
	QMRF identifier (JRC Inventory):	
	QMRF Title: <i>QSAR model for blood-brain barrier (BBB) partitioning</i>	
	Printing Date: <i>22.01.2010</i>	

1. QSAR identifier

1.1. QSAR identifier (title):

QSAR model for blood-brain barrier (BBB) partitioning

1.2. Other related models:

1.3. Software coding the model:

QSARModel 3.7.0 Molcode Ltd., Turu 2, Tartu, 51014, Estonia
<http://www.molcode.com>

2. General information

2.1. Date of QMRF:

19.10.2009

2.2. QMRF author(s) and contact details:

[1] Indrek Tulp Molcode Ltd. Turu 2, Tartu, 51014, Estonia
models@molcode.com <http://www.molcode.com>

[2] Tarmo Tamm Molcode Ltd. Turu 2, Tartu, 51014, Estonia
models@molcode.com <http://www.molcode.com>

[3] Gunnar Karelson Molcode Ltd. Turu 2, Tartu, 51014, Estonia
models@molcode.com <http://www.molcode.com>

[4] Dimitar Dobchev Molcode Ltd. Turu 2, Tartu, 51014, Estonia
models@molcode.com <http://www.molcode.com>

[5] Dana Martin Molcode Ltd. Turu 2, Tartu, 51014, Estonia
models@molcode.com <http://www.molcode.com>

[6] Kaido Tämm Molcode Ltd. Turu 2, Tartu, 51014, Estonia
models@molcode.com <http://www.molcode.com>

[7] Deniss Savchenko Molcode Ltd. Turu 2, Tartu, 51014, Estonia
models@molcode.com <http://www.molcode.com>

[8] Jaak Jänes Molcode Ltd. Turu 2, Tartu, 51014, Estonia
models@molcode.com <http://www.molcode.com>

[9] Eneli Härk Molcode Ltd. Turu 2, Tartu, 51014, Estonia
models@molcode.com <http://www.molcode.com>

[10] Andres Kreegipuu Molcode Ltd. Turu 2, Tartu, 51014, Estonia
models@molcode.com <http://www.molcode.com>

[11] Mati Karelson Molcode Ltd. Turu 2, Tartu, 51014, Estonia
models@molcode.com <http://www.molcode.com>

2.3. Date of QMRF update(s):

-

2.4. QMRF update(s):

-

2.5. Model developer(s) and contact details:

Molcode model development team Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com http://www.molcode.com

2.6.Date of model development and/or publication:

18.10.2009

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

[1]Karelson M, Dobchev D, Tamm T, Tulp I, Jänes J, Tämm K, Lomaka A, Savchenko D & Karelson G (2008). Correlation of blood-brain penetration and human serum albumin binding with theoretical descriptors. ARKIVOC 16, 38-60.

[2]Karelson M, Karelson G, Tamm T, Tulp I, Jänes J, Tämm K, Lomaka A, Savchenko D & Dobchev D (2009). QSAR study of pharmacological permeabilities. ARKIVOC 2, 218-238.

2.8.Availability of information about the model:

Model is proprietary, but the training and test sets are available.

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

None to date

3.Defining the endpoint - OECD Principle 1

3.1.Species:

rat

3.2.Endpoint:

5. Toxicokinetics 5.4 Blood-brain barrier penetration

3.3.Comment on endpoint:

The blood-brain barrier (BBB) is a complex membranous system of brain capillary endothelial cells, pericytes, astrocytes, and nerve endings that plays a central role in maintaining the homeostasis of the central nervous system by blocking the movement of molecules.

3.4.Endpoint units:

unitless

3.5.Dependent variable:

logP(blood-brain-barrier)

3.6.Experimental protocol:

The blood-brain barrier (BBB) is a complex membranous system of brain tissue. In drug discovery and development, the determination of BBB permeation is crucial to design new potential candidates for central nervous system and to avoid undesirable side effects of compounds acting in peripheral tissue. The blood-brain distribution is expressed as the ratio of the steady state molar concentration of a compound in the brain and in the blood:

$$BB=C_{\text{brain}}/C_{\text{blood}}$$

Blood-brain barrier partitioning data has been measured experimentally by several investigators, whereas the data were limited to in vivo measurements obtained from rats. In these protocols, the compound was administered via an iv mode; the rat was subsequently sacrificed. The concentration of the compound was measured in both the brain tissue and

in the blood.

3.7. Endpoint data quality and variability:

Data from a series of different experiments was used (as assembled in ref 2. in 9.2). The assembly is justified by the development of different models as published in refs 1-2.

Statistics: min value: -1.52; max value: 1.51; standard deviation: 0.724; skewness: 0.042

4. Defining the algorithm - OECD Principle 2

4.1. Type of model:

2D and 3D regression-based QSAR

4.2. Explicit algorithm:

multilinear regression QSAR

multilinear regression QSAR derived with BMLR (Best Multiple Linear Regression) method

$$\log P(\text{blood-brain-barrier}) = -3.03$$

$$+0.398 * \text{Number of halogenide groups}$$

$$-25.7 * \text{HA dependent HDCA-2/SQRT(TMSA)} \text{ (Zefirov) (all)}$$

$$+0.324 * \text{HOMO-1 energy (AM1)}$$

$$-0.00625 * \text{WFOSA Atomic charge (AM1) weighted FOSA}$$

$$-9.99 * \text{Max net atomic charge (Zefirov) for N atoms}$$

4.3. Descriptors in the model:

[1] Number of halogenide groups unitless count of halogenide (F, Cl, Br, I) functional groups in molecule

[2] HA dependent HDCA-2/SQRT(TMSA) (Zefirov) (all) au Area-weighted surface charge of hydrogen bonding donor atoms

[3] HOMO-1 energy (AM1) eV second highest occupied molecular orbital

[4] WFOSA Atomic charge (AM1) weighted FOSA au * Å charge weighted hydrophobic component of the SASA (solvent accessible surface area)

[5] Max net atomic charge (Zefirov) for N atoms au maximum (most positive) atomic partial charges over nitrogens in molecule

4.4. Descriptor selection:

Initial pool of ~1000 descriptors. Stepwise descriptor selection based on a set of statistical selection rules (one-parameter equations: Fisher criterion and R2 over threshold, variance and t-test value over threshold, intercorrelation with another descriptor not over threshold),

(two-parameter equations: intercorrelation coefficient below threshold, significant correlation with endpoint, in terms of correlation coefficient and t-test)

Stepwise trial of additional descriptors not significantly correlated to any already in the model.

4.5. Algorithm and descriptor generation:

1D, 2D, and 3D theoretical calculations. Quantum chemical descriptors derived from MMFFs (Merck Molecular Force Field) (vacuum) conformational search and AM1 calculation. Model developed

by using multilinear regression.

4.6. Software name and version for descriptor generation:

QSARModel 3.7.0

QSAR/QSPR package that will compute chemically meaningful descriptors and includes statistical tools for regression modeling

Molcode Ltd, Turu 2, Tartu, 51014, Estonia

<http://www.molcode.com>

4.7. Descriptors/Chemicals ratio:

0.093, (6 descriptors / 68 chemicals); 11.33 (68 chemicals/ 6 descriptors)

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Applicability domain based on training set:

a) by chemical identity: pharmaceuticals

b) by descriptor value range: The model is suitable for compounds that have the descriptors in the following range:

Number of halogenide groups: 0 – 3

HA dependent HDCA-2/SQRT(TMSA) (Zefirov) (all): 0.00283 – 0.0994

HOMO-1 energy (AM1): -11.5 – -8.77

WFOSA Atomic charge (AM1) weighted FOSA: 0 – 103

Max net atomic charge (Zefirov) for N atoms: -0.125 – 0.0083

5.2. Method used to assess the applicability domain:

Range of descriptor values in training set with $\pm 30\%$ confidence. Descriptor values must fall between maximal and minimal descriptor values of training set $\pm 30\%$.

5.3. Software name and version for applicability domain assessment:

QSARModel 3.7.0

QSAR/QSPR package that will compute chemically meaningful descriptors and includes statistical tools for regression modeling

Molcode Ltd, Turu 2, Tartu, 51014, Estonia

<http://www.molcode.com>

5.4. Limits of applicability:

See 5.1

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: No

Formula: Yes

INChI: No

MOL file: Yes

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:

54 data points

26 negative values

28 positive values

6.6.Pre-processing of data before modelling:

n/a

6.7.Statistics for goodness-of-fit:

$R^2 = 0.753$ (Correlation coefficient)

$s^2 = 0.382$ (Standard error of the estimate)

$F = 29.22$ (Fisher function)

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

$R^2_{CV} = 0.680$

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

$R^2_{CVMO} = 0.680$

6.10.Robustness - Statistics obtained by Y-scrambling:

n/a

6.11.Robustness - Statistics obtained by bootstrap:

n/a

6.12.Robustness - Statistics obtained by other methods:

ABC analysis (2:1 training : prediction) on sorted (in increased order of endpoint value) data divided into 3 subsets (A;B;C). Training set formed with 2/3 of the compounds (set A+B, A+C, B+C) and validation set consisted of 1/3 of the compounds (C, B, A).

average R^2 (fitting) = 0.753

average R^2 (prediction) = 0.739

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:Yes

Smiles:No

Formula:Yes

INChI:No

MOL file:Yes

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

All

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

All

7.5.Other information about the external validation set:

6 data points,
3 negative values,
3 positive values

7.6. Experimental design of test set:

From sorted data each 10th was subjected to the test set starting from 5th in order to assure the equality in distribution tails.

7.7. Predictivity - Statistics obtained by external validation:

$R^2 = 0.742$ (Correlation coefficient)

7.8. Predictivity - Assessment of the external validation set:

Descriptor value range (all in range of applicability domain):

Number of halogenide groups: 0 – 1

HA dependent HDCA-2/SQRT(TMSA) (Zefirov) (all): 0.0114 – 0.0424

HOMO-1 energy (AM1): -10.4 – -8.88

WFOSA Atomic charge (AM1) weighted FOSA: 0 – 81.9

Max net atomic charge (Zefirov) for N atoms: -0.114 – 0.00916

7.9. Comments on the external validation of the model:

The validation correlation coefficient (R^2) for the test set is significant, and it is close to that of the training set R^2 .

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

"HA dependent HDCA-2 (AM1) (all)" is related to the hydrogen donor ability of the compound, and the respective partial charge on the hydrogen surface area. Increasing the values of this variable leads to decreased logP(blood-brain-barrier) values. Hydrogen bonding is known to be one of the most important factors for the blood-brain barrier penetration.

"Max net atomic charge (Zefirov) for N atoms" reflects the influence of the presence of N atoms in the drug molecules on logP(blood-brain-barrier) while simultaneously accounting the highest charge of N atom. Negative sign of the coefficient indicates that positive charges on N atoms leads to decreased logP(blood-brain-barrier) values.

Descriptors "Number of halogenated groups", "HOMO-1 energy (AM1)" and "WFOSA Atomic charge (AM1) weighted FOSA" are connected to the short-range and hydrophobic interactions of the molecules with the membranous system while penetrating the blood brain barrier. It is known that halogenated compounds are more lipophilic and poorly soluble in water, which leads to favorable penetration of the such structures through the barrier. This fact is reflected by the positive sign of the coefficient of "Number of halogenated groups" in the model. "HOMO-1 energy (AM1)" describes the availability/reactivity of electron pairs (as H-bond acceptors); "WFOSA Atomic charge (AM1) weighted FOSA" is a measure of the polarity of the hydrophobic part of the solvent accessible surface area, as it is weighted by the partial charges of the hydrogen and sp³ hybridized carbon atoms.

8.2.A priori or a posteriori mechanistic interpretation:

a posteriori mechanistic interpretation

8.3.Other information about the mechanistic interpretation:

The mechanistic interpretation is consistent with published scientific interpretations of experiments and modelling (1-2 in section 9.2)

9.Miscellaneous information

9.1.Comments:

Experimental data is taken from: publication 2 (in 9.2)

9.2.Bibliography:

[1]Karelson M, Dobchev D, Tamm T, Tulp I, Jänes J, Tämm K, Lomaka A, Savchenko D, Karelson G, Correlation of blood-brain penetration and human serum albumin binding with theoretical descriptors. ARKIVOC (2008) 16, 38-60. <http://www.arkat-usa.org/get-file/26925/>

[2]Rose K., Hall L. H. and Kier L. B. Modeling Blood-Brain Barrier Partitioning Using the Electrotopological State, J. Chem. Inf. Comput. Sci. 2002, 42, 651-666. <http://dx.doi.org/10.1016/j.jmgm.2004.03.010>

9.3.Supporting information:

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

Karelson Arkivoc 2008	http://qsar.db.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf83_Karelson_Arkivoc_2008.pdf
Karelson Arkivoc 2009	http://qsar.db.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf83_Karelson_Arkivoc_2009.pdf

10.Summary (ECB Inventory)

10.1.QMRF number:

10.2.Publication date:

10.3.Keywords:

10.4.Comments: