	QMRF identifier (JRC Inventory): **To be entered by JRC**
	QMRF Title: **Mutagenicity (Ames test) model (CAESAR) - version 2.1.13**
	Printing Date: **15-04-2022**

1.QSAR identifier

1.1.QSAR identifier (title):

Mutagenicity (Ames test) model (CAESAR) - version 2.1.13

1.2.Other related models:

Two models have been created and validated using a large set of molecular structures accompanied by the respective mutagenic toxicity experimental test results on Salmonella test. Model A is based on data mining with support vector machines (SVM) and Model B is based on expert knowledge coded as structural alerts (SA). The final model C combines models A and B to achieve a better predictive performance.

Version 2.1.6

First official release published in the VEGA platform.

Version 2.1.7

Some minor code updates, mainly due to changes in the core. There are NO changes in prediction values and AD assessment.

Version 2.1.8

This version is updated with the new calculation core (1.0.26) where similarity algorithm is slightly changed. The new version considers halogen atoms are really similar, especially Chlorine and Bromine atoms are considered almost the same. The main difference with previous algorithm can be thus seen just for halogenated compounds.

A more precise check for similarity has been introduced for the extraction of experimental values, in order to avoid mismatches (as the similarity index is based on fingerprints, there are some rare cases in which a value equal to 1 does not points to a exactly isomorph compound).

Some minor bugs in the procedure for reading molecule structures have been fixed; some compounds, previously not loaded, could now be correctly processed.

There are NO changes in prediction values, but as similarity is changed some small differences in AD assessment can be found.

Version 2.1.9

This version is updated with the new calculation core (1.0.27), that generates a graphically renewed PDF report. In this version, the propositions for prediction and assessment are changed, but there are NO changes in their values.

Version 2.1.10

This version is updated with the new calculation core (1.0.30).

In this version, a major bug was fixed: molecule structures were modified while calculating some descriptors, thus when this model was executed together with other models (in VegaNIC), the results of the models executed after this one could be affected by some errors.

Version 2.1.12

This version is updated with the new calculation core (1.1.1) based on a new release of the CDK libraries (1.4.9). These updates can influence the calculation, so there could be some changes in the predictions produced.

The Structural Alerts implementation have been revised and corrected, there could be changes in the SAs found and thus also in the prediction of some compounds.

The new calculation core implements a new version of the algorithm used for calculating the similarity index. This means that the list of similar molecules given as part of the applicability domain evaluation will often be different from the ones produced by older releases of the model. Furthermore, the applicability domain index (ADI) itself and the final assessment could often be different.

Model statistics in the current guide have been updated with the new values.

Version 2.1.13

This version is updated with the new calculation core (1.2.8). This update can influence some calculation, in particular similarity evaluation, so there could be some changes in the applicability domain values produced.

Furthermore, some predictions can be different from previous versions as in the new calculation core Benigni/Bossa alerts have been completely revised (using ToxTree 2.6.13 as reference for their implementation)

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:

15-04-2022

2.2.QMRF author(s) and contact details:

[1] Alessio Gamba Istituto di Ricerche Farmacologiche Mario Negri - IRCSS Via Mario Negri 2, 20156 Milano, Italy alessio.gamba@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):

NA

2.4.QMRF update(s):

NA

2.5.Model developer(s) and contact details:

[1] Thomas Ferrari Department of Electronics and Information (DEI), Politecnico di Milano

[2] Alberto Manganaro RCCS-Istituto di Ricerche Farmacologiche Mario Negri Via La Masa 19, 20156 Milano, Italy alberto.manganaro@marionegri.it

2.6.Date of model development and/or publication:

The original version of the model was developed in 2010 [1]. The last version (2.1.13) was released in 2016.

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

[1] Ferrari T., Gini G. An open source multistep model to predict mutagenicity from statistical analysis and relevant structural alerts. Chemistry Central Journal (2010), 4 (Suppl 1):S2

This reference, as well as the statistics reported there, is directly related to the original version of the model.

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:

Salmonella typhimurium (Ames test)

3.2.Endpoint:

4.Human Health Effects 4.10.Mutagenicity

3.3.Comment on endpoint:

Mutagenic toxicity is the capacity of a substance to cause genetic mutations. This property is of high public concern because it has a close relationship with carcinogenicity and eventually reproductive toxicity: most of the mutagenic substances are suspected carcinogenic substance in case a genotoxic mechanism is considered. The Ames test is the basic in vitro assay to detect mutagens. The relevant test guideline covering this endpoint is OECD TG 471. The training set is based on test results from either the original version of the test guideline from 1983 or a newer version from 1997. The endpoint covers the DNA base-pair substitution and frameshift mutagenic mechanisms that are covered by the Ames tester strains: TA 1535, TA100, TA 98, and TA 1537 or TA97 or TA 97a. A part of the training set data additionally covers cross-linking mutagenic events measured by the inclusion of the E.coli WP2 or E.coli WP2 (pKM101) or TA 102 test strains. The test strains for DNA cross-links were included in the 1997 guideline update. As the training set does not systematically cover DNA cross-links, mutagenic substances acting by this mechanism may be under-predicted.

The endpoint is measured on the parent compound and the metabolites generated in vitro by the employed S9 mix of enzyme-induced rodent liver homogenates. In a few cases, liver homogenates from hamsters may have been used.

3.4.Endpoint units:

Adimensional

3.5.Dependent variable:

The dependent variable is mutagenic effect, as binary classification: 0 (non-mutagenic), 1 (mutagenic)

3.6.Experimental protocol:

Ames test: an in vitro model of chemical mutagenicity and carcinogenicity, and consists of a range of bacterial strains that together are sensitive to a large array of DNA-damaging agents

3.7.Endpoint data quality and variability:

For the development and the validation of the model, the Bursi Mutagenicity Dataset was used [ref.4, sect.9.2]. The estimated inter-laboratory reproducibility rate of Salmonella test data is 85% [ref.2, sect.9.2].

Experimental procedure of data according to OECD TG 471

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

An integrated model, Model C, was arranged by cascading the two models: Model A, a trained SVM classifier with an additional Model B for false negatives (FNs) correction based on SAs. The SVM classifier is the one described in the section 4.2 of the paper proposing the final model (see 2.7), while the rule base for the expert filter was extracted from the Benigni/Bossa SAs [ref.4; sect.9.2] set.

4.2.Explicit algorithm:

Data mining with SVM coupled with knowledge based SAs for the correction of FNs.

The model consists of a complex architecture based on support vector machines model revised by structural alerts. First, the SVM identifies mutagens. The predicted non-mutagens are then processed with the second

model, Model B, based on two sets of structural alerts. If an alert of the first set is found (see 4.3 descriptors from #26 to #37), the chemical is labelled "mutagen"; if an alert of the second set is found (see 4.3 descriptors from #38 to #41), the chemical is labelled "suspicious mutagen". Unaffected chemicals are finally labelled "non-mutagens".. An integrated model, Model C, was arranged by cascading the two models: Model A, a trained SVM classifier with an additional Model B for false negative (FN) removal based on SAs.

4.3.Descriptors in the model:

[1]SsCH3_acnt Count of all (– CH3) groups in molecule

[2]SdCH2_acnt Count of all (= CH2) groups in molecule

[3]SssCH2_acnt Count of all (– CH2 –) groups in molecule

[4]SdsCH_acnt Count of all (= CH –) groups in molecule

[5]SaaCH_acnt Count of all (CH) groups in molecule

[6]SsssCH_acnt Count of all (> CH –) groups in molecule

[7]SdssC_acnt Count of all (= C <) groups in molecule

[8]SaasC_acnt Count of all (CH) groups in molecule

[9]SaaaC_acnt Count of all (CH) groups in molecule

[10]SssssC_acnt Count of all (> C <) groups in molecule

[11]SsNH2_acnt Count of all (– NH2) groups in molecule

[12]StN_acnt Count of all (N) groups in molecule

[13]SdsN_acnt Count of all (= N –)groups in molecule

[14]SaaN_acnt Count of all (N)groups in molecule

[15]SsssN_acnt Count of all (> N –)groups in molecule

[16]SdaaN_acnt Count of all (N) groups in molecule

[17]SsOH_acnt Count of all (– OH) groups in molecule

[18]SdO_acnt Count of all (= O) groups in molecule

[19]SssO_acnt Count of all (– O –) groups in molecule

[20]SaaO_acnt Count of all (O) groups in molecule

[21]SHCHnX_Acnt Count of all CH or CH2 groups with a -F or -Cl also bonded to the carbon

[22]Gmin Smallest atom E-State value in molecule

[23]idwbar Bonchev-Trinajsti mean information content

[24]ALOGP (DRAGON) Ghose-Crippen octanol water coefficient (calculated by DRAGON)

[25]nrings Number of rings (cyclomatic number)in a molecular graph

[26]SA 1 Acyl halides

[27]SA 6 Propiolactones or propiosultones

[28]SA 12 Quinones

[29]SA 13 Hydrazine

[30]SA 14 Aliphatic azo and azoxy

[31]SA 16 alkyl carbamate and thiocarbamate

[32]SA 18 Polycyclic Aromatic Hydrocarbons

[33]SA 21 alkyl and aryl N-nitroso groups

[34]SA 22 Azide and triazene groups

[35]SA 25 Aromatic nitroso group

[36]SA 28bis Aromatic mono- and dialkylamine

[37]SA 29 Aromatic diazo

[38]SA 7 Epoxides and aziridines

[39]SA 8 Aliphatic halogens

[40]SA 19 Heterocyclic Polycyclic Aromatic Hydrocarbons

[41]SA 27 Nitro-aromatic

4.4.Descriptor selection:

For the SVM classifier, 254 molecular descriptors were initially calculated using the MDL QSAR commercial software. Then, a subset of 25 descriptors was selected by using the tools provided by the Weka 3.5.8 environment for data mining. The BestFirst algorithm was used as bidirectional search method in the descriptor subsets, using as subset evaluator the 5-fold cross-validation score on the training set (in short: BestFirst algorithm searches the space of attribute subsets by greedy hill climbing, considering all possible single attribute additions and/or deletions at a given point, with a backtracking facility to explore also non-improving nodes). The structural alerts were selected from the Benigni/Bossa set of 30 genotoxic alerts after an analysis of their individual effects, evaluated on the structures of the training set labelled non-mutagenic by 5-fold cross-validation of the model.

4.5.Algorithm and descriptor generation:

1D and 2D descriptors

4.6.Software name and version for descriptor generation:

[1] MDL_QSAR software

http://mdl.com

[2] Toxtree

SAs have been implemented by using SMARTS within CAESAR.

[3] Ideaconsult Ltd

https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/qsar_tools

[4] DRAGON for LOGP

http://www.talete.mi.it

[5] VEGA

https://www.vegahub.eu

4.7.Chemicals/Descriptors ratio:

3367 chemicals (training) / 1 descriptors = 82.1

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.
Indices are calculated on the first k = 3 most similar molecules, each having Sk similarity value with the target molecule.

Similarity index (IdxSimilarity) is calculated as:

$$\frac{\sum_k S_k}{k} \times (1 - Diam^2)$$

where *Diam* is the difference in similarity values between the most similar molecule and the *k*-th molecule.

Accuracy index (*IdxAccuracy*) is calculated as:

$$\frac{\sum_c log\ (1 + S_c)}{\sum_k log\ (1 + S_k)}$$

where the molecules with *c* index are the subset of the *k* molecules where the prediction of the model matches with the experimental value of the molecule.

Concordance index (*IdxConcordance*) is calculated as:

$$\frac{\sum_c log\,(1+S_c)}{\sum_k log\,(1+S_k)}$$

where the molecules with *c* index are the subset of the *k* molecules where the experimental value of the molecule matches with the prediction made for the target molecule.

ACF contribution (*IdxACF*) index is calculated as

$$ACF = rare \times missing$$

where: *rare* is calculated on the number of fragments found in the molecule and found in the training set in less than 3 occurences as following: if the number is 0, *rare* is set to 1.0; if the number is 1, *rare* is set to 0.6; otherwise *rare* is set to 0.4

missing is calculated on the number of fragments found in the molecule and never found in the training set as following: if the number is 0, *missing* is set to 1.0; if the number is 1, *missing* is set to 0.6; otherwise *missing* is set to 0.4

Descriptors Range (*IdxDescRange*) index is calculated as 1.0 if all molecular descriptors used in the prediction fall within the range of descriptors used in the whole training set, 0.0 otherwise.

AD final index is calculated as following:

$$ADI = (IdxSimilarity^{0.5} \times IdxAccuracy^{0.25} \times IdxConcordance^{0.25}) \times IdxACF \times IdxDescRange$$

If 1 ≥ AD index ≥ 0.9, the predicted substance is into the Applicability Domain of the model. It corresponds to good reliability of prediction.

If 0.9 > AD index ≥ 0.7, the predicted substance could be out of the Applicability Domain of the model. It corresponds to moderate reliability of prediction.

If AD index < 0.7, the predicted substance is out of the Applicability Domain of the model and corresponds to low reliability of prediction.

5.2. Method used to assess the applicability domain:

The chemical similarity is measured with the algorithm developed for VEGA. Full details are in the VEGA website (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.

5.3. Software name and version for applicability domain assessment:

VEGA

Included in the VEGA software and automatically displayed when running the model

emilio.benfenati@marionegri.it

https://www.vegahub.eu/

5.4. Limits of applicability:

The model is not applicable to inorganic chemicals and substances containing unusual elements (i.e., different from C, O, N, S, P, Cl, Br, F, I). Salts can be predicted only if converted to the neutralized form.

The model is suitable for compounds that have the descriptors in the

following ranges:

[1] SsCH3_acnt min 0 - max 16; SdCH2_acnt min 0 - max 3;

[2] SssCH2_acnt min 0 - max 39; SdsCH_acnt min 0 - max 18;

[3] SaaCH_acnt min 0 - max 20; SsssCH_acnt min 0 - max 26;

[4] SdssC_acnt min 0 - max 36; SaasC_acnt min 0 - max 18;

[5] SaaaC_acnt min 0 - max 12; SssssC_acnt min 0 - max 10;

[6] SsNH2_acnt min 0 - max 8; StN_acnt min 0 - max 4;

[7] SdsN_acnt min 0 - max 6; SaaN_acnt min 0 - max 5;

[8] SsssN_acnt min 0 - max 6; SdaaN_acnt min 0 - max 2;

[9] SsOH_acnt min 0 - max 14; SdO_acnt min 0 - max 31;

[10] SssO_acnt min 0 - max 14; SaaO_acnt min 0 - max 2;

[11] SHCHnX_Acnt min 0 - max 6; Gmin min -9.06 - max 2.25;

[12] idwbar min 0 - max 14.28; nrings min 0 - max 10;ALOGP min -12.9 - max 13.59;

The user has also to evaluate the ADI described in 5.1.

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: Yes

Formula: Yes

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:

The training set is composed of 3367 substances (1883 positive, 1484 negative)

The test set is composed of 837 substances (465 positive, 372 negative)

6.6.Pre-processing of data before modelling:

All chemical structures have been checked manually

6.7.Statistics for goodness-of-fit:

144 training set compounds were predicted as “Suspicious”.

If “Suspicious” predictions are omitted:

Training set number: 3253

Accuracy = 0.92

Sensitivity = 0.97

Specificity = 0.86

TP 1804, TN 1188, FP 200, FN 61

39 test set compounds were predicted as “Suspicious

If “Suspicious” predictions are omitted:

Test set number: 798

Accuracy = 0.83

Sensitivity = 0.90

Specificity = 0.74

TP 403, TN 260, FP 92, FN 43

13% of False Negatives in the SVM predictions are corrected by the first set of structural alerts. By applying even the second set of alerts (i.e., "suspicious" predictions are taken as "mutagenic") more than one-third of False Negatives is corrected (35%) boosting sensitivity to 90% without noticeably downgrading prediction accuracy

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:

NA

7.2.Available information for the external validation set:

The external validation set is composed of a set of data not in common with the training and the test set of the model. Those data were selected from a big dataset comprising public and proprietary data [5] [6].

7.3.Data for each descriptor variable for the external validation set:

NA

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

NA

7.5.Other information about the external validation set:

The external validation set is composed of 14517 substances, 2878 experimentally positive and 11639 experimentally negative on Ames test.

7.6.Experimental design of test set:

No selection of chemicals prior to experimentation

7.7.Predictivity - Statistics obtained by external validation:

Five compounds were not predicted (molecule error: unable to normalize SMILES string or model error: unable to calculate some molecular descriptors), then the available predictions for the statistical assessment were 14512. 834 compounds were predicted as “Suspicious”.

We applied AD index thresholds to perform predictions on the external validation set and if “Suspicious” predictions are omitted, the results are:

The predictions of 3106 substances are in AD. AD index >=0.9.

Sensitivity	Specificity	Accuracy	MCC
0.83	0.82	0.83	0.61

TP 670, TN1897, FP 406, FN 133

The predictions of 5028 substances could be out of the AD. 0.9> AD index >= 0.7

Sensitivity	Specificity	Accuracy	MCC
0.69	0.76	0.75	0.37

TP 584, TN 3197, FP 985, FN 262

The predictions of 5544 substances are out of the AD. AD index <0.7

Sensitivity	Specificity	Accuracy	MCC
0.63	0.60	0.60	0.17

TP 582, TN2766, FP 1857, FN 339

7.8.Predictivity - Assessment of the external validation set:

NA

7.9.Comments on the external validation of the model:

The distribution of the external validation dataset is unbalanced: the 80% of the compounds is non mutagenic experimentally.

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

The model includes SAs to identify toxic compounds, according to the mechanistic basis described by the Benigni-Bossa rules. In addition a stocastic model is included, to provide basis also for negative results.

8.2.A priori or a posteriori mechanistic interpretation:

A priori

8.3.Other information about the mechanistic interpretation:

NA

9.Miscellaneous information

9.1.Comments:

NA

9.2.Bibliography:

[1]Ferrari T & Gini G (2010) An open source multistep model to predict mutagenicity from statistical analysis and relevant structural alerts. Chemistry Central Journal , 4(Suppl 1):S2

http://www.journal.chemistrycentral.com/content/4/S1/S2

[2]Piegorsch WW & Zeiger E (1991) Measuring intra-assay agreement for the Ames salmonella assay. In Statistical Methods in Toxicology, Lecture Notes in Medical Informatics. Edited by Hotorn L. Springer-Verlag, 35-41

[3]Benigni R, Bossa C, Jeliazkova N, Netzeva T & Worth A (2008). The Benigni / Bossa rulebase for mutagenicity and carcinogenicity - a module of Toxtree. EUR 23241 EN. http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/1028/1/eur%20report%20benigni%20130208%20final.pdf

[4]Kazius J, Mcguire R & Bursi R (2005) Derivation and validation of toxicophores for mutagenicity prediction. Journal of Medicinal Chemistry, 48(1),312-320.

http://pubs.acs.org/doi/abs/10.1021/jm040835a

[5] Honma M, Kitazawa A, Cayley A, Williams RV, Barber C, Hanser T, Saiakhov R, Chakravarti S, Myatt GJ, Cross KP, Benfenati E, Raitano G, Mekenyan O, Petkov P, Bossa C, Benigni R, Battistelli CL, Giuliani A, Tcheremenskaia O, DeMeo C, Norinder U, Koga H, Jose C, Jeliazkova N, Kochev N, Paskaleva V, Yang C, Daga PR, Clark RD, Rathman J. Improvement of quantitative structure-activity relationship (QSAR) tools for predicting Ames mutagenicity: outcomes of the Ames/QSAR International Challenge Project. Mutagenesis. 2019 Mar 6;34(1):3-16. doi: 10.1093/mutage/gey031. PMID: 30357358; PMCID: PMC6402315.

[6] Cassano, A.; Raitano, G.; Mombelli, E.; Fernández, A.; Cester, J.; Roncaglioni, A.; Benfenati, E. Evaluation of QSAR Models for the Prediction of Ames Genotoxicity: A Retrospective Exercise on the Chemical

Substances Registered Under the EU REACH Regulation. J. Environ. Sci. Health C Environ. Carcinog. Ecotoxicol. Rev. 2014, 32, 273–298. DOI: 10.1080/10590501.2014.938955.

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