	QMRF identifier (JRC Inventory): **Q13-414-0057**
	QMRF Title: **Nonlinear QSAR: artificial neural network for classification of repeated dose toxicity**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

Nonlinear QSAR: artificial neural network for classification of repeated dose toxicity

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:

12.05.2010

2.2.QMRF author(s) and contact details:

Molcode model development team Molcode Ltd Turu 2, Tartu, 51014, Estonia
models@molcode.com 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 A R, Dobchev DA, Fara DC, Hur E, Tämm K, Kuruncz 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 and test sets are available. Model algorithm is available (snn file).

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

None to date.

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:

A combined repeated dose toxicity study with the reproduction/developmental toxicity screening test was performed using the OECD TG 422.

The method comprises the basic repeated dose toxicity study that may be used for chemicals on which a 90-day study is not warranted or as a preliminary study to a long-term study. It further comprises a reproduction/developmental toxicity screening test and therefore can also be used to provide initial information on possible effects on male and female reproductive performance such as gonadal function, mating behaviour, conception, development of the conceptus and parturition.

This test offers only limited means of detecting postnatal manifestations of prenatal exposure, or effects that may be induced during postnatal exposure. Due (amongst other reasons) to the selectivity of the end points, and the short duration of the study, this method will not provide evidence for definite claims of no reproduction/developmental effects [see 9.2, ref 1].

3.4.Endpoint units:

The variable is dimensionless: it is a positive or negative class (indicated as +1 or -1) with respect to the toxic or nontoxic effects

3.5.Dependent variable:

Repeated dose toxicity: presence or absence of toxicity (indicated as +1 and -1). These binary values were used to develop a classification ANN model

3.6.Experimental protocol:

The experimental guideline assumes oral administration of the test substance and is designed for use with the rat. In the test, the dosing period is longer than in a conventional 28-day repeated dose study. However, it uses fewer animals of each sex per group when compared with the situation where a conventional 28-day repeated dose study is conducted in addition to a reproduction/developmental toxicity screening test.

The test substance is administered in graduated doses to several groups of males and females. Males should be dosed for a minimum of four weeks, up to and including the day before scheduled kill (this includes a minimum of two weeks prior to mating, during the mating period and, approximately, two weeks post mating). In view of the limited pre-mating dosing period in males, fertility may not be a particularly sensitive indicator of testicular toxicity. Therefore, a detailed histological examination of the testes is essential. The combination of a pre-mating

dosing period of two weeks and subsequent mating/fertility observations with an overall dosing period of at least four weeks, followed by detailed histopathology of the male gonads, is considered sufficient to enable detection of the majority of effects on male fertility and spermatogenesis.

Females should be dosed throughout the study. This includes two weeks prior to mating (with the objective of covering at least two complete estrous cycles), the variable time to conception, the duration of pregnancy and at least four days after delivery, up to and including the day before scheduled kill. Duration of study, following acclimatization, is dependent on the female performance and is approximately 54 days [at least 14 days pre-mating, (up to) 14 days mating, 22 days gestation, 4 days lactation].

During the period of administration, the animals are observed closely each day 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 killed and necropsied [see 9.2, ref 1].

A database for computational analysis of developmental toxicity was created by combining subsets of information from the Teratogen Information System (TERIS) and the Food and Drug Administration (FDA) guidelines. This database was constructed by researchers in the Department of Environmental and Occupational Health at the University of Pittsburgh and is described elsewhere in detail. The database contains information about 292 chemicals and their disposition as to development toxicity. Forty-one percent, 116 chemicals were deemed to be developmentally active substances, while the remaining 176 chemicals showed no evidence of developmental toxicity [see 9.2, refs 3-5].

3.7.Endpoint data quality and variability:

The data for the model were taken mainly from refs 1 and 2 using relatively similar protocols.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

Neural Network

4.2.Explicit algorithm:

Neural Network

Nonlinear QSAR: Backpropagation Neural Network (Multilayer Perceptron) classification

The algorithm is based on neural network predictor with structure 9-9-8-1.

The precise explicit algorithm of the network is given in supplementary file ANN.snn.

4.3.Descriptors in the model:

[1]Highest coulombic interaction (AM1)

[2]Number of ether groups

[3]Max net atomic charge (AM1) for N atoms

[4]Positively Charged Part of Charged Surface Area (Zefirov)

[5]Negatively Charged Part of Partial Charged Surface Area (AM1)

[6]Topographic electronic index (Zefirov) (all atoms) all bonds

[7]Square root of Charged (Zefirov) Surface Area of O atoms

[8]Number of single bonds

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

4.4.Descriptor selection:

Initial pool of ~1000 descriptors. Stepwise descriptor selection based on a set of statistical selection rules as F statistic and p. The first highest F (low p) descriptors (9) were selected from the whole (~1000) descriptors. These 9 descriptors were used as inputs to the network. 23 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 998 epochs were used to train the final network with architecture depicted in 4.2. Optimization of the weights was performed with standard Backpropagation algorithm (learning Rate =0.01 and momentum=0.3) using linear and hyperbolic activation functions.

4.5.Algorithm and descriptor generation:

All descriptors were generated using QSARModel on structure optimized by the AM1 semiempirical quantum mechanical model. The QSARModel program is able to calculate number of descriptors that can be classified as: (i) constitutional, (ii) geometrical, (iii) topological, (iv) charge-related, (v) quantum chemical, and (vi) thermodynamic. The total number of descriptors for each property ranged between 600 and over 1000 per compound. Optimization of the compound consist of 1) applying molecular mechanics and 2) semiempirical AM 1 method for the final geometries. The stopping criterion was the gradient norm for both 1) and 2) 0.05kcal/mol.A.

4.6.Software name and version for descriptor generation:

QSARModel

http://www.molcode.com

4.7.Chemicals/Descriptors ratio:

118 chemicals / 9 descriptors = 20.1 chemicals per descriptor

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 ranges augmented with the confidence in 5.2:

Desc ID (See 4.3): 1 2 3 4 5 6 7 8 9

Min 4.499 0.000 -0.976 3.499 0.018 0.126 0.000 2.000 -17.046

Max 13.802 5.000 0.610 28.874 0.365 5.372 6.958 88.000 0.225

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

http://www.molcode.com

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: 181

6.6.Pre-processing of data before modelling:

Standardization and normalization by taking into account the mean and standard deviation (of the input variables)

6.7.Statistics for goodness-of-fit:

Training negatives Training positives Selection negatives Selection positives Test negatives Test positives

Total 107.000 74.000 28.000 12.000 18.000 22.000

Correct 101.000 68.000 20.000 9.000 14.000 14.000

Wrong 6.000 6.000 8.000 3.000 4.000 8.000

Correct (%) 94.393 91.892 71.429 75.000 77.778 63.636

Wrong (%) 5.607 8.108 28.571 25.000 22.222 36.364

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)= 0.434011; RMS (Selection)= 1.433205; RMS (Test) = 2.521522.

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: No

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 randomly selected validation sets – selection (40) and test (40)

7.6.Experimental design of test set:

Randomly selected data points: 40 and 40

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 predictivity 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 given in the classification matrix (see 6.7).

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

The mechanistic picture of the model is complicated due to the nature of the ANN (Artificial Neural Network). However, it can be concluded that the descriptors Highest coulombic interaction (AM1), Number of ether groups are very important for the property. It is evident from the descriptors that electrostatic interactions and structural characteristics (including the specific charged surfaces) of the compounds play an important role for the Index. The descriptor Number of single bonds shows roughly that compounds with large values of this descriptor are associated with non-toxic compounds. However, this

conjecture has to be considered simultaneously with the remaining descriptors. Also, lower values of the descriptor Highest coulombic interaction (AM1) are associated with non-toxic compounds.

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:

9.Miscellaneous information

9.1.Comments:

The 9-9-8-1.snn file (binary) includes the ANN model. In order to be used, the user must have Statistica 7 or higher with ANN modules.

9.2.Bibliography:

[1]OECD (1996). Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test (1996), OECD TG 422.

[2]Arena VC, Sussman NB, Mazumdar S, Yu S & Macina OT (2004) The Utility of Structure-Activity Relationship (SAR) Models for Prediction and Covariate Selection in Developmental Toxicity: Comparative Analysis of Logistic Regression and Decision Tree Models. SAR and QSAR in Environmental Research 15, 1-18. http://dx.doi.org/10.1080/1062936032000169633

[3]Ghanooni M, Mattiston DR, Zhang YP, Macina OT, Rosenkranz HS & Klopman G (1997). Structural determinants associated with risk of human developmental toxicity. American Journal of Obstetrics & Gynecology 176, 799–806.

[4]Briggs GG, Freeman RK & Yaffe SJ (1990). Drugs in Pregnancy and Lactation, 3rd Edn. Williams and Wilkens, Baltimore, MD, USA, p. 537.

[5]Shepard TH (1992). Catalog of Teratologic Agents, 5th Edn. Johns Hopkins University Press, Baltimore, MD, USA, p. 534.

9.3.Supporting information:

ANN train_181.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-414-0057/attachment/A740
ANN test_40.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-414-0057/attachment/A741

Test set(s)

10.Summary (JRC QSAR Model Database)

Supporting information

10.1.QMRF number:

Q13-414-0057

10.2.Publication date:

2013-07-01

10.3.Keywords:

Molcode;neural network;repeated dose;reproduction;developmental toxicity;rat;

10.4.Comments:

former Q17-10-31-264