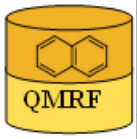	*QMRF identifier (ECB Inventory):* *QMRF Title:* *QSAR for mutagenicity (Salmonella typhimurium TA98 strain)* *Printing Date:*Dec 9, 2009	

1.QSAR identifier

1.1.QSAR identifier (title):

QSAR for mutagenicity (Salmonella typhimurium TA98 strain)

1.2.Other related models:

Same endpoint and dataset as present model:

Cash GG, Anderson B, Mayo K, Bogaczyk S & Tunkel J (2005). Mutation Research/Genetic Toxicology and Environmental Mutagenesis 585, 170-183.

1.3.Software coding the model:

QSARModel 3.3.8 Molcode Ltd., Turu 2, Tartu, 51014, Estonia http://www.molcode.com

2.General information

2.1.Date of QMRF:

10.05.2009

2.2.QMRF author(s) and contact details:

[1]Dimitar Dobchev 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]Indrek Tulp 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:

24.10.2008

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:

Salmonella typhimurium (TA98 strain)

3.2.Endpoint:

4.Human health effects 4.10.Mutagenicity

3.3.Comment on endpoint:

Mutagenicity, measured on Salmonella typhimurium TA98 with the addition of an exogenous metabolic activation system (S9).

3.4.Endpoint units:

log(reversions/nmol), log of the number of reversions per nanomole of substance

3.5.Dependent variable:

log(reversions/nmol), logR

3.6.Experimental protocol:

Mutagenicity: reverse mutation test using bacteria was carried out according to the OECD 471 (EU B.13/14) test guideline. The bacterial reversed mutation assay (Ames Test) is used to detect point mutations, which involve substitution, addition or deletion of one or a few DNA base pairs. The test uses bacterial culture of amino acid-dependent strain of Salmonella typhimurium (TA98 or TA100), which is exposed to the test substance in the presence and in the absence of an exogenous metabolic activation system [cofactor-supplemented post-mitochondrial fraction (S9) prepared from the liver]. Mutagenic compounds cause an increase in the number of revertant colonies relative to the background level. The mutagenic potencies of tested substances are expressed as log(revertants/nmol). For non-mutagenic compounds, the potency is arbitrarily coded as -100.

The test substances were with a high purity grade: 99.9% or higher. Solid test substances were dissolved or suspended in appropriate solvents and diluted if appropriate prior to treatment of the bacteria. Liquid test substances were added directly to the test systems and/or diluted prior to treatment.

References (related to experimental protocol):

1) Debnath AK, Debnath G,Shusterman AJ & Hansch C (1992). A QSAR investigation of the role of hydrophobicity in regulating mutagenicity in the Ames test, Part 1: Mutagenicity of aromatic and heteroaromatic amines in Salmonella typhimurium TA98 and TA100. Environmental and Molecular Mutagenesis 19, 37–52.

2) Glende C, Schmitt H, Erdinger L, Engelhardt G & Boche G (2001).Transformation of mutagenic aromatic amines into non-mutagenic species by alkyl substituents. I. Alkylation ortho to the amino function. Mutation Research 498, 19–37.

3) Glende C, Klein M, Schmitt H, Erdinger L & Boche G (2002). Transformation of mutagenic aromatic amines into non-mutagenic species by alkyl substituents. II. Alkylation far away from the amino function. Mutation Research 515, 15-38.

3.7.Endpoint data quality and variability:

Experimental data from different sources have been used.

Statistics:
max value: 3.97
min value: -3.32
standard deviation: 1.78
skewness: 0.44

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

multilinear regression QSAR

multilinear regression QSAR based on three descriptors - quantum chemical (3D), topological (2D) and constitutional (1D)

log(reversions/nmol) = -8.03 + 1.45* Average Information content (order 2) + 0.475*Number of aromatic bonds - 0.0837* Shadow plane YZ (AM1)

4.3.Descriptors in the model:

[1]Average Information content (order 2) Topological descriptor, based on the vertex equivalence of subgraphs of the molecular graph

[2]Number of aromatic bonds Count of benzene ring–type bonds

[3]Shadow plane YZ (AM1) squared angstroms Surface area of the projection of the molecule onto the Y-Z plane, obtained from semi-empirical AM1 optimization

4.4.Descriptor selection:

Initial pool of ~1000 descriptors. Stepwise descriptor selection based on a set of statistical selection rules (one-parameter equations: Fisher criterion and R2 over threshold, variance and t-test value over threshold, intercorrelation with another descriptor not over threshold),

(two-parameter equations: intercorrelation coefficient below 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 MMFFs (Merck Molecular Force Field) (vacuum) conformational search and AM1

calculation. Model developed by using multilinear regression.

