	QMRF identifier (JRC Inventory): **To be entered by JRC**
	QMRF Title: **ADMET Predictor - Bacterial mutagenicity model (MUT_Risk)**
	Printing Date: **Jul 22, 2021**

1.QSAR identifier

1.1.QSAR identifier (title):

ADMET Predictor - Bacterial mutagenicity model (MUT_Risk)

1.2.Other related models:

The bacterial mutagenicity panel within ADMET Predictor Toxicity Module features a series of 10 MUT_*** models that predict Ames Mutagenicity in 5 individual strains of Salmonella (and/or E.coli), with or without metabolic activation, i.e.: MUT_97+1537; MUT_m97+1537; MUT_98; MUT_m98; MUT_100; MUT_m100; MUT_102+wp2; MUT_m102+wp2; MUT_1535; MUT_m1535. The ten TOX_MUT* Artificial Neural Network Ensembles (ANNE) are qualitative models, predicting the mutagenicity of new compounds as “Positive” (i.e., mutagenic) or “Negative”. Two additional mutagenicity models are included: i) ADMET Risk™ rule file, called "MUT_Risk", which predicts overall mutagenicity by counting instances of “Positive"; ii) "MUT_NIHS", classification model based on the proprietary Ames database provided by the Division of Genetics and Mutagenesis, National Institute of Health Sciences of Japan (DGM/NIHS).

1.3.Software coding the model:

ADMET Predictor 10.0
ADMET property prediction and QSAR model-building application
Simulations Plus, Inc., 42505 10th Street West, Lancaster, 93534-7059, CA, USA.
https://www.simulations-plus.com/software/admetpredictor/

2.General information

2.1.Date of QMRF:

16 February 2021

2.2.QMRF author(s) and contact details:

Simona Kovarich TOXIT of S-IN Soluzioni S.r.l. Via Ferrari 14, I-36100 Vicenza (Italy)
simona.kovarich@s-in.it https://www.toxit.it/en/; http://www.s-in.it

2.3.Date of QMRF update(s):

n/a

2.4.QMRF update(s):

n/a

2.5.Model developer(s) and contact details:

Pankaj Daga Simulation Plus, Inc. 42505 10th Street West, Lancaster, 93534- 7059, CA, USA
bob@simulations-plus.com https://www.simulations-plus.com/

2.6.Date of model development and/or publication:

The model was developed in April 2019 and was first released in ADMET Predictor 9.5. The model is currently implemented in ADMET Predictor 10.0 (2020).

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

[1]ADMET Predictor 10.0 https://www.simulations-plus.com/software/admetpredictor/

[2]Ghosh J, Lawless MS, Waldman M, Gombar V, Fraczkiewicz R (2016) Modeling ADMET. Methods Mol Biol. 1425:63-83. https://pubmed.ncbi.nlm.nih.gov/27311462/

[3]Clark R.D., Daga P.R. (2019) Building a Quantitative Structure-Property Relationship (QSPR) Model. In: Larson R., Oprea T. (eds) Bioinformatics and Drug Discovery. Methods in Molecular Biology, vol 1939. Humana Press, New York, NY.

[4]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 U0, 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, 34:3-16 https://pubmed.ncbi.nlm.nih.gov/30357358/

2.8.Availability of information about the model:

The model is proprietary and implemented in the commercial software ADMET Predictor (by Simulation Plus).

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

No

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Salmonella typhimurium and Escherichia coli

3.2.Endpoint:

TOX 7.6.1. Genetic toxicity in vitro

3.3.Comment on endpoint:

Qualitative estimate of overall mutagenicity (Ames test) obtained by combining individual positive predictions computed by the 11 MUT_*** QSAR models, i.e.: MUT_97+1537; MUT_m97+1537; MUT_98; MUT_m98; MUT_100; MUT_m100; MUT_102+wp2; MUT_m102+wp2; MUT_1535; MUT_m1535; MUT_NIHS.

3.4.Endpoint units:

Unitless

3.5.Dependent variable:

Mutagenicity Risk score. Each individual “Mutagenic” prediction contributes 0.6 “vote” to the score, with the exception of TA98 and TA100. Results for this pair of tests overlap mechanistically (Mortelmans and Ziegler, 2000) [6], and so each rule for those strains gets vote of 0.3.

By default, a Mutagenicity Risk score greater than 1 is indicative of a mutagenic potential.

3.6.Experimental protocol:

Being the MUT_Risk model based on the outcome of the 11 individual MUT_*** QSAR models, for the experimental dataset and experimental protocol one should refers to those of the individual MUT_*** models.

EXPERIMENTAL DATASETS: 1) MUT_NIHS: The source of the dataset is a proprietary Ames database, established by the Division of Genetics and

Mutagenesis, National Institute of Health Sciences of Japan (DGM/NIHS), and consisting of 12,140 new chemical substances that have not been previously used for developing QSAR models. These chemicals have been classified into three classes: Strongly Positive (Class A), Positive (Class B) and Negative (Class C) (Honma et al. 2019) [4]. Simulation Plus chose to merge Class A and Class B into a single positive class. The curated data set to build the model released with ADMET Predictor 9.5 included 11,736 compounds. 2) remaining 10 MUT_*** models: The source of these datasets is Bacha et al. (2002) [5], according to which bacterial mutagenicity data were primarily obtained from the Chemical Carcinogenesis Research Information System (CCRIS). This toxicology data file is maintained by the National Cancer Institute and made public through the National Library of Medicine's Toxicology Data Network (TOXNET). It is a scientifically evaluated and fully referenced database with mutagenicity results for individual bacterial indicator strains (S.typhimurium TA97, TA1537, TA98, TA100, TA1535, and TA102 and E. coli WP2uVrA) with and without addition of rat liver microsomal preparation to measure metabolic activation. These data were supplemented both with information from the Genetic Activity Profile database maintained by the Environmental Protection Agency in association with the International Agency for Research on Cancer and with data from a series of literature references.

DATA CURATION:

Curation of chemical structures was performed automatically and/or manually within the ADMET Modeler/Predictor, and included the following: i) extraction of the active moiety from salts and other multicomponent compounds; ii) standardization of substructural representations (e.g., nitro groups); iii) standardization of tautomers (rule-based system that strikes a balance between consistency and accuracy; the microstate analysis tool was used to check cases where automatic tautomer assignments were questionable).

Curation of mutagenicity data was performed automatically and/or manually within the ADMET Predictor, and included the following: i) removal of duplicate entries (based on shared name or structure or based on tautomeric equivalence), eliminating all but one example that represents a consensus of the replicates; ii) handling of structures with conflicting results (positive and negative) from different data sources: data are further verified for correctness analysing the original data source(s) (e.g., journal articles); if the conflict can't be resolved, then the records are removed.

3.7.Endpoint data quality and variability:

Endpoint quality was dependent on the original literature. Experimental variability was not taken directly into account, but is known historically to be about 85% between labs.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

Risk model (based on multiple QSAR models)

4.2.Explicit algorithm:

Risk Model

MUT_Risk model yields a qualitative estimate of overall mutagenicity by combining individual positive predictions computed by 11 MUT_*** QSAR ANNE models.

Each individual “Mutagenic” (i.e., Positive) prediction contributes 0.6 “vote” to the score, with the exception of TA98 and TA100. Results for this pair of tests overlap mechanistically (Mortelmans and Ziegler, 2000) [6], and so each rule for those strains gets vote of 0.3. RULES: 1) S_97: MUT_97+1537 = Positive 2) m_97: MUT_m97+1537 = Positive AND NOT MUT_97+1537 = Positive 3) S_98: MUT_98 = Positive 4) m_98: MUT_m98 = Positive AND NOT MUT_98 = Positive 5) S100: MUT_100 = Positive 6) m100: MUT_m100 = Positive AND NOT MUT_100 = Positive 7) S102: MUT_102+wp2 = Positive 8) m102: MUT_m102+wp2 = Positive AND TOX_MUT_102+wp2 = Positive 9) S535: MUT_1535 = Positive 10) m535: MUT_m1535 = Positive AND NOT MUT_1535 = Positive 10) NIHS: MUT_NIHS = Positive. By default, a MUT_Risk score > 1 is indicative of mutagenic potential.The Risk score is accompanied by a mnemonic Code list that indicates which rules were violated.

4.3.Descriptors in the model:

Not applicable (list of descriptors for individual MUT_*** models is provided as Supporting Information attached to the respective QMRFs)

4.4.Descriptor selection:

The descriptors for MUT_Risk are the models used in the definition:

MUT_97+1537
MUT_m97+1537
MUT_98
MUT_m98
MUT_100
MUT_m100
MUT_102+wp2
MUT_m102+wp2
MUT_1535
MUT_m1535
MUT_NIHS

4.5.Algorithm and descriptor generation:

Not applicable for the MUT_Risk model (see section 4.5 of the QMRF of individual MUT_*** QSAR models)

4.6.Software name and version for descriptor generation:

ADMET Predictor 10.0

4.7.Chemicals/Descriptors ratio:

Not applicable for the MUT_Risk model (see section 4.7 of the QMRF of individual MUT_*** QSAR models)

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The applicability domain of the MUT_Risk model is determined by the Applicability Domain of individual 11 MUT_*** QSAR models, i.e:
DESCRIPTOR DOMAIN: applicability domain defined by the descriptor space of training set compounds (hypercubes in the model's standardized space). Predictions computed for compounds lying outside the applicability domain of the model should be assessed as low reliable.
RESPONSE DOMAIN: positive/negative
The Out-of-scope Factor, introduced in ADMET Predictor 10, helps the user to deal semi-quantitatively with the uncertainty derived by out-of-scope (i.e., out of domain) predictions (see section 5.4).

5.2.Method used to assess the applicability domain:

The applicability domain of individual MUT_*** QSAR models is defined by hypercubes in the model's standardized space: the range of training set values for each descriptor used in the model is mapped to the interval [0,1]. A compound for which any of those descriptors is below -0.1 or above 1.1 is flagged as "out of scope" - i.e., as lying outside the applicability domain of the model. The prediction for such a compound may be correct but it would be unwise to put much faith in it (i.e., low reliable prediction).

5.3.Software name and version for applicability domain assessment:

ADMET Predictor 10.0
ADMET property prediction and QSAR model-building application
Simulations Plus, Inc., 42505 10th Street West, Lancaster, 93534-7059, CA, USA.
https://www.simulations-plus.com/software/admetpredictor/

5.4.Limits of applicability:

The limits of the model's applicability domain of individual MUT_*** QSAR models are defined by the DESCRIPTOR SPACE of training set compounds (see section 5.2). For new compounds, the standardised modelling descriptors' values should fall within the interval [0,1]; if any of those descriptors is below -0.1 or above 1.1, the compound is flagged as "out-of-scope" (i.e. outside the applicability domain of the model).
In case of out-of-scope predictions, the MUT_Risk penalty (i.e. Out-of-scope Factor) is applied, even if the model prediction is negative. Information on the model providing the out-of-scope prediction and the predicted value (i.e., "–" or "+") is added to the MUT_Code. A better description of the Out-of-scope Factor is following provided.
The Out-of-scope Factor treats any rule dependent on a prediction that has gone out-of-scope as having been triggered if it could be triggered by that prediction being in scope, but then attenuates the resulting incremental increase in the Risk score by an Out-of-Scope Factor, which is 0.5 by default. More specifically,the MUT_Risk gets incremented regardless of what the predicted value is, but softened to the degree specified by the

Out-of-Scope Factor: by default, that would be i) 0.5 * 0.6 = 0.3 (for all MUT_*** models with the exception of TA98 and TA100); ii) 0.5 * 0.3 = 0.15 (for TA98 and TA100 models). A mnemonic Code is then added to the MUT_Code list, but with a "+" or "-" flag appended to indicate that it was triggered by an out-of-scope condition rather than by the prediction per se: "+" is appended if the predicted value would have triggered the rule in the absence of the out-of-scope condition and "-" is appended otherwise.

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

No

6.2.Available information for the training set:

CAS RN: No

Chemical Name: No

Smiles: No

Formula: No

INChI: No

MOL file: No

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

No

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

No

6.5.Other information about the training set:

Not applicable for the MUT_Risk model (see section 6.5 of the QMRF of individual MUT_*** QSAR models)

6.6.Pre-processing of data before modelling:

n/a

6.7.Statistics for goodness-of-fit:

Not applicable for the MUT_Risk model (see section 6.7 of the QMRF of individual MUT_*** QSAR models)

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

n/a

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

n/a

6.10.Robustness - Statistics obtained by Y-scrambling:

n/a

6.11.Robustness - Statistics obtained by bootstrap:

n/a

6.12.Robustness - Statistics obtained by other methods:

n/a

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

No

7.2.Available information for the external validation set:

CAS RN: No

Chemical Name: No

Smiles: No

Formula: No

INChI: No

MOL file: 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:

No

7.5.Other information about the external validation set:

A reference set of 2260 commercial drugs from the World Drug Index (WDI) was used to validate the MUT_Risk model.

Irrelevant classes compounds were removed from a 2008 version of the WDI. In particular, the following compounds/classes were removed: phosphates, antiseptics, insecticides, emollients, laxatives, etc., as well as any compound that did not have an associated United States Adopted Name (USAN) or International Non-proprietary Name (INN) identifier. The structure of the principal component in salts was extracted and neutralized, after which duplicate structures were removed. The final curated data set included 2260 molecules.

7.6.Experimental design of test set:

The validation reference set was assembled after the develompent of the MUT_Risk model.

7.7.Predictivity - Statistics obtained by external validation:

Not available

7.8.Predictivity - Assessment of the external validation set:

The Mutagenicity Risk score exceeds 1.0 for 15% of the reference WDI dataset and exceeds 1.2 for 9% of it.

7.9.Comments on the external validation of the model:

n/a

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

The mechanistic interpretation parallels that of the Ames assay - i.e., that a replicated positive against one strain implies mutagenicity.

The Ames assay reproducibility is ca 85%. MUT_Risk combines 11 models with about 90% accuracy,we cannot run replicates, and we need to allow for the possibility of false positives. This leads naturally to a calibrated threshold of >1.

8.2.A priori or a posteriori mechanistic interpretation:

Not applicable to the MUT_Risk model (see section 8.2 of the QMRF of individual MUT_*** QSAR models)

8.3.Other information about the mechanistic interpretation:

n/a

9.Miscellaneous information

9.1.Comments:

n/a

9.2.Bibliography:

[1]Simulation Plus. ADMET Predictor Manual. September 2020 ver. 10.0

[2]Ghosh J, Lawless MS, Waldman M, Gombar V, Fraczkiewicz R (2016) Modeling ADMET. Methods Mol Biol. 1425:63-83. https://pubmed.ncbi.nlm.nih.gov/27311462/

[3]Clark R.D., Daga P.R. (2019) Building a Quantitative Structure-Property Relationship (QSPR) Model. In: Larson R., Oprea T. (eds) Bioinformatics and Drug Discovery. Methods in Molecular Biology, vol 1939. Humana Press, New York, NY.

[4]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 U0, 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, 34:3-16 https://pubmed.ncbi.nlm.nih.gov/30357358/

[5]Bacha PA, Gruver HS, Den Hartog BK, Tamura SY, Nuttet RF (2002) Rule Extraction from a Mutagenicity Data Set Using Adaptively Grown Phylogenetic-like Trees. J. Chem. Inf. Comput. Sci. 42, 1104-1111.

[6]Mortelmans K and Ziegler E. "The Ames Salmonella/microsome mutagenicity assay." Mutation Research 2000; 455:29-60. https://pubmed.ncbi.nlm.nih.gov/11113466/

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