	QMRF identifier (JRC Inventory): **To be entered by JRC**
	QMRF Title: **In vivo Micronucleus activity (IRFMN) – v. 1.0.2**
	Printing Date: **June 7, 2022**

1.QSAR identifier

1.1.QSAR identifier (title):

In vivo Micronucleus activity (IRFMN) – v. 1.0.2

1.2.Other related models:

Other genotoxicity models (including models for carcinogenicity and mutagenicity) are implemented inside VEGA online platform, accessible at: https://www.vegahub.eu/.

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:

June 2022

2.2.QMRF author(s) and contact details:

[1] Giuseppa Raitano, Istituto di Ricerche Farmacologiche Mario Negri - IRCSS Via Mario Negri 2, 20156 Milano, Italy, giuseppa.raitano@marionegri.it, https://www.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):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

[1] Giuseppa Raitano, Istituto di Ricerche Farmacologiche Mario Negri – IRCSS Via Mario Negri 2, 20156 Milano, Italy, https://www.marionegri.it/, giuseppa.raitano@marionegri.it

[2] Alberto Manganaro, IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, 20156 Milano, Italy, https://www.marionegri.it/, alberto.manganaro@marionegri.i.it.

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

2.6.Date of model development and/or publication:

November 2020

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

1. Van Bossuyt M, Raitano G, Honma M, Van Hoeck E, Vanhaecke T, Rogiers V, Mertens B, Benfenati E, New QSAR models to predict chromosome damaging potential based on the in vivo micronucleus test (2020) Toxicology Letters, 329, 80-84, ISSN 0378-4274, https://doi.org/10.1016/j.toxlet.2020.04.016.

2. Ferrari T, Cattaneo D, Gini G, Golbamaki Bakhtyari N, Manganaro A and Benfenati E, Automatic knowledge extraction from chemical structures: the case of mutagenicity prediction, SAR and QSAR in Environmental Research (2013), vol. 24 issue 5, 365-83. http://dx.doi.org/10.1080/1062936X.2013.773376

3. Manganaro A, Pizzo F, Lombardo A, Pogliaghi A and Benfenati, E (2015). Predicting persistence in the sediment compartment with a new automatic software based on the k-Nearest Neighbor (k-NN) algorithm. Chemosphere. 144, 1624-1630

5. https://www.vegahub.eu/

2.8.Availability of information about the model:

The model is non-proprietary and the training set is available. In addition, complete documentation about the model, including training and test sets, is available on the guideline of the model, inside the VEGA application.

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:

Rodents

3.2.Endpoint:

TOX, 7.6.2. Genetic toxicity in vivo, OECD TG 474 [2]

3.3.Comment on endpoint:

Genotoxicity is a broader term and refers to processes which alter the structure, information content or segregation of DNA and are not necessarily associated with mutagenicity. Thus, tests for genotoxicity include tests which provide an indication of induced damage to DNA (but not direct evidence of mutation) via effects such as DNA strandbreaks, unscheduled DNA synthesis (UDS), sister chromatid exchange (SCE), DNA adduct formation or mitotic recombination, as well as tests for mutagenicity. The mammalian in vivo micronucleus test is especially relevant for assessing genotoxicity because, although they may vary among species, factors of in vivo metabolism, pharmacokinetics and DNA repair processes are active and contribute to the responses. An in vivo assay is also useful for further investigation of genotoxicity detected by an in vitro system.

3.4.Endpoint units:

Adimensional

3.5.Dependent variable:

Binary classification as: Genotoxic/ Non-genotoxic.

3.6.Experimental protocol:

The mammalian in vivo micronucleus test is used for the detection of damage induced by the test chemical to the chromosomes or the mitotic apparatus of erythroblasts. The test evaluates micronucleus formation in erythrocytes sampled either in the bone marrow or peripheral blood cells of animals, usually rodents.

3.7.Endpoint data quality and variability:

Newly formed micronucleated erythrocytes are identified and quantitated by staining followed by either visual scoring using a microscope, or by automated analysis. Counting sufficient immature erythrocytes in the peripheral blood or bone marrow of adult animals is greatly facilitated by using an automated scoring platform. Such platforms are acceptable alternatives to manual evaluation.

Comparative studies have shown that such methods, using appropriate calibration standards, can provide better inter- and intra-laboratory reproducibility and sensitivity than manual microscopic scoring.

Automated systems that can measure micronucleated erythrocyte frequencies include, but are not limited to, flow cytometers, image analysis platforms, and laser scanning cytometers.

The models have been developed using a set of 1228 compounds and their experimental results of in vivo micronucleus test, classified as genotoxic (378) and non-genotoxic (850).

This is the extended version of a selected dataset as described in Van Bossuyt M et al. 2020 [1].
Besides the updated versions of the public databases, this new version of the dataset has data from further sources: Bioassay Genetox Conclusion Dataset (https://manticore.niehs.nih.gov/datasets/search/trf) [3], EFSA genotoxicity database (https://data.europa.eu/euodp/data/dataset/database-pesticide-genotoxicity-endpoints) [4] and non-private data on botanicals from NCSTOX project (http://www.unitis.org/en/ncs-tox-project,378.html) [5].
The data were curated by deletion of duplicates, salts, mixtures, polymers and inorganic compounds. Substances with ambiguous or contradictory test results were also excluded. Next, in-house software was applied to neutralize the substances' SMILES strings based on the required format for the prediction model.
The dataset was split in training (1209 compounds) and test set (19 botanicals) considering the experimental distribution of the molecules within the whole dataset (378 genotoxic (31%) and 850 non-genotoxic (69%)) and the presence of botanicals.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

The model performs a consensus assessment based on the predictions of two single models: 1) SAR in python (SARpy) and 2) k-nearest neighbor (k-NN).

4.2.Explicit algorithm:

Consensus model:

- If both models' predictions have same reliability but discordant outcomes, the consensus model cannot predict.
- In all the other cases of discordant predictions, the consensus model reports the prediction with the higher reliability.
- If one of the two models' output is a missing value (this could happen for Sarpy if no alerts are found, and for k-NN if not enough similar compounds are found), the consensus prediction collapses to the unique available prediction and the reliability is assigned according to the VEGA method.
- Predictions based on concordant outcomes have higher reliability.

Single models description:

1) **SARpy model** was built as a set of rules (active and inactive) extracted automatically by SARpy software (http://sarpy.sourceforge.net/, SAR in python) from the training set.

Generally, the SARpy software extracts each possible fragment from a set of molecular structures and correlates these substructures with the activity of the molecules that contain them. As the last step, it selects fragments suitable to become SAs (structural alerts), based on their prediction performance on the training set.

Double SAs and those with low accuracy were removed. At the end of the selection, a total of 89 SAs were applied to predict by using a stepwise approach. This approach was selected since the performance on the training set were the best. If no alert is found in the structure of the target compound, the model does not give the prediction.

2) **k-NN model** was selected by testing several possible combinations using in-house software (istKNN) according to its performance in leave one out and coverage.

The model performs a read-across approach and it is based on the similarity index used inside the VEGA platform.

The algorithm for the prediction involves the following steps:

a) The first k molecules with the closest similarity to the target compound are extracted.
b) Molecules with a similarity index lower than a selected threshold (S1) are excluded.
c) If no molecules are left, no prediction is provided (missing value).
d) If only one molecule is left, it is used as prediction only if its similarity value is equal to or higher than a given threshold (S2), otherwise no prediction is provided (missing value).

e) In all other cases, the prediction is calculated as a weighted consensus of the experimental values among the remaining molecules. A score for each class is calculated as the sum of the weights of compounds experimentally belonging to the class itself.

f) Finally, the class with the highest score is chosen as the prediction to be provided. The weights (similarity values) can be raised to the power of a given value (E) called the enhance factor, as for integers larger than 1 the result is to enhance the role of molecules with higher similarity values in the prediction.

4.3.Descriptors in the model:

The predictions of the two single models (SARpy and k-NN) are the input for this model so no other types of descriptors are directly used by the consensus.

4.4.Descriptor selection:

Not applicable

4.5.Algorithm and descriptor generation:

Not applicable

4.6.Software name and version for descriptor generation:

Not applicable

4.7.Chemicals/Descriptors ratio:

Not applicable

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The model is suitable for heterogeneous organic, monocostituent chemicals. The applicability domain of predictions of the single models is assessed using an Applicability Domain Index (ADI) as implemented in VEGA that has values from 0 (worst case) to 1 (best case). The ADI is calculated by grouping several other indices, each one taking into account a particular issue of the applicability domain. Most of the indices are based on the calculation of the most similar compounds found in the training and test set of the model, calculated by a similarity index that consider molecule's fingerprint and structural aspects (count of atoms, rings and relevant fragments). For the consensus model, see 5.2.

If 1 ≥ ADI > 0.9, predicted substance is regarded in the Applicability Domain of the model and predictions are characterized by good reliability

If 0.9 ≥ ADI > 0.65 predicted substance could be out of the Applicability Domain of the model and predictions are characterized by moderate reliability

If ADI ≤ 0.65 predicted substance is regarded out of the Applicability Domain of the model and predictions are characterized by low reliability

5.2.Method used to assess the applicability domain:

The consensus algorithm uses the Applicability Domain assessment of each single model's prediction as its weight, so that the final assessment will be more influenced by the single models that produced more reliable predictions. The Applicability Domain assessment of each model is converted to a numerical value in the range [0..1] with the following scheme:

VEGA Applicability Domain qualitative assessment	Applicability Domain reference value
Good reliability	0.9
Moderate reliability	0.6
Low reliability	0.2

The consensus score (CS) is calculated separately for the two outcomes as follow:

$$\begin{cases} CS_M = \sum IR_M / nr.tot\ models \\ CS_{nM} = \sum IR_{nM} / nr.tot\ models \end{cases}$$

The final class assignment is done according to the value of the CS: the compound is assigned to the positive class if CSM≥ CSnM. If at least one experimental value is available, the CS is calculated as 1 if all values are concordant or as a ratio depending on the prevalence of the experimental responses. In this case, the reported number of models used refers only to the number of models having an experimental value.

The Applicability domain and 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 [5]. The VEGA 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 > 0.80, strongly similar compounds with known experimental value in the training set have been found

If 0.80 ≥ 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 index < 0.5, accuracy of prediction for similar molecules found in the training set is good

If 0.9 > index ≥ 0.5, accuracy of prediction for similar molecules found in the training set is not optimal

If index ≥ 0.9, 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 index < 0.5, molecules found in the training set have experimental values that agree with the target compound predicted value

If 0.9 > index ≥ 0.5, similar molecules found in the training set have experimental values that slightly disagree with the target compound predicted value

If index ≥ 0.9, similar molecules found in the training set have experimental values that completely disagree with the target compound predicted value

Maximum error of prediction between similar molecules:
This index takes into account the maximum error in prediction between the two most similar compounds. Values near 0 means that the predicted compounds fall in an area of the model's space where the model gives reliable predictions without any outlier value. Defined intervals are:

If index < 0.5, the maximum error in prediction of similar molecules found in the training set has a low value, considering the experimental variability

If 0.9 > index ≥ 0.5, the maximum error in prediction of similar molecules found in the training set has a moderate value, considering the experimental variability

If index ≥ 0.9, the maximum error in prediction of similar molecules found in the training set has a high value, considering the experimental variability

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 software

5.4.Limits of applicability:

Inorganic compounds, UVCB, mixtures, polymers and salts cannot be predicted.

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:

No

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

All

6.5.Other information about the training set:

More information about the training set is within the guide of the model.

6.6.Pre-processing of data before modelling:

The data were curated by deletion of duplicates, salts, mixtures, polymers and inorganic compounds. Substances with ambiguous or contradictory test results were also excluded. Next, in-house software was applied to neutralize the substances' SMILES strings based on the required format for the prediction model.

The dataset was split in training (1209 compounds) and test set (19 botanicals) considering the experimental distribution of the molecules within the whole dataset (378 genotoxic (31%) and 850 non-genotoxic (69%)) and the presence of botanicals.

6.7.Statistics for goodness-of-fit:

Not predicted: 2

N = 1207, Accuracy 0.99, Sensitivity 0.99, Specificity 1.00, MCC 0.99. TP 363, TN 839, FP 2, FN 3

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:

No

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

All

7.5.Other information about the external validation set:

More information about the test set is within the guide of the model.

7.6.Experimental design of test set:

The dataset was split in training (1209 compounds) and test set (19 botanicals) considering the experimental distribution of the molecules within the whole dataset (378 genotoxic (31%) and 850 non-genotoxic (69%)) and the presence of botanicals.

7.7.Predictivity - Statistics obtained by external validation:

Test set (19 chemicals)

Accuracy = 0.94 Sensitivity = 0.90 Specificity = 1.00 MCC = 0.9. TP 9, TN 9, FP 0, FN 1

7.8.Predictivity - Assessment of the external validation set:

See 6.6

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:

The model includes structural alerts to identify both genotoxic and non-genotoxic compounds. The VEGA system provides, in the final PDF report for the prediction, a set built with the most similar compounds found in the training and test set of the model, coupled to the assessment of the target. An expert-based analysis of these compounds like the predicted one, which are provided with their experimental activity, can lead to a further mechanistic interpretation of the results given by the model.

8.2.A priori or a posteriori mechanistic interpretation:

A posteriori, see 8.1

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

9.2.Bibliography:

[1] Van Bossuyt M, Raitano G, Honma M, Van Hoeck E, Vanhaecke T, Rogiers V, Mertens B, Benfenati E, New QSAR models to predict chromosome damaging potential based on the in vivo micronucleus test (2020) Toxicology Letters, 329, 80-84, ISSN 0378-4274, https://doi.org/10.1016/j.toxlet.2020.04.016.

[2] OECD guideline 474 (2016) OECD Guidelines for the Testing of Chemicals No. 474: Mammalian Erythrocyte Micronucleus Test. Organization for Economic Cooperation and Development; Paris, France. Available online at: https://www.oecd-ilibrary.org/environment/test-no-474-mammalian-erythrocyte-micronucleus-test_9789264264762-en

[3] https://manticore.niehs.nih.gov/datasets/search/trf

[4] https://data.europa.eu/euodp/data/dataset/database-pesticide-genotoxicity-endpoints

[5] http://www.unitis.org/en/ncs-tox-project,378.html

[6] Floris, M., Manganaro, A., Nicolotti, R. Medda, G. F. Mangiatordi, E. Benfenati (2014). A generalizable definition of chemical similarity for read-across, J. Cheminform., 6, 39.

Guidance on Information Requirements and Chemical Safety Assessment Chapter R.7a: Endpoint specific guidance. Version 6.0 July 2017.
Available online at: https://echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf

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