	QMRF identifier (JRC Inventory): **To be entered by JRC**
	QMRF Title: **Hepatotoxicity model (IRFMN) – v. 1.0.1**
	Printing Date: **Feb 14, 2022**

1.QSAR identifier

1.1. QSAR identifier (title):

Hepatotoxicity model (IRFMN) – v. 1.0.1

1.2. Other related models:

NA

1.3. Software coding the model:

VEGA (https://www.vegahub.eu/)

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

emilio.benfenati@marionegri.it

2.General information

2.1. Date of QMRF:

14/02/2022

2.2. QMRF author(s) and contact details:

[1] Domenico Gadaleta Laboratory of Environmental Chemistry and Toxicology, Istituto di Ricerche Farmacologiche "Mario Negri", IRCCS domenico.gadaleta@marionegri.it

[2] Emilio Benfenati Istituto di Ricerche Farmacologiche Mario Negri - IRCSS Via Mario Negri 2,20156 Milano, Italy emilio.benfenati@marionegri.it https://www.marionegri.it/

2.3. Date of QMRF update(s):

NA

2.4. QMRF update(s):

NA

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 https://www.marionegri.it/

[2] Fabiola Pizzo Laboratory of Environmental Chemistry and Toxicology, Istituto di Ricerche Farmacologiche "Mario Negri", IRCCS

[3] Anna Lombardo Laboratory of Environmental Chemistry and Toxicology Istituto di Ricerche Farmacologiche "Mario Negri", IRCCS anna.lombardo@marionegri.it

2.6. Date of model development and/or publication:

The model was developed in 2016.

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

[1] Pizzo, F., Lombardo, A., Manganaro, A., & Benfenati, E. (2016). A new structure-activity relationship (SAR) model for predicting drug-induced liver injury, based on statistical and expert-based structural alerts. Frontiers in pharmacology, 7, 442.https://www.frontiersin.org/articles/10.3389/fphar.2016.00442/full

[2]Ferrari, T., Cattaneo, D., Gini, G., Golbamaki Bakhtyari, N., Manganaro, A., and Benfenati, E.(2013). Automatic knowledge extraction from chemical structures: the case of mutagenicity prediction. SAR QSAR Environ. Res. 24, 365–383https://www.tandfonline.com/doi/abs/10.1080/1062936X.2013.773376

[3] Benfenati E, Roncaglioni A, Lombardo A, Manganaro A. Integrating QSAR, Read-Across, and Screening Tools: The VEGAHUB Platform as an Example. Advances in Computational Toxicology;Springer; 2019. p. 365-81.

2.8. Availability of information about the model:

The model is non-proprietary and the training set is available.

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

Another QMRF is not available.

3.Defining the endpoint - OECD Principle 1

3.1. Species:

Human

3.2. Endpoint:

QMRF 6. Other Drug Induced Liver Injury (DILI)

3.3. Comment on endpoint:

NA

3.4. Endpoint units:

Adimensional

3.5. Dependent variable:

Binary classification for hepatotoxicity (hepatotoxic / non-hepatotoxic)

3.6. Experimental protocol:

NA

3.7. Endpoint data quality and variability:

Only data on human from literature were considered:

- Fourches et al. (2010) [2], which contains 950 hepatotoxicity data (drugs) on humans, rodents and non-rodent species. We selected only data referring to humans (650 data) and eliminated the rest.
- United States Food and Drug Administration (US FDA) Human Liver Adverse Effects Database [3]. This contains 631 unique pharmaceuticals, 491 of which (non-proprietary data) have adverse drug reaction data for one or more of the 47 liver effects Coding Symbols for Thesaurus of Adverse Reaction (COSTAR) term endpoints. Since only two compounds were labeled as M (marginally active) we eliminated them in order to reduce the uncertainty of the data set.

The two datasets were merged: duplicates and compounds with contrasting experimental values were eliminated. Compounds with concordant experimental activity considered ones. The final data set was fairly balanced, with 510 compounds labeled as hepatotoxic and 440 non-hepatotoxic.

The final dataset was randomly splittd into a training set (760 mono constituent organic compounds) and a test set, *test set 1*, (190 mono constituent organic compounds)

The external validation set (*test set 2*) was retrieved in the Liver Toxicity Knowledge Base (LTKB) Benchmark Dataset developed by the US FDA [7]. 101 chemicals are selected (after elimination of compounds already present in the dataset), 69 labeled as hepatotoxicity and 32 labeled as non-hepatotoxicity.

The VEGA implemented model merged the *test set 1* and the *test set 2* (external validation set) and hence consisted of 291 number of substances (171 labelled hepatotoxic and 120 labeled non-hepatotoxic).

4.Defining the algorithm - OECD Principle 2

4.1. Type of model:

The VEGA model is a re-implementation of a previously developed model [1] and provides a qualitative prediction of hepatotoxicity (drug-induced liver injury) in human (adverse effects): hepatotoxic/ non hepatotoxic/ no prediction. It is implemented inside the VEGA online platform, accessible at: http://www.vegahub.eu

The model is a decision tree based on structural alerts.

4.2. Explicit algorithm:

Decision tree based on structural alerts (SAs)

Categories of hepatotoxicity were assigned based on the presence of structural alerts that were obtained both manually and automatically.

If no SAs are found for the target compound, no prediction is provided and the compound is labeled “unknown (not-predicted).” If one SA is identified, the prediction for the target compound is hepatotoxic or non-hepatotoxic depending on the SA. If more than one SAs is found, the prediction depends on the number of SAs: if more SAs for non-hepatotoxicity are found than those for hepatotoxicity, the target compound is predicted as non-hepatotoxic; otherwise (the number of SAs found for non-hepatotoxicity is lower or equal to the number of SAs found for hepatotoxicity) it is hepatotoxic

4.3. Descriptors in the model:

[1][n,c]1ccn[n,c]c1 Hepatotoxic

[2]NS(=O)(=O) c1ccccc1 Hepatotoxic

[3]OC(=O)C1[C,S][S,O,C] C2CC(=O)N Hepatotoxic

[4]O=C1N~CC =C[N,C]1C2 C~[S,C]CO2 Hepatotoxic

[5]C1[S,C,N,O] c2ccccc2[N, C,S,O]c3ccc cc13 Hepatotoxic

[6][N;!$([N+]);! $(NC=O);!$(N=[N,C,O])] [a] Hepatotoxic

[7]O=C1CCCC CCCCCCCC O1 Hepatotoxic

[8]Nc1[n,c]cc2 C(=O)C(=C Nc2[c,n]1)C(O)=O Hepatotoxic

[9]*N(*)CCC(c 1cccc[n,c]1)c 2cccc[n,c]2 Hepatotoxic

[10]CC=C(C)C= CC=C(C)C= C[R,a] Hepatotoxic

[11]CNC(=O)N(CCCl)N=O Hepatotoxic

[12]C1CC2CCC 3C(CC[C,c]4 [C,c][C,c][C, c][C,c][C,c]3 4)C2C1 non – hepatotoxic

[13]CC(=O)NC1 C2[S,O]CC= C(N2C1=O) C(O)=O non – hepatotoxic

4.4. Descriptor selection:

Manually obtained SAs were defined by a similarity-based clustering of active and inactive chemicals, so that similarity between chemicals in each cluster was minimized. Similarity was determined as described by Floris et al. (2014).

Clusters were further refined with a k-means iterative procedure, that moved compounds between clusters until the intra-cluster similarity was maximized and the inter-cluster similarity was minimized. Clusters that are not characterized by univocal common chemical structures and those with an inter-cluster similarity lower than 0.70 were disregarded. A SA in form of a SMART was then codified for each cluster.

SAs with a percentage of correctly predicted compounds (TN of TP) lower than 60% were discarded. Automated extraction of SAs was performed with SARPy software (Ferrari et al., 2013) that extracts sets of rules by automatically generating and selecting substructures without any a priori knowledge, solely based on their prediction. SAs with a percentage of correctly predicted compounds (TN of TP) lower than 60% were discarded. SAs obtained with the two methods were graphically compared, then in case of similar SAs that matched the same compounds in the training set, only the manually extracted one was kept and the other was eliminated

4.5. Algorithm and descriptor generation:

NA

4.6. Software name and version for descriptor generation:

SARPy

Software for automated extraction of fragments (SAs) flagging for toxicity/non-toxicity)https://www.vegahub.eu/portfolio-item/sarpy

4.7. Chemicals/Descriptors ratio:

760/13 = 58

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.

ADI is defined in this way for this QSAR model´s predictions:

If 1 ≥ AD index ≥ 0.8, the predicted substance is regarded in the Applicability Domain of the model. It corresponds to “good reliability” of prediction.

If 0.8 >AD index ≥ 0.6, the predicted substance could be out of the Applicability Domain of the model. It corresponds to “moderate reliability” of prediction.

If AD index < 0.6, the predicted substance is regarded out of the Applicability Domain of the model. It corresponds to “low reliability” of prediction.

No ADI threshold was used to provide performance calculations of the validation set

5.2. Method used to assess the applicability domain:

The AD and the chemical similarity are measured with the algorithm developed for VEGA. Full details are in the VEGA website (www.vegahub.eu), including the open access paper describing it [6]. 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.

These indices are defined in this way for this QSAR model:

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:

If 1 ≥ index > 0.8, strongly similar compounds with known experimental value in the training set have been found

If 0.8 ≥ index > 0.6, only moderately similar compounds with known experimental value in the training set have been found

If index ≤ 0.6, no similar compounds with known experimental value in the training set have been found

Accuracy (average error) of prediction for similar molecules:

This index takes into account the classification accuracy in prediction for the two most similar compounds found. Values near 1 mean that the predicted compounds fall in an area of the model's space where the model gives reliable predictions (no misclassifications), otherwise the lower is the value, the worse the model behaves. Defined intervals are:

If 1 ≥ index > 0.8accuracy of prediction for similar molecules found in the training set is good

If 0.8 ≥ index > 0.6, accuracy of prediction for similar molecules found in the training set is not optimal

If index ≤ 0.6, accuracy of prediction for similar molecules found in the training set is not adequate

Concordance for 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 disagrees with the values found in the model's space, thus the prediction could be unreliable. Defined intervals are:

If 1≥index >0.8, molecules found in the training set have experimental values that agree with the target compound predicted value

If 0.8 ≥ index > 0.6, similar molecules found in the training set have experimental values that slightly disagree with the target compound predicted value

If index ≤ 0.6, similar molecules found in the training set have experimental values that completely disagree with the target compound predicted value

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:

If index = 1, all atom centered fragment of the compound have been found in the compounds of the training set

If 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

If index < 0.7, a prominent number of atoms centered fragments of the compound have not been found in the compounds of the training set or are rare fragments

5.3. Software name and version for applicability domain assessment:

VEGA

The VEGA software provides QSAR models to predict tox, ecotox, environ, and phys-chemproperties of chemical substances. emilio.benfenati@marionegri.it https://www.vegahub.eu/

5.4. Limits of applicability:

The model is not applicable on inorganic chemicals and those including unusual elements (i.e., different from C, O, N, S, Cl, Br, F, I). Salts can be predicted only if stripped of the counter ion and converted to the neutralized form

6.Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS RN: No

Chemical Name: Yes

Smiles: Yes

Formula: No

INChI: No

MOL file: No

NanoMaterial: No

6.3. Data for each descriptor variable for the training set:

No

6.4. Data for the dependent variable for the training set:

All

6.5. Other information about the training set:

The training set includes 760 unique compounds with a balanced activity distribution.

6.6. Pre-processing of data before modelling:

Data referring to humans (650 data) were extracted from the dataset provided by Fourches et al. (2010), which contains 950 hepatotoxicity data (drugs) on humans, rodents and non-rodent species. In addition, chemicals included in the United States Food and Drug Administration (US FDA) Human Liver Adverse Effects Database were considered (Matthews et al., 2004). For each compound there is an overall activity category (A for active, I for inactive and M for marginally active) referring to five hepatic endpoints(alkaline phosphatase ALP, aspartate aminotransferase AST, alanine aminotransferase ALT, lactate dehydrogenase LDH and gamma-glutamyl transferase GGT). Two M chemicals were discarded, keeping the remaining 629. The two datasets were merged removing duplicates and compounds with contrasting experimental activity between the two datasets. The final dataset includes 950 unique mono-constituent organic substances. It was randomly split into a training set (760 compounds, 80%) and a test set (*test set 1*) (190 compounds, 20%)

6.7. Statistics for goodness-of-fit:

According to model described in the paper [1]

Out of 760 compounds that were present in the training set, 263 were not predicted by the model (unknown, non-predicted). 263 compounds were correctly predicted as hepatotoxic (TP) and 144 were correctly predicted as non-hepatotoxic (TN). 72 molecules experimentally non-hepatotoxic were identified by the model as hepatotoxic (FP) and only 18 compounds experimentally hepatotoxic were predicted as non-hepatotoxic (FN). Performance in the training set was the following: accuracy 81%, sensitivity 93%, specificity 67 and MCC 0.64

For the VEGA implemented model (see section 3.7):

Out of 760 compounds that were present in the training set, 265 were not predicted by the model (unknown, non-predicted). 261 compounds were correctly predicted as hepatotoxic (TP) and 144 were correctly predicted as non-hepatotoxic (TN). 72 molecules experimentally non-hepatotoxic were identified by the model as hepatotoxic (FP) and only 18 compounds experimentally hepatotoxic were predicted as non-hepatotoxic (FN). Performance of the training set was the following: accuracy 82%, sensitivity 94%, specificity 67%.

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

NA

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

NA

6.10. Robustness - Statistics obtained by Y-scrambling:

NA

6.11. Robustness - Statistics obtained by bootstrap:

NA

6.12. Robustness - Statistics obtained by other methods:

NA

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:

No

7.4. Data for the dependent variable for the external validation set:

All

7.5. Other information about the external validation set:

NA

7.6. Experimental design of test set:

For the design of test set, see 6.6.

For the design of the external validation set (*test set 2*), the Liver Toxicity Knowledge Base (LTKB) Benchmark Dataset developed by the US FDA was used (Chen et al., 2011) This dataset contains drugs labeled as most-, less- or no-DILI concern. Only those compounds labeled as most-DILI concern (hepatotoxic) or no-DILI-concern (non-hepatotoxic) were considered. Compounds already present in the training set were eliminated. A final dataset of 101 chemicals, 69 of which were labeled as hepatotoxic and 32 as non-hepatotoxic, was used for validation

7.7. Predictivity - Statistics obtained by external validation:

According to the model described in paper [1] in section 2.7:

For 99 compounds of the *test set 1* (190 molecules in total) the model provided predictions; 48 compounds were correctly identified as hepatotoxic (TP) and 15 as non-hepatotoxic (TN). The number of experimentally negative (non-hepatotoxic) compounds wrongly predicted as hepatotoxic (FP) was 30 and the number of positive compounds (hepatotoxic) wrongly predicted as negative (FN) was 6. Performances are the following: accuracy 63%, sensitivity 88%, specificity 33% and MCC 0.27. 91 compounds were not predicted (unknown, non-predicted)

In the external validation set (*test set 2*, 101 compounds in total), 59 chemicals were predicted by the model ; the numbers of TP and TN was 35 and 5 respectively. 10 compounds were wrongly classified as hepatotoxic (FP) and 9 as non-hepatotoxic (FN). Performances are therefore following: accuracy 68%, sensitivity 80%, specificity 33% and MCC 0.13. 42 compounds were not predicted (unknown, non-predicted),

For the VEGA implemented model (see section 3.7):

Out of 291 (171 labelled hepatotoxic and 120 labeled non-hepatotoxic) compounds that were present in the jointed test set (c.f. section 3.7), 157 were predicted by the model; 82 compounds were correctly predicted as hepatotoxic (TP) and 22 were correctly predicted as non-hepatotoxic (TN). 38 molecules experimentally non-hepatotoxic were identified by the model as hepatotoxic (FP) and 15 compounds experimentally hepatotoxic were predicted as non-hepatotoxic (FN). 134 compounds were not predicted by the model (unknown, non-predicted). Performance in the test set was therefore the following: accuracy 66%, sensitivity 85%, specificity 37%.

7.8. Predictivity - Assessment of the external validation set:

NA

7.9. Comments on the external validation of the model:

The model had according to the validations – in particular the external validation – a low specificity meaning a large number of false positives relative to the number of true positives for chemical universes with a low prevalence of hepatotoxic substances.

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:

A posteriori

8.3. Other information about the mechanistic interpretation:

NA

9.Miscellaneous information

9.1. Comments:

Based on the validations results and due to the type of model it is recommended to use positive predictions only as rough indications preferably together with other relevant information.

9.2. Bibliography:

[1]Chen, M. J., Vijay, V., Shi, Q., Liu, Z. C., Fang, H., and Tong, W. D. (2011). FDA approved drug labeling for the study of drug-induced liver injury. Drug Discov. Today 16, 697–703 doi:10.1016/j.drudis.2011.05.007

[2]Fourches, D., Barnes, J. C., Day, N. C., Bradley, P., Reed, J. Z., and Tropsha, A. (2010).Cheminformatics analysis of assertions mined from literature that described drug-induced liver injury in different species. Chem. Res. Toxicol. 23, 171–183 doi: 10.1021/tx900326k

[3]Matthews, E. J., Kruhlak, N. L.,Weaver, J. L., Benz, R. D., and Contrera, J. F. (2004). Assessment of the health effects of chemicals in humans: II. Construction of an adverse effects database for QSAR modeling. Curr. Drug Disc. Technol. 4, 243–254. doi: 10.2174/1570163043484789

[4]Pizzo, F., Lombardo, A., Manganaro, A., & Benfenati, E. (2016). A new structure-activity relationship (SAR) model for predicting drug-induced liver injury, based on statistical and expert-based structural alerts. Frontiers in pharmacology, 7, 442.

[5]Ferrari, T., Cattaneo, D., Gini, G., Golbamaki Bakhtyari, N., Manganaro, A., and Benfenati, E.(2013). Automatic knowledge extraction from chemical structures: the case of mutagenicity prediction. SAR QSAR Environ. Res. 24, 365–383. doi: 10.1080/1062936X.2013.773376

[6] Floris, M., Manganaro, A., Nicolotti, O. et al. A generalizable definition of chemical similarity for read-across. J Cheminform 6, 39 (2014). https://doi.org/10.1186/s13321-014-0039-1

[7] LTKB Benchmark Dataset | FDA.. Retrieved April 21, 2022, from https://www.fda.gov/science-research/liver-toxicity-knowledge-base-ltkb/ltkb-benchmark-dataset

9.3. Supporting information:

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

All available datasets are present in the model inside the VEGA software.

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