	QMRF identifier (JRC Inventory): **Q13-22a-0034**
	QMRF Title: **QSAR for abiotic degradation in air**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

QSAR for abiotic degradation in air

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:

30.04.2010

2.2.QMRF author(s) and contact details:

Molcode model development team 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:

28.04.2010

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:

Training, selection and test sets are available. Algorithm is defined.

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

None to date.

3.Defining the endpoint - OECD Principle 1

3.1.Species:

3.2.Endpoint:

2.Environmental fate parameters 2.2.a.Persistence: Abiotic degradation in air

3.3.Comment on endpoint:

The half-life is the time required for the concentration of a substance to halve its original value in a particular environmental medium. The half-lives of organic compounds are among the most commonly used

criteria for studying persistence [sect.9.2;ref.1]. The semi-quantitative data based on expert judgment and actual experimental values have already been suggested by Webster et al. [sect.9.2;ref.2] as preferable for half life identification, and are commonly used to develop widely applied multimedia models [sect.9.2;ref.3,4]. In addition, a simple QSPR regression model has been demonstrated to be a useful tool for the identification and prioritization of existing or not yet synthesized potential persistent organic pollutants [sect.9.2;ref.5].

3.4.Endpoint units:

The half-life values (55-55000 h) were transformed into logarhitmic form for modelling.

3.5.Dependent variable:

log T(0.5)

3.6.Experimental protocol:

3.7.Endpoint data quality and variability:

The dataset includes 250 organic compounds of known half-lives for transformation in air [sect.9.2;ref.6]. The dataset is structurally heterogeneous and highly representative of many classes of already defined problematic chemicalsA collection of experimental data from different labs were used. Semi-quantitative degradation half lives in air were organized in nine half-life categories.
Statistics: max value: 4.74; min value: 0.699; standard deviation: 0.832; skewness: 1.02.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Multilinear regression QSAR
Multilinear regression QSAR derived by the BMLR (Best Multiple Linear Regression) method
log T(0.5) = 0.273
-0.364*HOMO energy (AM1)
-0.219*Min valency (AM1) for O atoms
-4.085*Relative number of H atoms
+3.955E-002*count of H-donor sites (AM1)
-0.307*Square root of Charged (AM1) Surface Area of N atoms

4.3.Descriptors in the model:

[1]HOMO energy (AM1) eV Energy of highest occupied molecular orbital energy
[2]Min valency (AM1) for O atoms unitless minimum valency on oxygen atoms
[3]Relative number of H atoms unitless Relative number of H atoms
[4]count of H-donor sites (AM1) unitless count of H-donor sites (AM1)u

4.4.Descriptor selection:

Initial pool of ~1000 descriptors. Stepwise descriptor selection based on a set of statistical selection rules:
a) one-parameter equations: Fisher criterion and R^2over threshold, variance and t-test value over threshold, intercorrelation

with another descriptor not over threshold.
b) 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.Chemicals/Descriptors ratio:

33.4 (167 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: diverse set of organic pollutants (aromatic,alphatic and cyclic amines, ketones, alcohols, esters, etc)
b) by descriptor value range: The model is suitable for compounds that have the descriptors in the following minimal-maximal ranges:

HOMO energy (AM1: -12.2 - -7.92
Min valency (AM1) for O atoms: 0 - 2.39
Relative number of H atoms: 0.00 - 0.70
count of H-donor sites (AM1) : 0.00 - 19.0
Square root of Charged (AM1) Surface Area of N atoms: 0.00 - 3.60

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 4.0.4
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:

167 data points: 0 negative values; 167 positive values

6.6.Pre-processing of data before modelling:

6.7.Statistics for goodness-of-fit:

R^2= 0.714 (Correlation coefficient)

s_2= 0.453 (Standard error of the estimate)

F = 80.3 (Fisher function)

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

R^2_{CV}= 0.692

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

R^2_{CVMO}= 0.689

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 (in increasing 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) = 0.717

average R^2(prediction) = 0.660

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:

83 data points: 0 negative values; 83 positive values

7.6.Experimental design of test set:

From sorted source data, each 3rd was subjected to the test set

7.7.Predictivity - Statistics obtained by external validation:

R^2= 0.602 (Coefficient of determination)

7.8.Predictivity - Assessment of the external validation set:

Two compounds were left out because their descriptor "Square root of Charged (AM1) Surface Area of N atoms" values did not fit in the domain. These compounds are: 2,4-dinitrotoluene (CAS: 121-14-2)4,6-dinitro-o-cresol (CAS: 534-52-1)
The rest are in range of applicability domain:
HOMO energy (AM1): -12.4 - -7.86
Min valency (AM1) for O atoms: 1.80 - 2.29
Relative number of H atoms: 0.0455 - 0.706
count of H-donor sites (AM1) : 0.00 - 21.0
Square root of Charged (AM1) Surface Area of N atoms: 0.00 - 3.70

7.9.Comments on the external validation of the model:

The validation coefficient of determination (R^2) is significant and close to the coefficients of internal validation (R^2CV and R^2CVMO).

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

"HOMO energy (AM1)" reflects an ionization potential of molecule. "Relative number of H atoms" indirectly shows a magnitude of unsaturated bonds (including aromaticity). A smaller number of hydrogens in carbon chain indicates larger content of unsaturated bonds. These bonds include p-orbitals which are important for carbon chain degradation. Hydrogen bonding capability is covered by the descriptor "count of H-donor sites (AM1)". Others atom specific descriptors reflect the reactivity of these heteroatom sites.

8.2.A priori or a posteriori mechanistic interpretation:

A posteriori mechanistic interpretation, consistent with published scientific interpretations of experiments [sect.9.2; ref.5].

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

The data were gathered from the Physical-Chemical Properties and Environmental Fate Handbook [sect.9.2; ref.6], which includes data from different sources. Therefore the experimental protocol cannot be provided.

The data were also semi-quantitatively classified as proposed by Webster et al [sect.9.2; ref.2].

9.2.Bibliography:

[1]UNEP, Stockholm Convention on Persistent Organic Pollutants, United Nations Environment Program, Geneva, Switzerland, 2001 http://www.pops.int

[2]Webster E, Mackay D & Wania F (1998). Evaluating Environmental Persistence. Environmental Toxicologycal Chemistry 17, 2148–2158.

[3]Klasmeier J, Matthies M, MacLeod M, Fenner K, Scheringer M, Stroebe M, Le Gall A C, McKone T, Van De Meent D & Wania F (2006). Application of Multimedia Models for Screening Assessment of Long-Range Transport Potential and Overall Persistence. Environmental Science & Technology 40, 53-60.

[4]Fenner K, Scheringer M, Macleod M, Matthies M, McKone T, Stroebe M, Beyer A, Bonnell M, Le Gall A C, Klasmeier J, Mackay D, Van de Meent D, Pennington D, Scharenberg B, Suzuki N & Wania F (2005). Comparing Estimates of Persistence And Long-Range Transport Potential among Multimedia Models. Environmental Science and Technology 39, 1932-1942.

[5]Gramatica P & Papa E (2007). Screening and ranking of POPs for global half-life: QSAR approaches for prioritization based on molecular structure. Environmental Science and Technology 41, 2833–2839.

[6]Mackay D, Shiu WY & Ma KC (2000). Physical-Chemical Properties and Environmental Fate Handbook. CRCnet-BASE CD-ROM. Chapman and Hall/CRC, Boca Raton, FL, USA.

9.3.Supporting information:

Abiotic degradation air training_167.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-22a-0034/attachment/A694
Abiotic degradation air test_83.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-22a-0034/attachment/A695

Test set(s)

10.Summary (JRC QSAR Model Database)

Supporting information

10.1.QMRF number:

Q13-22a-0034

10.2.Publication date:

2013-06-27

10.3.Keywords:

Molcode;persistence;abiotic degradation;air;biodegradation;

10.4.Comments:

former Q8-10-30-266