	*QMRF identifier (JRC Inventory):* **To be entered by JRC**
	QMRF Title: **Adipose tissue:blood model (INERIS) - v. 1.0.0**
	Printing Date: **20-10-2020**

1.QSAR identifier

1.1.QSAR identifier (title):

Adipose tissue:blood model (INERIS) - v. 1.0.0

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:

07-10-2020

2.2.QMRF author(s) and contact details:

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):

No update

2.5.Model developer(s) and contact details:

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

2.6.Date of model development and/or publication:

2019

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

Cappelli CI, Manganelli,Toma C, Emilio Benfenati E,Mombelli E Prediction of the Partition Coefficient between Adipose Tissue and Blood for Environmental Chemicals: from Single QSAR Models to an Integrated Approach Mol. Inf. 2020, 39, 2000072, DOI: 10.1002/minf.202000072

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:

Rat

3.2.Endpoint:

TOX Adipose tissue:blood partition coefficient (K_{AB}).

3.3.Comment on endpoint:

The adipose tissue:blood partition coefficient (K_{AB}) is a key-endpoint to predict the bioaccumulation and the pharmacokinetics in humans and animals, since other organ:blood affinities can be estimated as a function of this parameter

3.4.Endpoint units:

K_{AB}

3.5.Dependent variable:

Log (K_{AB}).

3.6.Experimental protocol:

OECD guideline for this endpoint is OECD TG 417.

3.7.Endpoint data quality and variability:

NA

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

Model is based on Random Forest (RF) approach.

4.2.Explicit algorithm:

Machine Learning Algorithm for Regression

The number of trees selected for the RF were finely tuned by identifying the onset of the plateau of the curve describing the Q2LOO as a function of the number of trees.

4.3.Descriptors in the model:

PaDEL-Descriptors v. 2.21 PaDEL computes the 2D descriptors.

4.4.Descriptor selection:

"VSURF" v. 1.0.4 package was used.

4.5.Algorithm and descriptor generation:

Random Forest (RF)

4.6.Software name and version for descriptor generation:

The "randomForest" v. 4.6.14 in R v. 3.4.3

The "randomForest" package provides an R interface to the Fortran programs by Breiman and Cutlert (http://www.stat.berkeley.edu/users/breiman/).

4.7.Chemicals/Descriptors ratio:

63/6 = 6.05

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

NanoMaterial: 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:

Training set: n = 63

6.6.Pre-processing of data before modelling:

The RF model using the package “randomForest” v. 4.6.14 in R v. 3.4.3. The RF model was calibrated by using unscaled descriptors. The number of trees selected for the RF was finely tuned by identifying the onset of the plateau of the curve describing the Q2LOO as a function of the number of trees.

6.7.Statistics for goodness-of-fit:

We used only the leave-one-out approach

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

Q2 LOO = 0.73

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

Q2 LSO = 0.72

6.10.Robustness - Statistics obtained by Y-scrambling:

Scrambled Q2 LSO = - 0.17

6.11.Robustness - Statistics obtained by bootstrap:

NA

6.12.Robustness - Statistics obtained by other methods:

MAE LOO = 0.41

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: No

Smiles: Yes

Formula: No

INChI: No

MOL file: No

NanoMaterial: No

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:

Training set: n = 38

7.6.Experimental design of test set:

Test set relied on comparable proportions of acidic, basic, zwitterions and neutral chemicals between the two sets training and test sets. Number of neighbours (k) and Sahigara's thresholds (t) optimized on the training set by setting a minimum acceptable coverage to 70%.

7.7.Predictivity - Statistics obtained by external validation:

Q2 F3 = 0.87

7.8.Predictivity - Assessment of the external validation set:

MAE ext = 0.24

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:

No assumption on the mechanism is done.

8.2.A priori or a posteriori mechanistic interpretation:

NA

8.3.Other information about the mechanistic interpretation:

NA

9.Miscellaneous information

9.1.Comments:

NA

9.2.Bibliography:

[1] Cappelli CI, Manganelli,Toma C, Emilio Benfenati E,Mombelli E
Prediction of the Partition Coefficient between Adipose Tissue and Blood for Environmental Chemicals: from Single QSAR Models to an Integrated Approach Mol. Inf. 2020, 39, 2000072, DOI: 10.1002/minf.202000072

[2]Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32. https://link.springer.com/article/10.1023/A:1010933404324

[3]Genuer, R., Poggi, J. M., &Tuleau-Malot, C. (2018). VSURF: Variable Selection Using Random Forests. R package version 1.0.4. URL https://CRAN.R-project.org/package=VSURF https://journal.r-project.org/archive/2015/RJ-2015-018/RJ-2015-018.pdf

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