QMRF	*QMRF identifier (JRC Inventory):* *QMRF Title:* *QSAR model for Human Serum Albumin Binding (logK(HSA))* *Printing Date:*5.02.2010	QMRF

1.QSAR identifier

1.1.QSAR identifier (title):

QSAR model for Human Serum Albumin Binding (logK(HSA))

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:

06.10.2009

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:

03.10.2009

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:

In vitro assay (HPLC chromatographic retention index [logk(HSA)] on immobilized albumin)

3.2.Endpoint:

5. Toxicokinetics 5.9 Protein-binding

3.3.Comment on endpoint:

Drugs bind reversibly with varying degrees of association to human plasma proteins: serum albumin (HSA), alpha-1-acid glycoprotein (AGP), and lipoproteins. Since the drug-protein complex in the plasma acts as a reservoir for the drug, the degree of binding is an important parameter in pharmacokinetic profiling.

3.4.Endpoint units:

mM

3.5.Dependent variable:

logK(HSA)

3.6.Experimental protocol:

The set of compounds comprises many families of well-known compounds from many different therapeutic areas. These compounds were assayed for HSA binding through high-performance affinity chromatography by using an immobilized HSA column (ThermoHypersil, 150 mm x 4.6 mm size). This technique is well established as a fast and reliable method to obtain HSA binding constants (Kaliszan et al., 1992; Andrisano et al., 2000).

The mobile phase used was 25 mM Na2HPO4, 25 mM KH2PO4 (pH 7.0)/acetonitrile [85:15; v/v]. A flow rate of 0.8 mL min-1 was used throughout. Experiments were conducted at 25 ± 0.1 °C. A minimum of four different chromatograms were obtained for each compound to ensure

the reproducibility of the measurements and to estimate their errors. As is customary in protein binding studies by highperformance affinity chromatography, the binding constants were calculated in the logarithmic scale as log K'hsa=log((t - t0)/t0), where t and t0 are the retention times of the drug and NaNO3 (dead time of the column), respectively (Kaliszan et al., 1992; Andrisano et al., 2000).

The compounds span a wide range of K'hsa binding constants (3 orders of magnitude) corresponding to retention times between 2 and 56 min. It is therefore expected to be appropriate to model a wide range of druglike molecules.

3.7.Endpoint data quality and variability:

Source experimental data from one lab has been used (as published in ref 1, section 9.2). The data has also been successfully modelled previously (ref 2), supporting consistency.

Statistics:
max value: 1.34
min value: -1.39
standard deviation: 0.601
skewness: 0.071

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

logK(HSA) = 16.7
-7.53* Min net atomic charge (Zefirov) for any atom type
+0.150* Kier&Hall index (order 3)
-0.520* Topographic electronic index (Zefirov) all bonds
+0.366* Number of benzene rings
-35.2* Partial Charged (Zefirov) Surface Area of O atoms
+0.178* Highest atomic state energy (AM1) for C atoms

4.3.Descriptors in the model:

[1]Min net atomic charge (Zefirov) for any atom type au The most negative atomic charge over all atom types based on Zefirov's charge distribution
[2]Kier&Hall index (order 3) unitless three contiguous bond fragment Kier and Hall valence connectivity index
[3]Topographic electronic index (Zefirov) all bonds au/Å2 Topolographical electronic index calculated over all bonds between non-hydrogen atom in the molecule and based on Zefirov's charge distribution
[4]Number of benzene rings unitless Count of benzene rings in the molecule

[5]Partial Charged (Zefirov) Surface Area of O atoms Å2 sum of charged solvent accessible surface areas of oxygen atoms in the molecule

[6]Highest atomic state energy (AM1) for C atoms eV Highest carbon atom state energy obtained from semi-empirical AM1 calculations

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:

0.088, (6 descriptors / 68 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 and druglike molecules (organic structures with a variety of functionalities - alcohols, amides, amines, ketones, esters, ethers)

b) by descriptor value range: The model is suitable for compounds that have the descriptors
in the following range:
Min net atomic charge (Zefirov) for any atom type: -0.188 – -0.101

Kier&Hall index (order 3) : 0.759 – 15.9
Topographic electronic index (Zefirov) all bonds: 0.200 – 2.48
Number of benzene rings: 0.00 – 3.00
Partial Charged (Zefirov) Surface Area of O atoms: 0.000 – 0.0438
Highest atomic state energy (AM1) for C atoms: -102 – -99.2

5.2.Method used to assess the applicability domain:

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

descriptor values of training set ±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:

68 data points

39 negative values

29 positive values

6.6.Pre-processing of data before modelling:

n/a

6.7.Statistics for goodness-of-fit:

R2 = 0.835 (Correlation coefficient)

s2 = 0.258 (Standard error of the estimate)

F = 51.43 (Fisher function)

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

R2CV = 0.800

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

R2CVMO = 0.798

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) = 0.835
average R2 (prediction) = 0.798

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:

17 data points,
10 negative values,
7 positive values

7.6.Experimental design of test set:

From originl source data set sorted by endpoint values, each 5th was subjected to the test set starting from 3rd in order to assure the equality in distribution tails.

7.7.Predictivity - Statistics obtained by external validation:

R2 = 0.748 (Correlation coefficient)

7.8.Predictivity - Assessment of the external validation set:

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

Min net atomic charge (Zefirov) for any atom type: -0.181 – -0.0986

Kier&Hall index (order 3) : 2.32 – 8.45
Topographic electronic index (Zefirov) all bonds: 0.35 – 1.5
Number of benzene rings: 0.00 – 2.00
Partial Charged (Zefirov) Surface Area of O atoms: 0.000 – 0.0349

Highest atomic state energy (AM1) for C atoms: -102 – -99.7

7.9.Comments on the external validation of the model:

The validation correlation coefficient (R2) for the test set is good and it is close to that of the training set.

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

Current linear model comprise size and charge related descriptors. "Kier&Hall index (order 3)" is a valence connectivity index which reflects a molecule size and branching while taking into account also the differences among the atom types, hence heteroatoms. Together with "Number of benzene rings" these descriptors account a molecule size and hydrophobic interaction capabilities. Both descriptors have positive regression coefficients therefore they suggest hydrophobic interactions for HSA binding affinity. High negative charges ("Min net atomic charge (Zefirov) for any atom type") tend to form stronger interactions with protein positively charged side-chains. Another hand large negatively charged surface areas ("Partial Charged (Zefirov) Surface Area of O atoms") are not preferred. Descriptors "Topographic electronic index (Zefirov) all bonds" and "Highest atomic state energy (AM1) for C atoms" are related to reactivity of a molecule and they reflect a magnitude of electrophile/nucleophile interactions.

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:

Hall LM, Hall LH, Kier LB., Modeling drug albumin binding affinity with e-state topological structure representation., J Chem Inf Comput Sci. 2003 Nov-Dec;43(6):2120-8.

Kratochwil NA, Huber W, Müller F, Kansy M, Gerber PR., Predicting plasma protein binding of drugs: a new approach., Biochem Pharmacol. 2002 Nov 1;64(9):1355-74.

9.Miscellaneous information

9.1.Comments:

Same experimental data set was used for modelling in ref 2 (sec. 9.2)

9.2.Bibliography:

[1]Colmenarejo, G.; Alvarez-Pedraglio, A.; Lavandera J.-L. Cheminformatic Models To Predict Binding Affinities to Human Serum Albumin J. Med. Chem. 2001, 44, 4370-4378. http://dx.doi.org/10.1021/jm010960b

[2]Hall L.M., Hall L.H., Kier L.B., Modeling drug albumin binding affinity with e-state topological structure representation., J Chem Inf Comput Sci. 2003 Nov-Dec;43(6):2120-8. http://dx.doi.org/10.1021/ci030019w

[3]Kaliszan, R.; Noctor, T. A. G.; Wainer, I. W. Quantitative Structure-Enantioselective Retention Relationships for the Chromatography of 1! ,4-Benzo diazepines on a Human Serum Albumin Based HPLC Chiral Stationary Phase: An Approach to the Computational Prediction of Retention and Enantioselectivity. Chromatographia 1992, 33, 546-550. http://dx.doi.org/10.1007/BF02262246

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: