	QMRF identifier (JRC Inventory):Q13-26-0009
	QMRF Title:QSAR for the soil adsorption coefficient Koc
	Printing Date:Dec 11, 2019

1.QSAR identifier

1.1.QSAR identifier (title):

QSAR for the soil adsorption coefficient Koc

1.2.Other related models:

1.3.Software coding the model:

QSARModel 3.7.0

Molcode Ltd., Turu 2, Tartu, 51014, Estonia

http://www.molcode.com

2.General information

2.1.Date of QMRF:

04.09.2009

2.2.QMRF author(s) and contact details:

Molcode model development team Molcode Ltd. 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. Molcode Ltd., Turu 2, Tartu, 51014, Estonia

models@molcode.com http://www.molcode.com

2.6.Date of model development and/or publication:

03.09.2009

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.

Algorithm 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:

3.2.Endpoint:

2.Environmental fate parameters 2.6.Partition coefficient. Organic carbon-sorption partition
coefficient (organic carbon; Koc)

3.3.Comment on endpoint:

Koc values on soil and on sewage sludge were obtained using high performace liquid chromatography (HPLC). The adsorption coefficient is defined as the ratio between the concentration of the substance in the soil/sludge and the concentration of the substance in the aqueous phase at adsorption equilibrium.

3.4. Endpoint units:

unitless

3.5. Dependent variable:

logKoc

3.6. Experimental protocol:

Estimation of the adsorption coefficient (KOC) on soil was determined using the OECD Test Guideline TG121 (EU C.19).
The adsorption coefficient normalized to the organic carbon content of the soil KOC is a useful indicator of the binding capacity of a chemical on organic matter of soil and sewage sludge and allows comparisons to be made between different chemicals.KOC=Kd/fOC or KOC=Kf/fOC, where Kd is distribution coefficient, Kf is Freundlich adsorption coefficient and fOC is organic carbon content of a sorbent. The application of HPLC screening has become an accepted tool to estimate reliably soil adsorption coefficients of many organic chemicals. The method is based on the similitude between soil and a chromatographic column, and correlates soil adsorption coefficients with the HPLC-retention behaviour expressed as capacity factor (references 8,9).
Experimental soil adsorption coefficients KOC of 142 non-ionic organic pesticides (one compound (CAS 2303-16-4) were excluded because undefined isomer form) were taken from references 1,3 (in section 9.2),. KOC values measured in soils with low carbon content have been excluded, since interactions of a pesticide with soil inorganic matrix may become more important. KOC data for 20 other chemicals obtained from several sources (references 4-7, in 9.2) were used as external validation set (in addition to the 34, split from the training set). The experimental data for testing set is a collection of several scientific publications, training set is from ref (1) in section 9.2

3.7. Endpoint data quality and variability:

Statistics: max value: 5.15, min value: 0.42, standard deviation: 0.88, skewness: 0.70

4. Defining the algorithm - OECD Principle 2

4.1. Type of model:

QSAR

4.2. Explicit algorithm:

multilinear regression QSAR
logKoc = 0.96 -0.26*Polarity parameter (AM1) / distance +1.07E-002*ALFA polarizability (DIP) (AM1) -1.99*Max net atomic charge (AM1) for C atoms +1.30E-002*WNSA1 Weighted PNSA (PNSA1*TMSA/1000) (Zefirov)

4.3.Descriptors in the model:

[1]ALFA polarizability (DIP) (AM1) [au] Quatum chemically (AM1 method) obtained dipole-based polarizability of molecule

[2]Max net atomic charge (AM1) for C atoms [au] maximum partial charge on any carbon atom (from AM1 calculation)

[3]Polarity parameter (AM1) / distance [au/ Å] difference of maximum positive and negative partial charges (from AM1 calculations)

[4]WNSA1 Weighted PNSA (PNSA1*TMSA/1000) (Zefirov) [mÅ4] partial charge weighted partial negatively charged surface area of molecule

4.4.Descriptor selection:

Initial pool of ~1000 descriptors

A set of statistical selection rules applied, including Fisher

criterion, variance check, intercorrelation check, signifficance check

for 1- and 2-parameter correlations (in terms of correlation coefficient

and t-test)

4.5.Algorithm and descriptor generation:

Multiple linear regression applied to generate the model.

1D, 2D, and 3D descrptors were calculated, 3D descriptos were based on

the 3D molecular structures (optimized by AM1 method as included in

MOPAC 6.0)

4.6.Software name and version for descriptor generation:

QSARModel 3.7.0

quantum chemical descriptors derived from AM1 calculation

Molcode Ltd, Turu 2, Tartu, 51014, Estonia,

http://www.molcode.com

4.7.Chemicals/Descriptors ratio:

27.0 (108 chemicals / 4 descriptors)

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

By chemical identity: non-ionic organic heterogeneous pesticides

By descriptor value range:

ALFA polarizability (DIP) (AM1): 71.2 - 239

Max net atomic charge (AM1) for C atoms: -0.0812 - 0.546

Polarity parameter (AM1) / distance: 0.0426 - 2.58

WNSA1 Weighted PNSA (PNSA1*TMSA/1000) (Zefirov): 27.2 - 185

5.2.Method used to assess the applicability domain:

Range of descriptor values in training set with ±30% confidence (from

minimal value - 1/3 of total range to maximum value +1/3 of total range

of descriptor values)

5.3.Software name and version for applicability domain assessment:

QSARModel 3.7.0

Molcode Ltd, Turu 2, Tartu, 51014, Estonia

http://www.molcode.com

5.4.Limits of applicability:

no other limits than specified in 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:

108 data points: 0 negative values; 108 positive values

6.6.Pre-processing of data before modelling:

No other than tranforming to logarithmic scale

6.7.Statistics for goodness-of-fit:

R^2= 0.756 (Correlation coefficient);

s^2= 0.445 (Standard error of the estimate);

F = 79.76 (Fisher function);

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

R^2_{CV} = 0.726 (Cross-validated correlation

coefficient);

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

R^2_{CVMO}= 0.727 (80% / 20 %)

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 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) = 0.756; average R^2(prediction) = 0.755

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:

54 data points: 0 negative values; 54 positive values

7.6.Experimental design of test set:

From initial (sorted by endpoint value) dataset, each 3rd was subjected

to test set, additional 20 points from other publications.

7.7.Predictivity - Statistics obtained by external validation:

R^2= 0.754 (Correlation coefficient)

7.8.Predictivity - Assessment of the external validation set:

Descriptor value range:

ALFA polarizability (DIP) (AM1): 80.2 - 181

Max net atomic charge (AM1) for C atoms: -0.0229 - 0.475

Polarity parameter (AM1) / distance: 0.0367 - 2.51

WNSA1 Weighted PNSA (PNSA1*TMSA/1000) (Zefirov): 28.1 - 183

7.9.Comments on the external validation of the model:

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

Soil sorption is closely related to water solubility and hydrophobicity
(logPow). Therefore, the chemical features which determine the soil
sorption are similar to the ones which determine water solubility. In
particular equation the "ALFA polarizability (DIP) (AM1)" and the "WNSA1
Weighted PNSA (PNSA1*TMSA/1000) (Zefirov)" are size related descriptors
and larger compounds tend to have higher soil sorption because they do
have lower water solubility. The "Max net atomic charge (AM1) for C
atoms" and the "Polarity parameter (AM1) / distance" are related to
charges and to charge distribution. High charge on carbon atom show the
presence of active functional group next to carbon and leads to better
water solubility, likewise higher polarity leads to better water
solubility. Thus these descriptors have minus sign in QSPR equation and
their higher vales lower the soil sorption.

8.2.A priori or a posteriori mechanistic interpretation:

A posteriori mechanistic interpretation, consistent with published
scientific interpretations of experiments (references 1-3,7 in section
9.2)

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

Concerned to publication where the source data came from, some CAS
numbers were corrected:

116-03-3 should be 116-06-3 (Aldicarb)

12709-03-6 should be 12789-03-6 (Chlordane)

02425-10-7 should be 2425-10-7 (Xylylcarb - name also corrected).

And one compound was excluded from the original set because two
compounds (cis/trans isomers) had the same CAS (2303-16-4) with the same
endpoint value where the CAS corresponds to cis isomer.

9.2.Bibliography:

[1]Gramatica P, Corradi M, Consonni V, 2000. Modelling and prediction of soil sorption coefficients
of non-ionic organic pesticides by molecular descriptors. Chemosphere 41, 763-777.
http://dx.doi.org/10.1016/S0045-6535(99)00463-4

[2]Hance RJ, 1969. Empirical relationship between chemical structure and the sorption of some
herbicides by soil. J. Agric. Food Chem. 17, 667-668. http://dx.doi.org/10.1021/jf60163a032

[3]Sabljic A, Güsten H, Verhaar H, Hermens J, 1995. QSAR modeling of soil sorption. Improvements
and systematics of log KOC vs log KOW correlations. Chemosphere 31, 4489-4514.
http://dx.doi.org/10.1016/0045-6535(95)00327-5

[4]Meylan W, Howard PH, Boethling RS, 1992. Molecular topology/fragment contribution method for
predicting soil sorption coefficients. Environ. Sci. Technol. 26, 1560-1567.
http://dx.doi.org/10.1021/es00032a011

[5]Kanazawa J, 1989. Relationship between the soil sorption constants for pesticides and their
physicochemical properties. Environ. Toxicol. Chem. 8, 477-484.
http://dx.doi.org/10.1002/etc.5620080604

[6]Gawlik BM, Feicht EA, Karcher W, Kettrup A, Muntau H, 1998. Application of the european
reference soil set (EUROSOILS) to a HPLC-screening method for the estimation of soil adsorption
coefficients of organic compounds. Chemosphere 36, 2903-2919. http://dx.doi.org/10.1016/S0045-
6535(97)10247-8

[7]Lohninger H, 1994. Estimation of soil partition coefficients of pesticides from their chemical
structure. Chemosphere 29, 1611-1626. http://dx.doi.org/10.1016/0045-6535(94)90309-3

[8]Kördel W, Stutte J, Kotthoff G, 1993. HPLC-screening method for the determination of the
adsorption coefficient on soil - comparison of different stationary phases. Chemosphere 27, 2341-
2352. http://dx.doi.org/10.1016/0045-6535(93)90257-6

[9]Kördel W, Stutte J, Kotthoff G, 1995. HPLC-screening method to determine the adsorption
coefficient in soil - comparison of immobilized humic acid and clay mineral phases for cyanopropyl
columns. Sci. Tot. Environ. 162, 119-125. http://dx.doi.org/10.1016/0048-9697(95)04443-5

9.3.Supporting information:

Soil sorption_108_training.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-26-0009/attachment/A642
Soil sorption_54_test.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-26-0009/attachment/A643

Test set(s)

10.Summary (JRC QSAR Model Database)

Supporting information

10.1.QMRF number:

Q13-26-0009

10.2.Publication date:

2013-06-21

10.3.Keywords:

Molcode;soil adsorption;desorption;Koc;

10.4.Comments:

former Q2-10-26-179