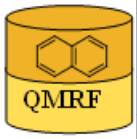	***QMRF identifier (ECB Inventory):****Q8-10-14-169* ***QMRF Title:*** *QSAR for honey bee acute contact toxicity (ether derivatives not containing amide groups)* ***Printing Date:****Feb 16, 2010*	

1.QSAR identifier

1.1.QSAR identifier (title):

QSAR for honey bee acute contact toxicity (ether derivatives not containing amide groups)

1.2.Other related models:

QSAR for honey bee acute contact toxicity (amide derivatives)

QSAR for honey bee acute contact toxicity (amine derivatives)

QSAR for honey bee acute contact toxicity (ester derivatives)

1.3.Software coding the model:

QSARModel 3.5.0 Molcode Ltd., Turu 2, Tartu, 51014, Estonia http://www.molcode.com

2.General information

2.1.Date of QMRF:

26.05.2009

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:

26.05.2009

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.

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

None to date.

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Honey bee

3.2.Endpoint:

3.Ecotoxic effects Honey bee acute contact toxicity

3.3.Comment on endpoint:

Honey bee acute contact toxicity.

3.4.Endpoint units:

log(μmol/org)

3.5.Dependent variable:

log(LD50)

3.6.Experimental protocol:

Honey bee acute contact toxicity was determined using the OECD Test Guideline 214 (EU C.17). The acute contact toxicity test is carried out to determine the inherent toxicity of pesticides and other chemicals to bees. Acute contact toxicity is presented as the median lethal dose (LD50) at which half of the population of honeybee died after exposure for 48 h.

3.7.Endpoint data quality and variability:

Toxicity data for EPA - Office of Pesticide Programs database (http://www.epa.gov/pesticides/) are drawn from several sources and then reviewed:

(i) Ecotoxicological studies conducted by commercial laboratories and submitted by pesticide companies in support of their products. EPAs Office of Compliance and Monitoring conducts periodic audits of these laboratories.

(ii) Studies conducted by US-EPA, USDA, and USFWS laboratories over the last 25 years.

(iii) Published data considered to meet their guideline criteria for acceptable data.

Inorganic compounds and mixtures in which components have different molecular weight or connectivity (i.e., substances with different chemical identity) were eliminated from the original EPA dataset. However, mixtures of stereoisomers were kept, because they are super imposable using common 2D descriptors. Data for LC50 48 h exposure of honeybee were then pruned as follows:

(i) Eliminating studies with an a.s.<85% purity.

(ii) Those identified as invalid where invalid studies were defined by the EPA as studies which may not be scientifically sound, or they were performed under conditions that deviated so significantly from the recommended protocols that the results will not be useful in a risk assessment.

(iii) Furthermore, only studies with actual values were kept discarding data given as higher or lower that values.

Statistics:

max value: 0.11

min value: -4.26

standard deviation: 1.33

skewness: -0.69

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

Multilinear regression QSAR

Log(LD50)= -136 -0.548 Highest total interaction (AM1) for C - C bonds +11.0*Min atomic orbital electronic population (AM1) -4.00*ZX Shadow / ZX Rectangle (AM1) -0.315*Highest e-n attraction (AM1) for C - O bonds

4.3.Descriptors in the model:

[1]Highest total interaction (AM1) for C - C bonds [eV]
[2]Min atomic orbital electronic population (AM1), [-]
[3]ZX Shadow / ZX Rectangle (AM1) [-]
[4]Highest e-n attraction (AM1) for C - O bonds [eV]

4.4.Descriptor selection:

Initial pool of ~1000 descriptors. Stepwise descriptor selection based on a set of statistical selection rules:

1-parameter equations: Fisher criterion and R2 over threshold, variance and t-test value over threshold, intercorrelation with another descriptor not over threshold.

2 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 were derived from Merck Molecular Force Field (MMFFs) (water) conformational search and AM1 calculation. Model developed by using multilinear regression

4.6.Software name and version for descriptor generation:

QSARModel 3.5.0
Molcode Ltd., Turu 2, Tartu, 51014, Estonia
http://www.molcode.com

4.7.Descriptors/Chemicals ratio:

11.8 (47 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: pesticides with ester functionality

b) by descriptor value range:

Highest total interaction (AM1) for C - C bonds (min: -15.9, max:-11.9)

Min atomic orbital electronic population (AM1) (min: 0.48, max: 0.81)

ZX Shadow / ZX Rectangle (AM1) (min: 0.49, max:0.78)

Highest e-n attraction (AM1) for C - O bonds (min: -393, max: -381)

5.2.Method used to assess the applicability domain:

Pesticides with presence of ester functional groups in the structures

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 3.5.0

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

http://www.molcode.com

5.4.Limits of applicability:

Compounds should contain the ester functional group.

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:No

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:

47 data points: 45 negative values; 2 positive values.

6.6.Pre-processing of data before modelling:

6.7.Statistics for goodness-of-fit:

R^2 = 0.819 (Correlation coefficient);

s^2 = 0.357 (Standard error of the estimate);

F = 47.4 (Fisher statistic);

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

$R^2cv = 0.770$ (LOO)

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

$R^2cv = 0.760$ (LMO)

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 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) = 0.83;

average R^2 (prediction) = 0.78;

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:No

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:

5 data points: 5 negative values; 0 positive values

7.6.Experimental design of test set:

The full experimental dataset was sorted according to increasing values of property and each tenth compound was assigned to the test set.

7.7.Predictivity - Statistics obtained by external validation:

$R^2 = 0.928$ (Correlation coefficient)

7.8.Predictivity - Assessment of the external validation set:

The descriptor values of the compounds in the test set are in the limit of applicability.

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:

ZX Shadow / ZX Rectangle (AM1) descriptor reflects how well is the compound packed in its length and height dimension. The descriptor is related to geometrical shape of the molecule. It describes not only on the conformation but also the orientation of the molecule. Quantum-chemical energy descriptors Highest total interaction (AM1) for C - C bonds and Highest e-n attraction (AM1) for C - O bonds are related to the stability/activity of the esters. Higher interaction energies in absolute scale (the values are negative) show higher stability of the compounds, and thus, they lead to lower toxicity, hence the negative sign in the model. Min atomic orbital electronic population (AM1) is related to electrophilicity. Lower values lead to higher toxicity because the compound forms stronger electrophilic interactions.

8.2.A priori or a posteriori mechanistic interpretation:

A priori - it was postulated that amines have different mechanism than other compounds. The rest of the interpretation – a posteriori.

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

Since bees pollinate plants and fruit trees, the influence of pesticides on bees is an important ecologic question.

9.2.Bibliography:

[1]EPA - Office of Pesticide Programs database http://www.epa.gov/pesticides/

[2]Toropov AA & Benfenati E (2007). SMILES as an alternative to the graph in QSAR modelling of bee toxicity. Computational Biology and Chemistry 31, 57–60.

[3]Roncaglioni A, Benfenati E, Boriani E & Clook M (2004). A Protocol to Select High Quality Datasets of Ecotoxicity Values for Pesticides. Journal of Environmental Science and Health. B39 (4), 641–652.

[4]Toropov AA & Benfenati E(2008). Additive SMILES-based optimal descriptors in QSAR modelling bee toxicity: Using rare SMILES attributes to define the applicability domain. Bioorganic & Medicinal Chemistry 16, 4801–4809.

9.3.Supporting information:

Training set(s)

Honey_bee_ethers_trainingset_56	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf169_Honey_bee_ethers_trainingset_56.sdf

Test set(s)

Honey_bee_ethers_testset_6	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf169_Honey_bee_ethers_testset_6.sdf

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

Q8-10-14-169

10.2.Publication date:

2010/01/25

10.3.Keywords:

Molcode, honey bee, acute contact toxicity

10.4.Comments: