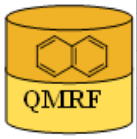	*QMRF identifier (JRC Inventory):* *QMRF Title:* *QSAR model for Toxicity to algae* *Printing Date:*26.02.2010	

1.QSAR identifier

1.1.QSAR identifier (title):

QSAR model for Toxicity to algae

1.2.Other related models:

1.3.Software coding the model:

QSARModel 4.0.4 Molcode Ltd., Turu 2, Tartu, 51014, Estonia http://www.molcode.com

2.General information

2.1.Date of QMRF:

14.01.2010

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:

11.12.2009

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:

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:

Pseudokirschneriella subcapitata

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 classification. The EC50 is the concentration (mM) that induces toxicity response halfway between the baseline and maximum after 96 h.

3.4.Endpoint units:

mM

3.5.Dependent variable:

log(1/EC50)

3.6.Experimental protocol:

The unicellular green algae (also known by its synonyms Selenastrum capricornutum and Raphidocelis subcapitata) is the most widely used freshwater organism to test alga toxicity. Growth rate inhibition was selected as the toxic end-point within 96 h.

The test alga was an unicellular green algal species Selenastrum capricornutum Printz (also known as Pseudokirschneriella subcapitata and Raphidocelis subcapitata) and the culture medium 10% Z 8. The inoculum was taken from a stock culture in the exponential growth phase. The initial algal density was 104 ± 10% cells/mL. The test algae were cultivated in 100-mL solutions in 250-mL sterile, foam-plugged Erlenmeyer flasks with three replicates of each concentration. In addition, there were two control cultures: Selenastrum cells in culture medium and in acetone series. There were also controls for chemicals

without algae.

The cultures were incubated at +22 ± 20 C in continuous illumination of approximately 72 µE m-2 s-1 (Airam L 40 W 35). The growth of cultures was followed by measuring the cell density after 24, 48, 72 and 96 hr by means of an electronic particle counter (Coulter Counter Z B). The effect of acetone on the growth of the cultures was eliminated by comparing the growth of test cultures with the growth of acetone-controls. The results, as percent of control, were calculated as a mean value of the cell density of the triplicates after one test series.

In Selenastrum assays, the EC50-values were estimated from semilogarithmic paper using cell density after 96 hr and areal comparison of growth curves during 0-96 hr incubation (ISO 1983).

Reference:

K. Kuivasniemi, V. Eloranta, and J. Knuutinen, Arch. Environ. Contam. Toxicol. 14, 43-49 (1985)

3.7.Endpoint data quality and variability:

Experimental data from a number of different publications was used, as assembled in publication listed in 9.2 The consistency additionally confirmed by previous QSAR treatment (Ref 1. in 9.2)

Statistics:

max value: 3.99

min value: -0.2

standard deviation: 1.02

skewness: 0.175

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) = 1.656

-9.940*Relative number of rings

-8.465E-002*WPSA3 Weighted PPSA (PPSA3*TMSA/1000) (AM1)

+1.111E-003*Gravitation index (all atom pairs) (AM1)

-2.543*Polarity parameter (AM1) / square distance

4.3.Descriptors in the model:

[1]Relative number of rings unitless number of rings divided by number of atoms

[2]WPSA3 Weighted PPSA (PPSA3*TMSA/1000) (AM1) mÅ4 surface weighted atomic charge weighted partial positively charged surface area

[3]Gravitation index (all atom pairs) (AM1) amu2/Å2 sum over masses of all pairs of atoms divided by interatomic distance

[4]Polarity parameter (AM1) / square distance au/ Å2 difference between

most positive and most negative atomic charge divided by squared distance

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:

0.1, (4 descriptors / 40 chemicals) 10 (40 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: diverse set of organic chemicals (alcohols, ketones, amines, ethers, halogeno compounds, etc)

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

in the following minimal-maximal range:

Relative number of rings: 0 - 0.13

WPSA3 Weighted PPSA (PPSA3*TMSA/1000) (AM1): 3.13 - 33.9

Gravitation index (all atom pairs) (AM1): 296 - 4920

Polarity parameter (AM1) / square distance: 0.0191 - 0.622

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 4.0.4

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:

40 data points
1 negative values
39 positive values

6.6.Pre-processing of data before modelling:

n/a

6.7.Statistics for goodness-of-fit:

R2 = 0.924 (Correlation coefficient)
s2 = 0.301 (Standard error of the estimate)
F = 106.0 (Fisher function)

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

R2CV = 0.881

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

R2CVMO = 0.877

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) = 0.930
average R2 (prediction) = 0.848

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:

5 data points,
0 negative values,
5 positive values

7.6.Experimental design of test set:

From sorted data each 8th was subjected to the test set.

7.7.Predictivity - Statistics obtained by external validation:

R2 = 0.852 (Coefficient of determination)

7.8.Predictivity - Assessment of the external validation set:

Descriptor value range (all except Octafonium chloride are in range of applicability domain):

Relative number of rings: 0.0241 - 0.0769
WPSA3 Weighted PPSA (PPSA3*TMSA/1000) (AM1): 4.88 - 58.1
Gravitation index (all atom pairs) (AM1): 962 - 5950
Polarity parameter (AM1) / square distance: 0.0915 - 0.514

Although Octafonium chloride "WPSA3 Weighted PPSA (PPSA3*TMSA/1000) (AM1)" descriptor value is out of model range the prediction of pEC50 is good. This shows that the model can be even extrapolated.

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:

Descriptor "Gravitation index (all atom pairs) (AM1)" is size related and represents non-specific interactions. Therefore it describes nonpolar narcosis. "Relative number of rings" is related to flexibility of the molecules. Higher flexibility leads to better membrane permeability and to higher interaction freedom. Descriptors "WPSA3 Weighted PPSA (PPSA3*TMSA/1000) (AM1)" and "Polarity parameter (AM1) / square distance" are directly charge dependent and thus they contribute too to

membrane permeability but more importantly they are related to polar narcosis.

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:

Similar mechanistic interpretation as in:

Prediction of Toxicity of Phenols and Anilines to Algae by Quantitative Structure-activity Relationship

Guang-Hua LU, Chao WANGXiao-Ling GUO

Biomedical and Environmental Sciences

Volume 21, Issue 3, February 2008, Pages 193-196

doi:10.1016/S0895-3988(08)60028-8

9.Miscellaneous information

9.1.Comments:

Data is taken from:

The importance of outlier detection and training set selection for reliable environmental QSAR predictions

Erik Furusjö, Anders Svenson, Magnus Rahmberg, Magnus Andersson

Chemosphere, Volume 63, Issue 1, March 2006, Pages 99-108

http://dx.doi.org/10.1016/j.chemosphere.2005.07.002

9.2.Bibliography:

The importance of outlier detection and training set selection for reliable environmental QSAR predictions http://dx.doi.org/10.1016/j.chemosphere.2005.07.002

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: