| | |
|---|---|
| | *QMRF identifier (JRC Inventory):* |
| | *QMRF Title:* QSAR model for Toxicity to algae of benzene derivatives |
| | *Printing Date:*25.02.2010 |
| | |

## 1.QSAR identifier

### 1.1.QSAR identifier (title):

QSAR model for Toxicity to algae of benzene derivatives

### 1.2.Other related models:

### 1.3.Software coding the model:

QSARModel 4.0.3 Molcode Ltd., Turu 2, Tartu, 51014, Estonia http://www.molcode.com

## 2.General information

### 2.1.Date of QMRF:

07.12.2009

### 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:**

09.01.2010

**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:**

All information in full detail 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:**

Chlorella vulgaris

**3.2.Endpoint:**

3.Ecotoxic effects 3.2.Short-term toxicity to algae (inhibition of the exponential growth rate)

**3.3.Comment on endpoint:**

C.3 in REACH reach classification. The EC50 is the concentration (mM) that induces toxicity response halfway between the baseline and maximum after 15 min.

**3.4.Endpoint units:**

mM

**3.5.Dependent variable:**

log(1/EC50)

**3.6.Experimental protocol:**

Toxicity data [log(1/EC50)] were determined in a biochemical assay utilizing the unicellular alga C. vulgaris. Algae in the logarithmic phase of their growth cycle were used. All toxicological analyses were performed in a buffer solution with a pH of 6.9 and a temperature between 25 and 30 °C. Assays were conducted following the protocol described by Worgan et al. (i) with a 15 min static design. The disappearance of FDA was accounted for by spectrofluorimetric measurement of fluorescein (the product of hydrolysis) (ii) at an excitation wavelength of 465 nm and an emission wavelength of 515 nm. Range-finding experiments were performed in order to determine the highest and lowest concentrations required to produce a dose-response relationship ranging from 100% inhibition of enzyme activity to no observed toxicological effect. Blank buffer solution was utilized as a control, and the relative responses to it were used to generate the

dose-response curve. The 50% effective concentration was estimated by Probit analysis using the SPSS ver. 10.0 (SPSS Inc., Chicago, IL) software. The average EC50 was taken from a minimum of three analyses.

Literature data from one data series and one publication (ref 1. in 9.2) was selected to ensure consistency.

References:

(i) Worgan, A. D. P., Dearden, J. C., Edwards, R., Netzeva, T. I., and Cronin, M. T. D. (2003) Evaluation of a novel short-term algal toxicity assay by the development of QSARs and inter-species relationships for narcotic chemicals. QSAR Comb. Sci. 22, 204-209.

(ii) Leszczynska, M., and Oleszkiewic, J. A. (1996) Application of fluorescein diacetate hydrolysis as an acute toxicity test. Environ. Technol. 17, 79-85.

### 3.7.Endpoint data quality and variability:

The toxicity data are from one lab and one experiment series. The consistency additionally confirmed by previous QSAR treatment (Ref 1. in 9.2)

Statistics:
max value: 3.1
min value: -4.06
standard deviation: 1.465
skewness: -0.422

## 4.Defining the algorithm - OECD Principle 2

### 4.1.Type of model:

2D and 3D regression-based QSAR

### 4.2.Explicit algorithm:

multilinear regression QSAR
multilinear regression QSAR derived with BMLR (Best Multiple Linear Regression) method

log(1/EC50) = -3.532
+0.371*Kier&Hall index (order 0)
+1.233*Number of benzene rings
-23.698*Difference (Pos - Neg) in Charged Part of Partial Charged Surface Area (Zefirov)
-22.954*HA dependent HDCA-2/SQRT(TMSA) (Zefirov) (all)

### 4.3.Descriptors in the model:

[1]Kier&Hall index (order 0) unitless zero order Kier and Hall valence connectivity index
[2]Number of benzene rings unitless Count of benzene rings in the molecule

[3]Difference (Pos - Neg) in Charged Part of Partial Charged Surface Area (Zefirov) Å2 total difference between the charged positive and negative

charged surface areas
[4]HA dependent HDCA-2/SQRT(TMSA) (Zefirov) (all) au Area-weighted surface charge of hydrogen bonding donor atoms

### 4.4.Descriptor selection:

Initial pool of ~1000 descriptors. Stepwise descriptor selection based on a set of statistical selection rules (one-parameter equations: Fisher criterion and R2 over threshold, variance and t-test value over threshold, intercorrelation with another descriptor not over threshold),

