	QMRF identifier (JRC Inventory): *To be entered by JRC*
	QMRF Title: *QSAR for toxicity to activated sludge*
	Printing Date: *30.01.2011*

1.QSAR identifier

1.1.QSAR identifier (title):

QSAR for toxicity to activated sludge

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:

18.08.2009

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]Dimitar Dobchev Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[4]Gunnar Karelson Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[5]Jaak Jänes 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]Eneli Härk Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[8]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:

18.08.2009

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:

Pseudomonad bacteria strain (Shk1)

3.2.Endpoint:

3.Ecotoxic effects 3.6.Microbial inhibition (activated sludge respiration inhibition, inhibition of nitrification, other)

3.3.Comment on endpoint:

C.11 in REACH classification, The EC50 refers to the concentration of a drug, antibody or toxicant which induces a response halfway between the baseline and maximum after some specified exposure time.

3.4.Endpoint units:

mM

3.5.Dependent variable:

log (1/EC50) - logarithm of the reciprocal of the half maximal effective concentration (EC50).

3.6.Experimental protocol:

Biodegradation - activated sludge respiration inhibition was determined using the EU Test Guideline C.11. This method described assesses the effect of a test substance on micro-organisms by measuring the respiration rate under defined conditions in the presence of different concentrations of the test substance. Activated sludge respiration inhibition test provides a rapid screening and identification of water soluble substances which may adversely affect aerobic microbial treatment plants. Toxicants can cause upsets in the operations of an activated sludge process (decrease in waste organics removal, reduction of solids separation efficiency etc). EC50 is the concentration of the test substance at which the respiration rate is 50% of that shown by the control under conditions described in this method.

The European Chemicals regulation REACH (registration, evaluation, authorisation of chemicals) describes possible applications of in vitro alternatives to replace expensive and time consuming test methods. Demanding validation procedures are applied by European Centre for the Validation of Alternative Methods (ECVAM) to assure the suitability of any new in vitro method for regulatory purposes.

An alternative method for monitoring of activated sludge respiration

inhibition uses an assay based on genetically modified luminescent bacterium Shk1 whose original strain was a Pseudomonad isolated from the activated sludge in an industrial wastewater treatment plant. Recent studies have been demonstrated that the response of Shk1 to toxicants, due to the microorganism's origin in a wastewater treatment plant, is closer to that of the activated sludge microbial community than Vibrio fischeri, a marine bacterium used in the Microtoxs assay [3]. Using Shk1 assay, EC50 is the concentration of the toxicant required to reduce the emission of the bioluminescence of 50%. The original experimental data of the present model represents that of the Shk1 assay.

The experimental setup consisted of a continuous stirred tank reactor (CSTR) operated at steady state. The operating conditions of the CSTR in this study were a dilution rate of 0.52 h-1 , pH=7 and T=220C. The CSTR was used to supply Shk1 cells continuously. The Shk1 cells were alternately mixed with buffer (pH=7) or test solutions at a 1:3 ratio (volume). The bioluminescence of Shk1 was measured after exposing Shk1 cells to buffer or test solutions for 5 min.

3.7.Endpoint data quality and variability:

The data has been selected from the EPA ECOTOX database [1], the experimental data originates from different laboratories, previous [2] and present successful modelling suport its consistency.

Statistics: max value: 1.5, min value: -3.10, standard deviation: 0.87, skewness: -0.86

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

multilinear regression QSAR

Log(1/EC50) = -2.96 + 0.43*log(Kow) + 0.36*Kier&Hall index(order 0)

4.3.Descriptors in the model:

[1]log(Kow) log of octaonl-water partition coefficient

[2]Kier&Hall index(order 0)

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 (0.1), variance and t-test value over threshold (1.5), intercorrelation with another descriptor not over threshold), (2 parameter equations: intercorrelation coefficient bellow threshold (0.3), significant correlation with endpoint in terms of correlation coefficient and t-test (0.15, 1.5). 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 MMFFs(vacuum) conformational search and AM1 calculation. All structures treated as neutral molecules. Model developed by using multilinear regression.

4.6.Software name and version for descriptor generation:

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

4.7.Descriptors/Chemicals ratio:

42 (84 chemicals / 2 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:

diverse set of organic compounds: halogenated compounds, alcohols, phenols, amines, carboxylic acids, etc.

By descriptor value range:

This model is suitable for compounds that have the descriptors in the following range: log(Kow)(min: -1.30 , max: 5.47), Kier&Hall index(order 0)(min: 1.02, max: 10.48)

5.2.Method used to assess the applicability domain:

Presence of functional groups in structures. 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 3.3.8
Molcode Ltd., Turu 2, Tartu, 51014, Estonia,
http://www.molcode.com

5.4.Limits of applicability:

none more than indicated in 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: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:

84 data points: 50 negative values; 34 positive values

6.6.Pre-processing of data before modelling:

original data in mg/L were converted to molar form (mmol/L) and expressed in logarithmic form

6.7.Statistics for goodness-of-fit:

R^2 = 0. 790 (Correlation coefficient); s = 0.170 (Standard error of the estimate); F = 152.41 (Fisher function);

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

R^2_{CV} = 0.770 LOO;

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

R^2_{CV} = 0.800 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) = 0.800

average R^2 (prediction) = 0.760

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:

9 data points: 6 negative values; 3 positive values

7.6.Experimental design of test set:

The full experimental dataset was sorted according to increasing endpoint value, and each tenth compound was assigned to the test set.

7.7.Predictivity - Statistics obtained by external validation:

R^2=0.840

7.8.Predictivity - Assessment of the external validation set:

The descriptors for the test set are in the limit of applicability domain.

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 sludge toxicity depends on the hydrophobic/hydrophilic character (represented by log(Kow)), as it relates to the capacity of an organic compound to penetrate into the animal tissue). The second descriptor that appears in the model of the toxicity is Kier&Hall index (order 0). This descriptor gives information about the valency and identity of atoms present in the molecule. The value of the Kier&Hall reflects also the size of the molecule skeleton. The toxicity increase with increased value of the descriptor Kier&Hall index (order 0) and so with the increased size of the molecule, as well as the increased complexity of the structure (presence of heteroatoms and functional groups).

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:

The descriptor Kier&Hall index (order 0) gives information about the valency and identity of atoms in the molecule. The Kier&Hall index (order 0) refers to the zero-order subgroups in the molecular graph. The number of subgroups of zero order is simply the number of skeletal atoms or vertices in the molecular graph. Each vertex has a property δ, which is the number of its electrons in sigma bonds to skeletal neighbors. When the molecule has a bigger structural skeleton (more functional groups in the molecule) the descriptor Kier&Hall index (order 0) has higher values. Log(1/EC50) increase with the increasing value of Kier&Hall index(order 0) and as a consequence with the increased skeleton size of the molecule. The descriptor log(Kow) gives information about the solubility of the compound in octanol relative to that in water. The toxicity increases with the solubility of the compound in octanol (higher hydrophobicity). The proposed mechanism based on the model agrees well with literature [2,3]

9.Miscellaneous information

9.1.Comments:

The activated sludge process (ASP) is widely used for the treatment of industrial and domestic wastewater. Toxicants in ASP influent can cause upsets in the operations of an ASP. These upsets include a decrease in waste organics removal, reduction of solids

separation efficiency, and modification of sludge compacting properties. Such upsets can be avoided if ASP influent wastewater is screened for toxicity and protective actions are taken.

9.2.Bibliography:

[1]EPA ECOTOX Database http://cfpub.epa.gov/ecotox/
[2]Ren S, Frymier PD (2002) Estimating the toxicities of organic chemicals to bioluminescent bacteria and activated sludge. Water Research, 36(17), pp. 4406-4414(9)
[3]Ren S. Development of a continuous bioluminescent bacteria-based system for POTW influent wastewater toxicity monitoring. Ph.D. Dissertation, University of Tennessee, Knoxville, 2001.

9.3.Supporting information:

Training set(s)

Act_sludge_tox_training.sdf	file:///D:\User-DATA\My Documents\a_QSAR\a_QSAR 2009\QMRF 2009\model submissions\1 new\2009-9-08_Activated_sludge_toxicity_E Haerk\Act_sludge_tox_training.sdf

Test set(s)

Act_sludge_tox_test.sdf	file:///D:\User-DATA\My Documents\a_QSAR\a_QSAR 2009\QMRF 2009\model submissions\1 new\2009-9-08_Activated_sludge_toxicity_E Haerk\Act_sludge_tox_test.sdf

Supporting information

10.Summary (JRC QSAR Model Database)

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