	QMRF identifier (JRC Inventory): **To be entered by JRC**
	QMRF Title: **Linear nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents. The model has been presented in the puclication "A Molecular-Based Model for Prediction of Solubility of C60 Fullerene in Various Solvents" Farhad Gharagheizi &Reza Fareghi Alamdari, Fullerenes, Nanotubes and Carbon Nanostructures, Volume 16, 2008 - Issue 1**
	Printing Date: **22-Apr-2019**

1.QSAR identifier

1.1.QSAR identifier (title):

Linear nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents. The model has been presented in the publication "A Molecular-Based Model for Prediction of Solubility of C60 Fullerene in Various Solvents" Farhad Gharagheizi & Reza Fareghi Alamdari, Fullerenes, Nanotubes and Carbon Nanostructures, Volume 16, 2008 - Issue 1 .

1.2.Other related models:

Neural Network nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents.

1.3.Software coding the model:

Jaqpot

Jaqpot is q web platform that support development, validation and sharing of QSAR models

Haralambos Sarimveis

apps.jaqpot.org

2.General information

2.1.Date of QMRF:

21 April 2019

2.2.QMRF author(s) and contact details:

Haralambos Sarimveis National Technical University of Athens hsarimv@central.ntua.gr

https://www.chemeng.ntua.gr/labs/control_lab/sarimveis.html

2.3.Date of QMRF update(s):

22 July 2019

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

[1]Farhad Gharagheizi Department of Chemical Engineering, Faculty of Engineering, University of Tehran, Tehran, Iran fghara@ut.ac.ir

[2]Reza Fareghi Alamdari Department of Chemistry, Faculty of Materials and Chemical Engineering, Malek-Ashtar University of Technology, Lavizan, Tehran, Iran

2.6.Date of model development and/or publication:

"A Molecular-Based Model for Prediction of Solubility of C60 Fullerene in Various Solvents" Farhad Gharagheizi &Reza Fareghi Alamdari, Fullerenes, Nanotubes and Carbon Nanostructures, Volume 16, 2008 - Issue 1.

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

2.8.Availability of information about the model:

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

3.Defining the endpoint - OECD Principle 1

3.1.Species:

3.2.Endpoint:

P-CHEM 4.9. Solubility in organic solvents

3.3.Comment on endpoint:

The solubility values are not given in weight units (e.g., mg/mL) but in terms of logarithmic values of molar fractions log(S) because the log(S) values correspond to the Gibbs free energy changes in the solvation process

3.4.Endpoint units:

Logarithmic values of molar fractions log(S)

3.5.Dependent variable:

Logarithmic values of molar fractions log(S)

3.6.Experimental protocol:

3.7.Endpoint data quality and variability:

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

nanoQSAR linear model

4.2.Explicit algorithm:

Multiple linear regression (MLR)

4.3.Descriptors in the model:

[1]piPC03 Molecular multiple path count of order 03

[2]ATS1m Broto-Mreau autocorrelation of a topological structure-lag 1/weighted by atomic masses

[3]SEigp Eigenvalue sum from polarizability weighted distance matrix

[4]More23e 3D-MORSE-signal 23/weighted by atomic sanderson electronegativities

[5]H1m H autocorrelation of lag 1/weighted by atomic masses

4.4.Descriptor selection:

The GA-MLR algorithm proposed by Leardi et al.

4.5.Algorithm and descriptor generation:

The full set contains 1664 molecular descriptors. After calculating molecular descriptors, the pool of molecular descriptors was reduced by removing descriptors that could not be calculated for every structure in the dataset, and those descriptors with an essentially constant value or all the structures. In this step, the pool of 1664 molecular descriptors reduced to a new pool of 1259 molecular descriptors. The GA-MLR algorithm was applied on this set of descriptors

4.6.Software name and version for descriptor generation:

Dragon

Dragon 7.0 calculates 5,270 molecular descriptors, organized in different logical blocks as in the previous versions. Blocks are further divided into sub-blocks to make management, selection, and analysis of descriptors easier.

chm@kode-solutions.net

https://chm.kode-solutions.net/products_dragon.php

4.7.Chemicals/Descriptors ratio:

128/5

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

5.2.Method used to assess the applicability domain:

5.3.Software name and version for applicability domain assessment:

5.4.Limits of applicability:

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS RN: No

Chemical Name: Yes

Smiles: No

Formula: No

INChI: No

MOL file: No

NanoMaterial: No

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:

4 solvents of the original dataset were removed because they are considered as outliers in the original paper

6.6.Pre-processing of data before modelling:

The data are scaled in the range [0,1]

6.7.Statistics for goodness-of-fit:

R2 score: 0.899

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

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

R2 scores for 5-fold cross validation: 0.91906039, 0.88995619, 0.90445436, 0.86506266, 0.62316459

6.10.Robustness - Statistics obtained by Y-scrambling:

6.11.Robustness - Statistics obtained by bootstrap:

6.12.Robustness - Statistics obtained by other methods:

External validation: R2_external: 0.904

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: No

Chemical Name: Yes

Smiles: No

Formula: No

INChI: No

MOL file: No

NanoMaterial: No

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:

The external validation set consists of 25% of the original data selected randomly

7.6.Experimental design of test set:

By randomly setting aside 25% of chemicals in the training data

7.7.Predictivity - Statistics obtained by external validation:

External validation: R2_external: 0.904

7.8.Predictivity - Assessment of the external validation set:

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:

8.2.A priori or a posteriori mechanistic interpretation:

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

9.2.Bibliography:

9.3.Supporting information:

Training set(s)

70_model_reduced.csv	https://app.jaqpot.org/dataset/cjCXljkX0kBbjelkHMZuNg

Test set(s)Supporting information

10.Summary (JRC QSAR Model Database)

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