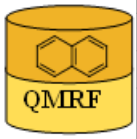


*QMRF identifier (ECB Inventory):*

*QMRF Title:* *QSAR for Relative Binding Affinity to Estrogen Receptor*

*Printing Date:*Jan 28, 2010


## 1.QSAR identifier

### 1.1.QSAR identifier (title):

QSAR for Relative Binding Affinity to Estrogen Receptor

### 1.2.Other related models:

### 1.3.Software coding the model:

QSARModel 3.5.0 Molcode Ltd., Turu 2, Tartu, 51014, Estonia http://www.molcode.com

## 2.General information

### 2.1.Date of QMRF:

07.09.2009

### 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:**

07.09.2009

**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:**

Estrogen receptor binding in rat uterine cytosol

**3.2.Endpoint:**

6.Other Relative Binding Affinity to Estrogen Receptor 6.6.Other

**3.3.Comment on endpoint:**

**3.4.Endpoint units:**

Estrogen Receptor Relative Binding Affinity (ER-RBA) is expressed in %

**3.5.Dependent variable:**

Log(ER-RBA) logarithm of the Estrogen Receptor Relative Binding Affinity (ER-RBA). The chemicals are screened at high concentrations to see if they compete with [3H]-Estradiol for Estrogen Receptor. The ER-RBA was calculated for each competitor by dividing the IC50 (inhibition concentration) of Estradiol by IC50 of the competitor and expressing the result as a percent.

**3.6.Experimental protocol:**

The estrogen receptor (ER) is activated by the hormone 17β-estradiol (estrogen). However, some structurally diverse steroidal and non-steroidal chemicals are believed to exert their effect via interaction with the estrogen receptor altering the structure or function(s) of the endocrine system. The Environmental Protection Agency (EPA) implements a screening strategy for assessing the risk associated with endocrine disrupting chemicals focusing on development of priority-setting approaches and Tier 1 screening methods, initially for assessing estrogenic activity, that would guide the more limited application of Tier 2 animal testing. Priority setting primarily refers to quantitative structure-activity relationship (QSAR) methods for assessing the potential estrogenic activity of chemicals for which test data are unavailable. The database is a structurally diverse set of natural, synthetic, and environmental estrogens covering most known estrogenic classes and spanning a wide range of biological activity. It represents the largest published ER binding database of same-assay results generated in a single laboratory. The Food and Drug Administration (FDA) National Center for Toxicological Research (NCTR) ER database consists of 232 chemicals. A chemical's binding activity was determined by competing with radiolabeled [3H]E2 for the ER in rat uterine cytosol. The IC50 (50% inhibition of [3H]E2 binding) for each competitor was determined. The relative binding affinity (RBA) for each competitor was calculated by

dividing the IC50 of E2 by the IC50 of the competitor and multiplying by 100 (E2 RBA = 100). The validated assay incubation conditions were 20 h at 4 °C using 17 mg of uterine tissue/mL (Bmax = 0.22 nM) with 1 nM [3H]E2. The competing chemical concentrations ranged from 1 nM to 1 mM. Chemicals that failed to compete for [3H]E2 binding to the ER were designated as "not active" (NA). Chemicals that exhibited binding, but did not reach 50% inhibition in the designed concentration range, were designated as "slight binders". All assays were repeated at least twice; the IC50 values of positive chemicals are the means of the replicate values. The standard deviation of IC50 for each chemical was reported, and only the mean RBA value was used for this study. The largest fold difference (10-fold) was found for nonylphenol from different commercial sources due to the impurity of the sample. Information on the percent purity and purchasing source for all chemicals can be obtained from the original NCTR ER Source database (1) and references (2-4) are listed in Section 9.

### 3.7.Endpoint data quality and variability:

Statistics: max value: 2.60; min value: -4.50; standard deviation: 1.79; skewness: 0.89

## 4.Defining the algorithm - OECD Principle 2

### 4.1.Type of model:

QSAR

### 4.2.Explicit algorithm:

Multilinear regression QSAR

LogER-RBA = -19.12 + 2.11*Average Information content (order 1)+ 0.80*Number of rings + 7.33*Relative ALFA polarizability (DIP) (AM1) - 13.83*Max net atomic charge (Zefirov) for O atoms + 0.84*logP

### 4.3.Descriptors in the model:

[1]Average Information content (order 1)
[2]Number of rings
[3]Relative ALFA polarizability (DIP)
[4]Max net atomic charge (Zefirov) for O atoms
[5]logP

### 4.4.Descriptor selection:

Initial pool of ~1000 descriptors. Stepwise descriptor selection based on a set of statistical selection rules:1-parameter equations: Fisher criterion and $R^2$ over threshold, variance and t-test value over threshold, intercorrelation with another descriptor not over threshold;2 parameter equations: intercorrelation coefficient bellow 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 3.5.0
Molcode Ltd, Turu 2, Tartu, 51014, Estonia
http://www.molcode.com

### 4.7.Descriptors/Chemicals ratio:

12.4 (62 chemicals/7 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. By chemical identity: diverse organics, phenols, organic acids, esters, ketones, halogenated compounds, etc. By descriptor value range: the model is suitable for compounds that have the descriptors in the following range: Average Information content (order 1)(min: 1.14, max: 4.76 ), Number of rings (min: 1, max: 7), Relative ALFA polarizability (DIP) (AM1) (min: 0.36, max: 1.12), Max net atomic charge (Zefirov) for O atoms (min: -0.23, max: 0.0), logP (min: -1.44, max: 11.82).

### 5.2.Method used to assess the applicability domain:

Presence of functional groups in structures. 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 3.5.0
Molcode Ltd, Turu 2, Tartu, 51014, Estonia
http://www.molcode.com

### 5.4.Limits of applicability:

## 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:No
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:

62 data points: 51 negative values; 11 positive values

### 6.6.Pre-processing of data before modelling:

No preprocessing done. Four structures were excluded as statistically significant outliers (as listed in 9.1)

### 6.7.Statistics for goodness-of-fit:

$R^2$ = 0. 80 (Correlation coefficient); s = 0.70 (Standard error of the estimate); F = 45.43 (Fisher function)

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

$R^2_{cv}$ = 0.73 LOO

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

$R^2_{cv}$ = 0.71 LMO

**6.10.Robustness - Statistics obtained by Y-scrambling:**

**6.11.Robustness - Statistics obtained by bootstrap:**

**6.12.Robustness - Statistics obtained by other methods:**

ABC analysis (2:1 training : prediction) on sorted 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) = 0.81; average $R^2$ (prediction) = 0.77

## 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:No
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: 6 negative values; 0 positive values

**7.6.Experimental design of test set:**

The full experimental dataset was sorted according to increasing values of logER-RBA and each tenth compound was assigned to the test set.

**7.7.Predictivity - Statistics obtained by external validation:**

$R^2$= 0.71

**7.8.Predictivity - Assessment of the external validation set:**

The descriptors for the test set are in the limits of the applicability domain.

**7.9.Comments on the external validation of the model:**

## 8.Providing a mechanistic interpretation - OECD Principle 5

**8.1.Mechanistic basis of the model:**

For ER-RBA, one of the important structural features is the presence of aromatic moieties in form of benzene rings, which is reflected by the correlation of the logER-RBA with the descriptor Number of rings. The increased length of the aliphatic chain in a compound is beneficial for binding to the ER. The increased length of the aliphatic chain is reflected in increased value of logP, which is directly correlated with logER-RBA. The shape of the molecule is also important for binding to ER, as reflected by increasing of logER-RBA with increasing values of the descriptor Average Information Content (order 1). The

descriptors Relative ALFA polarizability (DIP) (AM1) and Max net atomic charge (Zefirov) for O atoms bring additional correction to the model, providing information on the dipole-induced dipole and dipole-dipole capabilities of a molecule.

### 8.2.A priori or a posteriori mechanistic interpretation:

A posteriori mechanistic interpretation.

### 8.3.Other information about the mechanistic interpretation:

The partition coefficient logP is the ratio of the concentrations of a compound in the two phases of a mixture of two miscible solvents at equilibrium (usually water and octanol), and gives information about the hydrophobicity of the compound. Hydrophobicity increases with increase of number and length of aliphatic chains. The Average Information Content is a topological descriptor calculated on the basis of Shannon information theory and gives information about the connectivity of the atoms in the molecule. The descriptors Relative ALFA polarizability (DIP) (AM1) and Max net atomic charge (Zefirov) for O atoms represent information about partial charges (inducible and initially present) that exist in the compound; these charges can influence the binding of the compound to estrogen receptor.

## 9.Miscellaneous information

### 9.1.Comments:

The molecules: (3R,7R)-7,14,16-trihydroxy-3-methyl-3,4,5,6,7,8,9,10,11,12-decahydro-1H-2-benzoxacyclotetradecin-1-one (26538-44-3), 4H-1-Benzopyran-4-one, 5,7-dihydroxy-3-(4-methoxyphenyl)- (491-80-5) and 2-[bis(4-hydroxyphenyl)methyl]benzoic acid (81-90-3) have been considered as statistical errors and have been excluded, as compared to the source dataset.

### 9.2.Bibliography:

[1]Distributed Structure-Searchable Toxicity (DSSTox) Public Database Network http://www.epa.gov/NCCT/dsstox/sdf_nctrer.html

[2]Fang H, Tong W, Shi LM, Blair R, Perkins R, Branham W, Hass BS, Xie Q, Dial SL, Moland CL & Sheehan DM (2001). Structure-activity relationships for a large diverse set of natural, synthetic, and environmental estrogens. Chemical Research in Toxicology 14, 280-294.

[3]Blair RM, Fang H, Branham WS, Hass BS, Dial SL, Moland CL, Tong W, Shi L, Perkins R & Sheehan DM (2000). The estrogen receptor relative binding affinities of 188 natural and xenochemicals: Structural diversity of ligands. Toxicological Sciences 54, 138-153.

[4]Branham WS, Dial SL, Moland CL, Hass BS, Blair RM, Fang H, Shi L, Tong W, Perkins RG & Sheehan DM (2002). Binding of phytoestrogens and mycoestrogens to the rat uterine estrogen receptor. Jounral of Nutrition 132, 658-664.

### 9.3.Supporting information:

Training set(s)

| | |
|---|---|
| Estrogen_Receptor 62_train | http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf171_Estrogen_Receptor 62_train.sdf |

Test set(s)

| | |
|---|---|
| Estrogen_Receptor 6_test | http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf171_Estrogen_Receptor 6_test.sdf |

## 10.Summary (ECB Inventory)

### 10.1.QMRF number:

### 10.2.Publication date:

### 10.3.Keywords:

Molcode, rat uterine cytosol, estrogen receptor binding

### 10.4.Comments: