



| | | |
|---|--|---|
|  | QMRf identifier (JRC Inventory): |  |
| | QMRf Title: <i>QSAR model for Toxicokinetics, Transfer Index (TI)</i> | |
| | Printing Date: <i>16.02.2011</i> | |

1. QSAR identifier

1.1. QSAR identifier (title):

QSAR model for Toxicokinetics, Transfer Index (TI)

1.2. Other related models:

1.3. Software coding the model:

QSARModel 4.0.4 Molcode Ltd., Turu 2, Tartu, 51014, Estonia
<http://www.molcode.com>

2. General information

2.1. Date of QMRf:

16.02.2011

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

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

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

[10] 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:

03.02.2011

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:

placental transfer

3.2.Endpoint:

5.Toxicokinetics 5.5.Placental barrier penetration

3.3.Comment on endpoint:

B.36 in REACH reach classification

3.4.Endpoint units:

unitless

3.5.Dependent variable:

TI (transfer index)

3.6.Experimental protocol:

Data are collected from the literature. The index is in the form of a ratio, where antipyrine is used as a reference compound. Antipyrine is a small lipophilic compound that is known to be transported across the placenta via passive diffusion and is not bound to protein. The use of such a reference compound acts as an internal standardisation procedure and corrects for inter-placental variations such as blood flow, placenta weight, and surface area. In theory, as the figure calculated is a ratio using antipyrine as a standard, data gathered in multiple laboratories should be directly comparable and the problems with inter-laboratory error will be avoided [1].

3.7.Endpoint data quality and variability:

Statistics:

max value: 1.24

min value: 0.00

standard deviation: 0.322

skewness: -0.119

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

$$\text{Transfer index (TI)} = 3.367$$

$$+0.220 * \text{Balaban index}$$

$$-3.274E-003 * \text{HASA H-acceptors surface area (AM1)}$$

$$-1.017 * \text{Highest e-e repulsion (1-center) (AM1) for H atoms}$$

$$-29.427 * \text{Min electrophilic reactivity index (AM1) for N atoms}$$

4.3.Descriptors in the model:

[1]Balaban index [unitless] Average distance connectivity index J

[2]HASA H-acceptors surface area (AM1) [\AA^2] Solvent-accessible surface area of H-bonding acceptor atoms (from AM1 calculation)

[3]Highest e-e repulsion (1-center) (AM1) for H atoms [eV] Highest electron-electron repulsion energy for hydrogen atoms

[4]Min electrophilic reactivity index (AM1) for N atoms [1/eV] Lowest Fukui electrophilic reactivity index for N atoms

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 AM1 calculation. Model developed by using multilinear regression.

4.6.Software name and version for descriptor generation:

QSARModel 4.0.4

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:

0.095, (4 descriptors / 42 chemicals)

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: drugs

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

in $\pm 30\%$ the following minimal-maximal range:

Balaban index: 0.833 - 3.99

HASA H-acceptors surface area (AM1): 0 - 244

Highest e-e repulsion (1-center) (AM1) for H atoms: 2.44 - 2.99

Min electrophilic reactivity index (AM1) for N atoms: 0 - 0.0403

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 $\pm 30\%$ of descriptor distribution range of training set .

5.3.Software name and version for applicability domain assessment:

QSARModel 4.0.4

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:

42 data points

0 negative values

42 positive values

6.6.Pre-processing of data before modelling:

n/a

6.7.Statistics for goodness-of-fit:

$R^2 = 0.719$ (Correlation coefficient)

$s^2 = 0.182$ (Standard error of the estimate)

$F = 23.65$ (Fisher function)

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

R2CV = 0.633

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

R2CVMO = 0.590

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 R2 (fitting) = 0.730

average R2 (prediction) = 0.623

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:

10 data points,

0 negative values,

10 positive values

7.6. Experimental design of test set:

From sorted data each 5th was subjected to the test set. First was picked 4th, further 9th and etc. This was chosen in order to assure the test set equal division on data distribution tails.

7.7. Predictivity - Statistics obtained by external validation:

R2 = 0.782 (Coefficient of determination)

7.8. Predictivity - Assessment of the external validation set:

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

Balaban index: 1.26 - 2.45

HASA H-acceptors surface area (AM1): 26.8 - 184

Highest e-e repulsion (1-center) (AM1) for H atoms: 2.35 - 2.93

Min electrophilic reactivity index (AM1) for N atoms: 1.16E-8 - 5.28E-4

7.9. Comments on the external validation of the model:

The validation coefficient of determination (R^2) is better to those coefficients of internal validation (R^2_{CV} and R^2_{CVMO}). This is probably caused by division of structurally complicated compounds into training set. This fluctuation can easily happen with complicated structures in small data sets.

8. Providing a mechanistic interpretation - OECD Principle 5

8.1. Mechanistic basis of the model:

Obtained QSAR model possess one topological index and three quantum chemically derived descriptors. "Balaban index" is molecule shape and branching related descriptor. It can be related with hydrophobicity and favours membrane permeability. "HASA H-acceptors surface area (AM1)" and "Highest e-e repulsion (1-center) (AM1) for H atoms" can be attributed as hydrophilicity related. Specifically hydrogen bonding related, taking into account hydrogen bond donor and acceptor characteristics, respectively. "Min electrophilic reactivity index (AM1) for N atoms" specifically accounts quarterly amines. The latter three electrostatic interaction related descriptors unfavour membrane permeability, hence negative correlation coefficient sign in the QSAR model.

8.2. A priori or a posteriori mechanistic interpretation:

a posteriori mechanistic interpretation, consistent with published scientific interpretations of experiments

8.3. Other information about the mechanistic interpretation:

Interpretation in general agreement with literature given in references 1 and 2.

9. Miscellaneous information

9.1. Comments:

9.2. Bibliography:

[1] Structure-based modelling in reproductive toxicology: (Q)SARs for the placental barrier M. Hewitt; J. C. Madden; P. H. Rowe; M. T. D. Cronin SAR and QSAR in Environmental Research, Volume 18, Issue 1 & 2 January 2007, pages 57 - 76 <http://dx.doi.org/10.1016/j.bmcl.2004.11.030>

[2] Application of quantitative structure-activity relationships for modeling drug and chemical transport across the human placenta barrier: a multivariate data analysis approach Costas Giaginis, Athina Zira, Stamatios Theocharis and Anna Tsantili-Kakoulidou Journal of Applied Toxicology, Volume 29, Issue 8, pages 724-733, November 2009 <http://dx.doi.org/10.1002/jat.1466>

9.3. Supporting information:

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

| | |
|-----------------------|---|
| Karelson Arkivoc 2008 | http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf83_Karelson_Arkivoc 2008.pdf |
| Karelson Arkivoc 2009 | http://qsardb.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: