	QMRF identifier (JRC Inventory):**Q13-24a-0024**
	QMRF Title:**QSAR for the bioconcentration factor of non-ionic organic compounds**
	Printing Date:**Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

QSAR for the bioconcentration factor of non-ionic organic compounds

1.2.Other related models:

1.3.Software coding the model:

QSARModel 3.5.0

Molcode Ltd., Turu 2, Tartu, 51014, Estonia

http://www.molcode.com

2.General information

2.1.Date of QMRF:

06.11.2009

2.2.QMRF author(s) and contact details:

Molcode model development team Molcode Ltd. Turu 2, Tartu, 51014, Estonia

models@molcode.com http://www.molcode.com

2.3.Date of QMRF update(s):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

Molcode model development team Molcode Ltd. Turu 2, Tartu, 51014, Estonia

models@molcode.com http://www.molcode.com

2.6.Date of model development and/or publication:

06.11.2009

2.7.Reference(s) to main scientific papers and/or software package:

[1]Karelson M, Dobchev D, Tamm T, Tulp I, Jänes J, Tämm K, Lomaka A, Savchenko D & Karelson G (2008). Correlation of blood-brain penetration and human serum albumin binding with theoretical descriptors. ARKIVOC 16, 38-60.

[2]Karelson M, Karelson G, Tamm T, Tulp I, Jänes J, Tämm K, Lomaka A, Savchenko D & Dobchev D (2009). QSAR study of pharmacological permeabilities. ARKIVOC 2, 218–238.

2.8.Availability of information about the model:

Model is proprietary, but the training and test sets are available.

Algorithm is available.

2.9.Availability of another QMRF for exactly the same model:

None to date.

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Several fish: rainbow trout, guppy, fathead minnow, bluegill sunfish,

golden ide.

3.2.Endpoint:

2.Environmental fate parameters 2.4.a.Bioconcentration . BCF fish

3.3.Comment on endpoint:

Logarithm of the bioconcentration factor (BCF). The BCF for a particular chemical compound is defined as the equilibrium ratio of the concentration of a chemical inside an organism to the concentration in the surrounding environment.

3.4.Endpoint units:

unitless

3.5.Dependent variable:

logBCF

3.6.Experimental protocol:

3.7.Endpoint data quality and variability:

Statistics: max value: 4.84; min value: 0.18; standard deviation: 1.09; skewness: -0.0543

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Multilinear regression QSAR

Log(BCF)= -114.12 +29.39 * Max valency (AM1) -0.94 * Square root of Charged (Zefirov) Surface Area of C atoms -10.13 * Square root of Partial Surface Area of H atoms +0.57 * logP

4.3.Descriptors in the model:

[1]Max valency (AM1) none

[2]Square root of Charged (Zefirov) Surface Area of C atoms Å

[3]Square root of Partial Surface Area of H atoms Å

[4]logP none

4.4.Descriptor selection:

Initial pool of ~1000 descriptors. Stepwise descriptor selection based on a set of statistical selection rules:

1-parameter equations: Fisher criterion and R^2 over threshold, variance and t-test value over threshold, intercorrelation with another descriptor not over threshold;

2 parameter equations: intercorrelation coefficient bellow threshold, significant correlation with endpoint in terms of correlation coefficient and t-test.

Stepwise trial of additional descriptors not significantly correlated to any already in the model.

4.5.Algorithm and descriptor generation:

1D, 2D, and 3D theoretical calculations quantum chemical descriptors derived from AM1 calculation. Model developed by using multilinear regression.

4.6.Software name and version for descriptor generation:

QSARModel 3.5.0

Molcode Ltd. Turu 2, Tartu, 51014, Estonia

http://www.molcode.com

4.7.Chemicals/Descriptors ratio:

39 (156 chemicals/4 descriptors)

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

a) by chemical identity: halogenated benzenes, chlorinated biphenyls, chlorinated naphthalenes, chlorinated aliphatic hydrocarbons, monocyclic aromatic hydrocarbons, polycyclic aromatic hydrocarbons, chlorinated dibenzo-dioxins, chlorinated dibenzofurans, phenols, anilines, phthalated, carbonates, phosphates, esters, ethers, nitroaromatics.

b) by descriptor value range: this model is suitable for compounds that have the descriptors in the following range:

Max valency (AM1) (min: 2.72, max: 5.14)

Square root of Charged (Zefirov) Surface Area of C atoms (min: 0.00 , max: 2.89)

Square root of Partial Surface Area of H atoms (min: 0.00, max: 0.26)

logP (min: 0.50 , max: 10.15)

5.2.Method used to assess the applicability domain:

Presence of functional groups in structures;

Range of descriptor values in training set with ±30% confidence;

Descriptor values must fall between maximal and minimal descriptor values of training set ± 30%;

5.3.Software name and version for applicability domain assessment:

QSARModel 3.5.0

Molcode Ltd. Turu 2, Tartu, 51014, Estonia

http://www.molcode.com

5.4.Limits of applicability:

See 5.1

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS RN: Yes

Chemical Name: Yes

Smiles: No

Formula: Yes

INChI: No

MOL file: Yes

6.3.Data for each descriptor variable for the training set:

All

6.4.Data for the dependent variable for the training set:

All

6.5.Other information about the training set:

156 data points: 0 negative values; 156 positive values

6.6.Pre-processing of data before modelling:

6.7.Statistics for goodness-of-fit:

R^2= 0.77 (Correlation coefficient); s = 0.28 (Standard error of the estimate); F = 127.63 (Fisher function);

6.8.Robustness - Statistics obtained by leave-one-out cross-validation:

R^2_{cv}= 0.76 LOO;

6.9.Robustness - Statistics obtained by leave-many-out cross-validation:

R^2_{cv}= 0.74 LMO;

6.10.Robustness - Statistics obtained by Y-scrambling:

6.11.Robustness - Statistics obtained by bootstrap:

6.12.Robustness - Statistics obtained by other methods:

ABC analysis (2:1 training : prediction) on sorted data divided into 3 subsets (A;B;C). Training set formed with 2/3 of the compounds (set A+B, A+C, B+C) and validation set consisted of 1/3 of the compounds (C, B, A); average R^2(fitting) = 0.78; average R^2(prediction) = 0.76

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

Yes

7.2.Available information for the external validation set:

CAS RN: Yes

Chemical Name: Yes

Smiles: No

Formula: Yes

INChI: No

MOL file: Yes

7.3.Data for each descriptor variable for the external validation set:

All

7.4.Data for the dependent variable for the external validation set:

All

7.5.Other information about the external validation set:

17 data points: 0 negative values; 17 positive values

7.6.Experimental design of test set:

The full experimental dataset was sorted according to increasing values of log(LD50) and each tenth compound was assigned to the test set.

7.7.Predictivity - Statistics obtained by external validation:

R^2=0.763

7.8.Predictivity - Assessment of the external validation set:

The descriptors for the test set are in the limit of applicability.

7.9.Comments on the external validation of the model:

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

The bioconcentration factor increases with the increasing solubility of the chemical compounds in fish tissue. Solubility in fish tissue is

similar to solubility in octanol, so logP is a direct measure of the solubility of chemical compound in fish and is the most significant descriptor in this model. The solubility of a chemical compound in fish tissue decreases with the increasing charged surface area of the chemical compound and with the increasing partial surface area of H atoms. Chemical compounds having many charges and H atoms able to form H bonds are more soluble in polar solvents than in octanol. The bioconcentration also increases with the increasing valency of the chemical compound, maybe because an increased valency makes the compound more able to form chemical bonds with the fish tissue.

8.2.A priori or a posteriori mechanistic interpretation:

A posteriori mechanistic interpretation, consistent with published scientific interpretations of experiments.

8.3.Other information about the mechanistic interpretation:

The partition coefficient logP is the ratio of the concentrations of a compound in the two phases of a mixture of two immiscible solvents at equilibrium. Normally one of the solvents chosen is water while the second is hydrophobic such as octanol.

The descriptor Square root of Charged (Zefirov) Surface Area of C atoms encodes features responsible for polar interactions between molecules.

The descriptor Square root of Partial Surface Area of H atoms gives information about the capacity of the molecule to form H bonds.

9.Miscellaneous information

9.1.Comments:

9.2.Bibliography:

Xiaoxia Lu, Shu Tao, Hanyin Hu & Dawson RW (2000). Estimation of bioconcentration factors of nonionic organic compounds in fish by molecular connectivity indices and polarity correction factors. Chemosphere 41, 1675-1688.

9.3.Supporting information:

BCF3 training_156.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-24a-0024/attachment/A672
BCF3 test_17.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-24a-0024/attachment/A673

Test set(s)

10.Summary (JRC QSAR Model Database)

Supporting information

10.1.QMRF number:

Q13-24a-0024

10.2.Publication date:

2013-06-26

10.3.Keywords:

Molcode;BCF;bioconcentration factor;nonionic organic compound;rainbow trout;guppy;fathead minnow;bluegill;

10.4.Comments:

former Q8-10-14-207