	QMRF identifier (JRC Inventory): **Q13-22b-0032**
	QMRF Title: **QSAR for abiotic degradation in air (O_3radical reaction of volatile organic compounds)**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

QSAR for abiotic degradation in air (O_3radical 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:

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:

21.02.2010

2.7.Reference(s) to main scientific papers and/or software package:

2.8.Availability of information about the model:

Software is proprietary but model training and test sets provided.

Algorithm provided.

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

None to date.

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Not applicable - environmental fate parameter

3.2.Endpoint:

2.Environmental fate parameters 2.2.b.Persistence: Abiotic degradation in air (Phototransformation). Indirect photolysis (O3-radical reaction)

3.3.Comment on endpoint:

Rate constant for O_3 radical reaction (degradation).

The dominant chemical process of chemicals in the gasphase is their reaction with OH radicals, NO_3radicals and ozone.

3.4.Endpoint units:

$cm^3s^{-1}molecule^{-1}$

3.5.Dependent variable:

-logK (OH) (original rate constants were transformed into log scale and multiplied by -1 to reduce data range and obtain positive values)

3.6.Experimental protocol:

The selected data are for reactions at 25 °C and 1 atm. The gas-phase reaction rate constants of

O_3 radical and organic chemicals have been directly measured.

3.7.Endpoint data quality and variability:

The experimental data have been collected from different sources.

Experimental data originated from ref. 3.

Statistics for -logK (O_3):

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:

QSAR

4.2.Explicit algorithm:

Multilinear regression QSAR

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

-logK (NO3) = -7.355

+9.660E-002*HASA-2 (AM1) (all)

-2.070*HOMO energy (AM1)

+12.005*Relative number of aromatic bonds

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

4.4.Descriptor selection:

Initial pool of ~1000 descriptors for each structure calculated.

Stepwise descriptor selection was applied to reduce the pool based on a set of statistical selection rules.

For one-parameter equations: Fisher criterion and R^2over threshold, variance and t-test value over threshold, intercorrelation with another descriptor not over threshold)

Two parameter correlations developed from previously reduced pool, the statistical selection applied: 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. See refs 1-2 for explicit methodology.

4.5.Algorithm and descriptor generation:

1D, 2D, and 3D theoretical calculations. Descriptors derived from mol files. Quantum chemical descriptors from AM1 calculations. Model developed by using multilinear regression using ordinary least squares.

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 set of Volatile Organic Compounds (alphatic and aromatic hydrocarbons, alcohols, amines, halogenated compounds, etc)

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; 94 positive values

Original source dataset of 125 compounds split into training and testing sets - sorted by experimental value, each 4th structure subjected to testing set, others to training set.

6.6.Pre-processing of data before modelling:

No more than specified in 3.5

6.7.Statistics for goodness-of-fit:

R^2 = 0.878 (Correlation coefficient)

s_2= 0.811 (Standard error of the estimate)

F = 159.3 (Fisher function)

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

R^2_{CV}= 0.863

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

R^2_{CVMO}= 0.862 (80% : 20% , training : testing)

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 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) = 0.879; average R^2(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:

Original source dataset split into testing and training. From the original source data, sorted by endpoint value, each 4th was subjected to the test set.

7.7.Predictivity - Statistics obtained by external validation:

R^2= 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 (R^2) is close to the coefficients derived by internal validation ($R^2_{CV}$and R^2_{CVMO}).

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

The "HOMO energy (AM1)" is an indicator of the nucleophilicity of the molecule, also reactive molecules have relatively higher HOMO energy. "Lowest resonance energy (AM1)" is related to a molecule's stability against O_3radical. "Relative number of aromatic bonds" is represening an account of aromaticity which differentiates these compounds from aliphatic ones. "Topographic electronic index (Zefirov) (all atoms) all pairs" reflects charge distribution in the molecule as a whoile and is related to reactivity of a molecule. The descriptors in the model are presenting important molecular properties related to H-abstraction. For most compounds, H-abstraction is known to be the predominant pathway for reactions with O_3radicals. While HOMO energy and the electronic index descriptors have a negative sign in the equation, increase in the descriptor values leads to faster reaction rates.

Aromatic compounds have smaller rate constants, as indicated by the negative signs in the equation. Also compounds with high resonance energy tend to be more resistent towards reaction with O_3.

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:

Most published studies and models (see refs 4-5) indicate that the HOMO energy is the most important factor detrmining the rate constants for gas phase reactions with O_3radicals (as it is for $NO_3$and OH). Other descriptors depend on the training set and methodology

used but usually add corrections for structural variations (e.g. aromatics) or heteroatoms.

9.Miscellaneous information

9.1.Comments:

Source data is taken from ref 4., sect 9.2

9.2.Bibliography:

[1]Atkinson R & Carter WPL(1984). Kinetics and mechanisms of gas-phase reactions of ozone with organic compounds under atmospheric conditions. Chemical Reviews 84, 437-470.

[2]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.

[3]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.

[4]Gramatica P, Pilutti P & Papa E(2003). QSAR Prediction of ozone tropospheric degradation. QSAR & Combinatorial Science 22, 364-373.

[5]OECD (2004). 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/20 04doc.nsf/LinkTo/NT00009192/ $FILE/JT00176183.PDF

9.3.Supporting information:

Photolysis training _212.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-22b-0032/attachment/A688
Photolysis tests_211.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-22b-0032/attachment/A689

Test set(s)

10.Summary (JRC QSAR Model Database)

Supporting information

10.1.QMRF number:

Q13-22b-0032

10.2.Publication date:

2013-06-27

10.3.Keywords:

Molcode;abiotic degradation in air;O3 radical reaction;volatile organic compound;

10.4.Comments:

former Q8-10-30-224