	***QMRF identifier (JRC Inventory):*** **To be entered by JRC**
	QMRF Title: **Skin permeation (LogKp) model (ten Berge) v-1.0.1**
	Printing Date: November 2022

1.QSAR identifier

1.1.QSAR identifier (title):

Skin permeation (LogKp) model (ten Berge) 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:

November 2022

2.2.QMRF author(s) and contact details:

[1]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]Gianluca Selvestrel Istituto di Ricerche farmacologiche Mario negri - IRCCSgianluca.selvestrel@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] 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:

NA

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

[1] Wil ten Berge, A simple dermal absorption model: Derivation and application. 2009, 75, 1440–1445 https://www.sciencedirect.com/science/article/abs/pii/S004565350900232X

[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:

human skin in vitro from aqueous solutions

3.2.Endpoint:

TOX Skin permeability coefficient (PBPK)

3.3.Comment on endpoint:

The rate of a chemical penetrating across the skin.

3.4.Endpoint units:

cm/h

3.5.Dependent variable:

LogKp

3.6.Experimental protocol:

Based on the OECD 428 test guideline. [2]

3.7.Endpoint data quality and variability:

The model is based on a dataset of 271 compounds. Following the criteria reported in the OECD guideline 428, only data obtained in compliance with the following features have been kept:

All the data are retrieved from “in vitro” experiments

Data are collected from human skin experiments

Studies concerned skin application of chemicals dissolved in water, aqueous solution, water gel, PBS and distilled water

The buffer solution at a pH of 7.4

The permeation coefficients were measured under comparable circumstances.

The model is an application of the ten Berge equation to the entire dataset. For this reason, a splitting into training and test set is not provided.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

PBK/D QSAR

4.2.Explicit algorithm:

Physiological based kinetic/dynamic based parameter QSAR

The QSAR model used the empirical parameters and equation to derive the model

LogKp sk-water=Log[1/(1/Klip+Kpol)+1/(1/Kaq)]

Where: Klip = permeation coefficient lipid medium

Kpol = permeation coefficient corneocytes [proteins]

Kaq = permeation coefficient epidermis [aqueous]

Kow = octanol/water partition coefficient.

M w = molecular weight.

b1, b2, b3, b4, b5, b6, b7 = regression coefficients:

b1 = -2,694 b2 = 0,9809 b3 = -7,868*10 -3 b4 = 0,05523 b5 = 1,383 b6 = 1,121*10 3 b7 = 1,957

4.3.Descriptors in the model:

[1]Kpol = permeation coefficient corneocytes [proteins]

[2]K aq = permeation coefficient epidermis [aqueous]

[3]K ow = octanol/water partition coefficient.

[4]M w = molecular weight

4.4.Descriptor selection:

NA

4.5.Algorithm and descriptor generation:

NA

4.6.Software name and version for descriptor generation:

Iterative non-linear Gauss–Newton least-squares fit Non linear regression toolhttps://www.sciencedirect.com/science/article/abs/pii/0045653595000232?via%3Dihub

4.7.Chemicals/Descriptors ratio:

NA

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.

ADI is defined in this way for this QSAR model´s predictions:

If 1 ≥ AD index > 0.85, the predicted substance is regarded in the Applicability Domain of the model. It corresponds to “good reliability” of prediction.

If 0.85 ≥ AD index > 0.7, the predicted substance could be out of the Applicability Domain of the model. It corresponds to “moderate reliability” of prediction.

If AD index ≤ 0.7, the predicted substance is regarded out of the Applicability Domain of the model. It corresponds to “low reliability” of prediction.

5.2.Method used to assess the applicability domain:

The AD and 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 [3]. 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.

These indices are defined in this way for this QSAR model:

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 are:

If 1 ≥ index > 0.85, strongly similar compounds with known experimental value in the training set have been found

If 0.85 ≥ index > 0.7, only moderately similar compounds with known experimental value in the training set have been found

If index ≤ 0.7, 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 classification accuracy in prediction for the two most similar compounds found. Values near 1 mean that the predicted compounds fall in an area of the model's space where the

model gives reliable predictions (no misclassifications), otherwise the lower is the value, the worse the model behaves. Defined intervals are:

If index < 0.6, accuracy of prediction for similar molecules found in the training set is good

If 1.2 > index ≥ 0.6, accuracy of prediction for similar molecules found in the training set is not optimal

If index ≥ 1.2, accuracy of prediction for similar molecules found in the training set is not adequate

Concordance for 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 disagrees with the values found in the model's space, thus the prediction could be unreliable. Defined intervals are:

If index < 0.6, molecules found in the training set have experimental values that agree with the target compound predicted value

If 1.2 > index ≥ 0.6, similar molecules found in the training set have experimental values that slightly disagree with the target compound predicted value

If index ≥ 1.2, similar molecules found in the training set have experimental values that completely disagree with the target compound predicted value

Maximum error of prediction between similar molecules:

This index takes into account the maximum error in prediction between the two most similar compounds. Values near 0 means that the predicted compounds fall in an area of the model's space where the model gives reliable predictions without any outlier value. Defined intervals are:

If index < 0.6, the maximum error in prediction of similar molecules found in the training set has a low value, considering the experimental variability

If 1.2 > index ≥ 0.6, the maximum error in prediction of similar molecules found in the training set has a moderate value, considering the experimental variability

If index ≥ 1.2, the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability

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 are:

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

If index= False, the maximum error in prediction of similar molecules found in the training set has a moderate value, considering the experimental variability

Atom Centered Fragments similarity check:

This index takes into account the presence of one or more fragments that aren't found in the training set, or that are rare fragments. First order atom centered fragments from all molecules in the training set are calculated, then compared with the first order atom centered fragments from the predicted compound; then the index is calculated as following: a first index RARE takes into account rare fragments (those who occur less than three times in the training set), having value of 1 if no such fragments are found, 0.85 if up to 2 fragments are found, 0.7 if more than 2 fragments are found; a second index NOTFOUND takes into account not found fragments, having value of 1 if no such fragments are found, 0.6 if a fragments is found, 0.4 if more than 1 fragment is found. Then, the final index is given as the product RARE * NOTFOUND. Defined intervals are:

If index = 1, all atom centered fragment of the compound have been found in the compounds of the training set

If 1 > index ≥ 0.7, some atom centered fragment of the compound have not been found in the compounds of the training set or are rare fragments

If index < 0.7, a prominent number of atoms centered fragments of the compound have not been found in the compounds of the training set or are rare 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.

Model is applicable for the Molecular weight between 18<MW<750 and octanol water partition coefficient between (-3 to +6)

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: Yes

INChI: Yes

MOL file: Yes

NanoMaterial: null

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:

Training set: 271

6.6.Pre-processing of data before modelling:

NA

6.7.Statistics for goodness-of-fit:

Training set: n 271, R^2 0.62, RMSE 0.69

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:

RMSE 0.72

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

NO

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:

NA

7.8.Predictivity - Assessment of the external validation set:

NA

7.9.Comments on the external validation of the model:

The test set was not required for the PBK/D QSAR

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

PBK/D QSAR was based on mechanistic explanatory descriptors

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:

Training set is available in VEGA. The model was also applicable to human skin permeation

9.2.Bibliography:

[1] Wil ten Berge, A simple dermal absorption model: Derivation and application. 2009, 75, 1440–1445https://www.sciencedirect.com/science/article/abs/pii/S004565350900232X

[2] OECD, Test No. 428: Skin Absorption: In Vitro Method. Paris: Organisation for Economic Co-operation and Development, 2004. Accessed: Nov. 09, 2022. [Online]. Available: https://www.oecd-ilibrary.org/environment/test-no-428-skin-absorption-in-vitro-method_9789264071087-en

[3] M. Floris, A. Manganaro, O. Nicolotti, R. Medda, G. Mangiatordi, and E. Benfenati, ‘A generalizable definition of chemical similarity for read-across’, Journal of Cheminformatics, vol. 6, Oct. 2014, doi: 10.1186/s13321-014-0039-1.

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