	QMRF identifier (JRC Inventory): *To be entered by JRC*
	QMRF Title: *Nonlinear ANN QSAR Model for Repeated dose 90-day oral toxicity study in rodents*
	Printing Date: *19.05.2011*

1.QSAR identifier

1.1.QSAR identifier (title):

Nonlinear ANN QSAR Model for Repeated dose 90-day oral toxicity study in rodents

1.2.Other related models:

1.3.Software coding the model:

[1]QSARModel 3.3.8 Turu 2, Tartu, 51014, Estonia, http://molcode.com/software
[2]Statistica, StatSoft Ltd. www.statsoft.com

2.General information

2.1.Date of QMRF:

27.04.2011

2.2.QMRF author(s) and contact details:

Dimitar Dobchev, Tarmo Tamm, Gunnar Karelson, Indrek Tulp, Kaido Tämm, Eneli Härk, 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:

Molcode model development team Molcode Ltd Molcode Ltd Turu 2, Tartu, 51014, Estonia models@molcode.com www.molcode.com

2.6.Date of model development and/or publication:

18.04.2011

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

[1]Karelson, M.; Dobchev, D. A.; Kulshyn, O. V.; Katritzky, A. (2006). Neural Networks Convergence Using Physicochemical Data. Journal of Chemical Information and Modeling, 46, 1891 - 1897.
[2]Statistica www.statsoft.com

2.8.Availability of information about the model:

All data and modeling information is available

2.9.Availability of another QMRF for exactly the same model:

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Rat

3.2.Endpoint:

4.Human health effects 4.14.Repeated dose toxicity

3.3.Comment on endpoint:

Determination of repeated dose 90-day oral toxicity study in rodents

Repeated dose 90-day oral toxicity was determined using the OECD test guideline 408 (EU B.26). This test guideline has been designed to fully characterize test article toxicity by the oral route for a subchronic duration (90 days). It can provide an estimate of a no-observed-adverse-effect level (NOAEL) of exposure which can be used in selecting dose levels for chronic studies and for establishing safety criteria for human exposure. Groups of 10 male and 10 female rodents are used at each dose level. The test substance is orally administered daily (any other dosing regime, e.g., five days per week, needs to be justified) in graduated doses to several groups of experimental animals, one dose level per group. During the period of administration the animals are observed closely for signs of toxicity. Animals which die or are killed during the test are necropsied and, at the conclusion of the test, surviving animals are also killed and necropsied [1].

3.4.Endpoint units:

mmol/kg

3.5.Dependent variable:

log(1/NOAEL)

3.6.Experimental protocol:

The IUCLID Chemical Data Sheets were the sources of rat 90-day oral NOAELs. IUCLID 4 was used worldwide by chemical industry companies, EU Member State Competent Authorities, the OECD Secretariat, the US EPA, the Japan METI, and third party service providers. Any study utilizing rats conducted for 90 days (3 months, 13 weeks) was considered a subchronic study. When several NOAELs of a chemical were available in IUCLID Chemical Data Sheets Information System the value of latest study was preferred. NOAEL data were originally expressed in units of %, ppm or milligrams of chemical per kilogram (mg/kg). These values were converted to their corresponding mmol/kg in order to compare the respective correlations. Regressions were performed with 1/log(NOAEL) [2].

3.7.Endpoint data quality and variability:

Data from multiple sources, as collected from IUCLID4 database has been used.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

Nonlinear QSAR: Backpropagation Neural Network (Multilayer Perceptron) regression

4.2.Explicit algorithm:

The algorithm is based on neural network predictor with structure 8-6-5-1 Standard Backpropagation Neural Network (Multilayer Perceptron) regression

4.3.Descriptors in the model:

[1]LogP Log P was introduced as an external descriptor in this modeling. This is the calculated value based on a atom-additive method [3].
[2]Kier&Hall index (order 3)
[3]Number of halogenide groups
[4]Highest total interaction (AM1) for O - H bonds eV
[5]Average atom weight g/mol
[6]HACA-1/TMSA (Zefirov) C
[7]Max net atomic charge (Zefirov) for H atoms C
[8]LUMO+1 energy (AM1) eV

4.4.Descriptor selection:

Initial consisted of ~1000 descriptors per structure. All descriptors were correlated with the property and then the first 20 descriptors were selected which indicated the largest correlation coeficient. Further, 11 networks with different structures and iputs form the preselected 20 descriptors were tested in order to find the best ANN with lowest RMS (root-mean-squared error) for training, selection and test sets. The best ANN model had 8 inputs. Then 107 epochs were used to train the final network with architecture depicted in 4.2. Optimization of the weights was performed with Levenberg-Marquardt algorithm using linear(inputs) and hyperbolic(hidden) adn logistic(output) activation functions.

4.5.Algorithm and descriptor generation:

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

4.6.Software name and version for descriptor generation:

QSARModel
http://molcode.com/software

4.7.Chemicals/Descriptors ratio:

9

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 descriptor value range (between min and max values): The model is suitable for compounds (including ethers, esters, amides, amines,halides, aromatic, aliphatic functional groups) that have the descriptors in the following range augmented with the confidence in 5.2:

Desc	1	2	3	4	5	6	7	8
min	-2.65	0.00	0.00	-14.56	4.75	0.00	0.00	-0.87
max	17.58	15.77	6.00	0.00	24.80	0.07	0.10	3.64

5.2.Method used to assess the applicability domain:

1) Functional Groups - presence of functional groups in structures such as in 5.1

2) Quantitative approach - range of descriptor values in training set with augmented with ±30% confidence

Descriptor values must fall between maximal and minimal descriptor values (see5.1) of training set augmented by ±30%.

The reason for the 30% confidence is that the ANN model is sensitive to the inputs and in addition our practice showed that in this confidence limits the predictions are more correct.

5.3.Software name and version for applicability domain assessment:

QSARModel 3.3.8

http://molcode.com/software

5.4.Limits of applicability:

See 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: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:

data points: 73

6.6.Pre-processing of data before modelling:

The data for the ANN inputs were normalized using their min and max values

6.7.Statistics for goodness-of-fit:

	Training Log(1/NOEAL)	Selection Log(1/NOEAL)	Test Log(1/NOEAL)
Data Mean	0.40	0.56	0.31
Data S.D.	0.94	0.94	1.29
Error Mean	0.00	-0.48	0.34
Error S.D.	0.43	0.85	0.89
S.D. Ratio	0.46	0.91	0.69
Correlation	0.89	0.50	0.82

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

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)= 0.11
, RMS(Selection)= 0.25
, RMS(Test) = 0.24

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 randonly selected validation sets – selection (15) and test(15)
7.6.Experimental design of test set:
Randomly selected 15 and 15 data points
7.7.Predictivity - Statistics obtained by external validation:
see 6.7 and 6.12
7.8.Predictivity - Assessment of the external validation set:
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 over fit it) and the test set (used to test the external prediction of the net after training) are significant according to 6.7 and 6.12

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:
Subchronic oral toxicity is quite complex property and depends on many factors. In adddition the ANN mathematical structure does not allow easy interpretation of the mehcanistic picture. However, the descriptors that apeared as inputs are related to 1) lipophylicity (LogP) and reactivity (LUMO+1) of the compounds 2) halogenic and O

compounds (Highest total interaction (AM1) for O - H bonds, Number of halogenide groups)

which in turn can be connected to 1), charge distributions (HACA-1/TMSA (Zefirov), Max net atomic charge (Zefirov) for H atoms).

8.2.A priori or a posteriori mechanistic interpretation:

A posteriori

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

Supporting information for :Training set(s)

Selection set(s)

Test set(s)

9.2.Bibliography:

[1]Repeated dose 90-day oral toxicity study in rodents, OECD TG 408, 1998

[2]http://ecb.jrc.ec.europa.eu/esis/index.php?PGM=dat http://ecb.jrc.ec.europa.eu/esis/index.php?PGM=dat

[3]Renxiao Wang, Ying Gao, Luhua Lai. Calculating partition coefficient by atom-additive method, Perspectives in Drug Discovery and Design, 19: 47–66, 2000.

9.3.Supporting information:

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

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