	QMRF identifier (JRC Inventory): **Q13-34-0063**
	QMRF Title: **Nonlinear QSAR: artificial neural network for the Daphnia magna reproduction test**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

Nonlinear QSAR: artificial neural network for the Daphnia magna reproduction test

1.2.Other related models:

1.3.Software coding the model:

QSARModel 3.3.8

Turu 2, Tartu, 51014, Estonia

http://www.molcode.com

Statistica 7

StatSoft Ltd.

2.General information

2.1.Date of QMRF:

10.10.2010

2.2.QMRF author(s) and contact details:

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:

Molcode model development team Molcode Ltd Turu 2, Tartu, 51014, Estonia

models@molcode.com www.molcode.com

2.6.Date of model development and/or publication:

12.04.2010

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

[1]Katritzky AR, Dobchev DA, Fara DC, Hur E, Tämm K, Kurunczi L, Karelson M, Varnek A & Solov'ev VP (2006). Skin Permeation Rate as a Function of Chemical Structure. Journal of Medicinal Chemistry 49, 3305-3314.

[2]Karelson M, Dobchev DA, Kulshyn OV & Katritzky A (2006). Neural Networks Convergence Using Physicochemical Data. Journal of Chemical Information and Modeling 46, 1891- 1897.

2.8.Availability of information about the model:

Training, selection and test sets are 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:

Daphnia magna

3.2.Endpoint:

3.Ecotoxic effects 3.4.Long-term toxicity to Daphnia (lethality, inhibition of reproduction)

3.3.Comment on endpoint:

see 3.6

3.4.Endpoint units:

mmol/L

3.5.Dependent variable:

LogEC50

3.6.Experimental protocol:

The reproduction toxicity to Daphnia was determined using the OECD 211 (EU C.20) test guideline [ref 1, sect 9.2]. Young female Daphnia (the parent animals), aged less than 24 hours at the start of the test, are exposed to the test substance added to water at a range of concentrations. The test duration is 21 days. At the end of the test, the total number of living offspring produced per parent animal alive at the end of the test is assessed. This means that juveniles produced by adults that die during the test are excluded from the calculations. The reproductive output of the animals exposed to the test substance is compared to that of the control(s) in order to determine the median effective concentration EC50 (LC50). This is the concentration of the test substance dissolved in water that results in a 50% reduction in reproduction of Daphnia magna within 21days. The concentrations of the substances are given in mmol per litre (mmol/L).

D. magna was obtained from the National Institute for Environmental Studies (NIES), Tsukuba, Japan. The reproduction test was performed for 21 days according to the methods for survival and reproduction tests on D. magna proposed by the OECD. Females less than 24 h old were used as the founding females in each test. They were exposed to various concentrations of the test substance according to the OECD test conditions, then fed and observed daily for 21 days. Cultures were kept in an incubator at a temperature of 24±10C and a photoperiod of 14 h light/10 h dark. Six nominal concentrations of each test chemical, including a culture water control, were prepared by dilution with fresh culture water. All 21-day experiments were conducted with a dilution factor of 3 for test substances. Eight replicate glass jars (100 ml), each containing an individual D. magna female in 50 ml of media, were used for each concentration. The jars were covered with Teflon caps to prevent volatilization of the test chemicals. The water quality (pH and dissolved oxygen concentration) was measured every 2 days (right after changing of water). A suspension of 0.05 ml of Chlorella (4.3 • 108 cells ml/1) was added to each jar daily. Water hardness, pH, and dissolved oxygen concentration were 75–85 mgl/1, 7.0–7.5, and 80–99%, respectively. The medium was changed every 2 days, and neonates were removed from the jar every day and were counted by eye. The total number of neonates born over 21 days at each concentration of test chemical, as

well as the total number born to the control group, were calculated and compared [ref 2 – 3, sect 9.2].

3.7.Endpoint data quality and variability:

The data are taken from one source [ref 1, sect 9.2]. However, it is uncertain whether all experimental data points were obtained from a single laboratory.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

QSAR

Nonlinear QSAR: Backpropagation Neural Network (Multilayer Perceptron) regression

The algorithm is based on regression neural network predictor with structure 7-6-5-1

4.3.Descriptors in the model:

[1]Avg nucleophilic reactivity index (AM1) for H atoms

[2]Max Sigma-Sigma bond order (AM1)

[3]Relative number of H atoms

[4]Tot molecular 2-center resonance energy (AM1) / # of atoms

[5]Lowest atomic state energy (AM1) for H atoms

[6]Highest resonance energy (AM1) for C - H bonds

[7]No. of occupied electronic levels (AM1) / # atoms

4.4.Descriptor selection:

Initial pool of ~899 descriptors. Stepwise descriptor selection based on a set of statistical selection rules as F statistic and p. The first highest F (low p) descriptors (7) were selected from the total number of descriptors. These 7 descriptors were used as inputs to the network. 16 networks with different structures were tested in order to find the best ANN with lowest RMS (root-mean-squared error) and highest correct predictions (for training, selection and test sets). Then 555 epochs were used to train the final network with architecture depicted in 4.2. Optimization of the weights was performed with Levenberg-Marquardt algorithm encoded in the backpropagation scheme using linear and hyperbolic activation functions.

4.5.Algorithm and descriptor generation:

All descriptors were generated using QSARModel on structures optimized by AM1 semiempirical quantum mechanical model.

4.6.Software name and version for descriptor generation:

QSARModel 3.3.8

Turu 2, Tartu, 51014, Estonia

http://www.molcode.com

Statistica 7

StatSoft Ltd.

http://www.statsoft.com

4.7.Chemicals/Descriptors ratio:

28 (196 chemicals / 7 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)functional groups such as phenols, aldehydes, nitro, amino, alcohols, halides, aromatics, aliphatic functional groups

b)The model is suitable for compounds that have descriptors values in the following range:

Desc: 1 2 3 4 5 6 7

min: 0.000; 0.633; 0.077; -17.765; -8.052; -11.095; 1.000

max: 0.005; 0.925; 0.667; -8.194; -6.684; 0.000; 2.600

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 (see 5.1) of training set ±30%.

5.3.Software name and version for applicability domain assessment:

QSARModel 3.3.8

Turu 2, Tartu, 51014, Estonia

http://www.molcode.com

Statistica 7

StatSoft Ltd.

http://www.statsoft.com

5.4.Limits of applicability:

See 5.1, 5.2

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:

196 data points

6.6.Pre-processing of data before modelling:

Standardization and normalization of the inputs by taking into account the mean and standard deviation

6.7.Statistics for goodness-of-fit:

Abs E. Mean: 0.632; 1.384; 1.416

SD Ratio: 0.393; 1.232; 1.011

Correlation: 0.919; 0.527; 0.750

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

See 6.7

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

6.10.Robustness - Statistics obtained by Y-scrambling:

6.11.Robustness - Statistics obtained by bootstrap:

6.12.Robustness - Statistics obtained by other methods:

RMS(Training)=0.068; RMS(Selection)=0.207; RMS(Test)=0.189

In this ANN, 2 randomly chosen sets (50) were used to test the network – selection set and test set; see also 6.7

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:

The method used two validation sets: selection (50) and test (50)

7.6.Experimental design of test set:

Randomly selected 50 selection and 50 test set points

7.7.Predictivity - Statistics obtained by external validation:

See 6.7 and 6.12

7.8.Predictivity - Assessment of the external validation set:

The descriptors for the test set are in the limit of applicability; see 6.7 and 6.12

7.9.Comments on the external validation of the model:

Overall predictions for the selection set (used to stop the ANN training and not to overfit it) and the test set (used to test the external

prediction of the net after training) are significant according to the RMS error and the standard deviation ratio (SD ratio); see 6.7 and 6.12

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

Most of the descriptors are related to the reactivity of the compounds related to the C and H atoms. A rough estimation can be made based on their values. Regarding the descriptor Avg nucleophilic reactivity index (AM1), for H atoms, it can be noted that it has slight negative correlation with the modelled property. This might suggest that with the increase of this descriptor, the property would decrease. The same holds for the descriptor Relative number of H atoms (correl -0.5). In contrast, the descriptor No. of occupied electronic levels (AM1) / # atoms leads to larger LogEC50 values (correlation 0.5).

8.2.A priori or a posteriori mechanistic interpretation:

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

Supporting information for: training set(s), delection set(s), test set(s)

9.2.Bibliography:

[1]OECD (1998). Daphnia magna reproduction test. In: OECD Guidelines for Testing of Chemicals 211. OECD, Paris.

[2]Results of Eco-toxicity tests of chemicals conducted by Ministry of the Environment in Japan (March 2010). http://www.env.go.jp/chemi/sesaku/02e.pdf

[3]Tatarazako N, Oda S, Watanabe H, Morita M & Iguchi T (2003). Juvenile hormone agonists affect the occurrence of male Daphnia. Chemosphere 53, 827–833.

9.3.Supporting information:

Daphnia_magna_reprod_21d_training_196.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-34-0063/attachment/A756
Daphnia_magna_reprod_21d_test_50.sdf	http://qsardb.jrc.ec.europa.eu/qmrffile:///C:\Documents and Settings\casadja\My Documents\a_QSAR\a_QMRFdb\model submissions\7 review completed\aQ19-22-1-336_Nonlinear ANN QSAR Dm EC50, 21d_prev - Q17-10-1-316\Daphnia_magna_reprod_21d_test_50.sdf
Daphnia_magna_reprod_21d_selection_50.sdf	http://qsardb.jrc.ec.europa.eu/qmrffile:///C:\Documents and Settings\casadja\My Documents\a_QSAR\a_QMRFdb\model submissions\7 review completed\aQ19-22-1-336_Nonlinear ANN QSAR Dm EC50, 21d_prev - Q17-10-1-316\Daphnia_magna_reprod_21d_selection_50.sdf

10.Summary (JRC QSAR Model Database)

Supporting information

10.1.QMRF number:

Q13-34-0063

10.2.Publication date:

2013-07-02

10.3.Keywords:

Daphnia magna;reproduction;Molcode;neural network;

10.4.Comments:

former Q19-22-1-336