	QMRF identifier (JRC Inventory): **To be entered by JRC**
	QMRF Title: **Mutagenicity (Ames test) model (SarPy/IRFMN) (version 1.0.8)**
	Printing Date: **07-02-2022**

1.QSAR identifier

1.1.QSAR identifier (title):

Mutagenicity (Ames test) model (SarPy/IRFMN) (version 1.0.8)

1.2.Other related models:

The model has been built as a set of rules, extracted automatically with the SARpy software [1] from a large set of compounds and extends the previous version belonging to CAESAR model [2].

Both are implemented inside VEGA online platform, accessible at: http://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:

February 2022

2.2.QMRF author(s) and contact details:

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

2.3.Date of QMRF update(s):

NA

2.4.QMRF update(s):

NA

2.5.Model developer(s) and contact details:

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

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

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

[5] Dario Cattaneo Department of Electronics and Information (DEI), Politecnico di Milano

[6] Nazanin Golbamaki Istituto di Ricerche Farmacologiche Mario Negri - IRCSS Via Mario Negri 2, 20156 Milano, Italy https://www.marionegri.it/

2.6.Date of model development and/or publication:

2013

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

[1] T. Ferrari, D. Cattaneo, G. Gini, N. Golbamaki Bakhtyari, A. Manganaro, E. Benfenati. Benfenati, “Automatic knowledge extraction from chemical structures: the case of mutagenicity prediction”, SAR and QSAR in Environmental Research (2013), vol. 24 issue 5, 365-83.

[2] 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

[3] http://www.vegahub.eu/

[4] http://sarpy.sourceforge.net/

[5] Benfenati E, Manganaro A, Gini G. VEGA-QSAR: AI inside a platform for predictive toxicology Proceedings of the workshop "Popularize Artificial Intelligence 2013", December 5th 2013, Turin, Italy Published on CEUR Workshop Proceedings Vol-1107

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:

Mutagenicity bacterial reverse mutation test

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 to be carcinogenic substance in case a genotoxic mechanism is considered. Furthermore, mutagenicity in somatic cells causes concern for possible mutagenic effects in germ cell (heritable deceases).

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 is an in vitro model of chemical mutagenicity 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, a large set of compounds was used [3]. The estimated inter-laboratory reproducibility rate of S. typhimurium test data is 85% [4]

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

The Mutagenicity (Ames test) model (SarPy/IRFMN) (version 1.0.7 provides a qualitative prediction of mutagenicity on Salmonella typhimurium (Ames test). It is based on a set of rules extracted from set of compounds [3] by SARpy software without any 'a priori' knowledge.

The original work [1] has been extended, resulting in two sets of rules for mutagenicity (112 rules) and non-mutagenicity (93 rules).

4.2.Explicit algorithm:

The algorithm generates substructures of arbitrary complexity, and the fragment candidates to become structural alerts (SAs) are automatically selected based on their prediction performance on a training set. Fragmentation is done directly on the SMILES notation of structures.

If at least one mutagenicity rule is matching with the given compound, a "Mutagenic" prediction is given; if only one or more non-mutagenicity rule is matching, a "Non-Mutagenic" prediction is given; if no rules match with the given compound, a "Possible Non-Mutagenic" prediction is given.

4.3.Descriptors in the model:

Fragments authomatically generate and statistically evaluated. 112 rules for mutagenicity and 93

rules non-mutagenicity.

Following, the list of the 112 rules for mutagenicity, expressed as SMARTS strings:

SM 1: O=[N+]([O-])c1ccc2ccccc2c1

SM 2: O=NN(C)C

SM 3: n1ccc(N)c2ccc(cc12)

SM 4: c1oc(cc1)[N+](=O)[O-]

SM 5: O=[N+]([O-])c1ccc(c2ccccc2)c(c1)

SM 6: Nc4ccc(N)cc4N

SM 7: C1C=Cc2ccccc2C1

SM 8: N1CC1

SM 9: c1cc([N+](=O)[O-])sc1

SM 10: n1ccnc2c1ccc(N)c2

SM 11: c1ccc2c(c1)cc3ccc(cc3c2)C

SM 12: Nc1c(ncn1)

SM 13: n1cc(nc2ccc3c(ncn3C)c12)

SM 14: O(c1ccccc1)CC2OC2

SM 15: O=C1c2ccccc2C(=O)c3c(O)ccc(O)c13

SM 16: N(O)c1ccc(C=C)cc1

SM 17: c1ccc2ccc3ccc(cc3c2c1)N

SM 18: O(Cc1cccc2ccccc12)C

SM 19: O=C(c1ccc(cc1)NO)

SM 20: O=C(c1ccccc1)Cl

SM 21: C1=Cc2cccc3cccc1c23

SM 22: O1CC1CCc2ccc(cc2)

SM 23: C(O)C=CCl

SM 24: OCC(CBr)
SM 25: [N-]=[N+]
SM 26: n1c2ccc(cc2c(cc1))C
SM 27: O=CC1(OC1)C
SM 28: n1cccc2c1ccc3c2ncn3
SM 29: c2nc3C(=O)C=CC(=O)c3cc2
SM 30: O=Nc1ccc(OC)cc1
SM 31: SC(=CCl)Cl
SM 32: O=C1c2cccc(N)c2C(=O)c3ccccc13
SM 33: Nc1ccc2c(c1)c3ccccc3n2
SM 34: Oc1ccc2ccc3ccc(cc3c2c1)
SM 35: C(c1ccccc1)COC=C

SM 36: c1cc2cccc3c4cc(ccc4c(c1)c23)
SM 37: N(O)c1ccc(Oc2ccccc2)cc1
SM 38: c1ccc2c(c1)c3ccccc3n2
SM 39: O(Cc1cccc2ccccc12)
SM 40: c1ccc2c3ccccc3CCc2c1
SM 41: n1cc2ccccc2s1
SM 42: P(=O)(N)N(C)CC
SM 43: C(N)Cl
SM 44: c1ccc(C=Cc2ccc(N)cc2)cc1
SM 45: c1cc(ccc1NCCCl)
SM 46: N(c1ccc(N=Nc2ccccc2)cc1)C
SM 47: O=C(NCc1ccccc1)C
SM 48: c1cc2ccc3cccc4ccc(c1)c2c34
SM 49: Nc1cccc(c1)c2ccccc2
SM 50: O=[N+]([O-])c1cccc2cccc(c12)
SM 51: O=C(Nc1ccc(cc1)c2ccccc2)
SM 52: O=Cc1cccc(c1)[N+]
SM 53: O=[N+]([O-])c1cc(N)c(c(N)c1)
SM 54: c1ccc(Oc2ccc(N)cc2)cc1
SM 55: COC=CC=CC
SM 56: N(=N)NC
SM 57: ONc1ccc(cc1)S
SM 58: O1CC1Cc2ccc(cc2)
SM 59: O=C(c1ccccc1O)c2ccccc2
SM 60: Nc1ccc(cc1)c2ccccc2
SM 61: c1ccc2c(c1)ccc3c2cc4ccccc4c3
SM 62: c1ccc2c(c1)cc3ccc(cc3c2C)
SM 63: c1ccc2c(ccc3c4ccccc4ccc23)c1
SM 64: c1c2ccccc2nc3ccccc13
SM 65: O=CC=C(C(=O)c1ccccc1)
SM 66: n1cc(cc2c1ncn2)
SM 67: Nc1nccn1C
SM 68: C1C(C=C(C))C1(C)C

SM 69: Nc1ccc(cc1)[N+](=O)[O-]
SM 70: Nc1ccc(cc1N)
SM 71: N=CC=C
SM 72: O=[N+]([O-])c1ccc(cc1)CO
SM 73: CCNCCCl
SM 74: O=S(=O)(OCC)
SM 75: c1ccc2c3ccccc3Cc2c1
SM 76: c1c2ccccc2n(c1)C
SM 77: C(CBr)Br
SM 78: Nc1ccccc1F
SM 79: c1ccc(N)c(c1N)C
SM 80: c1ccc2c(c1)ccc3cc(ccc23)

SM 81: c1ccc2c(c1)cc3ccccc3c2
SM 82: c1ccc2ccccc2c1C
SM 83: Nc1ccc(cc1)Cc2ccccc2
SM 84: Oc1ccc2Cc3ccccc3Oc2c1
SM 85: C(Cl)(Cl)Cl
SM 86: O(c1ccccc1N)C
SM 87: NN(c1ccccc1)
SM 88: n1c(N)n(c2ccccc12)
SM 89: O=C(N(O))C
SM 90: n1ccnc2c1cccc2
SM 91: c1cc(c(N)cc1N)C
SM 92: OCC1OC1
SM 93: C(C)Br
SM 94: C(OCC)N
SM 95: Nc1cccc(N)c1
SM 96: c1c(nn(c1))
SM 97: C1OC1
SM 98: C(O)N
SM 99: c1ccc2cccnc2c1
SM 100: N=NC
SM 101: O=CC(=C)Cl
SM 102: n1cnc2c(ncn2)c1N
SM 103: NNCC
SM 104: Cc2ccc(N)cc2
SM 105: Nc1ccc(N)cc1
SM 106: CCCl
SM 107: C=NN1N=Nc2c([nH]c3ccccc23)C1=O
SM 108: NC([N+])
SM 109: n1c2ccc(cc2[s+]c3cc(N)c(cc13))N
SM 110: O=Nn1cc(c2ccccc12)CC
SM 111: O=CC(=CC)C=CC
SM 112: O=C1OCC1

Following, the list of the 93 rules for non mutagenicity, expressed as SMARTS strings:

SM 113: C(O)CCCCCCCC=CCC

SM 114: CCOc1ccc(Cl)cc1

SM 115: C(NC(C(=O))C)C(NC)

SM 116: c1(c(ccc(c1)CCCCCC))O

SM 117: c1c(c(C(C)(C)C)cc(c1)C)

SM 118: c1(c(C(=O)O)cccc1)C(=O)

SM 119: S(=O)(=O)(N)c1ccc(N)cc1

SM 120: n1c(nc(nc1))

SM 121: C(=O)(C(CCC(=O)O))O

SM 122: CCOCCOCCOCCO

SM 123: CC(=O)OCC(CC)CCCC

SM 124: P(O)OCCCCC

SM 125: N(c1ccccc1)CCCNC

SM 126: c1(C(=O)OC)c(N)cccc1

SM 127: C(C)(Oc1ccccc1)(C)C

SM 128: N(CCO)(CCCC)C

SM 129: C(=C)CCCl

SM 130: C(=O)(C(=C)C)OCCCCC

SM 131: Oc1ccc2C=C(COc2c1)c3ccccc3

SM 132: c1(nc2c(o1)cccc2)c3ccccc3

SM 133: c1(c(c(cc(c1)))O)c2c(c(cc(c2)))O

SM 134: n1c(cc(c2c1cccc2))CO

SM 135: c1c(c(ccc1C(O)CNC)O)

SM 136: O(C(=O))C(=O)

SM 137: C(=O)(CCC(=O)O)

SM 138: S(=O)(=O)(c1ccccc1)N

SM 139: c1c(c(cc(c1Cl)Cl)Cl)

SM 140: C(C=C)(CCC=C)

SM 141: c1(CCCCCC)ccccc1

SM 142: N(C)(CCCC)CCCC

SM 143: C(=O)(O)CCCCCCCC

SM 144: C(=O)(N(c1ccccc1))N

SM 145: C1(C(CCC(C1)C)C)O

SM 146: c1(C(=O)OCC)ccccc1

SM 147: c1c(C(F)(F)F)cccc1

SM 148: C(=O)CS

SM 149: O(c1ccccc1)CC(CNC(C))O

SM 150: C(F)(F)C

SM 151: c1cc(c(c(c1)Br)O)

SM 152: c1(Br)ccc(cc1)C

SM 153: SCCCC

SM 154: C(=O)CC(=O)C

SM 155: N(CCO)(CCO)C

SM 156: C(=O)N(C)CCC

SM 157: CCCCCCCCCCCC

SM 158: c1(c(ccc(c1)C=CC)O)
SM 159: c1cc(c(cc1)Cl)Cl
SM 160: C=CC=C(CCC)C
SM 161: c1ccc(NC(C)C)cc1
SM 162: C1CC(CC(C1))(C)C
SM 163: CCCCCCC
SM 164: C(C(OC)(C)C)
SM 165: N#CCC
SM 166: C(=C(CC)O)
SM 167: CN1CN=CC=C1
SM 168: S(=O)(=O)c1ccc(N)cc1
SM 169: CC(CC)CCC
SM 170: C(=O)(c1ccccc1)OC
SM 171: c1(F)ccc(cc1)C
SM 172: C(=Cc1ccccc1)C(=O)c2ccccc2
SM 173: [nH]1cc(c2c1cccc2)CC
SM 174: P(=S)(OCC)OC
SM 175: [N+](C)(C)C
SM 176: OCCN(C)C
SM 177: C(=O)CCCCC
SM 178: C(=O)(CC(C))OCC
SM 179: CC#N
SM 180: C(=C(CC)C)C=O
SM 181: O=C1NC=NC=C1
SM 182: CCCC(C)C
SM 183: c1(cc(ccc1)Cl)Cl
SM 184: c12c(cc(cc2)O)ccc(S)c1
SM 185: PC
SM 186: OCC(CO)(C)C
SM 187: OC(=O)C(=C)C
SM 188: C(=O)(C)OCCCC
SM 189: c1(C(=O)O)c(ccc(c1))O
SM 190: c1(c(Cl)cccc1)C
SM 191: Cc1cc(c(cc1)OC)OC
SM 192: C#N
SM 193: Cc1ccc(Cl)cc1
SM 194: Nc1ccc2C=CCOc2c1
SM 195: C(=O)(O)Cc1ccccc1
SM 196: c1(cc(ccc1CCC)O)O
SM 197: N(c1ccccc1)c2ccccc2
SM 198: C=NO
SM 199: C(=S)(N)
SM 200: c1(CC)cc(OC)ccc1
SM 201: c1(nc2c(s1)cccc2)
SM 202: OCCOCCO
SM 203: N(C)(CCN)c1ccccc1

SM 204: [N+]([O-])CC

SM 205: S(=O)(=O)(c1cc(c(cc1))N)O

4.4. Descriptor selection:

NA

4.5. Algorithm and descriptor generation:

NA

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.
Indices are calculated on the first *k* = 3 most similar molecules, each having S_k 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

AD final index is calculated as following:

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

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

If 0.9 > AD index ≥ 0.66, the predicted substance could be out of the Applicability Domain of the model. It corresponds to moderate reliability of prediction.

If AD index < 0.65, the predicted substance is regarded 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 [5]. 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 centered 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: Yes

INChI: Yes

MOL file: Yes

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 counts 3367 compounds

6.6.Pre-processing of data before modelling:

All chemical structures have been checked manually.

6.7.Statistics for goodness-of-fit:

Following, statistics obtained applying the model to its original dataset, if “Possible Non-Mutagenic” is taken as “Non-Mutagenic” but without taking the ADI into account:

Training set: n = 3367 (1883 positive, 1484 negative)

Accuracy = 0.82

Specificity = 0.77

Sensitivity = 0.86

TP 1613, TN 1148, FP 336, FN 270

Following, statistics obtained applying the model to its test set, if “Possible NON-Mutagenic” is taken as “NON-Mutagenic” without taking the ADI into account:

Test set: n = 837 (465 positive, 372 negative)

Accuracy = 0.81

Specificity = 0.76

Sensitivity = 0.86

TP 398, TN 283, FP 89, FN 18

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 [6] [7].

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:

Four compounds were not predicted (molecule error: unable to normalize SMILES string) then the available predictions for the statistical assessment were 14513.

We applied AD index thresholds to perform predictions on the external validation set and if “Possible NON-Mutagenic” is taken as “NON-Mutagenic” the results are:

The predictions of 3129 substances are in AD. AD index >=0.9.

Sensitivity	Specificity	Accuracy	MCC
0.81	0.86	0.85	0.63

TP 646, TN 2001, FP 331, FN 151

The predictions of 5712 substances could be out of the AD. 0.9> AD index >= 0.65

Sensitivity	Specificity	Accuracy	MCC
0.67	0.74	0.73	0.33

TP 676, TN3480, FP 1221, FN 335

The predictions of 5672 substances are out of the AD. AD index <0.65

Sensitivity	Specificity	Accuracy	MCC
0.55	0.52	0.53	0.05

TP 584, TN2399, FP 2203, FN 486

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 both toxic and non-toxic 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. 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: the fragments identified as statistically associated to the toxic or non toxic class can be investigated to explore the mechanistic basis of the model.

8.3.Other information about the mechanistic interpretation:

NA

9.Miscellaneous information

9.1.Comments:

NA

9.2.Bibliography:

[1] T. Ferrari, D. Cattaneo, G. Gini, N. Golbamaki Bakhtyari, A. Manganaro, E. Benfenati, “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

[2] 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 (accessed 9/21/09).

[3] 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

[4] 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

[5] 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

[6] 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.

[7] 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