1.QSAR identifier

1.1.QSAR identifier (title):

QSAR model for Acute Toxicity (Inhalation) for fluorinated compounds

1.2.Other related models:

1.3.Software coding the model:

QSARModel 4.0.4 Molcode Ltd., Turu 2, Tartu, 51014, Estonia http://www.molcode.com

2.General information

2.1.Date of QMRF:

17.12.2010

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]Kaido Tämm Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[6]Jaak Jänes Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[7]Eneli Härk Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com
[8]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:

16.12.2010

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:

Rat

3.2.Endpoint:

4.Human health effects 4.1.Acute inhalation toxicity

3.3.Comment on endpoint:

Inhalation toxicity (LC_{50})

3.4.Endpoint units:

mmol

3.5.Dependent variable:

$\log(1/LC_{50})$

3.6.Experimental protocol:

The experimental data on rat LC50 inhalation studies were collected from ChemID plus [1], which has compiled the data from various literature and patents. Since the biological activity data in the database was collected from diverse sources, various laboratories, and at different conditions, in a period of over 59 years of study (starting from 1950), the data needs to be carefully verified. We did not fully cross-check these data by comparing with the original literature as some of them pertain to very early works and are reported in journals, patents, and dockets published in different languages. But the ChemID plus data was verified as much as possible and filtered by performing principle component analysis (PCA) and by omitting the spurious compounds which could badly influence the regression models. Those compounds are mainly of unusually high or low activity values and were not included in the model development. Only those compounds with organic backbones were included in the data set. Although the inclusion of different per- and polyfluorinated chemicals helps in maintaining the heterogeneity of the chemicals under study, several chemicals with inorganic and halogen centered backbones were not included despite the availability of bioactivity data. Hydrates, ions, dimers, complex mixtures, and salts of the perfluorinated chemicals were not considered. Thus, cleaning the data became an important part of the data set

preparation. Some of the most important perfluorinated chemicals in their salts form with Li^+, Na^+, K^+, and $NH4^+$ had to be discarded (due to limitations of the approach of the software used). The cleaned data set contained 52 compounds for rat inhalation LC50 data. The inhalation data are expressed in log 1/LC50 units where LC50 values expressed in g/m^3 or ppb were converted into mmol first before converting into the inverse log scale [2].

3.7.Endpoint data quality and variability:

See 3.6 for discussion on data quality.

Statistics:

max value: 6.61

min value: 0.159

standard deviation: 1.68

skewness: 0.294

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

pLC50 = -30.986

+59.839*FPSA3 Fractional PPSA (PPSA-3/TMSA) (AM1)

-0.192*Lowest total interaction (AM1)

+34.078*Min atomic orbital electronic population (AM1)

-0.142*Difference (Pos - Neg) in Charged Part of Charged Surface Area (Zefirov)

4.3.Descriptors in the model:

[1]FPSA3 Fractional PPSA (PPSA-3/TMSA) (AM1) [unitless] Fractional atomic charge weighted partial positive surface area

[2]Lowest total interaction (AM1) [eV] Lowest total interaction energy between two bonded atomic species

[3]Min atomic orbital electronic population (AM1) [unitless] lowest electronic population of atomic orbitals

[4]Difference (Pos - Neg) in Charged Part of Charged Surface Area (Zefirov) [au/Å2] Difference between total charge weighted partial positive and negative surface areas

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 4.0.4

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:

10.5, (42 chemicals / 4 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: per- and polyfluorinated chemicals, excluding sulfur containing compounds

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

in the following range:

FPSA3 Fractional PPSA (PPSA-3/TMSA) (AM1): 0 - 0.0580

Lowest total interaction (AM1): -35.0 - -16.9

Min atomic orbital electronic population (AM1): 0.729 - 0.812

Difference (Pos - Neg) in Charged Part of Charged Surface Area (Zefirov): -41.8 - -8.81

5.2.Method used to assess the applicability domain:

Compounds should not conatin sulfur.

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 4.0.4

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:
42 data points
0 negative values
42 positive values
6.6.Pre-processing of data before modelling:
n/a
6.7.Statistics for goodness-of-fit:
$R^2 = 0.765$ (Correlation coefficient)
$s^2 = 0.870$ (Standard error of the estimate)
$F = 30.0$ (Fisher function)
6.8.Robustness - Statistics obtained by leave-one-out cross-validation:
$R^2_{CV} = 0.713$ (Cross-validated correlation coefficient)
6.9.Robustness - Statistics obtained by leave-many-out cross-validation:
$R^2_{CVMO} = 0.710$
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) = 0.782
average R^2 (prediction) = 0.611

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:
9 data points,

0 negative values,

9 positive values

7.6.Experimental design of test set:
Data were sorted according to experimenatl values and each 5th compounds   starting from 4th was subjected to the test set in order to assure equal   distribution among training and test set.

7.7.Predictivity - Statistics obtained by external validation:
R^2 = 0.719 (Correlation coefficient)

7.8.Predictivity - Assessment of the external validation set:
Descriptor value range (all in range of applicability domain):

FPSA3 Fractional PPSA (PPSA-3/TMSA) (AM1): 0 - 0.0622

Lowest total interaction (AM1): -25.9 - -17.0

Min atomic orbital electronic population (AM1): 0.735 - 0.788

Difference (Pos - Neg) in Charged Part of Charged Surface Area (Zefirov): -32.0 - -9.11

7.9.Comments on the external validation of the model:
Initial test set contained 10 compounds where was spotted one serious   outlier - Methanesulfonic acid, trifluoro-, 2,2,2-trifluoroethyl ester   (6226-25-1). The outlier behaviour is justified since training set does   not comprise any sulfur containing compound and it was removed.

Therefore the QSAR model is not useable for prediction of inhalation toxicity of sulfur containing compounds.

The validation correlation coefficient (R2) for the test set is good and it is close to that of the training set.

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:
Descriptor "FPSA3 Fractional PPSA (PPSA-3/TMSA) (AM1)" is related to   level of halogenicity of compound. Since halogenes possess negative   charge, the descriptor value for perhalogenated compounds is 0, less   halogenited compounds do have higher value. The rest of the descriptors,   "Lowest total interaction (AM1)", "Min atomic orbital electronic   population (AM1)", "Difference (Pos - Neg) in Charged Part of Charged   Surface Area (Zefirov)", are electron and charge related and they are   directly related to compound's stability and reactivity which are important for toxicity.

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 [1].

9.Miscellaneous information

9.1.Comments:

9.2.Bibliography:
[1]ChemIDplus advanced http://chem.sis.nlm.nih.gov/chemidplus/
[2]Per- and Polyfluoro Toxicity (LC50 Inhalation) Study in Rat and Mouse Using QSAR Modeling, B. Bhhatarai, P. Gramatica, Chem. Res. Toxicol., 2010, 23 (3), pp 528–539 http://dx.doi.org/10.1021/tx900252h

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: