1. QSAR identifier

1.1. QSAR identifier (title):
QSAR model for Soil adsorption logKoc

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:
Indrek Tulp, Tarmo Tamm, Dimitar Dobchev, Dana Martin, Gunnar Karelson, Jaak Jänes, Kaido Tämm, Eneli Härk, Deniss Savchenko, Andres Kreegipuu, 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:
03.09.2009

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

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:
No other QMRF available for the same model.

3. Defining the endpoint - OECD Principle 1

3.1. Species:
N/A
3.2. Endpoint:
Environmental fate parameters 2.7. Adsorption/Desorption in soil

3.3. Comment on endpoint:
Environmental fate parameters. Ecotoxic effects. Adsorption/Desorption in soil. C.19

3.4. Endpoint units:
unitless

3.5. Dependent variable:
logKoc

3.6. Experimental protocol:
C.19 Estimation of the adsorption coefficient (KOC) on soil and on sewage sludge using high performance liquid chromatography (HPLC)

Estimation of the adsorption coefficient (KOC) on soil was determined using the OECD Test Guideline TG121 (EU C.19).

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. 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.

\[ KOC = K_d/f_{OC} \text{ or } KOC = K_f/f_{OC}, \]

where \( K_d \) is distribution coefficient, \( K_f \) is Freundlich adsorption coefficient and \( f_{OC} \) 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. Experimental soil adsorption coefficients KOC of 143 non-ionic organic pesticides, used as the training set, were taken from the work of Sabljic et al. (1995). 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 by various authors were used as external validation set.

References


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:
   - 2D and 3D regression-based QSAR

4.2. Explicit algorithm:
   - Multilinear regression QSAR

4.3. Descriptors in the model:
   [1] ALFA polarizability (DIP) (AM1) [au]
   [2] Max net atomic charge (AM1) for C atoms [au]
   [3] Polarity parameter (AM1) / distance [au/Å]
   [4] WNSA1 Weighted PNSA (PNSA1*TMSA/1000) (Zefirov) [mÅ4]

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:
   - 1D, 2D, and 3D theoretical calculations

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. Descriptors/Chemicals ratio:
   - 0.037, (4/108)

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
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:
n/a

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:
- 

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:
$R^2_{CVMO} = 0.727$

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 sorted data each 3rd was subjected to test set.

#### 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:
The validation correlation coefficient ($R^2$) for the test set is very good.

### 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 ($\log$Pow). 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 values lower the soil sorption.
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:

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:

9.3. Supporting information:
   Training set(s)

|-------------------------------|---------------------------------------------------------------------------------------------------------------|

Supporting information


10. Summary (JRC Inventory)

10.1. QMRF number:
   To be entered by JRC

10.2. Publication date:
   To be entered by JRC

10.3. Keywords:
   To be entered by JRC

10.4. Comments:
   To be entered by JRC