	QMRF identifier (JRC Inventory): **Q17-24a-0010**
	QMRF Title: **VEGA BCF model (Meylan)**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

VEGA BCF model (Meylan)

1.2.Other related models:

The model is based on the method proposed by Meylan et al. as implemented in the EPI Suite BCFBAF module (http://www.epa.gov/oppt/exposure/pubs/episuite.htm)

1.3.Software coding the model:

BCF model (Meylan) v.1.0.3

The model provides a quantitative prediction of bioconcentration factor (BCF) in fish.

http://www.vega-qsar.eu/

2.General information

2.1.Date of QMRF:

01/06/2016

2.2.QMRF author(s) and contact details:

Emilio Benfenati IRCCS - Istituto di Ricerche Farmacologiche Mario Negri

emilio.benfenati@marionegri.it

2.3.Date of QMRF update(s):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

Alberto Manganaro Kode srl info@kode-solutions.net

2.6.Date of model development and/or publication:

The model was published in 2013.

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

Meylan,W.M.,Howard,P.H.,Boethling,R.S.,1999.Improved method for estimating bioconcentration/bioaccumulation factor from octanol/water partition coefficient. Environ.Toxicol.Chem.18,664–672.

2.8.Availability of information about the model:

The model has been released open source and is freely available through the portal of the VEGA platform (www.vega-qsar.eu). The training and test set are available (see 9.3).

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

Other QMRF for this model are not available

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Fish

3.2.Endpoint:

2.Environmental fate parameters 2.4.a.Bioconcentration . BCF fish

3.3.Comment on endpoint:

The bioconcentration factor (BCF) is the concentration of test substance in/on the fish or specified tissues thereof divided by the concentration of the chemical in the surrounding medium at steady state.

3.4.Endpoint units:

Continuous value expressed in Log(L/kg)

3.5.Dependent variable:

Log P (see 4.2)

3.6.Experimental protocol:

Bioconcentration factor (BCF) data are provided from tests that are conducted with respect to the OECD Test No. 305: Bioaccumulation in Fish. A procedure for characterising the bioconcentration potential of substances in fish.

3.7.Endpoint data quality and variability:

Data as from U. S. Environmental Protection Agency website (http://esc.syrres.com/interkow/EpiSuiteData.htm) were used.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR model with different regression equations or fixed values, selected on the basis of an initial classification between ionic and non-ionic compounds, and on the value of the predicted logP value corrected with factors depending on the presence/absence of 12 fragments.

4.2.Explicit algorithm:

Meylan BCF methodology

Compounds are classified as either ionic or non-ionic. Ionic compounds include carboxylic acids, sulfonic acids and salts of sulfonic acids, and charged nitrogen compounds (nitrogen with a valence such as quaternary ammonium compounds). All other compounds are classified as non-ionic.Methodology forNon-Ionic was to separate compounds into

three divisions by Log Kow value as follows:

Log Kow < 1.0

Log Kow 1.0 to 7.0

Log Kow > 7.0

For Log Kow 1.0 to 7.0 the derived QSAR estimation equation is:

Log BCF = 0.6598 Log Kow - 0.333 + correction factors

For Log Kow > 7.0 the derived QSAR estimation equation is:

Log BCF = -0.49 Log Kow + 7.554 + correction factorsFor Log Kow < 1.0 the derived QSAR estimation equation is:

All compounds with a log Kow of less than 1.0 are assigned an estimated log BCF of 0.50Ionic compoundsare predicted as follows: log BCF = 0.50 (log Kow < 5.0) log BCF = 1.00 (log Kow 5.0 to 6.0) log BCF = 1.75 (log Kow 6.0 to 8.0)

log BCF = 1.00 (log Kow 8.0 to 9.0)

log BCF = 0.50 (log Kow > 9.0)

The correction factors and their values are the following:Ketone (aromatic connection) -0.5851

Phosphate ester -0.825

Multi-halogenated biphenyl/PAH 0.586
Aromatic ring-CH-OH -0.2556Aromatic sym-triazine ring -0.5169
Tert-Butyl ortho-phenol type -0.222
Phenanthrene ring 0.6609
Cyclopropyl-C(=O)-O- ester -1.2591Alkyl chains (8+ CH2 groups) with logKow >4&<7.0 -1.3743
Alkyl chains (8+ CH2 groups)with logKow 7-10 -0.5965Disulfide (-S-S-) -1.3404

4.3.Descriptors in the model:

logKow (logP) logP prediction based on the original Meylan approach (available in the EPI Sutie application) as re-implemented in VEGA

4.4.Descriptor selection:

No selection applied.

4.5.Algorithm and descriptor generation:

The model is based on fragments to define different chemical classes. Different models apply to different classes. Log P is the descriptor used within each model to separate chemical classes.

4.6.Software name and version for descriptor generation:

4.7.Chemicals/Descriptors ratio:

Only one descriptor (log P) is used.

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The applicability domain of predictions is assessed using an Applicability Domain Index (ADI) that has values from 0 (worst case) to 1 (best case).

5.2.Method used to assess the applicability domain:

The ADI is calculated by grouping several other indices, each one taking into account a particular issue of the applicability domain. Most of the indices are based on the calculation of the most similar compounds found in the training and test set of the model, calculated by a similarity index that consider molecule's fingerprint and structural aspects (count of atoms, rings and relevant fragments). Note that when the experimental value for the given compound is found in the database of the model, the Applicability Domain indices are calculated only considering this value, without taking into account the first n similar compounds.

5.3.Software name and version for applicability domain assessment:

BCF model (Meylan) v.1.0.3
This is included in the stand alone version of VEGA: VEGANic v 1.1.1
http://www.vega-qsar.eu/

5.4.Limits of applicability:

Thresholds of the applicability domain index are given for all applicability domain components with their explanation and the intervals used:
- Similar molecules with known experimental value. This index takes into account how similar are the first two most similar compounds found. Values near 1 mean that the predicted compound is well represented in the dataset used to build the model, otherwise the prediction could be an extrapolation.

Defined intervals for the applicability are:
1 >= index > 0.9 : strongly similar compounds with known experimental value in the training set have been found
0.9 >= index > 0.75 : only moderately similar compounds with known experimental value in the training set have been found
index <= 0.75 : no similar compounds with known experimental value in the training set have been found
- Accuracy (average error) of prediction for similar molecules. This index takes into account the error in prediction for the two most similar compounds found. Values near 0 mean that the predicted compounds falls in an area of the model's space where the model gives reliable predictions, otherwise the greater is the value, the worse the model behaves.
Defined intervals for the accuracy are:
index < 0.5 : accuracy of prediction for similar molecules found in the training set is good
0.5 <= index <= 1.0 : accuracy of prediction for similar molecules found in the training set is not optimal
index > 1.0 : accuracy of prediction for similar molecules found in the training set is not adequate
- Concordance with similar molecules (average difference between target compound prediction and experimental values of similar molecules) .and experimental values of similar molecules) . This index takes into account the difference between the predicted value and the experimental values of the two most similar compounds. Values near 0 mean that the prediction made agrees with the experimental values found in the model's space, thus the prediction is reliable.
Defined intervals for concordance are:
index < 0.5 : similar molecules found in the training set have experimental values that agree with the target compound predicted value
0.5 <= index <= 1.0 : similar molecules found in the training set have experimental values that slightly disagree with the target compound predicted value
index > 1.0 : similar molecules found in the training set have experimental values that completely disagree with the target compound predicted value
- Maximum error of prediction among similar molecules. This index takes into account the maximum error in prediction among the two most similar compounds. Values near 0 means that the predicted compounds falls in an area of the model's space where the model gives reliable predictions without any outlier value.
Defined intervals for the maximum error are:
index < 0.5 : the maximum error in prediction of similar molecules found in the training set has a low value, considering the experimental variability
0.5 <= index <= 1.0 : the maximum error in prediction of similar molecules found in the training set has a moderate value, considering the experimental variability
index > 1.0 : the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability
- LogP reliability. This index takes into account the reliability of the logP value used in the model. Note that the Meylan BCF model is strongly based on the logP prediction of the compound, thus this index is highly relevant for the assessment of the final prediction. The reliability of the logP value comes from the assessment of the VEGA LogP model (that provides the used logP value), which is also provided in the “Prediction summary” section of the report.
Defined intervals for Log P reliability are:
index = 1 : reliability of logP value used by the model is good

index = 0.7 : reliability of logP value used by the model is not optimal

index = 0 : reliability of logP value used by the model is not adequate

- Model descriptors range check. This index checks if the descriptors calculated for the predicted compound are inside the range of descriptors of the training and test set. The index has value 1 if all descriptors are inside the range, 0 if at least one descriptor is out of the range.

Defined intervals for the descriptors range check are:

index = True : descriptors for this compound have values inside the descriptor range of the compounds of the training set

index = False : descriptors for this compound have values outside the descriptor range of the compounds of the training set

- Global AD Index. The final global index takes into account all the previous indices, in order to give a general global assessment on the applicability domain for the predicted compound.

Defined intervals for the global ADI are:

1 >= index > 0.85 : predicted substance is into the Applicability Domain of the model

0.85 >= index > 0.75 : predicted substance could be out of the Applicability Domain of the model

index <= 0.75 : predicted substance is out of the the Applicability Domain of the model

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: 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:

-

6.6.Pre-processing of data before modelling:

The original dataset from EPI Suite has been taken, then processed and cleared from duplicates and compounds provided with structure that had problems. The final dataset has 662 compounds.

6.7.Statistics for goodness-of-fit:

Training set: n = 516; R2 = 0.80; RMSE = 0.55

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

-

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

-

6.10.Robustness - Statistics obtained by Y-scrambling:

-

6.11.Robustness - Statistics obtained by bootstrap:

-

6.12.Robustness - Statistics obtained by other methods:

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:

No

7.5.Other information about the external validation set:

To test the model a test set of 148 compounds was used.

7.6.Experimental design of test set:

7.7.Predictivity - Statistics obtained by external validation:

External validation set: n =148; R2 = 0.395; RMSE = 0.914. Data with ADI > 0.85 : n =76; R2 = 0.775; RMSE = 0.425 M. I.Petoumenou, F. Pizzo, J. Cester, A. Fernndez, E. Benfenati, "Comparison between bioconcentration factor(BCF) data provided by industry to the European Chemicals Agency (ECHA) and data derived from QSAR models", Environmental Research 142(2015) pages: 529 - 534.

7.8.Predictivity - Assessment of the external validation set:

The predictivity of the model is better when compounds fall within the ADI of the model.

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:

Log P is considered to keep into account the basis of the transfer from water to the lipidic phase of the cell.

8.2.A priori or a posteriori mechanistic interpretation:

A priori

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

9.2.Bibliography:

Meylan,W.M.,Howard,P.H.,Boethling,R.S.,1999.Improved method for estimating bioconcentration/bioaccumulation factor from octanol/water partition coefficient. Environ.Toxicol.Chem.18,664–672.

9.3.Supporting information:

dataset_BCF_MEYLAN.txt	http://qsardb.jrc.ec.europa.eu/qmrffile:///C:\Users\MPetoumenou\Desktop\dataset_BCF_MEYLAN.txt

Test set(s)Supporting information

10.Summary (JRC Inventory)

10.1.QMRF number:

Q17-24a-0010

10.2.Publication date:

2017-09-21

10.3.Keywords:

To be entered by JRC;

10.4.Comments:

To be entered by JRC