

	<b>QMRF identifier (JRC Inventory): To be entered by JRC</b>
	<b>QMRF Title: Estrogen Receptor Relative Binding Affinity Model v- 1.0.1</b>
	<b>Printing Date: 13-nov-2020</b>

## 1. QSAR identifier

### 1.1. QSAR identifier (title):

Estrogen Receptor Relative Binding Affinity Model v- 1.0.1

### 1.2. Other related models:

### 1.3. Software coding the model:

VEGA

The VEGA software provides QSAR models to predict tox, ecotox, environ, and phys-chem properties of chemical substances.

[emilio.benfenati@marionegri.it](mailto:emilio.benfenati@marionegri.it)

<https://www.vegahub.eu/>

## 2. General information

### 2.1. Date of QMRF:

13-10-2020

### 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](mailto:emilio.benfenati@marionegri.it) <https://www.marionegri.it/>

[2] Alessandra Roncaglioni Istituto di Ricerche Farmacologiche Mario Negri - IRCSS Via Mario Negri 2, 20156 Milano, Italy [alessandra.roncaglioni@marionegri.it](mailto:alessandra.roncaglioni@marionegri.it) <https://www.marionegri.it/>

### 2.3. Date of QMRF update(s):

### 2.4. QMRF update(s):

### 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](mailto:alberto.manganaro@marionegri.it) <https://www.marionegri.it/>

[2] Alessandra Roncaglioni Istituto di Ricerche Farmacologiche Mario Negri - IRCSS Via Mario Negri 2, 20156 Milano, Italy [alessandra.roncaglioni@marionegri.it](mailto:alessandra.roncaglioni@marionegri.it) <https://www.marionegri.it/>

### 2.6. Date of model development and/or publication:

2008

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

A. Roncaglioni, N. Piclin, M. Pintore, E. Benfenati, "Binary classification models for endocrine disrupter effects mediated through the estrogen receptor", SAR and QSAR in Environmental Research (2008), 19, 7-8 <https://www.tandfonline.com/doi/abs/10.1080/10629360802550606>

### 2.8. Availability of information about the model:

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

## 3. Defining the endpoint - OECD Principle 1

### 3.1. Species:

Human

### 3.2. Endpoint:

Toxicology Estrogen Receptor Relative Binding Affinity

### **3.3.Comment on endpoint:**

Experimentally determined values of human ER alpha for receptor binding assay (RBA), expressed as percentage of activity using 17-estradiol as reference.

### **3.4.Endpoint units:**

percentage Activity (% Activity)

### **3.5.Dependent variable:**

Relative Binding Affinity (RBA)

### **3.6.Experimental protocol:**

METI, Ministry of Economy Trade and Industry, Japan. Current status of testing methods development for endocrine disrupters. 6th meeting of the task force on endocrine disrupters testing and assessment (EDTA), 24–25 June 2002, Tokyo, Japan, 2002.

The tested substance is added to a system where radio-labeled reference hormone binds to a prescribed quantity of hormone receptor. The chemical concentration that inhibits 50% of the binding of the reference hormone to the receptor is measured and defined as IC50. Then, Relative Binding Affinity (RBA) between IC50 values of the chemical and natural hormone (E2) is defined as the endpoint when the IC50 concentration of natural hormone is set at 100.

### **3.7.Endpoint data quality and variability:**

## **4.Defining the algorithm - OECD Principle 2**

### **4.1.Type of model:**

QSAR classification model based on a classification and regression tree (CART). Compounds with any detected activity were labeled as positive while compounds with lack of any activity were labeled as negative.

### **4.2.Explicit algorithm:**

Classification and regression tree (CART) uses the methodology of tree building as a hierarchical classification method. CART formulates simple 'if/then' rules for binary recursive partitioning of all the objects into smaller subgroups, where the compounds belonging to the dataset, represented by a 'node' in a decision tree, can be split into only two groups. Thus, each node can be split into two new 'branches'. The goal of this process is to maximize homogeneity of the values of the dependent variable Y in the various subgroups taking into account the fact that different relationships may hold among variables in different parts of the data. The cost for the errors was considered equally important for the two classes, while the a priori probability for the two classes was set as equal. Automatic stepwise selection of variables was used in CART. All splits are ranked and the variable that achieves the highest purity at root is selected. The Gini measure for node impurity was used. The performances on the validation set were used to efficiently select the well dimensioned tree; then, the best tree was evaluated on the test set. In this study we used the CART algorithm as implemented in Statistica software

### **4.3.Descriptors in the model:**

[1]X2v Valence connectivity index chi-2

[2]MATS6m Moran autocorrelation – lag 6 / weighted by atomic masses

[3]MATS8v Moran autocorrelation – lag 8 / weighted by atomic van der Waals volumes

- [4]MATS5p Moran autocorrelation – lag 5 / weighted by atomic polarizabilities
- [5]BEH2e Highest eigenvalue no. 2 of Burden matrix/weighted by atomic Sanderson electronegativities
- [6]BEH1p Highest eigenvalue no. 1 of Burden matrix/weighted by atomic polarizabilities
- [7]nArOH Number of aromatic hydroxyls
- [8]MlogP Moriguchi octanol-water partition coeff. (log P)

#### **4.4.Descriptor selection:**

CART

#### **4.5.Algorithm and descriptor generation:**

Statistica

#### **4.6.Software name and version for descriptor generation:**

Statistica, version 6.1,  
StatSoft Italia Srl, Vigonza, Italy, 2003  
<https://statistica.software.informer.com/6.1/>

#### **4.7.Chemicals/Descriptors ratio:**

82

### **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.

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

The chemical similarity is measured with the algorithm developed for VEGA. Full details in the VEGA website ([www.vegahub.eu](http://www.vegahub.eu)), including the open access paper describing it. 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. Full details in the VEGA website.

#### **5.3.Software name and version for applicability domain assessment:**

VEGA

The VEGA software provides QSAR models to predict tox, ecotox, environ, and phys-chem properties of chemical substances  
[emilio.benfenati@marionegri.it](mailto:emilio.benfenati@marionegri.it)  
<https://www.vegahub.eu/>

#### **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: No

Smiles: Yes  
Formula: No  
INChI: No  
MOL file: No  
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 n= 656

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

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

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

No

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

No

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

No

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

No

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

Accuracy = 0.86; Specificity = 0.85; Sensitivity = 0.87

**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: No  
Smiles: Yes  
Formula: No  
INChI: No  
MOL file: No  
NanoMaterial: null

**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:**

Test n= 150

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

Kohonen map

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

Accuracy = 0.84; Specificity = 0.88; Sensitivity = 0.78

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

**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:**

None

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

A posteriori based on descriptors analysis. nArOH Number of aromatic hydroxyls accounts for the ability of the molecules to form hydrogen bonds with aminoacids in the binding pocket of estrogen receptor while MlogP Moriguchi octanol-water partition coeff. is useful to model the affinity with hydrophobic regions of the binding acvity.

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

## **9.Miscellaneous information**

**9.1.Comments:**

**9.2.Bibliography:**

A. Roncaglioni, N. Piclin, M. Pintore, E. Benfenati, "Binary classification models for endocrine disrupter effects mediated through the estrogen receptor", SAR and QSAR in Environmental Research (2008), 19, 7-8 <https://www.tandfonline.com/doi/abs/10.1080/10629360802550606>

**9.3.Supporting information:**

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

## **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