
	QMRF identifier (JRC Inventory):	
	QMRF Title: <i>QSAR model for Persistence: Abiotic degradation in soil</i>	
	Printing Date: <i>3.05.2010</i>	

1. QSAR identifier

1.1. QSAR identifier (title):

QSAR model for Persistence: Abiotic degradation in soil

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:

23.04.2010

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:

19.04.2010

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:

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:

[1][2]2.Environmental fate parameters 3.Persistence: Biodegradation

2.3.a.Ready/not ready biodegradability

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

3.4.Endpoint units:

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

3.5.Dependent variable:

log T(0.5)

3.6.Experimental protocol:

The dataset of structurally heterogeneous and highly representative of many classes of already defined problematic chemicals includes 250 organic compounds of known half-lives for transformation into soil [6].

3.7. Endpoint data quality and variability:

A collection of experimental data from different labs has been used; Semiquantitative degradation half lives in soil were organized in seven half-life categories.

Statistics:

max value: 4.74

min value: 1.74

standard deviation: 0.808

skewness: 0.410

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

$$\log T(0.5) = 2.912$$

$$+1.742 * \text{HASA-1/TMSA (AM1) (all)}$$

$$+0.320 * \text{Number of rings}$$

$$-22.969 * \text{HACA-1/TMSA (Zefirov)}$$

$$-3.422 * \text{Polarity parameter (Zefirov)}$$

4.3. Descriptors in the model:

[1] HASA-1/TMSA (AM1) (all) unitless relative solvent-accessible surface area of H-bonding acceptor atoms (from AM1 calculation)

[2] Number of rings unitless count of cycles in the molecule

[3] HACA-1/TMSA (Zefirov) unitless sum of relative solvent-accessible surface area of H-bonding acceptor atoms, selected by threshold charge

[4] Polarity parameter (Zefirov) au difference between the most positive and the most negative atomic partial charges in the 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 4.0.4

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:

41.8 (167 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 organic pollutants (aromatic, aliphatic 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 range:

HASA-1/TMSA (AM1) (all): 0.00 - 1.00

Number of rings: 0 - 5

HACA-1/TMSA (Zefirov): 0.00 - 0.0476

Polarity parameter (Zefirov): 0.0384 - 0.326

5.2. Method used to assess the applicability domain:

Range of descriptor values in training set with $\pm 30\%$ confidence.

Descriptor values must fall between maximal and minimal descriptor values of training set $\pm 30\%$.

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:

n/a

6.7.Statistics for goodness-of-fit:

$R^2 = 0.762$ (Correlation coefficient)

$s^2 = 0.400$ (Standard error of the estimate)

$F = 130$ (Fisher function)

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

$R^2_{CV} = 0.750$

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

$R^2_{CVMO} = 0.748$

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) = 0.762

average R^2 (prediction) = 0.742

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:

$R^2 = 0.824$ (Coefficient of determination)

7.8.Predictivity - Assessment of the external validation set:

All are in range of applicability domain:

HASA-1/TMSA (AM1) (all): 0.00 - 1.00

Number of rings: 0 - 5

HACA-1/TMSA (Zefirov): 0.00 - 0.0312

Polarity parameter (Zefirov): 0.0385 - 0.324

7.9.Comments on the external validation of the model:

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

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

Model consists two hydrogen bonding related descriptors - "HASA-1/TMSA (AM1) (all)" and "HACA-1/TMSA (Zefirov)". These descriptors together are considering charge and the surface area of hydrogen acceptors. Thus, the strength of possible hydrogen bondings. "Number of rings" is related to the flexibility of the compounds. Also, in most cases the ring is aromatic, it represents aromaticity. "Polarity parameter (Zefirov)" represents the charge differences. Higher electrostatic interactions would lead to bigger reactivities.

8.2.A priori or a posteriori mechanistic interpretation:

a posteriori mechanistic interpretation, consistent with published scientific interpretations of experiments and models [3-5]

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

The data are gathered from handbook (Physical-Chemical Properties and Environmental Fate Handbook) which includes data from different sources. Therefore the experimental protocol cannot be provided. The data were also semiquantitatively classified as proposed by Mackay [1]. [1] Webster, E.; Mackay, D.; Wania, F. Evaluating Environmental Persistence. Environ. Toxicol. Chem. 1998, 17, 2148-2158.

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. Evaluating Environmental Persistence, Environ. Toxicol. Chem. 1998, 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. Application of Multimedia Models for Screening Assessment of Long-Range Transport Potential and Overall Persistence, Environ. Sci. Technol. 2006, 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. Comparing Estimates of Persistence And Long-Range Transport Potential among Multimedia Models, Environ. Sci. Technol. 2005, 39, 1932-1942.

[5]Gramatica, P.; Papa, E. Screening and ranking of POPs for global half-life: QSAR approaches for prioritization based on molecular structure, Environ. Sci. Technol. 2007, 41, 2833-2839.

[6]Mackay, D.; Shiu, W. Y.; Ma, K. C. Physical-Chemical Properties and Environmental Fate Handbook, CRCnet-BASE CD-ROM; Chapman and Hall/CRC: Boca Raton, FL, 2000.

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_KarelsonArkivoc2008.pdf
Karelson Arkivoc 2009	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf83_KarelsonArkivoc2009.pdf

10.Summary (ECB Inventory)

10.1.QMRF number:

10.2.Publication date:

10.3.Keywords:

10.4.Comments: