ProtoICH

EXPERT RULE-BASED MODEL DOSSIER

ProtoQSAR expert rule-based model for Mutagenicity (Ames test)

www.protoqsar.com

Centro Europeo de Empresas Innovadoras (CEEI)
Parque Tecnológico de Valencia
Avda. Benjamin Franklin 12
46980 Paterna (Valencia, Spain)

protopred@protoqsar.com

+34 962 021 811

ProtoQSAR
Computational toxicology: fast, economical and ethical

ProtoQSAR expert rule-based model for Mutagenicity (Ames test)

1. Model identifier

1.1. Model identifier (title):

ProtoQSAR expert rule-based model for mutagenicity in bacteria (Ames test)

1.2. Other related models:

None

1.3. Software coding the model:

ProtoQSAR proprietary software

https://protoqsar.com

2. General information

2.1. Date of Report:

January 2022

2.2. Report author(s) and contact details:

ProtoQSAR SL

+34 96 202 18 11

info@protoqsar.com

2.3. Date of Report update(s):

Not applicable

2.4. Report update(s):

Not applicable

2.5. Model developer(s) and contact details:

[1] Gómez-Ganau, S.

[2] Roca-Martínez, J.

[3] Barigye, S. J.

[4] Serrano-Candelas, E.

[5] Gozalbes, R.

info@protoqsar.com

2.6. Date of model development and/or publication:

March 2020

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

Not applicable

2.8. Availability of information about the model:

The model is proprietary and the algorithm is confidential. The training and test sets are available upon request.

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

None to date

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Bacteria (Salmonella Typhimurium)

3.2. Endpoint:

Human health effects: Mutagenicity. OECD 471: Bacterial reverse mutation test.

3.3. Comment on endpoint:

N/A

3.4. Endpoint units:

N/A

3.5. Dependent variable:

Binary classification: positive (mutagenic) / negative (non-mutagenic).

3.6. Experimental protocol:

Histidine enable bacterial colony growth on a medium deficient in histidine (revertants). A compound is classified Ames positive if it significantly induces revertant colony growth in at least one of out of five strains.

3.7. Endpoint data quality and variability:

The data for developing the model was extracted from Hansen K et al research publication. However, all the information related to the model is owned by ProtoQSAR SL

4. Defining the algorithm - OECD Principle 2

4.1. Type of model:

Expert rule-based system

4.2. Explicit algorithm:

Expert rule-based system. Set of 48 rules (structural alerts) related to mutagenicity activity in Ames test. These rules are expressed SMARTS representing molecular fragments (reported in section 4.3). If at least one rule is matching with the target compound, an “ALERT” prediction is given. Otherwise, a “NO ALERT” prediction is given.

4.3. Descriptors in the model:

- Acyl halides
- Alkyl (C<5) or benzyl ester of sulphonic or phosphonic acid
- N-methylol derivatives
- Monohaloalkene
- S or N mustard
- Propiolactones and propiosultones
- Epoxides and aziridines
- Aliphatic halogens
- Alkyl nitrite
- a,b unsaturated carbonyls
- Simple aldehyde
- Quinones
- Hydrazine
- Aliphatic azo and azoxy
- Isocyanate and isothiocyanate groups
- Alkyl carbamate and thiocarbamate
- Polycyclic Aromatic Hydrocarbons
- Heterocyclic Polycyclic Aromatic Hydrocarbons
- Alkyl and aryl N-nitroso groups
- Azide and triazene groups
- Aliphatic N-nitro
- a,b unsaturated alkoxy
- Aromatic nitroso group
- Aromatic ring N-oxide
- Nitro aromatic
- Primary aromatic amine, hydroxyl amine and its derived esters
- Aromatic mono- and dialkylamine
- Aromatic N-acyl amine
- Aromatic diazo
- Coumarins and Furocoumarins
- Pyrrolizidine Alkaloids
- Alkenylbenzenes

- Steroidal estrogens
- Aromatic diazo with sulphonic groups
- Derived aromatic amines
- DNA Intercalating Agents with a basic side chain
- Haloalkene cysteine S-conjugates
- Xanthones, Thioxanthones, Acridones
- Flavonoids
- Alkyl hydroperoxides
- N-acyloxy-N -alkoxybenzamides?
- N-aryl-N-acetoxyacetamides
- Hydroxamic acid derivatives
- Halofuranones
- Anthrones
- Triphenylimidazole and related
- 9,10 - dihydrophenanthrenes
- Fluorinated quinolines

4.4. Descriptor selection:

The descriptor selection has been adapted from Benigni R, Bossa C (2011) Chem Rev 111: 2507–2536, Benigni R, Bossa C, Tcheremenskaia O (2013) Chem. Rev. 2013, 113, 5, 2940-2957 and Benigni R, Bossa C, Jeliazkova, N, Netzeva T, and Worth A. The Benigni / Bossa rulebase for mutagenicity and carcinogenicity - a module of Toxtree. European Commission report EUR 23241.

4.5. Algorithm and descriptor generation:

The 48 rules (structural alerts) are expressed as SMARTS representing molecular fragments, as described in Benigni R, Bossa C (2011) Chem Rev 111: 2507–2536, Benigni R, Bossa C, Tcheremenskaia O (2013) Chem. Rev. 2013, 113, 5, 2940-2957 and Benigni R, Bossa C, Jeliazkova, N, Netzeva T, and Worth A. The Benigni / Bossa rulebase for mutagenicity and carcinogenicity - a module of Toxtree. European Commission report EUR 23241. Nevertheless, additional scripts have been developed for some of them.

4.6. Software name and version for descriptor generation:

ProtoQSAR proprietary software v4.0

4.7. Chemicals/Descriptors ratio:

N/A

5. Defining the applicability domain - OECD Principle 3

5.1. Description of the applicability domain of the model:

Absence of alerts may be associated to absence of reactivity on the Ames test, or may be due to a lack of mechanistic knowledge. Therefore, the 'No structural alert' flag is not equivalent to a negative prediction.

5.2. Method used to assess the applicability domain:

This method works as structural alerts, by searching specific substructures in the molecule. If the molecule matches a defined structure in some alert, the system flags this molecule as "putative mutagenic" in this specific alert.

5.3. Software name and version for applicability domain assessment:

N/A

5.4. Limits of applicability:

The model was built only for discrete organic chemicals.

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Available in Toxtree v.2.1.0 and higher (not attached)

6.2. Available information for the training set:

CAS RN: No

Chemical Name: No

SMILES: Yes

Formula: No

INChI: No

MOL file: No

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

Not reported

6.4. Data for the dependent variable (response) for the training set:

Yes

6.5. Other information about the training set:

The training set is comprised of more than 7000 molecules, with around 3500 positive calls and 2800 negatives, as reported in Toxtree software

6.6. Pre-processing of data before modelling:

Information available in Toxtree v2.1.0 and higher

6.7. Statistics for goodness-of-fit:

Overall accuracy of 0.78 (reported in Benigni R, Bossa C, Jeliazkova, N, Netzeva T, and Worth A. The Benigni / Bossa rulebase for mutagenicity and carcinogenicity - a module of Toxtree. European Commission report EUR 23241.)

Parameters	Value
Accuracy (ACC)	0.78
Sensitivity, recall or true positive rate (TPR)	0.85
Specificity or true negative rate (TNR)	0.72

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

Not reported

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

Not reported

6.10. Robustness – Statistics obtained by Y-scrambling:

Not reported

6.11. Robustness – Statistics obtained by bootstrap:

Not reported

6.12. Robustness – Statistics obtained by other methods:

Not reported

7. External validation - OECD Principle 4

7.1. Availability of the external validation set:

It is available but not attached.

7.2. Available information for the external validation set:

CAS RN: No

Chemical Name: No

SMILES: Yes

Formula: No

INChI: No

MOL file: No

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

Not reported

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

Yes

7.5. Other information about the external validation set:

The external validation set is comprised of 6492 compounds from a raw dataset of 6500 compounds extracted from the publication of Hansen et al. (see references in section 9.2).

7.6. Experimental design of test set:

Not reported

7.7. Predictivity - Statistics obtained by external validation:

Experimental values	QSAR predictions		
	Negative	**Positive**	
Negative	1975	1025	**65.8% (TNR)**
Positive	558	2934	**84.0% (TPR)**
Total (%)	78.0 (% pred NEG)	74.1 (% pred POS)	**75.6% (ACC)**

Parameters	Validation
Accuracy (ACC)	0.76
Sensitivity, recall or true positive rate (TPR)	0.84
Specificity or true negative rate (TNR)	0.66
Precision or positive predictive value (PPV)	0.74
Negative predictive value (NPV)	0.78
Miss rate or false negative rate (FNR)	0.16
Fall-out or false positive rate (FPR)	0.34
False discovery rate (FDR)	0.26
False omission rate (FOR)	0.22
F-score	0.79
Matthews Correlation Coefficient (MCC)	0.51
Critical Success Index (CSI)	0.65
Area under the ROC (AUC)	0.75

7.8. Predictivity – Assessment of the external validation set:

Not reported

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 presented model identifies specific molecular substructures, described to be relevant to mutagenicity

8.2. A priori or a posteriori mechanistic interpretation:

A posteriori mechanistic interpretation: The identified chemical substructures may serve as starting point for a posteriori mechanistic interpretation.

8.3. Other information about the mechanistic interpretation:

N/A

9. Miscellaneous information

9.1. Comments:

The model can be applied to estimate Mutagenicity (Ames test).

ProtoPRED provides prediction for more than 25 endpoints, including physicochemical, toxicological and ecotoxicological, by using proprietary QSAR models. All ProtoPRED models meet OECD criteria and are valid for regulatory purposes.

9.2. Bibliography:

[1] OECD: Test No. 471: Bacterial Reverse Mutation Test. https://www.oecd-ilibrary.org/environment/test-no-471-bacterial-reverse-mutation-test_9789264071247-en

[2] Hansen K et al. Benchmark data set for in silico prediction of Ames mutagenicity. J Chem Inf Model. 2009 Sep;49(9):2077-81.

9.3. Supporting information:

N/A