

	<b>QMRF identifier (ECB Inventory):</b> Q8-10-14-150	
	<b>QMRF Title:</b> QSAR model for acute toxicity to rainbow trout	
	<b>Printing Date:</b> Feb 16, 2010	

## 1. QSAR identifier

### 1.1. QSAR identifier (title):

QSAR model for acute toxicity to rainbow trout

### 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:

10.06.2009

### 2.2. QMRF author(s) and contact details:

Molcode model development team Molcode Ltd. Turu 2, Tartu, 51014, Estonia  
[models@molcode.com](mailto: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](mailto:models@molcode.com) <http://www.molcode.com>

### 2.6. Date of model development and/or publication:

10.06.2009

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

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

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

### 2.8. Availability of information about the model:

All information in full detail is available.

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

No other QMRF available for the same model.

## 3. Defining the endpoint - OECD Principle 1

### 3.1. Species:

Fish (rainbow trout)

### 3.2. Endpoint:

3. Ecotoxic effects 3.3. Acute toxicity to fish (lethality)

### 3.3. Comment on endpoint:

Acute toxicity to fish (rainbow trout), 96h LC<sub>50</sub>

### 3.4. Endpoint units:

log( $\mu\text{g/L/MW}$ )

### 3.5. Dependent variable:

log(LC<sub>50</sub>)

### 3.6. Experimental protocol:

The acute toxicity for fish (96h, rainbow trout) 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 (LC<sub>50</sub>) 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 ( $\mu\text{mol/L}$ ).

### 3.7. Endpoint data quality and variability:

The data were selected from the EPA ECOTOX database, which contains data measured in different laboratories. 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. The toxicity values were translated to logarithmic scale (logLC<sub>50</sub>) to reduce the range of the data.

The toxicity data in the EPA - AQUIRE database were 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 super imposable using common 2D descriptors. Data for LC<sub>50</sub> 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: 3.46 min value: -1.72  
standard deviation: 1.01 skewness: -0.13

## 4. Defining the algorithm - OECD Principle 2

### 4.1. Type of model:

QSAR

### 4.2. Explicit algorithm:

Multilinear regression QSAR

$\text{Log(LD50)} = 3.73 - 2.15 \cdot \text{Kier\&Hall index (order 2)} - 2.66 \cdot \text{Relative number of C atoms} + 0.0415 \cdot \text{HDCA H-donors charged surface area (AM1)} - 0.00557 \cdot \text{Negatively Charged Surface Area (Zefirov)}$

### 4.3. Descriptors in the model:

[1] Kier&Hall index (order 2) [unitless]

[2] Relative number of C atoms [unitless]

[3] HDCA H-donors charged surface area (AM1) [ $\text{\AA}^2$ ]

[4] Negatively Charged Surface Area (Zefirov) [ $\text{\AA}^2$ ]

### 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 were derived from optimized structures and 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. Descriptors/Chemicals ratio:

40.25 (161 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:

a) by chemical identity: various organic compounds including saturated and unsaturated aliphatic, aromatic and heteroaromatic amines, alcohols, carbocyclic acids, carbonyl compounds, halogeno and nitro derivatives, etc.

b) by descriptor value range:

Kier&Hall index (order 2) (min: 0.5, max: 11.56)

Relative number of C atoms (min: 0.14, max: 0.56)

HDCA H-donors charged surface area (AM1) (min: 0, max: 61.8)

Negatively Charged Surface Area (Zefirov) (min: 15.4, max: 398)

## 5.2. Method used to assess the applicability domain:

By chemical identity (functional groups among those present in training set).

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

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:

161 data points: 28 negative values; 133 positive values.

## 6.6. Pre-processing of data before modelling:

## 6.7. Statistics for goodness-of-fit:

$R^2 = 0.860$  (Correlation coefficient);

$s^2 = 0.148$  (Standard error of the estimate);

$F = 239$  (Fisher function);

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

$R^2_{cv} = 0.852$  (LOO)

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

$R^2_{cv} = 0.852$  (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) = 0.86

Average  $R^2$  (prediction) = 0.85

#### 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:

40 data points: 6 negative values; 34 positive values.

##### 7.6.Experimental design of test set:

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

##### 7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.825$  (Correlation coefficient)

##### 7.8.Predictivity - Assessment of the external validation set:

The descriptor values of the compounds in 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:

Relative number of C atoms and Kier&Hall index (order 2) are related to complexity and flexibility of the structure. HDCA H-donors charged surface area (AM1) and Negatively Charged Surface Area (Zefirov) are charged surface areas related to ability of charge transfer and hydrogen bonding.

##### 8.2.A priori or a posteriori mechanistic interpretation:

A posteriori interpretation.

##### 8.3.Other information about the mechanistic interpretation:

#### 9.Miscellaneous information

##### 9.1.Comments:

## 9.2. Bibliography:

EPA – ECOTOX Database <http://cfpub.epa.gov/ecotox/>

## 9.3. Supporting information:

Training set(s)

training set	<a href="http://qsardb.jrc.ec.europa.eu:80/qmrf/download_attachment.jsp?name=qmrf150_Molcode_rainbow_trout_training_161.sdf">http://qsardb.jrc.ec.europa.eu:80/qmrf/download_attachment.jsp?name=qmrf150_Molcode_rainbow_trout_training_161.sdf</a>
--------------	---

Test set(s)

test set	<a href="http://qsardb.jrc.ec.europa.eu:80/qmrf/download_attachment.jsp?name=qmrf150_Molcode_rainbow_trout_test_40.sdf">http://qsardb.jrc.ec.europa.eu:80/qmrf/download_attachment.jsp?name=qmrf150_Molcode_rainbow_trout_test_40.sdf</a>
----------	---

## 10. Summary (JRC QSAR Model Database)

### 10.1. QMRF number:

Q8-10-14-150

### 10.2. Publication date:

2009/12/18

### 10.3. Keywords:

Molcode, acute fish toxicity, rainbow trout

### 10.4. Comments: