

	QMRF identifier (ECB Inventory): To be entered by JRC	
	QMRF Title: QSAR model for toxicity to <i>Tetrahymena Pyriformis</i> of non hydrogen-bonding aromatic compounds	
	Printing Date: Jan 19, 2012	

1. QSAR identifier

1.1. QSAR identifier (title):

QSAR model for toxicity to *Tetrahymena Pyriformis* of non hydrogen-bonding aromatic compounds

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:

10.06.2010

2.2. QMRF author(s) and contact details:

Dana Martin, Tarmo Tamm, Gunnar Karelson, Indrek Tulp, Kaido Tämm, Deniss Savchenko, Jaak Jänes, Eneli Härk, Andres Kreegipuu, Dimitar Dobchev, Mati Karelson. Molcode model development team 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:

Dana Martin, Tarmo Tamm, Gunnar Karelson, Indrek Tulp, Kaido Tämm, Deniss Savchenko, Jaak Jänes, Eneli Härk, Andres Kreegipuu, Dimitar Dobchev, Mati Karelson. 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:

10.06.2010

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

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

[2] Mati Karelson, Gunnar Karelson, Tarmo Tamm, Indrek Tulp, Jaak Jänes, Kaido Tämm, Andre Lomaka, Deniss Savchenko, and Dimitar Dobchev QSAR study of pharmacological permeabilities, *ARKIVOC* 2, 2009, pp. 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:

Tetrahymena Pyriformis

3.2. Endpoint:

3. Ecotoxic effects 3.10. Toxicity to soil invertebrates (survival, growth, reproduction)

3.3. Comment on endpoint:

The TETRATOX assay is a short-term, static protocol using Tetrahymena Pyriformis, with the 50% impairment growth concentration (IGC50) of the recorded endpoint.

3.4. Endpoint units:

mmol/L

3.5. Dependent variable:

$\log(1/IGC50)$

logarithm of 1/ IGC50.

3.6. Experimental protocol:

Tetrahymena are ciliated freshwater protozoa about 50 μm long. They are very common in aquatic habitats and have a short generation time. Following 40 h of incubation of tetrahymena cells at $27\pm 1^\circ$ in the media of test substances, the median growth inhibition (IGC50) was determined. The concentrations of test substances are given in mmol per litre (mmol/L).

The dataset of monoaromatic homologues each containing either a nitro or a cyano moiety was compiled from the TETRATOX database.

Tests were conducted in foam-stoppered 250 mL Erlenmeyer flasks containing 50 mL of sterile, semi-defined proteose peptone-based medium. Test cultures were inoculated with log growth phase culture to an initial density of about 2,500 cells/mL. The endpoint, population density, of this static 40-h assay was measured spectrophotometrically at 540 nm. Each compound was tested in a range finder prior to testing in duplicate for three additional replicates. Two controls, one without test chemical but inoculated with T. pyriformis and one with a blank with neither test chemical nor ciliates, were used to provide a measure of the test acceptability and as a basis for interpretation of treatment data. Each definitive test replicate consisted of six to eight different concentrations with duplicate flasks of each concentration. Only replicates with control-absorbency values between 0.60 and 0.75 were used in the analyses.

3.7. Endpoint data quality and variability:

Experimental data from database source (of probably different experimental test series and labs) has been used, previous successful modelling supports consistency.

Statistics:

max value: 2.66

min value: -0.82

standard deviation: 0.79

skewness: 0.45

4. Defining the algorithm - OECD Principle 2

4.1. Type of model:

2D and 3D regression-based QSAR Multilinear regression model based on 3-D quantum chemical descriptors.

4.2. Explicit algorithm:

multilinear regression QSAR

QSAR model based on eq.

$$\text{Log}(1/\text{IGC50}) = -7.96 - 2.56 * \text{Max bonding contribution of one MO(AM1)} - 6.35 \text{E-}002 * \text{Highest e-n attraction 1-center(AM1) for C atoms} + 0.38 * \text{Kier shape index(order 1)}$$

4.3. Descriptors in the model:

[1] Max bonding contribution of one MO(AM1)

[2] Highest e-n attraction 1-center(AM1) for C atoms N

[3] Kier shape index(order 1)

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 3.5.0

Molcode Ltd., Turu 2, Tartu, 51014, Estonia

<http://www.molcode.com>

4.7. Chemicals/Descriptors ratio:

33 (99 chemicals/3 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: aromatic compounds which do not form H bonds

By descriptor value range: This model is suitable for compounds that have the descriptors in the following range: Max bonding contribution of one MO(AM1) (min: 1.10 , max: 2.80), Highest e-n attraction 1-center(AM1) for C atoms (min: -229.96 , max: -207.94), Kier shape index(order 1) (min: 3.35 , max: 15.96)

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

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: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:

data points: 99 , negative: 12, positive values: 87

6.6.Pre-processing of data before modelling:

6.7.Statistics for goodness-of-fit:

$R^2 = 0.87$ (Correlation coefficient);

$s^2 = 8.38 \text{ E-}002$ (Standard error of the estimate);

$F = 212.13$ (Fisher function);

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

$R^2_{cv} = 0.86$ LOO;

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

$R^2_{cv} = 0.92$ 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) = 0.87

average R^2 (prediction) = 0.88

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:

data points: 11, negative: 1, positive values: 10

7.6.Experimental design of test set:

The full experimental dataset was sorted according to increasing values of $\log(1/IGC50)$ and each tenth compound was assigned to the test set.

7.7.Predictivity - Statistics obtained by external validation:

$R^2=0.801$

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:

The validation R^2 for the test set is good.

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

The toxicity to *Tetrahymena Pyriformis* depends on the shape and the stability of the molecule. The shape of the molecule is reflected by the descriptor Kier shape index (order 1). The toxicity increases with the increased value of the descriptor Kier shape index (order 1). The second descriptor that appear in the model of the toxicity is Max bonding contribution of one MO(AM1). The toxicity increases with decreasing stability of the molecule (decreasing values of Max bonding contribution of one MO(AM1))because a less stable molecule can better react with the animal tissue. The descriptor Highest e-n attraction 1-center(AM1) for C atoms brings additional correction to the stability of the molecule.

8.2.A priori or a posteriori mechanistic interpretation:

a posteriori mechanistic interpretation, consistent with published scientific interpretations of experiments [4]

8.3.Other information about the mechanistic interpretation:

Kier shape index (order 1) capture different aspects of model shape and encodes the count of atoms and the presence of cycles. The descriptor Max bonding contribution of one MO(AM1) reflects the stability of the molecule and increased values of the descriptor reflect an increased stability. The descriptor Highest e-n attraction 1-center (AM1) for C atoms gives information about the stability and polarity of the molecule. This descriptor reflects energy partitioning on atoms in the molecule. Carbon atoms with a low e-n attraction energy are usually part of a polar functional group(carboxyl, carbonyl).

9.Miscellaneous information

9.1.Comments:

9.2.Bibliography:

[1]Schultz T.W. TETRATOX: The Tetrahymena pyriformis population growth impairment endpoint – A surrogate for fish lethality, Toxicol. Methods 7 (1997) 289–309.

[2]TETRATOX <http://www.vet.utk.edu/TETRATOX/index.php>

[3]Cronin M. T. D., Manga N., Seward J.R., Sinks G.D., Schultz T.W. Parameterization of electrophilicity for the prediction of the toxicity of aromatic compounds, Chem. Res. Toxicol. 14 (2001) 1498 – 1505.

[4]Laszlo T. and Beteringhe A. QSAR Studies Related to Toxicity of Aromatic Compounds on Tetrahymena Pyriformis, QSAR Comb. Sci. 25 No.10 (2006) 944 – 951.

9.3.Supporting information:

Training set(s)

T_pyriformis_trainingset.sdf	http://qsardb.jrc.ec.europa.eu:80/qmrf/download_attachment.jsp?name=qmrf248_T_pyriformis_trainingset.sdf
------------------------------	---

Test set(s)

T_pyriformis_testset.sdf	http://qsardb.jrc.ec.europa.eu:80/qmrf/download_attachment.jsp?name=qmrf248_T_pyriformis_testset.sdf
--------------------------	---

10.Summary (JRC Inventory)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

To be entered by JRC

10.4.Comments:

To be entered by JRC