	QMRF identifier (JRC Inventory): **To be entered by JRC**
	QMRF Title: **48 hour Daphnia Magna LC50 Model version**
	Printing Date: **14-feb-2020**

1.QSAR identifier

1.1.QSAR identifier (title):

48 hour Daphnia Magna LC50 Model version

1.2.Other related models:

1.3.Software coding the model:

48 hour Daphnia Magna LC50 model version 1.0.4

The model is based on the OECD 202 data and provides a qualitative evaluation of ready toxicity properties.

https://www.vegahub.eu

https://www.vegahub.eu

2.General information

2.1.Date of QMRF:

14/02/2020

2.2.QMRF author(s) and contact details:

Emilio Benfenati Istituto di Ricerche Farmacologiche Mario Negri - IRCCS

emilio.benfenati@marionegri.it

2.3.Date of QMRF update(s):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

[1]Emilio Benfenati Istituto di Ricerche Farmacologiche Mario Negri - IRCCS

emilio.benfenati@marionegri.it

[2]Nicolas Amaury BioChemics Consulting SAS, Orléans, France

[3]Elena Borian ebor@food.dtu.dk

[4]Mosè Casalegno

[5]Antonio Chana

[6]Qasim Chaudhry qasim.chaundhry@fera.qsi.gov.uk

[7]Jacques R. Chrétien

[8]Jane Cotterill

[9]Frank Lemke

[10]Nadège Piclin

[11]Marco Pintore

[12]Chiara Porcelli

[13]Nicholas Price

[14]Alessandra Roncaglioni Istituto di Ricerche Farmacologiche Mario Negri - IRCCS

alessandra.roncaglioni@marionegri.it

[15]Andrey Toropov Istituto di Ricerche Farmacologiche Mario Negri - IRCCS

andrey.toropov@marionegri.it

2.6.Date of model development and/or publication:

28/09/2007

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

2.8.Availability of information about the model:

Freely available through the VEGA web site together with a file describing the models. The training, test and validation sets are also available (see 9.3).

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

Other QMRF for this model are not available

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Daphnia magna

3.2.Endpoint:

Short-term toxicity to Daphnia (immobilization). Daphnids toxixity OECD 202 Test

3.3.Comment on endpoint:

Number and percentage of daphnids that were immobilised or showed any adverse effects (including abnormal behaviour) in the controls and in each treatment group

3.4.Endpoint units:

The model provides a quantitative prediction for Daphnia Magna LC50 (48 hour), given in -log(mol/l) and its conversion in mg/L.

3.5.Dependent variable:

The model is a linear regression made on 16 molecular descriptors associate to Daphnia sp. toxicity

3.6.Experimental protocol:

OECD 202 Test. It measures the immobilized daphnids after 48h of exposure to a substance

3.7.Endpoint data quality and variability:

The regression coefficients have been calculated on the DEMETRA project original dataset,[1] that contains 263 compounds extracted from various databases, split in 220 compounds for the training and 43 for the test set.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

Hybrid model based on multiple linear regressions, using on 16 molecular descriptors.

4.2.Explicit algorithm:

See [7]

4.3.Descriptors in the model:

[1]BEHm1 highest eigenvalue n. 1 of Burden matrix / weighted by atomic masses

[2]Eig1p Leading eigenvalue from polarizability weighted distance matrix

[3]IC2 information content index (neighborhood symmetry of 2-order)

[4]IDE mean information content on the distance equality

[5]MLOGP Moriguchi octanol-water partition coeff. (logP)

[6]Mp mean atomic polarizability (scaled on Carbon atom)

[7]MW molecular weight

[8]nHAcc number of acceptor atoms for H-bonds (N O F)

[9]nNR2Ph number of tertiary amines (aromatic)