(two-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 4.0.3
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:

0.055, (4 descriptors / 73 chemicals) 18.25 (73 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: benzene derivatives with one aromatic core

b) by descriptor value range: The model is suitable for compounds that have the descriptors
in the following minimal-maximal range:
Kier&Hall index (order 0): 1.45 - 13.9
Number of benzene rings: 0 - 2
Difference (Pos - Neg) in Charged Part of Partial Charged Surface Area (Zefirov): -0.0593 - 0.00616
HA dependent HDCA-2/SQRT(TMSA) (Zefirov) (all): 0 - 0.0655

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

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 4.0.3
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:
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:
73 data points
34 negative values
39 positive values
6.6.Pre-processing of data before modelling:
n/a
6.7.Statistics for goodness-of-fit:
R2 = 0.921 (Correlation coefficient)
s2 = 0.427 (Standard error of the estimate)
F = 197.8 (Fisher function)
6.8.Robustness - Statistics obtained by leave-one-out cross-validation:
R2CV = 0.904
6.9.Robustness - Statistics obtained by leave-many-out cross-validation:
R2CVMO = 0.903
6.10.Robustness - Statistics obtained by Y-scrambling:
n/a
6.11.Robustness - Statistics obtained by bootstrap:
n/a
6.12.Robustness - Statistics obtained by other methods:
ABC analysis (2:1 training : prediction) on sorted (in increased 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 R2 (fitting) = 0.923
average R2 (prediction) = 0.900

## 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:

18 data points,
9 negative values,
9 positive values

### 7.6.Experimental design of test set:

From sorted data each 5th was subjected to the test set starting from 3rd in order to assure the equality in distribution tails.

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

R2 = 0.887 (Coefficient of determination)

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

Descriptor value range (all in range of applicability domain):
Kier&Hall index (order 0): 3.57 - 12.9
Number of benzene rings: 0 - 2
Difference (Pos - Neg) in Charged Part of Partial Charged Surface Area (Zefirov): -0.0236 - 0.00995
HA dependent HDCA-2/SQRT(TMSA) (Zefirov) (all): 0.00548 - 0.0543

### 7.9.Comments on the external validation of the model:

The validation coefficient of determination (R2) is good and close to those coefficients of internal validation (R2CV and R2CVMO).

## 8.Providing a mechanistic interpretation - OECD Principle 5

### 8.1.Mechanistic basis of the model:

Descriptors "Kier&Hall index (order 0)" and "Number of benzene rings" are defining a nonpolar narcosis. They cover the toxicity baseline that is usually modelled with logP. Descriptors "Difference (Pos - Neg) in Charged Part of Partial Charged Surface Area (Zefirov)" and "HA dependent HDCA-2/SQRT(TMSA) (Zefirov) (all)" are related to ractivity of the compounds and they are representing a polar narcosis

part of the toxicity.

### 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:

Assessment and Modeling of the Toxicity of Organic Chemicals to Chlorella vulgaris: Development of a Novel Database

Mark T. D. Cronin, Tatiana I. Netzeva, John C. Dearden, Robert Edwards, and Andrew D. P. Worgan

Chem. Res. Toxicol., 2004, 17 (4), pp 545–554

http://dx.doi.org/10.1021/tx0342518

## 9.Miscellaneous information

### 9.1.Comments:

Data is taken from:

Assessment and Modeling of the Toxicity of Organic Chemicals to Chlorella vulgaris: Development of a Novel Database

Mark T. D. Cronin, Tatiana I. Netzeva, John C. Dearden, Robert Edwards, and Andrew D. P. Worgan

Chem. Res. Toxicol., 2004, 17 (4), pp 545–554

http://dx.doi.org/10.1021/tx0342518

### 9.2.Bibliography:

Assessment and Modeling of the Toxicity of Organic Chemicals to Chlorella vulgaris: Development of a Novel Database http://dx.doi.org/10.1021/tx0342518

### 9.3.Supporting information:

Training set(s)Test set(s)Supporting information

| | |
|---|---|
| Karelson Arkivoc 2008 | http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf83_Karelson Arkivoc 2008.pdf |
| Karelson Arkivoc 2009 | http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf83_Karelson Arkivoc 2009.pdf |

## 10.Summary (ECB Inventory)

### 10.1.QMRF number:

### 10.2.Publication date:

### 10.3.Keywords:

### 10.4.Comments: