(Q)SAR Model Reporting Format (QMRF)

(The present QMRF v.2.1 is prepared in accordance with (Q)SAR Assessment Framework (QAF) document developed by OECD)
(https://one.oecd.org/document/ENV/CBC/MONO(2023)32/ANN1/en/pdf**)**

Welcome

Model version: Ames Mutagenicity S9 activated kinetic v.06.06
Platform version: OASIS TIMES 2.34.1
Name: Ames mutagenicity
Author: LMC, University "Prof. As. Zlatarov", Bourgas, Bulgaria
Date: 05 June 2025
E-mail: ovanes.mekenyan@oasis-lmc.org
www: http://www.oasis-lmc.org/

Section 1. QSAR identifier

1.1. QSAR identifier (title)

In vitro Ames Mutagenicity kinetic model with S9 metabolic activation v. 06.06

1.2. Other related models

Not applicable

1.3. Software coding the model

Model version: Ames Mutagenicity S9 activated kinetic v.06.06
Platform version: OASIS TIMES 2.34.1
Name: *In vitro* Ames Mutagenicity with S9 metabolic activation kinetic model
Developer: LMC, University "Prof. As. Zlatarov", Bourgas, Bulgaria
Coding language: Delphi 10.2

Section 2. General information

2.0. Abstract

The *in vitro* Ames kinetic model identifies chemicals which are able to elicit mutagenicity as a result of interactions with DNA taking into account their metabolic activation and the quantity of formed adducts with macromolecules.

The training set of the model consists of 4268 chemicals with experimental Ames data separated in two groups: 2092 positive as parents and after S9 metabolic activation (302 are proprietary), and 2176 negative as parents and after S9 metabolic activation (1428 are proprietary). Chemicals with proprietary data are used for deriving alert boundaries and estimating performance of the model (and its domain) but are not disclosed for public.

The model is based on an alerting group approach addressing mutagenicity of parents and their generated metabolites in *in vitro* liver S9 metabolic system taking into account the kinetics of metabolism. The model allows relating the level of mutagenic potency to the amount of formed DNA adducts under the assumption that the presence of alert is necessary but not sufficient reason for predicting positive mutagenicity.

2.1. Date of QMRF

05 June 2025

2.2. QMRF author(s) and contact details

Name: Laboratory of Mathematical Chemistry
Affiliation: Laboratory of Mathematical Chemistry, University "Prof. As. Zlatarov", "Yakimov" St. #1, 8010 Bourgas, BULGARIA
URL: http://www.oasis-lmc.org
E-mail: ovanes.mekenyan@oasis-lmc.org

2.3. Date of QMRF update(s)

29 November 2021, 24 March 2023, 11 April 2024, 05 June 2025

2.4. QMRF update(s)

Information which has been modified:

Sections 1.1 QSAR identifier (title)**; Sections 1.3** Software coding the model; **Section 2.** General information; **Sections 2.1** Date of QMRF; **Section 2.2** QMRF author(s) and contact details; **Sections 2.3** Date of QMRF update(s); **Sections 2.5** Model developer(s) and contact details; **Section 4.4.** Descriptor section**; Section 4.6.** Software name and version for descriptor generation; **Section 5.4.** Limits of applicability; **Section 6.7** Statistics for goodness-of-fit; **Section 7.7** Predictivity – Statistics obtained by external validation; **Section 7.9** Comment on the external validation of the model

2.5. Model developer(s) and contact details

Name: P. Petkov, A. Chapkanov, C. Kuseva, H. Ivanova, E. Kaloyanova, G. Dimitrova, D. Yordanova, R. Serafimova, M. Todorov, T. Pavlov, S. Kotov, E. Jacob, A. Aptula, O. Mekenyan

Affiliation: Laboratory of Mathematical Chemistry, University "Prof. As. Zlatarov", "Yakimov" St. #1, 8010 Bourgas, BULGARIA

URL: http://www.oasis-lmc.org
E-mail: ovanes.mekenyan@oasis-lmc.org

2.6. Date of model development and/or publication

Date of the model development: 2020

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

1. R. Serafimova, M. Todorov, T. Pavlov, S. Kotov, E. Jacob, A. Aptula, O. Mekenyan. 2007. Identification of the structural requirement for mutagenicity by incorporating molecular flexibility and metabolic activation of chemicals. II. General Ames mutagenicity model. *Chem. Res. Toxicol*., 662-676.
2. O. Mekenyan, S. Dimitrov, T. Pavlov, G. Dimitrova, M. Todorov, P. Petkov & S. Kotov. 2012. Simulation of chemical metabolism for fate and hazard assessment. V. Mammalian hazard assessment, *SAR and QSAR in Environmental Research*, Vol. 23, 553-606

2.8. Availability of information about the model

In vitro Ames Mutagenicity kinetic model with S9 metabolic activation is proprietary and its use is subject of licence agreement.
Information that cannot be disclosed:

- External validation sets,
- Proprietary chemicals,
- Source code.

For more details, please contact Professor Ovanes Mekenyan: ovanes.mekenyan@oasis-lmc.org
Details of the model is provided in the sections bellow.

2.9. Availability of another QMRF for exactly the same model

Not available.

Section 3. Defining the endpoint – OECD Principle 1

3.1. Species

Chemicals included in the training set of the TIMES Ames model are collected according to the recommendation in the OECD technical guideline 471 addressing the number of *Salmonella typhimurium* strains (TA100, TA98, TA1535, TA1537 (TA97, TA97a), TA102) and/or *E. coli*, associated with each data:

- For negative effect, all five *Salmonella* strains must show simultaneously negative data as described in the corresponding OECD guideline for testing of chemicals: http://www.oecd-ilibrary.org/environment/test-no-471-bacterial-reverse-mutation-test_9789264071247-en.
- For positive effect, positive data in a single Salmonella strain/ E. coli would be enough.

3.2. Endpoint

Bacterial Reverse Mutation Test

According to JRC pre-classification list of endpoints:
No. 207 QMRF Human Health Effects, QMRF 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 - simple and inexpensive *in vitro* assay is the basic in vitro assay to detect mutagens. The Ames test detects single nucleotide base change, base insertion or deletion in different *Salmonella* strains. All *Salmonella* strains carry some type of defective (mutant) gene that prevents them from synthesizing the amino acid histidine. In the presence of mutagenic chemicals, the defective gene may be mutated back to the functional state allowing the bacterium to grow. The relevant test guideline covering this endpoint is OECD TG 471. 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 endpoint is measured on the parent compound and the metabolites generated in vitro by the employed S9 mix of enzyme-induced rodent liver homogenates.

3.4. Endpoint units

Qualitative – positive/ negative

3.5. Dependent variable

Observed Mutagenicity with S9

3.6. Experimental protocol

OECD technical guideline 471: Bacterial Reverse Mutation Assay (e.g. Ames test). https://www.oecd-ilibrary.org/environment/test-no-471-bacterial-reverse-mutation-test_9789264071247-en

3.7. Endpoint data quality and variability

References associated with each documented mutagenicity data (except for proprietary data) included in the training set of the model are provided in Appendix 1.

Section 4. Defining the algorithm – OECD Principle 2

4.1. Type of model

Structural alerts based model

4.2. Explicit algorithm

Prediction of Bacterial (Ames) mutagenicity is based on modelling of the two events deemed to be crucial for the effect – interaction of the chemicals with DNA and their activation as a result of liver S9 metabolism.

In the new modelling concept the presence of alerts is necessary but not sufficient reason for predicting a positive effect. It requires taking into account the kinetics of metabolism. In the new (kinetic) in vitro S9 metabolic simulators:

- Experimental kinetic data (clearances) are used to calculate the probability of transformations as a function of time already, i.e. $P = (1-\exp(-Cl*t))$.
- Clearance data have been used to optimize the probability of metabolic transformations.

In addition, expert information for the stability of chemicals in rat S9 mix is also used.

As a result, the transformations in the metabolic simulator are modified to simulate the formation of DNA adducts (not existing in the original models). Mutagenic potency is associated with concentration of the formed DNA adducts which is estimated over time. The magnitude of these adducts is assumed to correspond to the level of damage of macromolecules and potency effect, respectively. As a result empirical thresholds of the formed DNA adducts are derived to distinguish positive from negative chemicals.

Details about the alerts included in the model are provided in the next sections.

4.3. Descriptors in the model

Descriptors in the model are structural alerts related to interactions with DNA. Alerts in the TIMES Ames kinetic model (+S9) constitute expertly-derived sets of structural fragments incorporating knowledge for the interactions of chemicals (parents and metabolites) with DNA.

Description of these alerts is provided in the next sections.

4.4. Descriptor section

The main characteristics of each DNA alert in TIMES_Ames (+S9) kinetic model are as follows:

- Alert name (corresponding to the name of the chemical class which is addressed);
- Performance of alert (correct/incorrect predictions) which is estimated based on proportion of observed positive chemicals from all chemicals captured by the alert. Performance of each alert is provided with its confidence range. As smaller is the size of local training sets as wider are the confidence ranges and vice versa.
- P-values addressing the reliability of alert performance estimation and taking into account possible bias of positive/negative chemicals in the training set of the model. Low p-values could be obtained only if both are satisfied:
    - The number of chemicals in local training set is high enough;

- The alert performance is significantly higher than the proportion of positive/negative chemicals in the model training set, i.e. so-called naïve alert.

Analogically, high p-values could be obtained in case of:

- Small number of local training set chemicals (1-2 chemicals); or
- Performance comparable to the performance of the naïve alert.

High performance associated with low *p-values* indicates for High Reliability of alerts. Full list of the alerts in TIMES_Ames (+S9) kinetic model with defined quantity thresholds is given in Table 1:

Table 1. List with DNA alerts in the TIMES Ames kinetic model (+S9).

No	Alert Name	Defined Quantity threshold for	
		parents	metabolites
1	1,4-Diazabutadiene Derivatives	0	0
2	4,4'-Bipyridinium Salts and N-Oxides	0	0
3	Acridone, Thioxanthone, Xanthone, Phenazine and Other Fused-Ring Heterocyclic DNA Intercalators	0.49	0.082
4	Acyclic Triazenes	0	0
5	Acyl Halides	0	0.38
6	Acyl Halides	0	0.38
7	Aliphatic saturated monoaldehydes	0	0
8	Alkyl Xanthate Esters	0	0
9	Alkylated nitrosoureas and nitrosoguanidines	0	0
10	Alkylnitrites	0	0.036
11	Alkylphosphates, Alkylthiophosphates and Alkylphosphonates	0.73	0.25
12	Alpha,Beta-Unsaturated Aldehydes	0	0.041
13	alpha,beta-Unsaturated Carbonyls and Related Compounds	0	0
14	alpha,beta-Unsaturated Carboxylic Acids and Esters	0	0
15	alpha,omega-Dihaloalkanes	0	0
16	alpha-Activated Benzyls	0	0
17	alpha-Activated Haloalkanes	0	0
18	Alpha-Beta Conjugated Alkene Derivatives with Geminal Electron-Withdrawing Groups	0.97	0.098
19	Alpha-Haloethers	0.48	0.55
20	Amidoxime Esters and Amidoximes	0.8	0.6
21	Amino Anthraquinones	0.5	0.125

22	Aminoacridine DNA Intercalators	0	0
23	Aminophenoxazinone Derivatives	0	0
24	Amphetamine derivatives	0	0
25	Anthrones	0	0.67
26	Antibiotic Aminoglycoside Derivatives	0	0
27	Arenecarboxylic Acid Esters	0	0
28	Arenediazonium and Diazonium Salts	0	0.017
29	Arenesulfonamides	0	0
30	Aromatic ester hydroxylamine	0	0
31	Atrazine derivatives	0	0
32	Azoalkanes with Activating Electron-Withdrawing Groups	0	0
33	Azodicarbonamides	0	0
34	Azoxyalkanes	0	0.038
35	Benzanthrone Derivatives	0	0
36	Benzofuranyl Carbamate Derivatives	0	0
37	Benzoquinoline and Acridine derivatives	0	0
38	Benzoyl Cyclohexanedione Derivatives	0	0
39	Bipyridilium Herbicides	0	0
40	Bleomycin and Structurally Related Compounds	0	0
41	C-Nitroso Compounds	0	0
42	Carbamates	0	0
43	Carboxylic Acid Amides	0	0
44	Carboxylic Acid Anhydrides	0	0
45	Chlorinated Diphenylmethane and Benzophenone Derivatives	0.7	0.2
46	Conjugated Benzoylethene Derivatives	0	0
47	Conjugated Nitroalkenes and Five-Membered Aromatic Nitro- and Amino Heterocycles	0.68	0.36
48	Coumarins and Thiocoumarins	0	0
49	Coumarins and Thiocoumarins	0	0
50	Coumarins and Thiocoumarins	0	0
51	Cyanohydrins	0	0
52	Cyclic maleic acid derivatives	0	0
53	Dialkyl Alkylphosphonates	0	0
54	Diazenes	0	0.021
55	Diazoalkanes	0.93	0.16
56	Dicarbonyl compounds	0.5	0.18
57	Dichlorophosphine and Dichlorophosphonium Derivatives	0	0
58	Dithianes	0	0
59	DNA Intercalators with Carboxamide and Aminoalkylamine Side Chain	0	0
60	Epoxides, Aziridines, Thiiranes and Oxetanes	0.53	0.105
61	Ethenyl Pyridines	0	0
62	Flavonoids	0	0
63	Fluoro bis-benzothiazole derivatives	0	0

64	Formaldehyde Releasers	0.81	0.13
65	Four- and Five-Membered Lactones	0	0
66	Fused-Ring Conjugated Lactones	0	0
67	Fused-Ring Nitroaromatics	0.39	0.2
68	Fused-Ring Primary Aromatic Amines	0	0.07
69	Gallic Acid Esters	0	0
70	Geminal Polyhaloalkane Derivatives	0	0
71	Haloalcohols	0	0
72	Haloalkane Derivatives Containing Chain Heteroatom	0	0
73	Haloalkane Derivatives with Labile Halogen	0	0
74	Haloalkene Cysteine S-Conjugates	0	0
75	Haloalkene Derivatives with Electron-Withdrawing Groups	0.58	0.33
76	Haloazaarene and Fused-Ring Haloquinoline Derivatives	0.87	0.45
77	Halofuranones	0	0
78	Halogenated Oxetanes and Haloepoxides	0	0
79	Halogenated Vicinal Hydrocarbons	0	0
80	Haloisothiazolinones	0	0
81	Heteroarene Sulfenamides	0	0
82	Heterocyclic Aromatic Amines	0	0
83	Heterocyclic N-Hydroxylamines	0	0.03
84	Heterocyclic nitro compounds	0	0
85	Heterocyclic Nitroso compounds	0	0
86	Heterocyclic urea derevatives	0	0
87	Hexahydrotriazine Derivatives	0	0
88	Hydrazine Derivatives	0.28	0.07
89	Hydroxamic Acids	0	0.31
90	Hydroxybenzophenone Derivatives	0	0.14
91	Hydroxylated Phenols	0	0
92	Hypoxanthine Derivatives	0.5	0.36
93	Imidazolinone derivatives	0	0
94	Isocyanates and Diisocyanates	0	0
95	Isothiocyanates	0	0
96	Lactones	0	0
97	Monohaloalkanes	0	0
98	N,N-Dialkyldithiocarbamate Derivatives and Azaarene Dithiocarbamates	0	0
99	N-acetoxyamines	0	0
100	N-Acyloxy(Alkoxy) Arenamides	0	0
101	N-Alkyl-N-nitrosocarbamates	0	0
102	N-Alkylindolinium and N-Alkylbenzothiazolium Salts	0	0
103	N-Aryl-N-Acetoxy(Benzoyloxy) Acetamides	0	0
104	N-Hydroxyethyl Lactams	0	0
105	N-Hydroxylamines	0.51	0.055
106	N-methylol derivatives	0	0
107	N-Nitrosamines	0.55	0.08

108	N-Nitroso Compounds	0.55	0.08
109	N-Nitrosoamine derivatives	0	0
110	N-Substituted Aromatic Amines	0	0
111	N-Trihalomethyl Imides	0	0
112	Nitroalkanes – Mononitroalkanes	0	0
113	Nitroaniline Derivatives	0	0.391
114	Nitroarenes with Other Active Groups	0.5	0.13
115	Nitroazoarenes and p-Monosubstituted Azobenzene Derivatives	0.5	0.05
116	Nitrobiphenyls and Bridged Nitrobiphenyls	0.84	0.16
117	Nitrogen and Sulfur Mustards	0	0
118	Nitrogen mustards	0	0
119	Nitrophenols, Nitrophenyl Ethers and Nitrobenzoic Acids	0.6	0.1
120	Non-aromatic conjugated systems with electron-withdrawing groups	0	0
121	Non-Aromatic Hydroxylamine Derivatives	0	0
122	Non-Cyclic Alkyl Phosphoramides and Thionophosphoramides	0	0
123	Organic Azides	0	0
124	Organic Diselenides and Ditellurides	0	0
125	Organic Peroxy Compounds	0	0
126	p-Aminobiphenyl Analogs	0	0
127	p-Substituted Mononitrobenzenes	0	0
128	PAH Benzylic Alcohol Esters	0	0
129	Perfluorinated Hypofluorites	0	0
130	Peroxyacyl Nitrates	0	0
131	Polarized Haloalkene Derivatives	0	0
132	Polycyclic Aromatic Hydrocarbon, Naphthaleneimide and Carbazole Derivatives	0	0.54
133	Polyethylene Polyamines	0	0
134	Polynitroarenes	0	0.5
135	Propargyl Alcohol Derivatives	0	0
136	Propyne Derivatives	0	0
137	Pyrazolone and Pyrazolidine-3,5-dione Derivatives	0	0
138	Pyrrolizidine Derivatives	0	0
139	Quinoline Derivatives	0.47	0
140	Quinolone Derivatives	0	0
141	Quinone methides	0.7	0
142	Quinoneimine, Thionine and Phenoxazinium Derivatives	0	0.05
143	Quinoneimines protein binding	0	0
144	Quinones and Trihydroxybenzenes	0	0.013
145	Quinoxaline-Type 1,4-Dioxides	0	0.07
146	S-Activated Cysteine Derivatives	0	0
147	Short-Chain Alkyltin and Alkylgermanium Halides	0	0
148	Single-Ring Substituted Primary Aromatic Amines	0.15	0.2
149	Specific 5-Substituted Uracil Derivatives	0	0

150	Specific Acetate Esters	0	0
151	Specific Imine and Thione Derivatives	0.47	0.4
152	Sterically Hindered Piperidine Derivatives	0	0
153	Substituted Anilines	0	0
154	Substituted Benzoindoline and Indole Derivatives	0	0
155	Substituted Chlorophenylalkylurea Derivatives	0	0
156	Substituted Nitropyridines, Aminopyridines and N-Oxides	0.4	0.08
157	Substituted Phenols	0	0
158	Sulfonates and Sulfates	0.81	0.045
159	Sulfonyl Azides	0	0
160	Sulfonyl Halides	0	0
161	Sultones	0	0
162	Tertiary aromatic amine	0	0
163	Thiadiazole-dioxide derivatives	0	0
164	Thiazolidinediones	0.8	0.4
165	Thiocarbonyl S,S-dioxides	0	0
166	Thiols	0.962	0.45
167	Tri-Methylindole derivatives	0	0
168	Triarylimidazole and Structurally Related DNA Intercalators	0.895	0
169	Triazinone derivative	0	0
170	Trifluoromethyl Benzamide Derivatives	0	0
171	Trifluoromethylpyridinone Derivatives	0	0.02
172	Vicinal Dihaloalkanes	0	0

4.5. Algorithm and descriptor generation

The structural boundaries of the alerts are derived from the chemicals included in the local training sets. For derivation of each alert mechanistically justifiable structural fragments for interaction with DNA are identified from the chemicals having positive data in the local training set. Additional structural fragments from the other parts of the molecules which could affect (enhance or reduce) the mutagenicity effect are also introduced to complete definition of most alerts.

4.6. Software name and version for descriptor generation

Ames Mutagenicity S9 activated kinetic model version 06.06

4.7. Chemicals/Descriptors ratio

Not applicable

Section 5. Defining the applicability domain of the model – OECD Principle 3

5.1. Description of the applicability domain of the model

The domain consists of the following sub-domain layers:

1. General parametric requirements.

The variations of molecular parameters that may affect the quality of the measured endpoint significantly are included here (such as molecular weight, etc.). The domain of general parametric includes the range of variation of hydrophobicity (log K_{OW}) and Molecular weight (MW) of chemicals in training set.

2. Structural domain.

The structural component of the model is based on the structural similarity between chemicals in the training set which were correctly predicted by the model. The structural neighborhood of atom-centered fragments (accounting for the first neighbours) extracted from correctly and incorrectly predicted parent structures from the training set is used to determine this similarity.

The target chemical could contain the following types of ACF:

- Fragments present in correctly predicted training chemicals only (i.e. correct fragments),
- Fragments found both in correctly and non-correctly predicted training chemicals (i.e. fuzzy fragments). These fragments are treated as correct fragments,
- Fragments present in non-correctly predicted training chemicals only (i.e. incorrect fragments),
- Fragments not present in the training chemicals (i.e. unknown fragments).

A chemical belongs to the structural domain of the model if it could be partitioned only on correct fragments. The user is able to analyse how important are unknown and incorrect fragments (if present in the target) and to make a decision about their effect on the quality of prediction. The distribution of structural characteristics of the target chemical and accepted thresholds is used as a criterion to determine how well the target is represented in the structural space of correctly predicted chemicals. The accepted domain thresholds for Mutagenicity are as follows:

- Correct = 100%
- Incorrect = 0%

A chemical is considered In Domain if it is classified to belong to all sub-domain levels. The information implemented in the applicability domain is extracted from the correctly predicted training chemicals used to build the model and in this respect the applicability domain determines practically the interpolation space of the model.

5.2. Method used to assess the applicability domain

The approach used to determine and assess the domain is described in:

Dimitrov S, Dimitrova G., Pavlov T., Dimitrova N., Patlewicz G., Niemela J., Mekenyan O., A stepwise approach for defining the applicability domain of SAR and QSAR models, *J. Chem. Inf. Model.*, 45, 839-849 (2005).

5.3. Software name and version for the applicability domain assessment

The LMC software OASIS Domain Manager v.1.13 (which is embedded in OASIS platform) is used to determine the applicability domain.
http://oasis-lmc.org/products/software/domain-manager.aspx

5.4. Limits of applicability

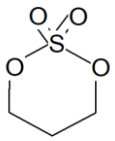
Applicability domain of the Ames kinetic model (+S9) includes three sub-domain layers: general parametric requirements, structural features and alerts reliability.

- General properties requirements:

As described in the Section 5.1.1, parametric domain of the model is derived based on Log *Kow* and *MW*. Example demonstrating belonging of a training set chemical to the parametric layer of the model domain is provided below:

Example chemical:

- CAS: 1073-05-8
- Name: 1,3, 1,3,2-dioxathiane 2,2-dioxide
- 2D Depiction:



Property	Domain	Example chemical
log K_{OW}	[-18.859; 35.185]	-0.410
MW, Da	[31.025; 2368.504]	138.137

* K_{OW} is calculated by EPIWin

The values of *logKow* and *MW* of the example chemical are within the ranges of these parameters extracted from the whole training set of the model. Hence, with respect to the general parametric requirements, the example chemical is estimated to be *In Domian*.

- Structural features

  Structural domain of the model contains:
  - 28683 correct fragments,
  - 3343 fuzzy fragments (treated as correct fragments),
  - 2250 incorrect fragments.

- Alerts reliability

Reliability of alerts is estimated based on:

- Alert performance of the local training set chemicals (AP);
- Number of the local training sets (N);
- Mechanistic justification (M).

According to these criteria, there are four reliability estimates for the alerts in the models:

- High reliability alerts (AP>0.6, N>10, M);
- Low reliability alerts (AP<0.6, N>10, M);
- Undetermined alerts (N<10, M);
- Undetermined theoretical alerts (M).

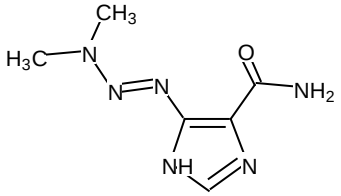
Example chemical belonging to alert with “High reliability”.

Chemical ID:

CAS: 4342-03-4

Name: Dacarbazine

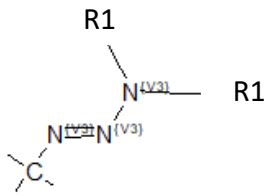
Depiction:



Belonging to alert:

Name: Acyclic Triazenes

Structural boundaries:



R_1 = H; CH_3; C_2H_5; $CH(CH_3)_2$; CH_2-C_6H_5

Reliability:

"High reliability" based on AP=1; N=15 and M.

Currently, information for alerts reliability is provided in the model reports.

Section 6. Defining goodness-of-fit and robustness (internal validation) – OECD Principle 4

6.1. Availability of the training set

Training set of the TIMES Ames kinetic model (+S9) includes 4268 organic compounds from different chemical classes.

6.2. Available information for the training set

CAS numbers, Chemical names, SMILES, documented data, literature sources and strain information are available for each compound in the model training set.

6.3. Data for each descriptor variable for the training set

Not applicable

6.4. Data for the dependent variable for the training set

The training set of 4268 chemicals includes:

- 2092 chemicals have positive observed Ames data
- 2176 chemicals have negative observed Ames chemicals.

Distribution of positive/negative chemicals in the training set of model is used for estimating performance and confidence range of the so-called *naïve alert* which is 0.490 (0.475 ÷ 0.505)[1)].

1) Confidence range is calculated at 95% confidence level

6.5. Other information about the training set

The training set is compiled according to the recommendations described in the OECD TG471 for availability of all five *Salmonella* strains (*E. coli*) for the Ames negative chemicals and at least one strain with positive data for the Ames positive chemicals.

6.6. Pre-processing of data before modelling

Not available

6.7. Statistics for goodness-of-fit

Statistics of the model:

- Sensitivity = (predicted positive/observed positive) = 88%
- Specificity = (predicted negative/observed negative) = 94%
- Concordance = (correct predicted positive and negative chemicals in respect to all training set chemicals) = 91%

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

Not performed

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

Not performed

6.10. Robustness - Statistics obtained by Y-scrambling

Not performed

6.11. Robustness - Statistics obtained by bootstrap

Not performed

6.12. Robustness - Statistics obtained by other methods

Not performed

6.13. Comment on the internal validation of the model

Not performed

Section 7. Defining predictivity (external validation) – OECD Principle 4

7.1. Availability of the external validation set

10 188 external chemicals are available to examine performance of the model.

7.2. Available information for the external validation set

According to the OECD TG 471, the external validation set addresses the five *Salmonella* strains (TA100, TA90, TA1535, TA1537 and *E. coli* WP2 uvrA).

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

Not available

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

Not available

7.5. Other information about the external validation set

The list with 10 188 chemicals with Ames negative experimental data (Class C) are provided by the Division of Genetics and Mutagenesis of National Institute of Health Sciences of Japan. Details for the data used in the current external validation are available in the corresponding publication:

M. Honma, A. Kitazawa, A. Cayley, R. Williams, C. Barber, T. Hanser, R. Saiakhov, S. Chakravarti, G. Myatt, K. Cross, E. Benfenati, G. Raitano, O. Mekenyan, P. Petkov, C. Bossa, R. Benigni, C. Battistelli, A. Giuliani, O. Tcheremenskaia, C. DeMeo, U. Norinder, H. Koga, C. Jose, N. Jeliazkova, N. Kochev, V. Paskaleva, C. Yang, P. Daga, R. Clark, J. Rathman. 2019. Improvement of quantitative structure-activity relationship (QSAR) tools

for predicting Ames mutagenicity: outcomes of Ames/QSAR International Challenge Project. Mutagenesis, Vol. 34, pp. 3-16.

7.6. Experimental design of test set

The external validation set contains only AMES negative substances.

7.7. Predictivity – Statistics obtained by external validation

- 9 161 out of 10 188 are predicted as Negatives by TIMES *in vitro* Ames kinetic model and these predictions are consistent with the experimental data, thus, the Specificity of the external set is 90% (9161/10188)
- 1 027 false positives
- Applicability domain – 9 441 out of 10 188 chemicals are *In domain*

7.8. Predictivity – Assessment of the external validation set

The study reports of the Ames tests were peer reviewed by the ANEI-HOU committee comprising several Ames experts from academia and National Institutes and the results were authorised.

7.9. Comment on the external validation of the model

Performance of the TIMES_Ames kinetic model (+S9) with respect to non-mutagenic chemicals is quite high: the Specificity of the external set is 90% (9161/10188).

Section 8. Providing a mechanistic interpretation – OECD Principle 5

8.1. Mechanistic basis of the model

Only alerts extracted from the local training sets having clear interpretation of the molecular mechanism causing the mutagenicity effect are included in the model. Mechanistic rational of each alert is provided by experts based on significant reference support from the literature.

8.2. *A priori* or *a posteriori* mechanistic interpretation

The model building followed the traditional approach:

a. Building a hypothesis for the modelled event,

b. Defining the alerting groups based on parent structures,
c. Defining empirically the appropriate adduct quantity thresholds,
d. Fitting of model variable to the observed data,
e. Verification of model quality,
f. Depending on the results found in step *e* model building could continue with step *a*, *b*, *c* or *g*,
g. Determination of the applicability domain and practical application of the model.

8.3. Other information about the mechanistic interpretation

Additional information about the mechanistic interpretation could be fund in Section 2 (2.7).

Section 9. Miscellaneous information

9.1. Comments

Model predictions are fully transparent. The user is able to analyse the whole prediction process and to verify whether it concise with his/her knowledge or purposes.
For other related models, see Section 1 (1.2).

9.2. Bibliography

Additional references are not provided.

9.3. Supporting information

Additional supporting information is not provided.