1.QSAR identifier

1.1.QSAR identifier (title):

QSAR model for Acute toxicity- Acute toxic Class Method (LD50) for benzene derivatives

1.2.Other related models:

-

1.3.Software coding the model:

QSARModel 4.0.3 Molcode Ltd., Turu 2, Tartu, 51014, Estonia
http://www.molcode.com

2.General information

2.1.Date of QMRF:

18.11.2009

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:
15.11.2009

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:
All information in full detail is 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:
Rat

3.2.Endpoint:
4.Human health effects 4.2.Acute oral toxicity

3.3.Comment on endpoint:
B.1 tris, in REACH classification

3.4.Endpoint units:
M

3.5.Dependent variable:
log(1/LD50)

3.6.Experimental protocol:
Acute oral toxicity – acute toxic class method was determined using the OECD Test Guideline 423 (EU B.1 tris). Acute oral toxicity testing allows obtaining information on the biologic/toxic activity of a chemical. Currently, the basis for toxicologic classification of chemicals is the median lethal dose (LD50, mg/kg b.w.), which is defined as the statistically derived dose required to kill half the members of a tested population. Animals are observed individually after dosing at least once during the first 30 minutes, periodically during the first 24 hours, with special attention given during the first 4 hours and daily thereafter, for a total of 14 days. The LD50 values of tested substances were translated to logarithmic scale (logLD50) to reduce the range of the data.

The dataset has been collected from a series of compounds that include different aromatic compounds with halogen-containing groups, nitrogen-containing groups accompanied by hydroxyl and amino-groups. The dataset is limited only to benzene derivatives, and number of studied compounds is limited by available acute toxicity data for rats.

The activities of the studied compounds are expressed in terms of oral LD50 dose for rats. All original LD50 toxicity data (mg/kg) has been converted to molar Log(1/LD50) response variables.

3.7.Endpoint data quality and variability:

The experimental data is from the TOXNET database (http://toxnet.nlm.nih.gov/) - not from one lab. The same dataset has been used in publication 1 (section 9.2)

Statistics:
max value: 3.994
min value: 0.989
standard deviation: 0.599
skewness: 0.73

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

2D and 3D regression-based QSAR

4.2.Explicit algorithm:

multilinear regression QSAR

multilinear regression QSAR derived with BMLR (Best Multiple Linear Regression) method

$-\log(LD50) = 3.78$

+0.521*Difference (Pos - Neg) in Charged Partial Surface Area (AM1)

-1.72*Max bond order (AM1)

+26.4*Min nucleophilic reactivity index (AM1) for N atoms

+5.93*Relative number of O atoms

+16.7*Min net atomic charge (Zefirov) for H atoms

4.3.Descriptors in the model:

[1]Difference (Pos - Neg) in Charged Partial Surface Area (AM1) Å2 Difference between positively and negatively charged partial surface areas

[2]Max bond order (AM1) unitless shows highest bond order in the molecule

[3]Min nucleophilic reactivity index (AM1) for N atoms 1/eV lowest Fukui atomic nucleophilic reactivity index for N atom

[4]Relative number of O atoms unitless count of O atoms in respect of all atoms

[5]Min net atomic charge (Zefirov) for H atoms au Minimum (lowest) atomic partial charge over hydrogens in molecule

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 AM1 calculation. Model developed by using multilinear regression.

4.6.Software name and version for descriptor generation:

QSARModel 4.0.3

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:

0.096, (5 descriptors / 52 chemicals); 10.40 (52 chemicals / 5 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: benzene derivatives with one aromatic core and various functionalities

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

in the following minimal-maximal range:

Difference (Pos - Neg) in Charged Partial Surface Area (AM1): -0.731 - 0.679

Max bond order (AM1): 1.33 - 1.84

Min nucleophilic reactivity index (AM1) for N atoms: 0 - 0.0561

Relative number of O atoms: 0 - 0.294

Min net atomic charge (Zefirov) for H atoms: 0 - 0.0947

5.2.Method used to assess the applicability domain:

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 4.0.3

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:

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:
52 data points
0 negative values
52 positive values
6.6.Pre-processing of data before modelling:
n/a
6.7.Statistics for goodness-of-fit:
$R^2 = 0.748$ (Correlation coefficient)
$s^2 = 0.319$ (Standard error of the estimate)
$F = 27.38$ (Fisher function)
6.8.Robustness - Statistics obtained by leave-one-out cross-validation:
$R^2CV = 0.670$
6.9.Robustness - Statistics obtained by leave-many-out cross-validation:
$R^2CVMO = 0.662$
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 R^2 (fitting) = 0.760
average R^2 (prediction) = 0.640

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:

5 data points,

0 negative values,

5 positive values

7.6.Experimental design of test set:

From sorted data each 10th was subjected to the test set.

7.7.Predictivity - Statistics obtained by external validation:

R^2 = 0.596 (Coefficient of determination)

7.8.Predictivity - Assessment of the external validation set:

Taken the relatively small size of the source data, subjecting 10% of the data to the validation set is about the limit that can be done without losing all predictivity of the model. The data were selected from all enpoint value regions of the source dataset.

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

Difference (Pos - Neg) in Charged Partial Surface Area (AM1): -0.0567 - 0.583

Max bond order (AM1): 1.42 - 1.94

Min nucleophilic reactivity index (AM1) for N atoms: 3.65E-6 - 0.0231

Relative number of O atoms: 0.0714 - 0.143

Min net atomic charge (Zefirov) for H atoms: 0.0220 - 0.0319

7.9.Comments on the external validation of the model:

The validation coefficient of determination (R^2) is close to those coefficients of internal validation (R2CV and R2CVMO).

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

Bearing in mind that the training set is quite homogenously defined the obtained linear model do not include any specific size characterizing descriptors. The descriptors in model are related to charge distributions and to reactivities. Particularly "Difference (Pos - Neg) in Charged Partial Surface Area (AM1)" reflects charged surface area distribution and positive sign of the descriptors shows that big positively charged surface areas will increase the toxicity. Also other atom specific descriptors "Min nucleophilic reactivity index (AM1) for N atoms", "Relative number of O atoms" and "Min net atomic charge (Zefirov) for H atoms" tend to increase the toxicity. "Min nucleophilic reactivity index (AM1) for N atoms" reflects nucleophilic reactivity of nitro and amino groups in benzenes. "Relative number of O atoms" is

related to amount of oxygen atoms in the benzene derivatives. "Min net atomic charge (Zefirov) for H atoms" reflects the stability or activity of the benzene rings. "Max bond order (AM1)" is the only descriptor in the model which deacrease the toxicity and it is related to saturation of double and delocalized bonds (bonds in nitro grous and in benzene core).

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:

Interpretation in general agreement with literature (ref 1-2 in 9.2). It must been understood, that mammalian toxicity is very complex in nature and hardly any model can take into account all possible mechanisms of action.

9.Miscellaneous information

9.1.Comments:

-

9.2.Bibliography:

[1]QSAR modeling of acute toxicity by balance of correlations http://dx.doi.org/10.1016/j.bmc.2008.04.055
[2]Review of (Q)SARs for Mammalian Toxicity http://ecb.jrc.ec.europa.eu/documents/QSAR/QSAR_Review_Mammalian_Toxicity.pdf

9.3.Supporting information:

Training set(s)Test set(s)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:

10.4.Comments: