
	QMRF identifier (ECB Inventory):	
	QMRF Title: <i>QSAR for acute toxicity to fathead minnow</i>	
	Printing Date: <i>Feb 16, 2010</i>	

1. QSAR identifier

1.1. QSAR identifier (title):

QSAR for acute toxicity to fathead minnow

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:

03.09.2009

2.2. QMRF author(s) and contact details:

[1] Indrek Tulp Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com
<http://www.molcode.com>

[2] Tarmo Tamm Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com
<http://www.molcode.com>

[3] Gunnar Karelson Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com
<http://www.molcode.com>

[4] Dimitar Dobchev Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com
<http://www.molcode.com>

[5] Dana Martin Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com
<http://www.molcode.com>

[6] Kaido Tämm Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com
<http://www.molcode.com>

[7] Deniss Savchenko Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com
<http://www.molcode.com>

[8] Jaak Jänes Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com
<http://www.molcode.com>

[9] Eneli Härk Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com
<http://www.molcode.com>

[10] Andres Kreegipuu Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com
<http://www.molcode.com>

[11] Mati Karelson 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:

03.09.2009

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:

Fathead Minnow

3.2. Endpoint:

3. Ecotoxic effects C.1. Acute toxicity for fish (Fathead minnow) 3.3. Acute toxicity to fish (lethality)

3.3. Comment on endpoint:

EU test method C.1. Acute toxicity for fish (Fathead minnow)

3.4. Endpoint units:

mg/MolWeight

3.5. Dependent variable:

$\log(\text{LC}_{50})$ - logarithm of the median lethal concentration (LC_{50}). The LC_{50} is the concentration that will kill 50% of the subjects after some specified exposure time.

3.6. Experimental protocol:

Acute toxicity to fish was determined using the EU Test Method C.1. The acute toxicity for fish is a method for investigating the discernible adverse effects induced in an organism within a short time (days) of exposure to a substance. Acute toxicity is expressed as the median lethal concentration (LC_{50}), that is the concentration in water which kills 50% of a test batch of fish within 96h. The concentrations of the test substance are given in millimoles per litre (mmol/L). The EPA Fathead Minnow Acute Toxicity database was generated by the U.S. EPA Mid-Continental Ecology Division (MED) for the purpose of developing an expert system to predict acute toxicity from chemical structure based on mode of action considerations. Hence, an important and unusual characteristic of this toxicity database is that the 617 tested industrial organic chemicals were expressly chosen to serve as a useful training set for development of predictive quantitative structure-activity relationships (QSARs). A second valuable aspect of this database, from a QSAR modeling perspective, is the inclusion of general mode-of-action (MOA) classifications of acute toxicity response for individual chemicals derived from study results. Each chemical was classified into one of eight modes of action: base-line narcosis or narcosis I, polar narcosis or narcosis II, ester narcosis or narcosis III, oxidative phosphorylation uncoupling, respiratory inhibition, electrophile/proelectrophile reactivity, AChE inhibition, or several mechanisms of CNS seizure responses. A detailed description of the biological and chemical test protocols used for these exposures has been published [Brooke LT et al. (1984), Geiger DL et al. (1985)].

Briefly, all tests were conducted using Lake Superior water at 25 ± 10 C. Aqueous toxicant concentrations were measured in all tests with quality assurance criteria requiring 80% agreement between duplicate samples and 90 to 110% spike recovery. Flow-through exposures were conducted using cycling proportional, modified Benoit, or electronic diluters. Tests conducted on the Benoit and electronic diluters did not have replicate tank exposures. Median lethal concentrations (LC50s) were calculated using the Trimmed Spearman–Karber Method, with 95% confidence intervals being calculated when possible. Information can be obtained from the EPA Fathead Minnow Acute Toxicity Database (1) and references (2-4) are listed in Section 9.

3.7. Endpoint data quality and variability:

Statistics: max value: 2.96, min value: -6.38, standard deviation: 1.40, skewness: -0.14

4. Defining the algorithm - OECD Principle 2

4.1. Type of model:

QSAR

4.2. Explicit algorithm:

multilinear regression QSAR

$\text{Log}(\text{LC50}) = 0.97 - 3.48 * \text{Average bond order (AM1)} - 0.32 * \text{Highest total interaction (AM1)} - 2.21\text{E-}003 * \text{LPSA Low polarity (AM1) part of SASA} - 0.16 * \text{count of H-acceptor sites (AM1) (all)} - 0.64 * \text{logP}$

4.3. Descriptors in the model:

[1] Average bond order (AM1)

[2] Highest total interaction (AM1)

[3] LPSA Low polarity (AM1) part of SASA

[4] count of H-acceptor sites (AM1) (all)

[5] logP

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. Descriptors/Chemicals ratio:

84.6 (423 chemicals / 5 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: diverse set of organic compounds: amines, nitro derivatives, nitriles, halogenated compounds, alcohols, phenols, organic acids, aromatic compounds. By descriptor value range: the model is suitable for compounds that have the descriptors in the following range: Average bond order (AM1) (min: 0, max: 2.09), Highest total interaction (AM1) (min: -18.19, max: 0), LPSA Low polarity (AM1) part of SASA (min: 0, max: 713.28), count of H-acceptor sites (AM1) (all) (min: 0, max: 10), logP (min: -2.38, max: 9.80).

5.2. Method used to assess the applicability domain:

Presence of functional groups in structures. Range of descriptor values in training set with $\pm 30\%$ confidence. Descriptor values must fall between maximal and minimal descriptor values of training set $\pm 30\%$.

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:

423 data points: 312 negative values; 111 positive values

6.6. Pre-processing of data before modelling:

6.7. Statistics for goodness-of-fit:

$R^2 = 0.76$ (Correlation coefficient); $s = 0.47$ (Standard error of the estimate); $F = 269.30$ (Fisher function);

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

$$R^2_{cv} = 0.75 \text{ LOO};$$

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

$$R^2_{cv} = 0.76 \text{ LMO};$$

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) = 0.76 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: 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:

46 data points: 34 negative values; 12 positive values

7.6. Experimental design of test set:

The full experimental dataset was sorted according to increasing values of $\log(\text{LC50})$ and each tenth compound was assigned to the test set.

7.7. Predictivity - Statistics obtained by external validation:

$$R^2 = 0.70$$

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 acute toxicity to Fathead Minnow increases with the solubility of the compound in octanol ($\log P$), this being a measure of the organic compound penetration in the animal tissue. The acute toxicity to Fathead Minnow also increases with increasing values of descriptor Count of H-acceptor sites (AM1) (all). The presence of H acceptor sites makes possible the binding of the molecule to the fish tissue and in this way increased toxicity. The toxicity is

further increased with the limited polarity of the molecule (reflected by the descriptor LPSA Low polarity (AM1) part of SASA). The increased unsaturation (reflected by descriptor Average bond order (AM1)) and increased 2-center interaction (electrons and nuclei) in the molecule (reflected by the descriptor Highest total interaction (AM1)) have as result an increased acute toxicity.

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 miscible solvents at equilibrium (usually water and octanol). The descriptor Count of H acceptor sites (AM1) is a measure of the ability of the compound to form H bonds. The limited polarity of the molecule (LPSA Low polarity (AM1) part of SASA) is an indication of mostly hydrophobic, but slightly polar compounds, and increases the possibility of binding the molecule to the fish tissue. An increased unsaturation (Average bond order (AM1)) and an increased 2-center interaction (Highest total interaction (AM1)), indicate strong (multiple) bonds in the molecule, causing some reactivity, and as a result render the molecule more toxic.

9.Miscellaneous information

9.1.Comments:

9.2.Bibliography:

[1] EPA FHM: EPA Fathead Minnow Acute Toxicity Database http://www.epa.gov/NCCCT/dsstox/sdf_epafhm.html

[2] Russom CL, Bradbury SP, Broderius SJ, Hammermeister DE & Drummond RA (1997). Predicting modes of toxic action from chemical structure: acute toxicity in the Fathead Minnow (*Pimephales Promelas*). *Environmental Toxicology and Chemistry* 16 (5), 948–967.

[3] Brooke LT, Call DJ, Geiger DL and Northcott CE (1984). *Acute Toxicities of Organic Chemicals to Fathead Minnows (Pimephales promelas)*, Vol. 1, Center for Lake Superior Environmental Studies, University of Wisconsin, Superior, WI, USA

[4] Geiger DL, Northcott CE, Call DJ and Brooke LT (1985). *Acute Toxicities of Organic Chemicals to Fathead Minnows (Pimephales promelas)*, Vol. 2., Center for Lake Superior Environmental Studies, University of Wisconsin, Superior, WI, USA.

9.3.Supporting information:

Training set(s)

Fathead_Minnow training 423	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf174_Fathead_Minnow_training_423.sdf
-----------------------------	---

Test set(s)

Fathead_Minnow test 46	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf174_Fathead_Minnow_test_46.sdf
------------------------	---

10.Summary (ECB Inventory)

10.1.QMRF number:

10.2.Publication date:

10.3.Keywords:

acute toxicity, fathead minnow, Molcode

10.4.Comments: