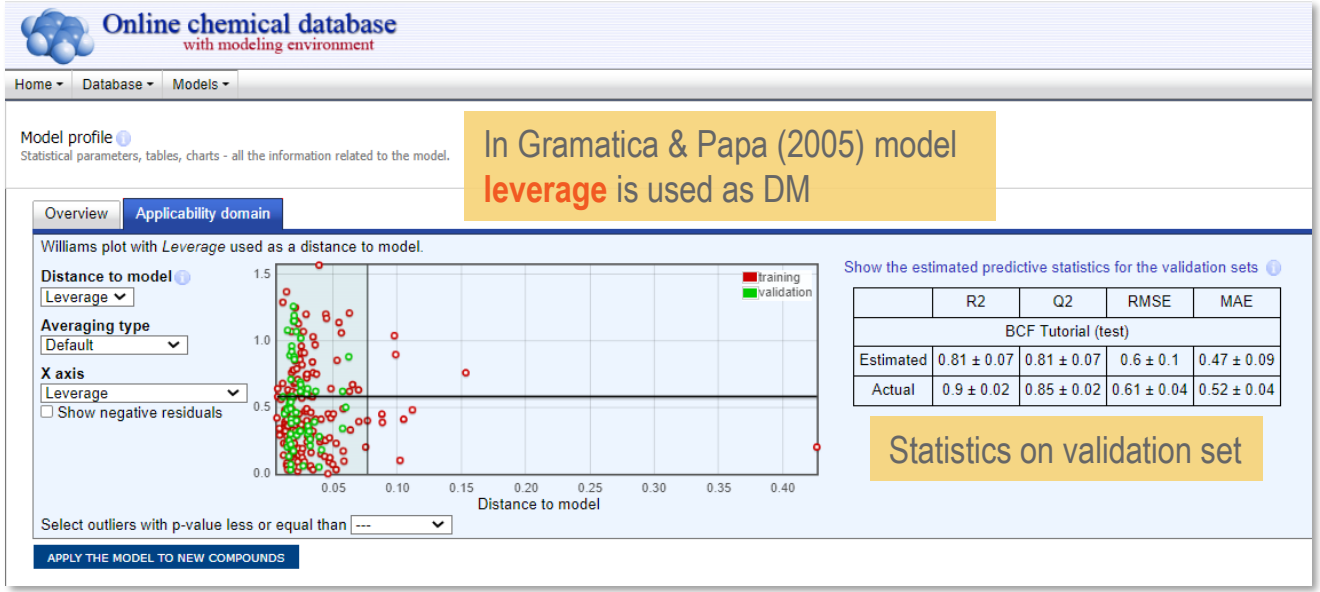
Descriptors refer to molecular size, hydrogen bonding capacity, electronic properties (atomic polarizability and electronegativity), molecular complexity

Observed-vs-predicted chart

Prediction generation

Model page layout – “Applicability domain” tab

- The AD assessment in OCHEM is based on **distance to model (DM)**, i.e., any numerical measure of the prediction uncertainty for a given compound by the model
- DM assesses how “far” is the compound from the model: **compounds with larger values of DM are expected to have lower prediction accuracy than compounds with smaller DM**
- DMs estimate the **reliability** of predictions.



	R2	Q2	RMSE	MAE
BCF Tutorial (test)				
Estimated	0.81 ± 0.07	0.81 ± 0.07	0.6 ± 0.1	0.47 ± 0.09
Actual	0.9 ± 0.02	0.85 ± 0.02	0.61 ± 0.04	0.52 ± 0.04

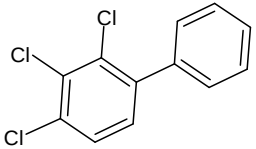
In Gramatica & Papa (2005) model **leverage** is used as DM

Statistics on validation set

- **High leverage** values indicate that one starts extrapolating outside the training set range and it is **no guaranteed that the model is valid and applicable**
- Compounds with leverage exceeding a warning threshold h* are often outside the AD of the model.

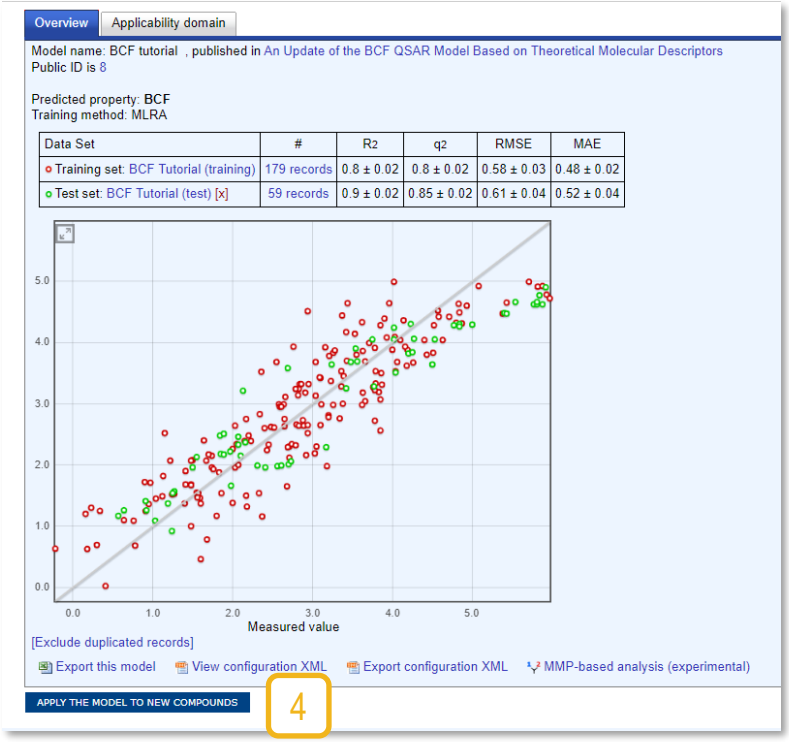
Tetko, I.V. et al, J. Chem. Inf. Mod. 2008, 48(9), 1733-1746
Sushko I., Applicability Domain of QSAR models. Doctoral work. 2011. http://mediatum.ub.tum.de/node?id=1004002

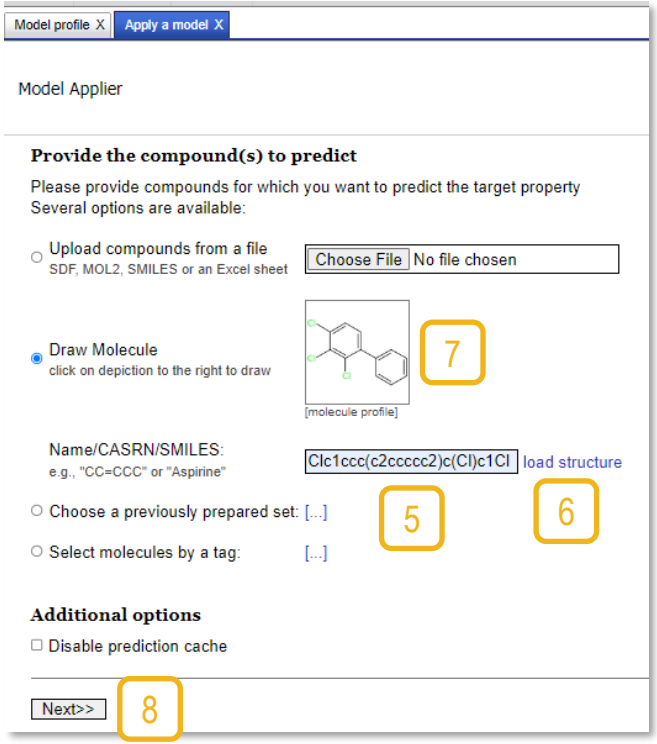
Predict the target chemical



Name: 2,3,4-Trichlorobiphenyl

SMILES: Clc1ccc(c2ccccc2)c(Cl)c1Cl





Assessment: scientific validity of the QSAR model (Gramatica & Papa, 2005)

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

9.3.2. Bioaccumulation in aquatic species, preferably fish

All | VEGA AND ToxRead | DANISH QSAR DATABASE | AMBIT | OCHEM

End Point	Model	Type	Dataset size	Training set size	Test set size	Cross-validation procedure	Platform	Remarks
Bioaccumulation in aquatic species, preferably fish	Linear	Continuous		179	59		OCHEM	

Full Papers

QSAR & Combinatorial Science QCS

An Update of the BCF QSAR Model Based on Theoretical Molecular Descriptors

Paola Gramatica* and Ester Papa

QSAR and Environmental Chemistry Research Unit, Department of Structural and Functional Biology, University of Insubria, via Dunant 3, 21100 Varese (Italy); http://www.qsar.it; E-mail: paola.gramatica@uninsubria.it

Keywords: BCF; Molecular descriptors; Genetic Algorithm; Validation; Hydrogen bonding; Molecular size

Received: February 22, 2005; Accepted: April 18, 2005

DOI: 10.1002/qsar.200530123

QMRF

QMRF identifier (JRC Inventory): Q13-24a-0012
QMRF Title: QSAR for bioconcentration factor in fish
Printing Date: Dec 11, 2019

1.QSAR identifier

1.1.QSAR identifier (title):
QSAR for bioconcentration factor in fish

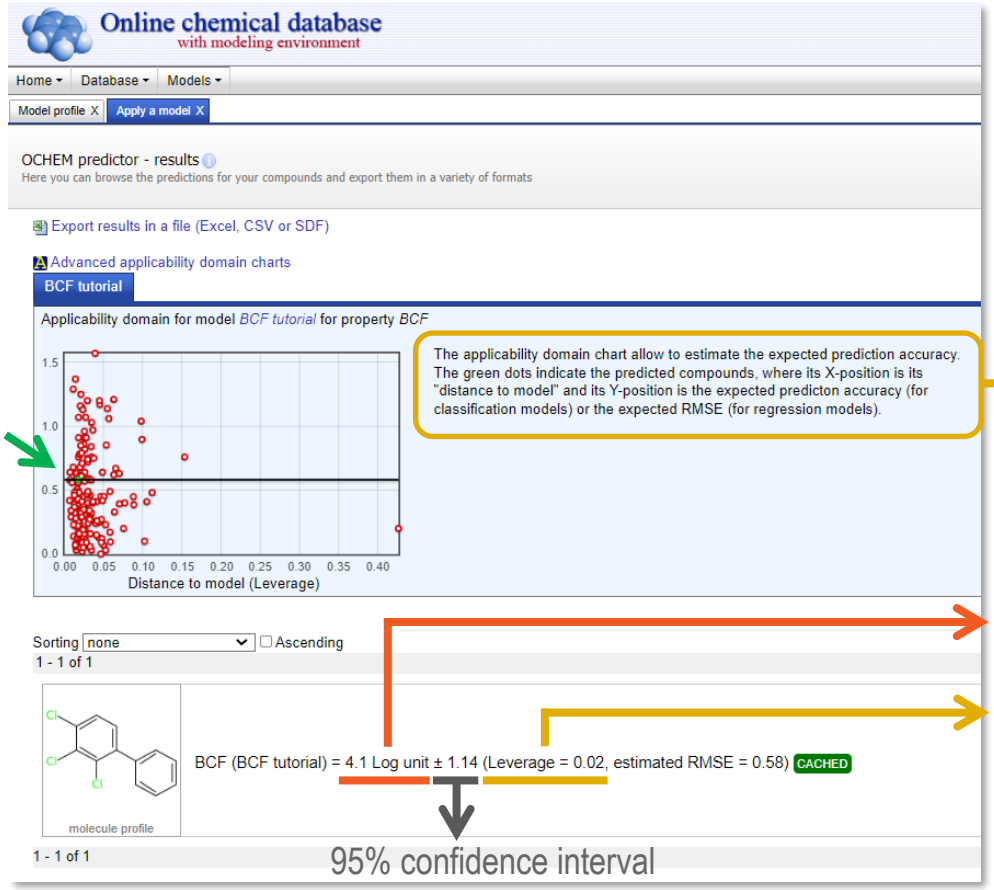
1.2.Other related models:
Gramatica P & Papa E (2003). QSAR Modeling of Bioconcentration Factor by theoretical molecular descriptors. QSAR & Combinatorial Science, 22, 374-385.

QMRF available at QsarDB

https://qsardb.org/repository/handle/10967/110

- Quantitative prediction of **BCF in fish (log of L/kg)**
- **Multiple linear regression (Ordinary Least Square regression method)**
- The **AD** was checked by **leverage approach** (high leverage compounds: hat value > 0.1). **Outliers** for the response were identified by analysis of the standardized residuals (values > 2.5 standard deviation units)
- **Goodness-of-fit**
  - n training set = 179; R2 = 80.7%; RMSE = 0.56
- **Robustness - leave many out cross validation**
  - $Q2_{LMO(25\%)}$=79%; $Q2_{LMO(50\%)}$=78.2%
- **Predictivity - external validation**
  - n test set = 59; Q2ext = 86.4%; R2ext = 90.5%; RMSE = 0.57

Assessment: QSAR model applicability to the query chemical



Target chemical compliant with the AD of the model

Compound predicted as bioaccumulative (LogBCF > 3.3)

DM value

Assessment: QSAR model applicability to the query chemical

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

Online chemical database
with modeling environment

Home ▾ Database ▾ Models ▾

Prediction results X | Prediction neighbors X

Prediction neighbors explorer
The training set compounds nearest to the selected prediction

molecule profile

BCF (BCF tutorial) = 4.1 Log unit ± 1.14 (Leverage = 0.02, estimated RMSE = 0.58) CACHED [prediction neighbors]

9

The predicted compound

molecule profile

BCF (BCF tutorial) = 4.1 Log unit ± 1.14 (Leverage = 0.02, estimated RMSE = 0.58) CACHED

Nearest training set neighbours

Similarity measure: Structural similarity

Similarity: 1.00
molecule profile
● BCF = 0.64 (in Log unit)
Predicted value: 1.10 (in Log unit)
Leverage: 0.03
Gramatica,P. Papa, E.
An Update of the BCF QSAR Model Based on Theoretical Molecul...
QSAR Comb. Sci. 2005; 24 (8) 953-960
Benzene , 71-43-2
MoleculeID: M663937
[open in browser] [prediction neighbors]
Dataset = Train
itetko / published

Experimental and predicted data, leverage and similarity of the most similar compounds

Sources

Similarity: 1.00
molecule profile
● BCF = 3.28 (in Log unit)
Predicted value: 3.87 (in Log unit)
Leverage: 0.02
Gramatica,P. Papa, E.
An Update of the BCF QSAR Model Based on Theoretical Molecul...
QSAR Comb. Sci. 2005; 24 (8) 953-960
4,4'-Dichlorobiphenyl , 2050-68-2
MoleculeID: M13799
[open in browser] [prediction neighbors]
Dataset = Train
itetko / published

Similarity: 1.00
molecule profile
● BCF = 5.88 (in Log unit)
Predicted value: 4.92 (in Log unit)
Leverage: 0.03
Gramatica,P. Papa, E.
An Update of the BCF QSAR Model Based on Theoretical Molecul...
QSAR Comb. Sci. 2005; 24 (8) 953-960
68194-17-2 , 2,2',3,3',4,5,5',6-OCTACHLOROBIPHENYL
MoleculeID: M44945
[open in browser] [prediction neighbors]
Dataset = Train
itetko / published

Two compounds from the training set:

- are similar to the target
- are moderately well predicted
- have experimental values in line with target prediction (logBCF = 4.1 ± 1.14)

Assessment: adequacy of the QSAR result

Highly reliable → based on model validity, AD compliance and information to derive it

- Strongly similar compounds to the target
- Training set analogues experimental values: in line with target prediction
- Training set analogues prediction accuracy: moderately good.

Relevant → based on the purpose

- The target fulfils B and vB criteria.

Adequate

Klimisch 2 - results derived from a valid QSAR model and falling into its applicability domain, with adequate and reliable documentation/justification.

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH

LIFE17 GIE/IT/000461

TABLE OF CONTENTS

Conclusion

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

- CAESAR and Gramatica & Papa (2005) models are documented in **QMRF** and provide supporting information (e.g., **AD compliance**, **similar molecules**), allowing **expert evaluation**
- In the present case study, CAESAR and Gramatica & Papa (2005) models provide **consistent**, **reliable** and **adequate predictions** and the target molecule 2,3,4-Trichlorobiphenyl can be assessed as B and vB (logBCF = 4.33 as a worst case).

The result may be used in the context of REACH:

- To cover the endpoint fully
- Together with other information (e.g., experimental data) as supporting data or part of WoE

CREDITS

Acknowledgements:

- knoell Academy team
- Igor V. Tetko (BigChem GmbH) and all partners of the LIFE CONCERT REACH project
- Antje Gerloff-Elias and the QSAR team at knoell
- The speakers of today and of 17/05

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

LIFE CONCERT REACH – Web-seminars on practical examples on using (Q)SAR for REACH

knoell
worldwide registration

Thanks for your attention

www.knoell.com

Think globally, act locally

Adequacy and reliability of documentation: IUCLID

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

- Results of (Q)SARs may be used instead of testing when the conditions set in REACH Annex XI (1.3) are met:
  - (i) a (Q)SAR model where the scientific validity has been established should be used;
  - (ii) the substance should fall within the applicability domain of the (Q)SAR model;
  - (iii) the prediction should be fit for the regulatory purpose; and
  - (iv) the information should be well documented.

Section 02

Section 03

Appendix

IUCLID 6

Compile one Robust Study Summary (RSS) for each QSAR result (at least two RSS of the predicted endpoint of interest for the target molecule, which are related to the two most reliable predictions derived from different models).

→ ECHA Practical guide "How to use and report (Q)SARs" Version 3.1 – July 2016

Adequacy and reliability of documentation: administrative data

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

Administrative data — None — EU: REACH

Endpoint
bioaccumulation in aquatic species, other

Type of information
(Q)SAR

Adequacy of study
weight of evidence

Adequacy of study — Please select
- key study
- supporting study
- weight of evidence ✓
- disregarded due to major methodological deficiencies
- other information

☑ Robust study summary

☐ Used for classification

☐ Used for SDS

Study period
None

Reliability
2 (reliable with restrictions)

Rationale for reliability incl. deficiencies
results derived from a valid (Q)SAR model and falling into its applicability domain, with adequate and reliable documentation / justification

Data waiving
None

Justification for data waiving
None

- results derived from a valid (Q)SAR model and falling into its applicability domain, with adequate and reliable documentation / justification -[Reliability 1 or 2] ✓
- results derived from a valid (Q)SAR model and falling into its applicability domain, with limited documentation / justification -[Reliability 2, 3 or 4]
- results derived from a valid (Q)SAR model, but not (completely) falling into its applicability domain, with adequate and reliable documentation / justification -[Reliability 2 or 3]
- results derived from a (Q)SAR model, with limited documentation / justification, but validity of model and reliability of prediction considered adequate based on a generally acknowledged source -[Reliability 2 or 3]
- results derived from a valid (Q)SAR model, but not (completely) falling into its applicability domain, and documentation / justification is limited -[Reliability 3 or 4]
- results derived from a (Q)SAR model, with limited documentation / -[Reliability

If the molecule is compliant with the applicability domain of the model, the QSAR result can be used as “key study”, “weight of evidence” or “supporting study”, depending on your case. When the molecule does not (completely) fit into the applicability domain, the QSAR result should be used only within a “weight of evidence” approach or as “supporting study”.

Reliability and its justification is case by case decision. As a general rule, all (Q)SAR entries should be of RL2, because RL1 is reserved for high quality experimental studies and RL3 or RL4 can be considered only in exceptional cases.

Adequacy and reliability of documentation: administrative data

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

Justification for type of information

1. SOFTWARE
VEGA v1.2.3

2. MODEL (incl. version number)
CAESAR v2.1.15

3. SMILES OR OTHER IDENTIFIERS USED AS INPUT FOR THE MODEL
Clc1ccc(c2ccccc2)c(Cl)c1Cl

4. SCIENTIFIC VALIDITY OF THE (Q)SAR MODEL
[[Explain how the model fulfils the OECD principles for (Q)SAR model validation. Consider attaching the QMRF and/or QPRF or providing a link]
- Defined endpoint:
- Unambiguous algorithm:
- Defined domain of applicability:
- Appropriate measures of goodness-of-fit and robustness and predictivity:
- Mechanistic interpretation:

5. APPLICABILITY DOMAIN
[Explain how the substance falls within the applicability domain of the model]
- Descriptor domain:
- Structural domain:
- Mechanistic domain:
- Similarity with analogues in the training set:
- Other considerations (as appropriate):

6. ADEQUACY OF THE RESULT
[Explain how the prediction fits the purpose of classification and labelling and/or risk assessment]

Attached justification + New item Import file

#	Attached justification	Reason / purpose
1	QMRF_BCF_CAESAR.pdf	(Q)SAR model reporting (QMRF)
2	report_BCF_CAESAR_trichlorobiphenyl.pdf	(Q)SAR: supporting information

Adequacy and reliability of documentation: data source

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

Data source

Reference
VEGA v1.2.3 | 2023

Data access
data published

Data protection claimed
None

General information

Reference Type
None

Title*
VEGA v1.2.3

Author
None

Year
2023

Bibliographic source
CAESAR BCF Model v2.1.15

Adequacy and reliability of documentation: materials and methods

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

Materials and methods

Test guideline New item Import file

#	Qualifier	Guideline	Version / remarks	Deviations	Actions
1	according to guideline	other: REACH Guidance R.6 on QSARs and grouping of chemicals	None	None	

Principles of method if other than guideline

- Software tool(s) used including version: Vega v1.2.3
- Model(s) used: CAESAR BCF Model version 2.1.15

Full reference and details of the used formulas can be found in:

<list of literature references>

- Model description: see field 'Justification for non-standard information', 'Attached justification' and 'any other information on Material and methods'
- Justification of QSAR prediction: see field 'Justification for type of information', 'Attached justification' and/or 'overall remarks'

GLP compliance

no

Test material

Test material information

2,3,4-trichlorobiphenyl_QSAR | 2,3,4-trichlorobiphenyl | 1,2,3-trichloro-4-phenylbenzene | 55702-46-0

Edit 2,3,4-trichlorobiphenyl_QSAR

Go to source

This IUCLID information is a re-usable data element. Note that any modification will impact all associated data.

Name*

2,3,4-trichlorobiphenyl_QSAR

Composition

Composition New item Import file

#	Type	Reference sub...	Concentration	Remarks	Actions
1	Constituent	2,3,4-trichlorobiphenyl | 1,2,3-trichloro-4-phenylbenzene | 55702-46-0	None	None	

Composition / purity: other information

not applicable for in silico study

Other characteristics

Test material form

None

Details on test material

None

Confidential details on test material

- SMILES: c1ccc(cc1)c2ccc(c(c2Cl)Cl)Cl

Adequacy and reliability of documentation: materials and methods

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

Test organisms

Test organisms (species)
other: Fish

Details on test organisms
None

Study design

Route of exposure
aqueous

Justification for method
minimised test method used to support BCF estimates based on QSAR

Test type
other: calculation

Water / sediment media type
natural water: freshwater

Details on estimation of bioconcentration
BASIS FOR CALCULATION OF BCF
- Estimation software: Vega version 1.2.3, CAESAR v 2.1.15
- Result based on measured OR calculated log Pow of: <value>

Any other information on materials and methods incl. tables

Supporting data for "Justification of type of information", e.g. tables.

Adequacy and reliability of documentation: results and discussion

CONCERTREACH
CONCERTING EXPERIMENTAL DATA AND IN SILICO MODELS FOR REACH
LIFE17 GIE/IT/000461

Results and discussion

Lipid content New item Import file

#	Lipid content	Time point	Remarks on result	Actions

Bioaccumulation factor New item Import file

#	Key result	Conc. / dose	Temp.	pH	Type	Value	Basis	Time of plateau	Calculation basis	Remarks on result	Actions
1		None	None	None	other: Log BCF	4.33 dimensionless	whole body w.w.	None	None	None	
2		None	None	None	BCF	21600 L/kg	whole body w.w.	None	None	None	

Any other information on results incl. tables

The performance of the model on similar substances as given in the attached report is summarized in the table:

Name and/or CAS No.	SMILES	Index of similarity to the test compound	Experimental result (log BCF)	Predicted result (log BCF)	Quality of prediction*
15862-07-4	c1ccc(cc1)c2cc(cc2Cl)Cl)Cl	0.998	4.22	4.33	Good
16606-02-3	c1cc(ccc1c2cc(ccc2Cl)Cl)Cl	0.986	3.95	4.33	Good
7012-37-5	c1cc(ccc1c2ccc(cc2Cl)Cl)Cl	0.979	4.33	4.32	Good

*Absolute difference (experimental-predicted): < 0.5 good, 0.5-1.0 moderate, >1.0 poor
The performance of the model on similar molecules is... (characterise: good, moderate, etc)

Overall remarks, attachments

Overall remarks

E.g. conclusion about adequacy of the result for the regulatory purpose under REACh regulation (EC) No 1907/2006.

This section includes information on the most similar substances to the target, as provided by the VEGA model (experimental and predicted data, similarity index). Quality of prediction is assigned by the user, as indicated. If the most similar substances are not provided automatically by the software, related information can be searched by the user, e.g., using the OECD QSAR Toolbox.

In this section the user has to conclude on applicability domain compliance of the target and validity of the prediction.