	QMRF identifier (JRC Inventory): **To be entered by JRC**
	QMRF Title: **ALog P model v. 1.0.1**
	Printing Date: 30/05/2022

1.QSAR identifier

1.1.QSAR identifier (title):

ALog P model v. 1.0.1

1.2.Other related models:

NA

1.3.Software coding the model:

VEGA (https://www.vegahub.eu/)

The VEGA software provides QSAR models to predict tox, ecotox, environ, phys-chem and toxicokinetic properties of chemical substances.

emilio.benfenati@marionegri.it

2.General information

2.1.Date of QMRF:

May 2022

2.2.QMRF author(s) and contact details:

[1]Domenico Gadaleta Laboratory of Environmental Chemistry and Toxicology, Istituto di Ricerche Farmacologiche "Mario Negri", IRCCS domenico.gadaleta@marionegri.it

[2]Emilio Benfenati Laboratory of Environmental Chemistry and Toxicology, Istituto di Ricerche Farmacologiche "Mario Negri", IRCCS emilio.benfenati@marionegri.it

2.3.Date of QMRF update(s):

NA

2.4.QMRF update(s):

NA

2.5.Model developer(s) and contact details:

[[1]Arup K. Ghose Department of Molecular Structure and Design, Amgen Inc., 1840 DeHaVillandDriVe, Thousand Oaks, California 91320

[2]Vellarkad N. Viswanadhan Department of Molecular Structure and Design, Amgen Inc., 1840DeHaVilland DriVe, Thousand Oaks, California 91320

[3]Gordon M. Crippen Department of Molecular Structure and Design, Amgen Inc., 1840DeHaVilland DriVe, Thousand Oaks, California 91320

[4]Alberto Manganaro Kode srl info@kode-solutions.net

2.6.Date of model development and/or publication:

1998

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

[1] A.K. Ghose, V.N. Viswanadhan, J.J. Wendoloski, J.Phys.Chem. A 1998, 102, 3762-3772

[2] Benfenati E, Manganaro A, Gini G. VEGA-QSAR: AI inside a platform for predictive toxicology Proceedings of the workshop "Popularize Artificial Intelligence 2013", December 5th 2013, Turin, Italy Published on CEUR Workshop Proceedings Vol-1107

2.8.Availability of information about the model:

The model is non-proprietary and the training set is available.

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

NA

3.Defining the endpoint - OECD Principle 1

3.1.Species:

NA

3.2.Endpoint:

QMRF 1. 6. Octanol-water partition coefficient (Kow) EC A.8 Partition Coefficient (EU methodincludes both shake flask and HPLC)

3.3.Comment on endpoint:

NA

3.4.Endpoint units:

Adimensional

3.5.Dependent variable:

Logarithm of octanol/water partition coefficient (log P)

3.6.Experimental protocol:

EC A.8 Partition Coefficient OECD 123 Partition Coefficient (nOctanol/Water): Slow-Stirring Method

OECD 117 Partition Coefficient (n-octanol/water) HPLC Method

OECD 107 Partition Coefficient (noctanol/water); Shake Flask Method

3.7.Endpoint data quality and variability:

The dataset was collected mainly from Star list Database; Biobyte Corporation: Pomona, CA, 1997

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

Regression equations

4.2.Explicit algorithm:

Regression equation based on the hydrophobicity contribution of 120 atom typesThe AlogP models in VEGA 1.4.4 implement the Ghose-Crippen-Viswanadhan LogP (ALogP) and consists of a regression equation based on the hydrophobicity contribution of 120 atom types. The classification of the atoms is made to differentiate (i) the electron distribution around the atom, and (ii) the approachability of the solvent to the atom. Each atom in every structure is classified into one of the 120 atom types. Then, estimated logP for any compound is given by: AlogP = Sumi(n i a i) where n i is the number of atom of type i and a i is the corresponding hydrophobicity constant..The AlogP model implemented in Dragon wasevaluated by the aid of a set of 3568 compounds with known experimental logP taken from the NCI Open DataBase. The resulted determination coefficient r2 was 0.931. Moreover, on our internal logP data set comprised of 9834 compounds the correlation coefficient rbetween experimental and calculated logP was 0.932

4.3.Descriptors in the model:

The contribution of the 120 atom types. Five, unused atom types are not listed.