[10]nP number of Phosphorous atoms

[11]O-057 phenol / enol / carboxyl OH

[12]O-060 Al-O-Ar / Ar-O-Ar / R..O..R / R-O-C=X

[13]S-107 R2S / RS-SR; Class: atom-centred fragments

[14]SRW05 self-returning walk count of order 05

[15]T(F..Cl) sum of topological distances between F..Cl

[16]WA mean Wiener index

4.4.Descriptor selection:

<html><body>See [2] </body></html>

4.5.Algorithm and descriptor generation:

See [2] and [3]

4.6.Software name and version for descriptor generation:

See [5] and [7]

4.7.Chemicals/Descriptors ratio:

220 chemicals (training set) / 16 descriptors = 13.75

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

5.2.Method used to assess the applicability domain:

The Applicability Domain of the model is defined by considering several parameters as described below: 1. Similar molecules with known experimental value. This index takes into account how similar are the first two most similar compounds found. Values near 1 mean that the predicted compound is well represented in the dataset used to build the model, otherwise the prediction could be an extrapolation. Defined intervals are: 1 >= index > 0.75 strongly similar compounds with known experimental value in the training set have been found, 0.75 >= index > 0.65 only moderately similar compounds with known experimental value in the training set have been found, index
<= 0.65 no similar compounds with known experimental value in the training set have been found. 2. Accuracy (average error) of prediction for similar molecules. This index takes into account the error in prediction for the two most similar compounds found. Values near 0 mean that the predicted compounds falls in an area of the model's space where the model gives reliable predictions, otherwise the greater is the value, the worse the model behaves. Defined intervals are: index
< 0.7 accuracy of prediction for similar molecules found in the training set is good, 0.7
<= index
<= 1 accuracy of prediction for similar molecules found in the training set is not optimal, index > 1 accuracy of prediction for similar molecules found in the training set is not adequate. 3. Concordance with similar molecules (average difference between target compound prediction and experimental values of similar molecules). This index takes into

account the difference between the predicted value and the experimental values of the two most similar compounds. Values near 0 mean that the prediction made agrees with the experimental values found in the model's space, thus the prediction is reliable. Defined intervals are: index < 0.7 similar molecules found in the training set have experimental values that agree with the target compound predicted value, 0.7 <= index <= 1 similar molecules found in the training set have experimental values that slightly disagree with the target compound predicted value, index > 1 similar molecules found in the training set have experimental values that completely disagree with the target compound predicted value. 4. Maximum error of prediction among similar molecules. This index takes into account the maximum error in prediction among the two most similar compounds. Values near 0 means that the predicted compounds falls in an area of the model's space where the model gives reliable predictions without any outlier value. Defined intervals are: index < 0.7 the maximum error in prediction of similar molecules found in the training set has a low value, considering the experimental variability, 0.7 <= index < 1 the maximum error in prediction of similar molecules found in the training set has a moderate value, considering the experimental variability, index >= 1 the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability. 5. Atom Centered Fragments similarity check. This index takes into account the presence of one or more fragments that aren't found in the training set, or that are rare fragments. First order atom centered fragments from all molecules in the training set are calculated, then compared with the first order atom centered fragments from the predicted compound; then the index is calculated as following: a first index RARE takes into account rare fragments (those who occur less than three times in the training set), having value of 1 if no such fragments are found, 0.85 if up to 2 fragments are found, 0.7 if more than 2 fragments are found; a second index NOTFOUND takes into account not found fragments, having value of 1 if no such fragments are found, 0.6 if a fragments is found, 0.4 if more than 1 fragment is found. Then, the final index is given as the product RARE * NOTFOUND. Defined intervals are: index = 1 all atom centered fragment of the compound have been found in the compounds of the training set, 1 > index >= 0.7 some atom centered fragment of the compound have not been found in the compounds of the training set or are rare fragments, index < 0.7 a prominent number of atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments. 6. Model descriptors range check. This index checks if the descriptors calculated for the predicted compound are inside the range

of descriptors of the training and test set. The index has value 1 if all descriptors are inside the range, 0 if at least one descriptor is out of the range. Defined intervals are: index = True descriptors for this compound have values inside the descriptor range of the compounds of the training set, index = False descriptors for this compound have values outside the descriptor range of the compounds of the training set 7. Global AD Index. The final global index takes into account all the previous indices, in order to give a general global assessment on the applicability domain for the predicted compound. Defined intervals are: 1 >= index > 0.8 predicted substance is into the Applicability Domain of the model, 0.8 => index > 0.65 predicted substance could be out of the Applicability Domain of the model, index <= 0.65 predicted substance is out of the Applicability Domain of the model.

5.3.Software name and version for applicability domain assessment:

5.4.Limits of applicability:

The model was built only for organic chemicals. The user should verify the ADI.

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: No

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:

6.6.Pre-processing of data before modelling:

All the chemical structures were manually checked deleting doubtful compounds, mixture, inorganic compounds and tautomers.

6.7.Statistics for goodness-of-fit:

The statistic for linear regression model is following Training set (220 chemicals) R2 = 0.75 RMSE = 0.1.37

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

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

6.10.Robustness - Statistics obtained by Y-scrambling:

6.11.Robustness - Statistics obtained by bootstrap:

6.12.Robustness - Statistics obtained by other methods:

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: No

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:

7.6.Experimental design of test set:

The dataset were randomly split into training and test set with respectively the 84% and the 16% of the compounds.

7.7.Predictivity - Statistics obtained by external validation:

The statistic for linear regression model is following Training set (43 chemicals) R2 = 0.68 RMSE = 1.49

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:

NA

8.2.A priori or a posteriori mechanistic interpretation:

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

9.2.Bibliography:

[1]Chapter 2 Databases for pesticide ecotoxicity Emilio Benfenati, Elena Boriani, Marian Craciun, Ladan Malazizi, Daniel Neagu, Alessandra Roncaglioni Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes Edited by Emilio Benfenati © 2007 Elsevier B.V. All rights reserved http://ir-library.mmarau.ac.ke:8080/bitstream/handle/123456789/1494/QUANTITATIVE%20STRUCTURE%20ACTIVITY%20RELATIONSHIP%20for%20PESTICIDE.pdf?sequence=1&isAllowed=y

[2]CHAPTER 5 Hybrid systems Nicolas Amaury, Emilio Benfenati, Severin Bumbaru, Antonio Chana, Marian Craciun, Jacques R. Chrétien, Giuseppina Gini, Gongde Guo, Frank Lemke, Viorel Minzu, Johann-Adolf Müller, Daniel Neagu, Marco Pintore, Silviu Augustin Stroia, Paul Trundle Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes Edited by Emilio Benfenati © 2007 Elsevier B.V. All rights reserved http://ir-

library.mmarau.ac.ke:8080/bitstream/handle/123456789/1494/QUANTITATIVE%20STRUCTURE%20ACTIVITY%20RELATIONSHIP%20for%20PESTICIDE.pdf?sequence=1&isAllowed=y
[3]CHAPTER 7 Results of DEMETRA models Nicolas Amaury, Emilio Benfenati, Elena Boriani, Mosè Casalegno, Antonio Chana, Qasim Chaudhry, Jacques R. Chrétien, Jane Cotterill, Frank Lemke, Nadège Piclin, Marco Pintore, Chiara Porcelli, Nicholas Price, Alessandra Roncaglioni, Andrey Toropov Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes Edited by Emilio Benfenati © 2007 Elsevier B.V. All rights reserved http://ir-library.mmarau.ac.ke:8080/bitstream/handle/123456789/1494/QUANTITATIVE%20STRUCTURE%20ACTIVITY%20RELATIONSHIP%20for%20PESTICIDE.pdf?sequence=1&isAllowed=y
[4]Topliss, J. G., and Edwards, R. P. 1979. Chance factors in Studies of Quantitative Structure-Activity Relationships. Journal of Medicinal Chemistry 22 (10):1238-1244. https://pubs.acs.org/doi/abs/10.1021/jm00196a017

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