	***QMRF identifier (JRC Inventory):*** **To be entered by JRC**
	QMRF Title: **Mutagenicity ISS Model - v. 1.0.2**
	Printing Date: **30-05-2018**

1.QSAR identifier

1.1.QSAR identifier (title):

Mutagenicity ISS Model (version 1.0.2)

1.2.Other related models:

This is the description of the VEGA model that implements the “In vitro mutagenicity (Ames test) alerts by ISS” as present in the software ToxTree v. 2.6

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:

30-05-2018

2.2.QMRF author(s) and contact details:

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

Azadi Golbamaki IRCCS-Istituto di Ricerche Farmacologiche Mario Negri Via La Masa 19, 20156 Milano, Italy azadi.golbamaki@marionegri.it https://www.marionegri.it/

Kristijan Vukovic IRCCS-Istituto di Ricerche Farmacologiche Mario Negri Via La Masa 19, 20156 Milano, Italy kristijan.vukovic@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]Romualdo Benigni Istituto Superiore di Sanità, Department of Environment and Primary Prevention Rome, Italy romualdo.benigni@iss.it

[2]Cecilia Bossa Istituto Superiore di Sanità; Department of Environment and Primary Prevention Rome, Italy cecilia.bossa@iss.it.

[3]Nina jeliazkova IDEA Consult Joseph II straat 40 B1, 1000 Brussels, Belgium jeliazkova.nina@gmail.com

[4]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:

2015

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

[1]The Benigni / Bossa rulebase for mutagenicity and carcinogenicity - a module of Toxtree, (2008) by R. Benigni, C. Bossa, N. Jeliazkova, T. Netzeva, and A. Worth. European Commission report EUR 23241 EN

[2]R. Benigni, C. Bossa, T. Netzeva, A. Rodomonte, and I. Tsakovska (2007) Mechanistic QSAR of aromatic amines: new models for discriminating between mutagens and nonmutagens, and validation of models for carcinogens. Environ mol mutag 48:754-771

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:

Histidine-dependent strains of *Salmonella typhimurium* (Ames test)

3.2.Endpoint:

TOX 7.6.1. Genetic toxicity in vitro

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.

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:

Based on the OECD 471 test guideline. The details about the dataset compilation are reported on the ISS website. Ames mutagenicity data is described in Benigni R. Bossa C. Richard A. M. Yang C. 2008 A novel approach: chemical relational databases, and the role of the ISSCAN database on assessing chemical carcinogenicity. Ann. Ist. Super. Sanità 44, 48–56

3.7.Endpoint data quality and variability:

NA

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

Structured-based model

4.2.Explicit algorithm:

The Mutagenicity ISS Model is based on In vitro mutagenicity (Ames test) alerts by ISS module of Toxtree: which is an open source application and the algorithm is available on (http://toxtree.sourceforge.net/ames.html)

4.3.Descriptors in the model:

If at least one mutagenicity rule matches with the given compound, it is classified as mutagenic, otherwise, it is non-mutagenic.

The following fragments are encoded as SA for matching mutagenic compounds:

SA1 Acyl halides

SA2 Alkyl (C <5) or benzyl ester of sulphonic or phosphonic acid

SA3 N-methylol derivatives

SA4 Monohaloalkene

SA5 S or N mustard

SA6 Propiolactones and propiosultones
SA7 Epoxides and aziridines
SA8 Aliphatic halogens
SA9 Alkyl nitrite
SA10 alfa, beta unsaturated carbonyls
SA11 Simple aldehyde
SA12 Quinones
SA13 Hydrazine
SA14 Aliphatic azo and azoxy
SA15 Isocyanate and isothiocyanate groups
SA16 Alkyl carbamate and thiocarbamate
SA18 Polycyclic Aromatic Hydrocarbons
SA19 Heterocyclic Polycyclic Aromatic Hydrocarbons
SA21 Alkyl and aryl N-nitroso groups
SA22 Azide and triazene groups
SA23 Aliphatic N-nitro
SA24 alfa,beta unsaturated alkoxy
SA25 Aromatic nitroso group
SA26 Aromatic ring N-oxide
SA27 Nitro aromatic
SA28 Primary aromatic amine, hydroxyl amine and its derived esters (with restrictions)
SA28bis Aromatic mono- and dialkylamine
SA28ter Aromatic N-acyl amine
SA29 Aromatic diazo
SA30 Coumarins and Furocoumarins
SA37 Pyrrolizidine Alkaloids
SA38 Alkenylbenzenes
SA39 Steroidal estrogens
SA57 DNA Intercalating Agents with a basic side chain
SA58 Haloalkene cysteine S-conjugates
SA59 Xanthones, Thioxanthones, Acridones
SA60 Flavonoids
SA61 Alkyl hydroperoxides
SA62 N-acyloxy-N –alkoxybenzamides
SA63 N-aryl-Nacetoxyacetamides
SA64 Hydroxamic acid derivatives
SA65 Halofuranones
SA66 Anthrones
SA67 Triphenylimidazole and related
SA68 9,10 – dihydrophenanthrenes
SA69 Fluorinated quinolones

4.4.Descriptor selection:

NA

4.5.Algorithm and descriptor generation:

The model has been built as a set of rules, taken from the work of Benigni and Bossa (ISS) as implemented in the software ToxTree version 2.6 (http://toxtree.sourceforge.net). The model implement all the rules

related to mutagenicity and does not implement the full decision tree used by ToxTree. If at least one mutagenicity rule is matching with the given compound, a “mutagen” prediction is given; otherwise, a “non-mutagen” prediction is given. The training set for the model has been extracted from ToxTree, and consists of 670 compounds.

4.6.Software name and version for descriptor generation:

NA

4.7.Chemicals/Descriptors ratio:

NA

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.

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.

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: 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:

Data for each descriptor variable for the training set can be obtained by running the VEGA model on the training set

6.6.Pre-processing of data before modelling:

NA

6.7.Statistics for goodness-of-fit:

This is not a formal QSAR model developed through statistical methods so no specific internal and external validation has been performed. The SA have been tested on the ISSCAN dataset who served as training set of the ToxTree module and the performances were the following:

Training set: n = 670; Accuracy = 0.79; Specificity = 0.68; Sensitivity = 0.89

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:

No

7.2.Available information for the external validation set:

NA

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:

NA

7.6.Experimental design of test set:

NA

7.7.Predictivity - Statistics obtained by external validation:

NA

7.8.Predictivity - Assessment of the external validation set:

NA

7.9.Comments on the external validation of the model:

NA

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.

8.2.A priori or a posteriori mechanistic interpretation:

NA

8.3.Other information about the mechanistic interpretation:

NA

9.Miscellaneous information

9.1.Comments:

NA

9.2.Bibliography:

NA

9.3.Supporting information:

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

All available dataset 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