4.4.Descriptor selection:

NA

4.5.Algorithm and descriptor generation:

NA

4.6.Software name and version for descriptor generation:

NA

4.7.Chemicals/Descriptors ratio:

8364/120 = 69.7

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The Applicability Domain (AD) is assessed using the original algorithm implemented within VEGA. An overall AD index is calculated, based on a number of parameters, which relate to the results obtained on similar chemicals within the training and test sets.

5.2.Method used to assess the applicability domain:

The chemical similarity is measured with the algorithm developed for VEGA. Full details are in the VEGA website (www.vegahub.eu), including the open access paper describing it. The AD also evaluates the correctness of the prediction on similar compounds (accuracy), the consistency between the predicted value for the target compound and the experimental values of the similar compounds, the range of the descriptors, and the presence of unusual fragments, using atom centred fragments.

5.3.Software name and version for applicability domain assessment:

VEGA

Included in the VEGA software and automatically displayed when running the model

emilio.benfenati@marionegri.it

https://www.vegahub.eu/

5.4.Limits of applicability:

The model is not applicable to inorganic chemicals and substances containing unusual elements (i.e., different from C, O, N, S, P, Cl, Br, F, I). Salts can be predicted only if converted to the neutralized form.

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: No

Smiles: Yes

Formula: No

INChI: No

MOL file: 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:

NA

6.6.Pre-processing of data before modelling:

NA

6.7.Statistics for goodness-of-fit:

The model coefficients are taken from Ghose et al, J.Phys.Chem. A 1998, 102, 3762-3772. They were estimated on the basis of a training set of 8364 compounds. The statistical parameters of the AlogP model are: r = 0.95; s = 0.55; predictive r 2= 0.90

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

NA

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

NA

6.10.Robustness - Statistics obtained by Y-scrambling:

NA

6.11.Robustness - Statistics obtained by bootstrap:

NA

6.12.Robustness - Statistics obtained by other methods:

NA

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: No

Smiles: Yes

Formula: No

INChI: No

MOL file: No

7.3.Data for each descriptor variable for the external validation set:

No

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 counts 726 compounds

7.6.Experimental design of test set:

The training set of the Meylan LogP model (9,961 compounds) from from EPI Suite KowWin module was used as test set during the implementation

7.7.Predictivity - Statistics obtained by external validation:

Viswanadhan et al., 1993 reports the following statistics obtained on a dataset of 47 nucleosides and bases:

n = 47; r = 0.842; SD = 0.51

Ghose et al., 1998 reports the following statistics in validation:

n = 931; r = 0.95; s = 0.55; predictive r 2= 0.90.

On the pruned training set from EPI Suite KowWin module (9,961 compounds), the logP model has the following statistics: n = 9961; R2 = 0.84; RMSE = 0.72

7.8.Predictivity - Assessment of the external validation set:

NA

7.9.Comments on the external validation of the model:

NA

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

The classification of the atoms is made to differentiate (i) the electron distribution around the atom, and (ii) the approachability of the solvent to the atom

8.2.A priori or a posteriori mechanistic interpretation:

A priori

8.3.Other information about the mechanistic interpretation:

NA

9.Miscellaneous information

9.1.Comments:

NA

9.2.Bibliography:

[1]A.K. Ghose and G.M. Crippen, J. Comput. Chem. 1986, 7, 565-577

[2]V.N. Viswanadhan et al., J. Comput. Chem. 1993,14, 1019-1026

[3]A.K. Ghose, V.N. Viswanadhan, J.J. Wendoloski, J.Phys.Chem. A 1998, 102, 3762-3772

9.3.Supporting information:

Training set(s)Test set(s)Supporting information:

All available dataset are present in the model inside the VEGA software.

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