4.6.Software name and version for descriptor generation:

QSARModel 3.3.8, Molcode Ltd, Turu 2, Tartu, 51014, Estonia
http://www.molcode.com

4.7.Descriptors/Chemicals ratio:

31.0 (93 chemicals/3 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) by chemical identity: aromatic and heteroaromatic amines

b) by descriptor value range: The model is suitable for compounds that have the descriptors

in the following range:

Average Information content (order 2): 1.84 - 5.36

Number of aromatic bonds: 4.2 - 27.3

Shadow plane YZ (AM1): 12.4 - 65.0

5.2.Method used to assess the applicability domain:

Presence of functional groups in structures (amino groups and aromatic moieties)

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

5.4.Limits of applicability:

See 5.1

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:

93 data points

53 negative values

40 positive values

6.6.Pre-processing of data before modelling:

n/a

6.7.Statistics for goodness-of-fit:

R2 = 0.722 (Correlation coefficient);

s = 0.906 (Standard error of the estimate);

F = 77.1 (Fisher function);

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

R2cv = 0.692 (Cross-validated correlation coefficient);

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

R2cv= 0.683

6.10.Robustness - Statistics obtained by Y-scrambling:

n/a

6.11.Robustness - Statistics obtained by bootstrap:

n/a

6.12.Robustness - Statistics obtained by other methods:

ABC analysis (2:1 training : prediction) on sorted (in increased order of endpoint value) 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 R2 (fitting) = 0.731

average R2 (prediction) = 0.698

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:Yes

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:

31 data points: 17 negative values, 14 positive values

7.6.Experimental design of test set:

From sorted data of the initial set (by endpoint value) each 4th was subjected to test set, starting from 3rd data point, the rest composing the training set.

7.7.Predictivity - Statistics obtained by external validation:

R2 = 0.647 (Correlation coefficient)

7.8.Predictivity - Assessment of the external validation set:

Descriptor value range (all in range of applicability domain):

Average Information content (order 2): 2.61 - 3.72

Number of aromatic bonds: 6 - 19

Shadow plane YZ (AM1): 18.0 - 44.0

7.9.Comments on the external validation of the model:

Correlation coefficient value of test set close to that of the training set.

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

"Average Information content (order 2)" shows how information rich is the molecule. It shows the branching of the molecule and simultaneously takes into account also different atom types, hence different electonegativities. "Number of aromatic bonds" counts all aromatic bonds in the molecule. "Shadow plane YZ (AM1)" reflects to the thickness of the molecule.

"Average Information content (order 2)" and "Number of aromatic bonds" are favorable descriptors, higher descriptor value leads to higher property value. Conversely "Shadow plane YZ (AM1)" is unfavorable to the property.

8.2.A priori or a posteriori mechanistic interpretation:

a posteriori mechanistic interpretation, consistent with published scientific interpretations of experiments (no logP/hydrophobicity dependence has been observed, same as in literature, importance of configuration and conformation of the molecule and other structural features, as described by the information content and shadow area descriptors. Importance of the amount of aromatic moieties, as described by the number of aromatic bonds descriptor. However, no HOMO/LUMO dependence observes, as some literature sources suggest.

8.3.Other information about the mechanistic interpretation:

Interpretation in general agreement with literature: Benigni R. Structure-activity relationship studies of chemical mutagens and carcinogens: Mechanistic investigations and prediction approaches. Chem Rev 105:1767–1800 (2005)

9.Miscellaneous information

9.1.Comments:

Experimental data is taken from Cash et al (2005).

9.2.Bibliography:

Predicting genotoxicity of aromatic and heteroaromatic amines using electrotopological state indices. Cash GC, Anderson B, Mayo K, Bogaczyk S & Tunkel J (2005). Mutation Research/Genetic Toxicology and Environmental Mutagenesis 585, 1-2, 170-183 http://dx.doi.org/10.1016/j.mrgentox.2005.05.001

9.3.Supporting information:

Training set(s)

Mutagenicity training	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf83_Mutagenicity training.sdf

Test set(s)

mutagenicity testset	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf83_mutagenicity-2_testset.sdf

Supporting information

Karelson Arkivoc 2008	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf83_Karelson Arkivoc 2008.pdf
Karelson Arkivoc 2009	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf83_Karelson Arkivoc 2009.pdf

10.Summary (ECB Inventory)

10.1.QMRF number:

10.2.Publication date:

10.3.Keywords:

mutagenicity, Salmonella typhimurium, TA98, Molcode

10.4.Comments: