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

1.QSAR identifier

1.1.QSAR identifier (title):

MLog P model v. 1.0.1

1.2.Other related models:

No

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 Istituto di Ricerche Farmacologiche Mario Negri - IRCSS Via Mario Negri 2, 20156 Milano, Italy emilio.benfenati@marionegri.it https://www.marionegri.it/

2.3.Date of QMRF update(s):

No update

2.4.QMRF update(s):

NA

2.5.Model developer(s) and contact details:

[1] Alberto Manganaro Istituto di Ricerche Farmacologiche Mario Negri - IRCSS Via Mario Negri 2, 20156 Milano, Italy alberto.manganaro@marionegri.it https://www.marionegri.it/

[2] Ikuo Moriguchi School of Pharmaceutical Sciences, Kitasato University, 5-9-1 Shirokane, Minato-ku,Tokyo 108, Japan.

2.6.Date of model development and/or publication:

1992

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

[1] I.Moriguchi, S.Hirono, Q.Liu, I.Nakagome, and Y.Matsushita, Chem.Pharm.Bull. 1992, 40, 127-130

[2] I.Moriguchi, S.Hirono, I.Nakagome, H.Hirano, Chem.Pharm.Bull. 1994, 42, 976-978.

[3] Benfenati E, Roncaglioni A, Lombardo A, Manganaro A. Integrating QSAR, Read-Across, and Screening Tools: The VEGAHUB Platform as an Example. Advances in Computational Toxicology; Springer; 2019. p. 365-81.

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:

Another QMRF is not available.

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 method includes 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:

NA

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

Regression equation based on structural parameters

4.2.Explicit algorithm:

Linear regression

The MlogP models in VEGA 1.4.4 implement the multiple linear regression developed by Morigouchi et al. (1992; 1995) that relates 13 structural parameters with the experimental log P values of 1230 compounds with different structures and including C, H, N, O, S, P, F, Cl, Br.

The equation is the following: MlogP = -1.041 + 1.244(CX) 0.6- 1.017(NO) 0.9 + 0.406(PRX) - 0.145(UB) 0.8+ 0.511(HB) + 0.268(POL) - 2.215(AMP) + 0.912(ALK) -0.392(RNG) -3.684(QN) + 0.474(NO2) + 1.582(NCS)+ 0.773(BLM)

The model variables are frequencies (denoted by N) or presence/absence (denoted by D) of some molecular features. Their description is reported in the table below:

C (Summation of weighted numbers of carbon and halogen atoms; the weights are: 0.5 fro F, 1.0 for C and Cl, 1.5 for Br, and 2.0 for I)

NO (Total number of Ns and Os)

PR (Proximity effects of N/O: 2 for X-Y and 1 X-A-Y. X.Y:N and/or O; A: C, S, or P; -: saturated or unsaturated bond) with a correction (-1) for -CON < and -SO2N <)

U (Number of unsaturated bonds including semi-polar bonds such as N-oxides and sulfoxides, except those in NO2=)

H (Dummy variable for the presence of intramolecular hydrogen bond as ortho-OH and -CO-R, -OH and -NH2, -NH2 and -COOH, or 8-OH/NH2 in quinolines, 5 or 8-OH/NH2 in quinoxalines, etc.)

POL (Number of aromatic polar substituents (aromatic substituents excluding Ar-C(X)(Y)- and Ar-C(X)=C; X, Y: C and/or H). Upper limit = 4)

AMP (Amphoteric property; a-aminoacid = 1, aminobenzoic acid = 0.5, pyridinecarboxylic acid = 0.5).

ALK (Dummy variable for alkane, alkene, cycloalkane, cycloalkene (hydrocarbons with 0 or 1 double bond) or hydrocarbon chain with at least 7 carbon atoms).

RNG (Dummy variable for the presence of ring structures except benzene and its condensed rings (aromatic, heteroaromatic, and hydrocarbon rings)

QN (Quaternary nitrogen >N+ <: 1; N-oxide: 0.5).

NO2 (Number of nitro groups).

NCS (Isothiocyanate (-N=C=S): 1.0; thiocyanate (-S-C#N): 0.5)

BLM (Dummy variable for the presence of ß-lactam)

4.3.Descriptors in the model:

The model variables are frequencies (denoted by N) or presence/absence (denoted by D) of some molecular features. Their description is reported in the table above (see 4.2)

4.4.Descriptor selection:

Details on the selection procedure of the 13 parameters and on the development of the multiple linear regression are reported in Morigouchi et al., 1992

4.5.Algorithm and descriptor generation:

See 4.2

4.6.Software name and version for descriptor generation:

NA

4.7.Chemicals/Descriptors ratio:

1200/12 = 92.3

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 (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: Yes

Chemical Name: No

Smiles: Yes

Formula: No

INChI: No

MOL file: No

6.3.Data for each descriptor variable for the training set:

NA

6.4.Data for the dependent variable for the training set:

NA

6.5.Other information about the training set:

The 1200 compounds used by Morigouchi et al. to derive the multiple linear regression were cited from Hansh & Leo, “Substituent Constant for Correlation Anaysis in Chemistry and Biology”, John Wiley and Sons, New York, 1979.The training set of the Meylan LogP model (9,961 compounds) was used as training set during the implementation.

6.6.Pre-processing of data before modelling:

NA

6.7.Statistics for goodness-of-fit:

Performances on the training set are reported in Morigouchi et al., 1992:

n = 1230, r = 0.952, s = 0.411, F 0(13,1216) = 900.4

On the pruned training set from EPI Suite KowWin module (9,961 compounds), the logP model has the following statistics:

Test set: n = 9961; R2 = 0.73; RMSE = 0.96

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:

NA

7.2.Available information for the external validation set:

NA

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

NA

7.4.Data for the dependent variable for the external validation set:

NA

7.5.Other information about the external validation set:

NA

7.6.Experimental design of test set:

NA

7.7.Predictivity - Statistics obtained by external validation:

Test set: R2 = 0.84; RMSE = 0.34.

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 two main parameters in the equation are CX (i.e., the summation of empirically weighted numbers of carbon and halogen atoms) accounting the hydrophobic contribution to log P, and NO (i.e. the total number of nitrogen and oxygen atoms) accounting for the hydrophilic contribution to log P. The proximity effect of nitrogen and oxygen atoms was also considered important as a correction for the electronic structures and implemented with PRX parameter. The remaining parameters account for the effect of various substructures.

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] Moriguchi, S.Hirono, Q.Liu, I.Nakagome, and Y.Matsushita, Chem.Pharm.Bull. 1992, 40, 127-130.

[2] I.Moriguchi, S.Hirono, I.Nakagome, H.Hirano, Chem.Pharm.Bull. 1994, 42, 976-978.

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