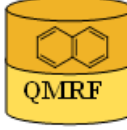
	QMRF identifier (ECB Inventory):	
	QMRF Title: <i>QSAR for acute toxicity to fish (Danio rerio)</i>	
	Printing Date: <i>Feb 16, 2010</i>	

1. QSAR identifier

1.1. QSAR identifier (title):

QSAR for acute toxicity to fish (Danio rerio)

1.2. Other related models:

1.3. Software coding the model:

QSARModel 3.3.8 Molcode Ltd., Turu 2, Tartu, 51014, Estonia <http://www.molcode.com>

2. General information

2.1. Date of QMRF:

29.06.2009

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:

20.06.2009

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.

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

None to date

3.Defining the endpoint - OECD Principle 1**3.1.Species:**

Fish (Danio rerio)

3.2.Endpoint:

3.Ecotoxic effects C.1 Acute toxicity to Danio rerio 3.3.Acute toxicity to fish (lethality)

3.3.Comment on endpoint:**3.4.Endpoint units:**

µM

3.5.Dependent variable:

log(LC50) logarithm of the median lethal concentration (LC50). The LC50 is the concentration that will kill 50% of the subjects after some specified exposure time.

3.6.Experimental protocol:

The acute toxicity for fish was determined using the EU Test Guideline 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 (LC50), 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 micromoles per litre (umol/L). Toxicity values were translated to logarithmic scale (logLC50) to reduce the range of the data.

References:

1. Dave et al (1981)
2. Roderer (1990)

3.7.Endpoint data quality and variability:

Experimental data selected from EPA AQUIRE database (<http://cfpub.epa.gov/ecotox/>). The US EPA AQUIRE database includes lethal, sublethal and residue toxic effects data on all aquatic species including plants and animals and freshwater and saltwater species. Toxicity data for EPA - AQUIRE database are drawn from several sources and then reviewed:

(i) Ecotoxicological studies conducted by commercial laboratories and submitted by pesticide companies in support of their products. EPA's Office of Compliance and Monitoring conducts periodic audits of these laboratories;

(ii) Studies conducted by US-EPA, USDA, and USFWS laboratories over the last 25 years;

(iii) Published data considered to meet their guideline criteria for acceptable data.

Inorganic compounds and mixtures in which components have different molecular weight or connectivity (i.e. substances with different chemical identity) were eliminated from the original EPA dataset. However, mixtures of stereoisomers were kept, because they are superimposable using common 2D descriptors. Data for LC50 96 h exposure of fish were then pruned as follows:

(i) Eliminating studies with an a.s.<85% purity;

(ii) Those identified as invalid where invalid studies were defined by the EPA as studies which may not be scientifically sound, or they were performed under conditions that deviated so significantly from the recommended protocols that the results will not be useful in a risk assessment;

(iii) Furthermore, only studies with actual values were kept discarding data given as higher or lower than values.

Statistics:

max value: 3.81

min value: -2.56

standard deviation: 1.43 skewness: -0.45

4. Defining the algorithm - OECD Principle 2

4.1. Type of model:

QSAR

4.2. Explicit algorithm:

multilinear regression QSAR

$$\text{Log(LC50)} = 2.98 + 7.75 * \text{Image of the Onsager-Kirkwood solvation energy (AM1)} - 60.24 * \text{Min electrophilic reactivity index (AM1) for N atoms} - 0.15 * \text{Number of aromatic bonds} + 1.54E-002 * \text{PNSA2} + 0.13 * \text{Lowest total interaction (AM1) for N - H bonds} - 1.64 * \text{Number of carbonyl groups}$$

4.3. Descriptors in the model:

[1] Image of the Onsager-Kirkwood solvation energy (AM1) [D2 mol/g]

[2] Min electrophilic reactivity index (AM1) for N atoms [1/eV]

[3] Number of aromatic bonds

[4] PNSA2 Total charge weighted PNSA (Zefirov) [Å²]

[5] Lowest total interaction (AM1) for N - H bonds [eV]

[6] Number of carbonyl groups

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 R² 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 AM1 calculation. Model developed by using multilinear regression.

4.6. Software name and version for descriptor generation:

QSARModel 3.3.8

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:

10.2 (61 chemicals/6 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:

a) by chemical identity: diverse organic chemicals (aromatic, cyclic and aliphatic alcohols, phenols, halogenoderivatives, carboxylic acids, esters, amines, etc)

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

Image of the Onsager-Kirkwood solvation energy (AM1) (min: 0 , max: 0.286), Min electrophilic reactivity index (AM1) for N atoms (min: 0 , max: 0.03), Number of aromatic bonds (min: 0 , max: 23), PNSA2 Total charge weighted PNSA (Zefirov) (min: -309.517, max: 0), Lowest total interaction (AM1) for N - H bonds (min: -18.096 , max: 0), Number of carbonyl groups (min: 0 , max: 3).

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 3.3.8

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:

61 data points: 12 negative values; 49 positive values

6.6.Pre-processing of data before modelling:

6.7.Statistics for goodness-of-fit:

$R^2 = 0.804$ (Correlation coefficient);

$s = 0.443$ (Standard error of the estimate);

$F = 36.891$ (Fisher function);

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

$R^2_{cv} = 0.745$ LOO;

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

$R^2_{cv} = 0.810$ 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 (in increasing order of endpoint value) 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) = 0.810

average R^2 (prediction) = 0.785

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:

6 data points: 1 negative values; 5 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:

$$R^2=0.847$$

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 acute toxicity to *Danio rerio* depends on the solvation energy of the compound. The toxicity decreases with increased solubility of the chemical compound in water and decreased solubility in organic tissue. The toxicity increases with the increased number of carbonyl functional groups and the increased number of aromatic bonds.

8.2. A priori or a posteriori mechanistic interpretation:

A posteriori mechanistic interpretation.

8.3. Other information about the mechanistic interpretation:

The descriptor Image of the Onsager-Kirkwood solvation energy (AM1) gives information about the solubility of the molecule in water. The solubility in water increases with the increasing of the values of the descriptor Image of the Onsager-Kirkwood solvation energy (AM1). The $\log(\text{LC50})$ increases (toxicity decrease) with the increasing solubility in water and decreased solubility in organic tissue. The descriptors Number of aromatic bonds and Number of carbonyl groups indicate that the acute toxicity to *Danio rerio* increases with the increasing in number of these functional groups. The charges in the molecule reflected by the descriptor PNSA2 Total charge weighted PNSA (Zefirov) influence also the toxicity. The increased number of electrical charges present in the chemical compound makes this less soluble in organic tissue and as a consequence less toxic. The toxicity has low values when the reactivity of the N atoms and N-H bonds is low, as reflected by the values of the descriptors: Min electrophilic reactivity index (AM1) for N atoms and Lowest total interaction (AM1) for N-H bonds.

9. Miscellaneous information

9.1. Comments:

9.2. Bibliography:

[1] US EPA ECOTOX Database <http://cfpub.epa.gov/ecotox/>

[2] Dave G, Andersson K, Berglund R & Hasselrot B (1981). Toxicity of eight solvent extraction chemicals and of cadmium to water fleas, *Daphnia magna*, rainbow trout, *Salmo gairdneri*, and zebrafish, *Brachydanio rerio*. *Comparative Biochemistry & Physiology C* 69, 83–98.

[3] Roderer G (1990). *Testung Wassergefährdender Stoffe als Grundlage für Wasserqualitätsstandards, Testbericht: Wassergefährdende Stoffe*. Fraunhofer-Institut für Umweltchemie und Ökotoxikologie, Schmallenberg (OECD Data File).

9.3.Supporting information:

Training set(s)

Danio rerio training_61	http://qsardb.jrc.ec.europa.eu:80/qmrf/download_attachment.jsp?name=qmrf136_Danio_rerio_training_61.sdf
-------------------------	---

Test set(s)

Danio rerio test_6	http://qsardb.jrc.ec.europa.eu:80/qmrf/download_attachment.jsp?name=qmrf136_Danio_rerio_test_6.sdf
--------------------	---

10.Summary (ECB Inventory)

10.1.QMRF number:

10.2.Publication date:

10.3.Keywords:

acute toxicity, fish, Danio rerio, QSARModel 3.3.8, LC50

10.4.Comments: