	QMRF identifier (JRC Inventory): **Q13-24a-0022**
	QMRF Title: **QSAR for bioconcentration (flow through fish test) of pesticides**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

QSAR for bioconcentration (flow through fish test) of pesticides

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:

28.09.2009

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:

28.09.2009

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:

Bluegill (Lepomis macrochirus)

3.2.Endpoint:

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

3.3.Comment on endpoint:

3.4.Endpoint units:

3.5.Dependent variable:

logBCF 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.

3.6.Experimental protocol:

Bioconcentration: flow-through fish test was determined using the OECD Test Guideline 305 (EU C.13). Bioconcentration is the process of accumulation of chemicals by fish through non-dietary routes under flow-through conditions and is one of the key steps through which chemicals are able to enter into the biosphere from physical surroundings. Bioconcentration factor (BCF) is the concentration of a particular chemical in a tissue per concentration of chemical in water (BCF=Ct/Cw), being dimensionless. The U.S. Environmental Protection Agency (U.S. EPA) Environmental Fate and Effects Division (EFED) has developed an online database that reports many results from regulatory environmental fate studies. The database contains detailed compound information for herbicides, insecticides, and fungicides including descriptive BCF data. This database was determined to be of great value for the development and evaluation of a structural BCF model because it contains high-quality data generated by following the same protocol and using the same test species. The BCF studies were all conducted following EPA guideline 165-4, Laboratory Studies of Pesticide Accumulation in Fish (now OPPTS 850.1730). The criteria for the compounds used in the dataset were (A) whole fish BCF values were measured, (B) bluegill was the test species, and (C) the test was conducted for about 28 days. The guideline protocol requires that the fish receive chemical exposure using a radiolabeled test substance under flow-through tank conditions. Because the entire tank volume is replaced many times daily, the animals are principally exposed to parent compound only. For each residue determination interval, the fish are separated into edible and viscera portions for determination of BCF.

3.7.Endpoint data quality and variability:

Experimental data from different sources has been validated as consistent (Jackson et al, 2009). Statistics: max value: 4.00; min value: -0.92; standard deviation: 1.03; skewness: -0.71

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Multilinear regression QSAR

logBCF = 5.78E-002 – 1.16 * HA dependent HDCA-2 (Zefirov)+ 8.40E-002 * Lowest exchange energy (AM1) for C - O bonds + 0.40 * Lowest coulombic interaction (AM1) for C - H bonds + 0.49 * logP_calc

4.3.Descriptors in the model:

[1]HA dependent HDCA-2 (Zefirov)

[2]Lowest exchange energy (AM1) for C - O bonds

[3]Lowest coulombic interaction (AM1) for C - H bonds

[4]logP_calc

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 below 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 MMFFs (vacuum) conformational search and AM1 calculation. Model developed by using multilinear regression.

4.6.Software name and version for descriptor generation:

QSARModel 3.5.0

QSAR/QSPR package that will compute chemically meaningful descriptors and includes statistical tools for regression modeling

Molcode Ltd, Turu 2, Tartu, 51014, Estonia

http://www.molcode.com

4.7.Chemicals/Descriptors ratio:

19.5 (78 chemicals/ 4 descriptors)

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Applicability domain based on training set: By chemical identity: pesticides. By descriptor value range: the model is suitable for compounds that have the descriptors in the following range: HA dependent HDCA-2 (Zefirov)(min: 0.00 , max: 1.66), Lowest exchange energy (AM1) for C - O bonds (min: -13.61 , max: 0.00), Lowest coulombic interaction (AM1) for C - H bonds (min: 0.00 , max: 4.95), logP calc(min: -1.25 , max: 8.31).

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

QSAR/QSPR package that will compute chemically meaningful descriptors and includes statistical tools for regression modeling

Molcode Ltd, Turu 2, Tartu, 51014, Estonia

http://www.molcode.com

5.4.Limits of applicability:

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: No

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:

78 data points: 3 negative values; 75 positive values

6.6.Pre-processing of data before modelling:

6.7.Statistics for goodness-of-fit:

R^2= 0. 83 (Correlation coefficient); s = 0.19 (Standard error of the estimate); F = 88.84 (Fisher function);

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

R^2_{cv}= 0.81 LOO;

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

R^2_{cv}= 0.89 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). Aaverage R^2(fitting) = 0.83; average R^2(prediction) = 0.83.

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: No

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:

8 data points: 0 negative values; 8 positive values

7.6.Experimental design of test set:

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

7.7.Predictivity - Statistics obtained by external validation:

R^2 = 0.91

7.8.Predictivity - Assessment of the external validation set:

The descriptors for the test set are in the limits of applicability domain.

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 process of bioaccumulation can be described as a phenomenon of crossing hydrophobic membranes and as a result the descriptors involved in the model are related to molecular hydrophobicity and charges. The descriptor that is best correlated with bioaccumulation is logP. From mechanistic point of view, logP descriptor is a predictor of the molecules tendency to partition to lipids. The second descriptor of importance is HA-dependent HDCA-2 (Zefirov). The descriptor appears in the equation with negative coefficient which indicates that bioaccumulation decrease with increasing hydrogen-bond donor capabilities of molecules. The descriptors: Lowest coulombic interaction (AM1) for C-H bonds and Lowest exchange energy (AM1) for C-O bonds bring additional correction to the model, indicating the importance of some specific interactions.

8.2.A priori or a posteriori mechanistic interpretation:

A posteriori mechanistic interpretation.

8.3.Other information about the mechanistic interpretation:

The partition coefficient logP is the ratio of 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. HA-dependent HDCA-2 (zefirov) is a hydrogen bonding descriptor based on solvent-accessible area of hydrogen-bond donor atoms. The lowest coulombic interaction (AM1) for C-H bonds is an electrostatic interaction energy, represented by the sum of electron-electron, electron-nuclear, and nuclear-nuclear interactions. The exchange energy between two atoms reflects the change in Fermi correlation energy between two electrons localized on the two atoms. It can be of importance in determining the conformational changes of the molecule and its spin properties.

9.Miscellaneous information

9.1.Comments:

The molecules: clethodim (99129-21-2), cyhalothrin (68085-85-8), diquat dibromide (85-00-7), glyphosate (1071-83-6), mepiquat chloride (24307-26-4), fluroxypyr-MHE (81406-37-3), quinclorac (84087-01-4) have been considered statistical errors and have been excluded, as compared to the source dataset.

9.2.Bibliography:

[1]Jackson SH, Cowan-Ellsberry CE & Thomas G (2009). Use of Quantitative Structural Analysis To Predict Fish Bioconcentration Factors for Pesticides. Journal of Agricultural and Food Chemistry 57 (3), 958-967.

[2]US EPA Pesticide Fate Database http://cfpub.epa.gov/pfate/home.cfm

[3]US EPA OPPTS Harmonized Test Guidelines http://www.epa.gov/opptsfrs/publications/OPPTS_Harmonized/850_Ecological_Effects_Test_Guidelines/Drafts/

9.3.Supporting information:

BCF pesticides training_78.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-24a-0022/attachment/A668
BCF pesticides test_8.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-24a-0022/attachment/A669

Test set(s)

10.Summary (JRC QSAR Model Database)

Supporting information

10.1.QMRF number:

Q13-24a-0022

10.2.Publication date:

2013-06-25

10.3.Keywords:

BCF;bioconcentration factor;flow through test;fish;bluegill;Molcode;

10.4.Comments:

former Q8-10-14-175