QMRF identifier (JRC Inventory):	
QMRF Title: *QSAR model for Persistence - Abiotic degradation in air, O3-radical reaction of Volatile Organic Compounds*	
*Printing Date:*10.03.2010	

1.QSAR identifier

1.1.QSAR identifier (title):

QSAR model for Persistence - Abiotic degradation in air, O3-radical reaction of Volatile Organic 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:

26.02.2010

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]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:

21.02.2010

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:

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:

n/a

3.2.Endpoint:

2.Environmental fate parameters 2.Persistence: Abiotic degradation in air (Phototransformation) 2.2.b.Indirect photolysis (OH-radical reaction, ozone-radical reaction, other)

3.3.Comment on endpoint:

O3-radical reaction; The dominant chemical process of chemicals in the gasphase is their reaction with OH radicals, NO3 radicals, and ozone.

3.4.Endpoint units:

cm3 s-1 molecule-1

3.5.Dependent variable:

-logK(O3)

3.6.Experimental protocol:

The selected data are for reactions at 25 °C and 1 atm; The gasphase reaction rate constants of O_3 radical and organic chemicals have been directly measured.

3.7.Endpoint data quality and variability:

The experimental data has been collected from different sources. The data has been successfully modelled before, somewhat supporting consistency.

Statistics:
max value: 25.3
min value: 13.1
standard deviation: 2.26
skewness: 1.02

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

-logK (O3) = -0.475
-2.338*HOMO energy (AM1)
+0.206*Lowest resonance energy (AM1)
+0.682*Number of aromatic bonds
-1.195*Topographic electronic index (Zefirov) (all atoms) all pairs

4.3.Descriptors in the model:

[1]HOMO energy (AM1) [eV] Energy of highest occupied molecular orbital energy

[2]Lowest resonance energy (AM1) [eV] Lowest resonance energy between two atomic species

[3]Relative number of aromatic bonds [unitless] Relative number of aromatic bonds

[4]Topographic electronic index (Zefirov) (all atoms) all pairs [au/Å2] Topolographical electronic index calculated over all atom pairs in the molecule and based on Zefirov's charge distribution

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.Chemicals/Descriptors ratio:

23.5, (94 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: Diverse Volatile Organic Compounds

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

in the following minimal-maximal range:

HOMO energy (AM1): -13.1 - -8.44

Lowest resonance energy (AM1): -31.0 - -13.0

Number of aromatic bonds: 0 - 6

Topographic electronic index (Zefirov) (all atoms) all pairs: 0.182 - 2.05

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:

94 data points

0 negative values

83 positive values

6.6.Pre-processing of data before modelling:

n/a

6.7.Statistics for goodness-of-fit:

R2 = 0.878 (Correlation coefficient)

s2 = 0.811 (Standard error of the estimate)

F = 159.3 (Fisher function)

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

R2CV = 0.863

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

R2CVMO = 0.862

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) = 0.879

average R2 (prediction) = 0.849

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,

0 negative values,

31 positive values

7.6.Experimental design of test set:

From sorted data each 4th was subjected to the test set.

7.7.Predictivity - Statistics obtained by external validation:

R2 = 0.831 (Coefficient of determination)

7.8.Predictivity - Assessment of the external validation set:

All are in range of applicability domain:

HOMO energy (AM1): -11.9 - -8.66

Lowest resonance energy (AM1): -29.7 - -13.6

Number of aromatic bonds: 0 - 6

Topographic electronic index (Zefirov) (all atoms) all pairs: 0.0835 - 1.10

7.9.Comments on the external validation of the model:

The validation coefficient of determination (R2) is good and 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:

"HOMO energy (AM1)" reflects an ionization potential of molecule. "Lowest resonance energy (AM1)" is related to molecule's stability against O3 radical. "Relative number of aromatic bonds" is represening an account of aromaticity which differs these compounds from aliphatic ones. "Topographic electronic index (Zefirov) (all atoms) all pairs" reflects charge distribution in the molecule and is related to reactivity of a molecule. The descriptors in the model are presenting important molecular properties related to H-abstraction.

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:

OECD SERIES ON TESTING AND ASSESSMENT, Number 49, THE REPORT FROM THE EXPERT GROUP ON (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIPS [(Q)SARs] ON THE PRINCIPLES FOR THE VALIDATION OF (Q)SARs, November 2004

9.Miscellaneous information

9.1.Comments:

Data is taken from:

OECD SERIES ON TESTING AND ASSESSMENT, Number 49, THE REPORT FROM THE EXPERT GROUP ON (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIPS [(Q)SARs] ON THE PRINCIPLES FOR THE VALIDATION OF (Q)SARs, November 2004

9.2.Bibliography:

OECD SERIES ON TESTING AND ASSESSMENT, Number 49, THE REPORT FROM THE EXPERT GROUP ON (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIPS [(Q)SARs] ON THE PRINCIPLES FOR THE VALIDATION OF (Q)SARs http://www.olis.oecd.org/olis/2004doc.nsf/LinkTo/NT00009192/$FILE/JT00176183.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: