QMRF	*QMRF identifier (ECB Inventory):* To be entered by JRC
	QMRF Title: Nonlinear ANN QSAR model for Chronic toxicity LOAEL (rat)
	Printing Date: Oct 19, 2010

1.QSAR identifier

1.1.QSAR identifier (title):

Nonlinear ANN QSAR model for Chronic toxicity LOAEL (rat)

1.2.Other related models:

-

1.3.Software coding the model:

QSARModel 3.3.8; Statistica 7, StatSoft Ltd. Turu 2, Tartu, 51014, Estonia http://www.molcode.com

2.General information

2.1.Date of QMRF:

10.10.2010

2.2.QMRF author(s) and contact details:

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

12.04.2010The methodology and software (QSARModel) used to create the present model were

applied also to obtain the results published in these papers.

1)Katritzky, A. R.; Dobchev, D. A.; Fara, D. C.; Hur, E.; Tämm, K.; Kurunczi, L.; Karelson, M.; Varnek, A.; Solov'ev, V. P. (2006). Skin Permeation Rate as a Function of Chemical Structure . Journal of Medicinal Chemistry, 49(11), 3305 - 3314.

2)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.

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

[1]Katritzky, A. R.; Dobchev, D. A.; Fara, D. C.; Hur, E.; Tämm, K.; Kurunczi, L.; Karelson, M.; Varnek, A.; Solov'ev, V. P. (2006). Skin Permeation Rate as a Function of Chemical Structure . Journal of Medicinal Chemistry, 49(11), 3305 - 3314.

[2]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.
[3]Statistica 7 www.statsoft.com

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:

rat

3.2.Endpoint:

4.Human health effects LOAEL 4.14.Repeated dose toxicity

3.3.Comment on endpoint:

Toxicity to other above-ground organisms. Rat chronic lowest observed adverse effect levels (LOAEL). Particular interest in the estimation of chronic low-est observed adverse effect level (LOAEL) has been raised recently due to its environmental implications. The presence of toxic substances in high trophic levels can affect humans. LOAEL was defined by the IUPAC as the lowest concentration or amount of a substance, found by experiment or observation, which causes an adverse alteration of morphology, functional ca-pacity, growth, development, or life span of a target organism distinguishable from normal (control) organisms of the same species and strain under defined conditions of exposure.

3.4.Endpoint units:

μmol/kg

3.5.Dependent variable:

Log (LOEAL)

3.6.Experimental protocol:

see 3.7 and [1].

3.7.Endpoint data quality and variability:

EPA data base of LOEAL was used to collect 99 compounds.

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 regression neural network predictor with structure 6-6-5-1

4.3.Descriptors in the model:

[1]Lowest total interaction (AM1) for C - Cl bonds
[2]Max bond order (AM1) for Cl atoms
[3]Max valency (AM1) for Cl atoms
[4]Michalic MTI' of Schultz triple weighted D matrix

[5]Lowest resonance energy (AM1) for C - Cl bonds
[6]Highest e-e repulsion (AM1) for O - P bonds

4.4.Descriptor selection:

Initial pool of ~997 descriptors. Stepwise descriptor selection based on a set of statistical selection rules as F statistic and p. The first highest F (low p) descriptors (6) were selected from the whole (~997) descriptors. These 6 descriptors were used as inputs to the network. 39 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 150 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 structure optimized by AM1 semiempirical quantum mechanical model.

4.6.Software name and version for descriptor generation:

QSARModel 3.3.8; Statistica 7, StatSoft Ltd.
Turu 2, Tartu, 51014, Estonia
http://www.molcode.com

4.7.Chemicals/Descriptors ratio:

11.5

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 grups as phenols, aldehydes, nitro, amino, Cl containing, P-conataining compounds and others

b)The model is suitable for compounds that
have descriptors values in the followin range;

Desc123456
min-13.398300820-13.76250
max01.0317241.15591523847680132.1219

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 (see5.1) of training set ±30%.

5.3.Software name and version for applicability domain assessment:

QSARModel 3.3.8; Statistica 7, StatSoft Ltd.
Turu 2, Tartu, 51014, Estonia
http://www.molcode.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: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: 69

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:

	Training Log(LOEAL)	Selection Log(LOEAL)	Test Log(LOEAL)
Data Mean	1.9854	2.3953	2.0420
Data S.D.	1.1606	1.2345	1.3391
Error Mean	-0.0164	-0.3579	-0.0284
Error S.D.	0.8055	0.5775	0.8303
Abs E. Mean	0.6318	0.5474	0.7137
S.D. Ratio	0.6940	0.4677	0.6200
Correlation	0.7201	0.8987	0.8028

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)= 0.1546,, RMS(Selection)= 0.130, RMS(Test) = 0.159,

In this ANN were used 2 sets randomly chosen (15) 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 (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:

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 (S.D.Ration), see 6.7 and 6.12

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

Since the ANN is a more complex predictor than a linear model, it is difficult to analyze the relation between the property and the descriptors. The important descriptors are related to the

Cl atoms. It seems that the descriptor Lowest total interaction (AM1) for C - Cl bonds leads to higher values of Log(LOEAL). In contrary the Highest e-e repulsion (AM1) for O - P bonds

increasing its value leads to lower property values (more toxic).

8.2.A priori or a posteriori mechanistic interpretation:

8.3.Other information about the mechanistic interpretation:

Structural characteristics of the compounds are also important for LOEAL as sugested by [1] and our descriptors Michalic MTI' of Schultz triple weighted D matrix and Max bond order (AM1) for Cl atoms,Max valency (AM1) for Cl atoms.

9.Miscellaneous information

9.1.Comments:

Supporting information for :Training set(s)

Selection set(s)

Test set(s)

6-6-5-1.snn file -includes the ANN model, in order to be used the user must have statistica 7 or higher with ANN modules to make predictions.

9.2.Bibliography:

9.3.Supporting information:

Training set(s)

LOAEL_trainingset.sdf	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf274_LOAEL_trainingset.sdf

Test set(s)

LOAEL_testset.sdf	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf274_LOAEL_testset.sdf

Supporting information

6-6-5-1.snn	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf274_6-6-5-1.snn
LOAEL_selectionset.sdf	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf274_LOAEL_selectionset.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