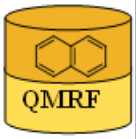	*QMRF identifier (ECB Inventory):* *QMRF Title:* *QSAR for female rat carcinogenicity (TD50) of nitro compounds* *Printing Date:*Feb 16, 2010	

1.QSAR identifier

1.1.QSAR identifier (title):

QSAR for female rat carcinogenicity (TD50) of nitro compounds

1.2.Other related models:

1.3.Software coding the model:

QSARModel 3.5.0 Molcode Ltd., Turu 2, Tartu, 51014, Estonia http://www.molcode.com

2.General information

2.1.Date of QMRF:

21.08.2009

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]Gunnar Karelson Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[4]Dimitar Dobchev Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[5]Dana Martin 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]Deniss Savchenko Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[8]Jaak Jänes Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[9]Eneli Härk Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[10]Andres Kreegipuu Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[11]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:

21.08.2009

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:

Model is proprietary, but the training 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:

Female rat

3.2.Endpoint:

4.Human health effects 4.12.Carcinogenicity

3.3.Comment on endpoint:

3.4.Endpoint units:

mg/kg body wt/day

3.5.Dependent variable:

log(TD50) - TD50 is the median toxic dose, the dose that produces a toxic effect in 50% of the population.

3.6.Experimental protocol:

Carcinogenicity was determined using the EU Test Guideline B.32. This method describes the administration of test substance normally seven days per week, by an appropriate route, to several groups of experimental animals, one dose per group, for a major portion of their lifespan, and the daily observation of experimental animals for detection of signs of toxicity, particularly the development of tumours. Chemical carcinogens have been categorized as either genotoxic or non-genotoxic. The former are DNA reactive and the latter act by a variety of other mechanisms. The toxicological property of interest was the carcinogenic potency, expressed as TD50 value. The TD50 value for a given target site (s) in the absence of tumors in control animals, was taken to be the chronic dose (in mg/kg body wt/day) which induced tumors in half of the test animals at the end of a standard lifespan for the species (Gold et al., 1999). The TD50 value used for each compound was selected by taking into account the lowest carcinogenic potency value reported for each chemical in all the positive reports for that chemical.

3.7.Endpoint data quality and variability:

A data set of aromatic and aliphatic nitro-compounds was collected from the Carcinogenic Potency Database (CPDB) established by Gold and Zeiger (1997). The CPDB is a single standardized resource of information from many chronic, long-term bioassays. It contains a large diversity of chemical structures, and includes tumour data reproduced from all of the NCI/NTP rodent bioassay Technical Reports, as well as additional data extracted from over 1200 literature sources (CPDB-Lit) subjected to extensive review (Gold et al.,

1999). The data from the female rat of the CPDB-Lit reports were used. According to Gottmann et al. (2001), the tests on rats seem to be far more reproducible than the mouse and experiments on female rats had a better concordance than those on male rats. The experimental data are most probably not from one laboratory but the data fare considered as consistent (Morales et al, 2006).Statistics:max value: 3.770min value: -0.590standard deviation: 0.820skewness: 1.014

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Multilinear regression QSAR

Log(TD50) = - 0.69 - 9.44E-002 * 1 X GAMMA polarizability(DIP) (AM1) + 0.13 * Lowest coulombic interaction (AM1) - 48.44 * FNSA 3 Fractional PNSA (PNSA-3/TMSA) (Zefirov) + 0.29 * WFOSA Atomic charge (Zefirov) weighted FOSA

4.3.Descriptors in the model:

[1]1 X GAMMA polarizability(DIP) (AM1)
[2]Lowest coulombic interaction (AM1)
[3]FNSA 3 Fractional PNSA (PNSA-3/TMSA) (Zefirov)
[4]WFOSA Atomic charge (Zefirov) weighted FOSA

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, variance and t-test value over threshold, intercorrelation with another descriptor not over threshold;2 parameter equations: intercorrelation coefficient bellow threshold, significant correlation with endpoint in terms of correlation coefficient and t-test). 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 AM1 calculation. Model developed by using multilinear regression.

4.6.Software name and version for descriptor generation:

QSARModel 3.3.5
QSAR/QSPR package that will compute chemically meaningful descriptors and includes statistical tools for regression modeling
Molcode Ltd, Turu 2, Tartu, 51014, Estonia
http://www.molcode.com

4.7.Descriptors/Chemicals ratio:

(42 chemicals / 4 descriptors) = 10.5

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

By chemical identity: nitrocompounds By descriptor value range: The model is suitable for compounds that have the descriptors in the following ranges: 1 X GAMMA polarizability(DIP) (AM1) (min: 0.714 , max: 12.844); Lowest coulombic interaction (AM1) (min: -18.941, max: 5.018);FNSA 3 Fractional PNSA (PNSA-3/TMSA) (Zefirov) (min: -6.400•10-2, max: -1.190•10-2);WFOSA Atomic charge (Zefirov) weighted FOSA (min:

0.000, max: 8.700).

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 3.3.5

QSAR/QSPR package that will compute chemically meaningful descriptors and includes statistical tools for regression modeling

Molcode Ltd, Turu 2, Tartu, 51014, Estonia

http://www.molcode.com

5.4.Limits of applicability:

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:

42 data points: 4 negative values; 38 positive values

6.6.Pre-processing of data before modelling:

6.7.Statistics for goodness-of-fit:

The model has been trained with 46 chemicals. R^2 = 0.730 (Correlation coefficient); s = 0.200 (Standard error of the estimate); F = 25.180 (Fisher function).

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

R^2cv = 0.660 LOO

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

R^2cv = 0.757 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) = 0.740Average R^2 (prediction) = 0.686

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:

4 data points: 0, negative values; 4 positive values

7.6.Experimental design of test set:

The full experimental dataset was sorted according to increasing values of logTD50 and compounds 5, 15, 35, 45 in the list have been assigned to the test set.

7.7.Predictivity - Statistics obtained by external validation:

$R^2cv = 0.942$

7.8.Predictivity - Assessment of the external validation set:

The descriptors for the test set are in the limits of applicability.

7.9.Comments on the external validation of the model:

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

The model reflects structural features which are characteristic to carcinogenicity.The toxic effect of the compounds increase with increased polarizability and increased negative patyial charges values in the molecule. The increased polarizability of the molecule and the increased values of the charges enable stronger interaction of the molecules with the DNA, and as a result, an increased carcinogenicity.

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 carcinogenic effect increases with increased values of the descriptors 1X GAMMA polarizability (DIP) (AM1) and FNSA 3 Fractional PNSA (PNSA-3/TMSA)(Zefirov). The descriptor Lowest coulombic interaction (AM1) and the descriptor WFOSA Atomic charge (Zefirov) weighted FOSA have higher values for less toxic compounds. The latter descriptoris a measure of the polarity of the hydrophobic part of the solvent accessible surface area, as it is weighted by the partial charges of the hydrogen and sp3 hybridized carbon

atoms.

9.Miscellaneous information

9.1.Comments:

The data from the female rat were used. According to Gottmann et al (2001), the tests on rats seem to be far more reproducible than the mouse and experiments on female rats had a better concordance than those on male rats.

9.2.Bibliography:

[1]Gold LS & Zeiger E (1997). Handbook of Carcinogenic Potency and Genotoxicity Databases. CRC Press, Boca Roca, FL.

[2]Gold database http://potency.berkeley.edu/cpdb.html

[3]old LS, Manley NB, Slone TH & Rohrbach L (1999). Supplement to the Carcinogenic Potency Database (CPDB): results of animal bioassays published in the general literature in 1993-1994 and by the National Toxicology Program in 1995-1996. Environmental Health Perspectives 107 (Suppl. 4), 527-602.

[4]Gottmann E, Kramer S, Pfahringer B & Helma C (2001). Data quality in predictive toxicology: reproducibility of rodent carcinogenicity experiments. Environmental Health Perspectives 109, 1-11.

[5]Morales AH, Cabrera MA, Combes RD & Gonzalez MP (2006). Quantitative structure activity relationship for the computational prediction of nitrocompounds carcinogenicity. Toxicology 220, 51-62.

9.3.Supporting information:

Training set(s)

training set	http://qsardb.jrc.ec.europa.eu:80/qmrf/download_attachment.jsp?name=qmrf153_Molcode rat carc_training_42.sdf

Test set(s)

test set	http://qsardb.jrc.ec.europa.eu:80/qmrf/download_attachment.jsp?name=qmrf153_Molcode rat carc_test_4.sdf

10.Summary (ECB Inventory)

10.1.QMRF number:

10.2.Publication date:

10.3.Keywords:

Molcode, female rat carcinogenicity, TD50

10.4.Comments: