(Q)SAR Model Reporting Format (QMRF)

(The present QMRF is prepared according the fields and *Help* recommendations of JRC implemented in QMRF v 1.3 and QMRF Editor v.2.0.0 (https://sourceforge.net/apps/mediawiki/qmrf/index.php?title=Main_Page**)**

Welcome

Model version: Ames mutagenicity v.18.18
Platform version: OASIS TIMES 2.31.2
Name: Ames mutagenicity
Author: LMC, University "Prof. As. Zlatarov", Bourgas, Bulgaria
Date: 23 Nov 2021
E-mail: omekenya@btu.bg
www: http://www.oasis-lmc.org/

Section 1. QSAR identifier

1.1. QSAR identifier (title)

In vitro Ames Mutagenicity with S9 metabolic activation

1.2. Other related models

Not applicable

1.3. Software coding the model

Model version: Ames mutagenicity v.18.18
Platform version: OASIS TIMES 2.31.2
Name: *In vitro* Ames Mutagenicity with S9 metabolic activation
Developer: LMC, University "Prof. As. Zlatarov", Bourgas, Bulgaria
Coding language: Delphi 10.2

Section 2. Date of QMRF

2.1. Date of QMRF

23 Nov 2021

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: omekenya@btu.bg

2.3. Date of QMRF update(s)

20 November 2014; 12 June 2015, 11 May 2016; 12 July 2016; 31 August 2016; 26 May 2017; 20 July 2018; 22 Jan 2020; 23 Nov 2021

2.4. QMRF update(s)

Information which has been modified:

Sections 2.8. Availability of information about the model; **Section 3.7.** Endpoint data quality and variability; **Section 4.2.** Explicit algorithm; **Section 4.3** Descriptors in the model; **Section 4.4.** Descriptor section**; Section 4.6.** Software name and version for descriptor generation; **Section 5.3.** Software name and version for the applicability domain assessment; **Section 5.4.** Limits of applicability; **Section 6.4** Data for the dependent variable for the training set; **Section 6.7** Statistics for goodness-of-fit; **Section 6.9** Robustness - Statistics obtained by leave-many-out cross-validation; **Section 6.11** Robustness - Statistics obtained by bootstrap; Section **6.13** Comment on the internal 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: omekenya@btu.bg

2.6. Date of model development and/or publication

2006/2012

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

TIMES_Ames mutagenicity model (+S9) is derived for identification of chemicals capable to interact with DNA. Training set of the model includes 4129 chemicals (part of which are proprietary data) from different literature sources. The model is based on an alerting group approach addressing mutagenicity of parents and their generated metabolites *in vitro* liver S9 metabolic system. Details of the model is provided in the sections bellow as well as in the following link:
http://oasis-lmc.org/products/models/human-health-endpoints/mutagenicity-(ames).aspx

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, *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 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

To detect and measure potency of DNA mutagens, a short-term, simple and inexpensive *in vitro* assay is used, such as the Bacterial reverse mutation (Ames) test. 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.

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).

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.

Ames mutagenicity predictions are obtained using an alerting group approach. Only alerts having clear interpretation of mechanisms leading to DNA mutagenicity are included in the model. To obtain predictions, a set of alerts (116) is applied on parents and each of the generated *in vitro* rat liver S9 metabolites. The *in vitro* S9 metabolic simulator is trained to reproduce documented maps for mammalian liver metabolism for 438 chemicals. Match of alerts either on parents or metabolites is sufficient for obtaining positive prediction. Chemicals are predicted to be mutagenic as parents only, parents and metabolites, or as metabolites only.

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 model (+S9) constitute expertly-derived sets of structural fragments incorporating knowledge for the interactions of chemicals (parents and metabolites) with DNA. Application of the alerts on the training set of the model forms fractions of representative chemicals for the alerts, i.e. so-called 'local' training sets. All chemicals captured by the alerts are considered as validation sets of the introduced expert knowledge addressing reactivity of chemicals with DNA. The procedure for obtaining local training sets includes applying the structural boundaries of the alert searching among all chemicals from the training set of the model after application of S9 metabolic simulator. According to this, local training sets contain parent chemicals in which general fragments are:

- found in their structures;
- not found in the parent structures but found in their metabolite(s).

Description of these alerts is provided in the next sections.

4.4. Descriptor section

Table 1 summarizes the main characteristics of each DNA alert in TIMES Ames (+S9):

- 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* indicate for High Reliability of alerts.

The above statistical measures along with the underlying mathematical formalisms are discussed in details in **Section 6** (Internal validation).

Table 1. Main characteristics of the DNA alerts in the TIMES Ames model (+S9).

No	Alert Name	Correct	Incorrect	Performance[1)]	p-value[1)]
1	Polynitroarenes	50	0	0.981 (0.943 ÷ 1.000)	$< 1.0 \times 10^{-10}$
2	N-Nitroso Compounds	41	0	0.979 (0.938 ÷ 1.000)	$< 1.0 \times 10^{-10}$
3	Nitrogen and Sulfur Mustards	41	0	0.977 (0.931 ÷ 1.000)	$< 1.0 \times 10^{-10}$
4	Nitrophenols, Nitrophenyl Ethers and Nitrobenzoic Acids	84	5	0.977 (0.883 ÷ 0.980)	$< 1.0 \times 10^{-10}$
5	Nitroarenes with Other Active Groups	36	0	0.974 (0.922 ÷ 1.000)	$< 1.0 \times 10^{-10}$
6	Heterocyclic nitro compounds	32	0	0.971 (0.913 ÷ 1.000)	2.4×10^{-10}

7	Fused-Ring Nitroaromatics	61	1	0.969 (0.927 ÷ 0.999)	$< 1.0 \times 10^{-10}$
8	Haloalkene Derivatives with Electron-Withdrawing Groups	30	0	0.969 (0.908 ÷ 1.000)	9.7×10^{-10}
9	N-Acyloxy(Alkoxy) Arenamides	30	0	0.969 (0.908 ÷ 1.000)	9.7×10^{-10}
10	Aminoacridine DNA Intercalators	28	0	0.967 (0.902 ÷ 1.000)	3.8×10^{-9}
11	Conjugated Nitroalkenes and Five-Membered Nitro- and Amino Heterocycles	56	1	0.966 (0.920 ÷ 0.966)	$< 1.0 \times 10^{-10}$
12	N-Acetoxyamines	26	0	0.964 (0.895 ÷ 1.000)	1.5×10^{-8}
13	N-Hydroxylamines	457	16	0.964 (0.947 ÷ 0.980)	$< 1.0 \times 10^{-10}$
14	Haloalkane Derivatives Containing Chain Heteroatom	99	3	0.962 (0.924 ÷ 0.993)	$< 1.0 \times 10^{-10}$
15	DNA Intercalators with Carboxamide and Aminoalkylamine Side Chain	22	0	0.958 (0.878 ÷ 1.000)	2.4×10^{-7}
16	Halofuranones	19	0	0.952 (0.861 ÷ 1.000)	1.9×10^{-6}
17	p-Substituted Mononitrobenzenes	38	1	0.951 (0.886 ÷ 0.999)	7.9×10^{-11}
18	Sulfonates and Sulfates	57	2	0.951 (0.897 ÷ 0.995)	$< 1.0 \times 10^{-10}$
19	Acridone, Thioxanthone, Xanthone, Phenazine and Other Fused-Ring Heterocyclic DNA Intercalators	35	1	0.947 (0.877 ÷ 0.999)	5.7×10^{-10}
20	Quinolone Derivatives	17	0	0.947 (0.847 ÷ 1.000)	7.6×10^{-6}
21	Organic Peroxy Compounds	33	1	0.944 (0.871 ÷ 0.999)	2.1×10^{-10}
22	Nitrobiphenyls and Bridged Nitrobiphenyls	32	1	0.943 (0.867 ÷ 0.999)	4.1×10^{-9}
23	Acyclic Triazenes	14	0	0.938 (0.810 ÷ 1)	0.0001
24	Nitroazoarenes and p-Monosubstituted Azobenzene Derivatives	59	3	0.938 (0.878 ÷ 0.988)	$< 1.0 \times 10^{-10}$

25	Geminal Polyhaloalkane Derivatives	51	3	0.929 (0.861 ÷ 0.986)	$< 1.0 \times 10^{-10}$
26	Haloalkane Derivatives with Labile Halogen	179	13	0.928 (0.891 ÷ 0.962)	$< 1.0 \times 10^{-10}$
27	Specific Imine and Thione Derivatives	24	1	0.926 (0.829 ÷ 0.998)	7.8×10^{-7}
28	Benzanthrone Derivatives	11	0	0.923 (0.779 ÷ 1.000)	0.0005
29	Diazoalkanes	11	0	0.923 (0.779 ÷ 1.000)	0.0005
30	Fused-Ring Primary Aromatic Amines	71	5	0.923 (0.864 ÷ 0.976)	$< 1.0 \times 10^{-10}$
31	N-methylol derivatives	11	0	0.923 (0.779 ÷ 1.000)	0.0005
32	Hydrazine Derivatives	57	7	0.917 (0.762 ÷ 1.000)	$< 1.0 \times 10^{-10}$
33	p-Aminobiphenyl Analogs	21	1	0.917 (0.809 ÷ 0.998)	5.5×10^{-10}
34	Epoxides, Aziridines, Thiiranes and Oxetanes - renamed	186	17	0.912 (0.873 ÷ 0.949)	$< 1.0 \times 10^{-10}$
35	Quinoneimine, Thionine and Phenoxazinium Derivatives	30	2	0.912 (0.818 ÷ 0.990)	1.3×10^{-7}
36	Alkylnitrites	9	0	0.909 (0.741 ÷ 1.000)	0.0019
37	Four- and Five-Membered Lactones	9	0	0.909 (0.741 ÷ 1.000)	0.019
38	Polycyclic Aromatic Hydrocarbon, Naphthaleneimide and Carbazole Derivatives	29	2	0.909 (0.812 ÷ 0.989)	2.4×10^{-7}
39	Heterocyclic N-Hydroxylamines	37	3	0.905 (0.816 ÷ 0.981)	1.0×10^{-8}
40	Coumarins and Thiocoumarins	8	0	0.900 (0.717 ÷ 1.000)	0.0039
41	N,N-Dialkyldithiocarbamate Derivatives and Azaarene Dithiocarbamates	8	0	0.900 (0.717 ÷ 1.000)	0.0039
42	Organic Azides	8	0	0.900 (0.717 ÷ 1.000)	0.0039
43	Polarized Haloalkene Derivatives	44	4	0.900 (0.817 ÷ 0.973)	8.4×10^{-10}

44	Quinone Methides	7	0	0.900 (0.717 ÷ 1.000)	0.0039
45	Thiols	7	0	0.900 (0.717 ÷ 1.000)	0.0039
46	Alkylphosphates, Alkylthiophosphates and Alkylphosphonates	16	1	0.895 (0.761 ÷ 0.997)	0.0001
47	Alpha-Haloethers	16	1	0.895 (0.761 ÷ 0.997)	0.0001
48	Amino Anthraquinones	23	2	0.889 (0.772 ÷ 0.986)	9.8×10^{-6}
49	Arenediazonium and Diazonium Salts	15	1	0.889 (0.748 ÷ 0.997)	0.0003
50	Quinones and Trihydroxybenzenes	74	9	0.882 (0.813 ÷ 0.946)	$< 1.0 \times 10^{-10}$
51	Hydroxamic acid	10	0	0.879 (0.224 ÷ 1.000)	0.001
52	Haloalkene Cysteine S-Conjugates	6	0	0.0.875 (0.652 ÷ 1.000)	0.015
53	N-Aryl-N-Acetoxy(Benzoyloxy) Acetamides	6	0	0.0.875 (0.652 ÷ 1.000)	0.015
54	Haloalcohols	55	7	0.875 (0.794 ÷ 0.950)	1.4×10^{-10}
55	Haloazaarene and Fused-Ring Haloquinoline Derivatives	12	2	0.875 (0.627 ÷ 0.973)	6.4×10^{-10}
56	Flavonoids	6	0	0.875 (0.652 ÷ 1.000)	0.015
57	Substituted Benzoindoline and Indole Derivatives	6	0	0.875 (0.652 ÷ 1.000)	0.015
58	Substituted Chlorophenylalkylurea Derivatives	6	0	0.875 (0.652 ÷ 1.000)	0.015
59	4,4’-Bipyridinium Salts and N-Oxides	5	0	0.857 (0.607 ÷ 1.000)	0.031
60	Pyrrolizidine derivatives	5	0	0.857 (0.607 ÷ 1.000)	0.031
61	Substituted Nitropyridines, Aminopyridines and N-Oxides	5	0	0.857 (0.607 ÷ 1.000)	0.031
62	Nitroaniline Derivatives	124	17	0.874 (0.819 ÷ 0.926)	$< 1.0 \times 10^{-10}$

63	Acyl Halides	25	3	0.867 (0.746 ÷ 0.972)	1.4×10^{-5}
64	Formaldehyde Releasers	12	1	0.867 (0.701 ÷ 0.996)	0.0017
65	Single-ring Substituted Primary Aromatic Amines	292	47	0.859 (0.822 ÷ 0.895)	$< 1.0 \times 10^{-10}$
66	C-Nitroso Compounds	266	10	0.833 (0.549 ÷ 1.000)	$< 1.0 \times 10^{-10}$
67	Specific Acetate Esters	9	1	0.833 (0.632 ÷ 0.994)	1.1×10^{-2}
68	Vicinal Dihaloalkanes	14	2	0.833 (0.666 ÷ 0.977)	2.1×10^{-3}
69	Alpha-Beta Conjugated Alkene Derivatives with Geminal Electron-Withdrawing Groups	4	0	0.833 (0.549 ÷ 1.000)	0.062
70	Chlorinated Diphenylmethane and Benzophenone Derivatives	4	0	0.833 (0.549 ÷ 1.000)	0.062
71	Non-Aromatic Hydroxylamine Derivatives	4	0	0.833 (0.549 ÷ 1.000)	0.062
72	Sultones	4	0	0.833 (0.549 ÷ 1.000)	0.062
73	Tri-Methylindole derivatives	4	0	0.833 (0.549 ÷ 1.000)	0.062
74	Triarylimidazole and Structurally Related DNA Intercalators	8	1	0.818 (0.602 ÷ 0.993)	1.9×10^{-3}
75	Alpha,Beta-Unsaturated Aldehydes	30	6	0.816 (0.693 ÷ 0.929)	3.5×10^{-5}
76	Diazenes	3	0	0.800 (0.473 ÷ 1.000)	0.124
77	Quinoxaline-Type 1,4-Dioxides	3	0	0.800 (0.473 ÷ 0.974)	0.124
78	Specific 5-Substituted Uracil Derivatives	3	0	0.800 (0.473 ÷ 0.974)	0.124
79	Sulfonyl Halides	3	0	0.800 (0.473 ÷ 1.000)	0.124
80	Dicarbonyl Compounds	54	13	0.797 (0.702 ÷ 0.888)	2.5×10^{-7}
81	Heterocyclic Nitroso compounds	13	3	0.778 (0.570 ÷ 0.947)	1.1×10^{-2}

82	Monohaloalkanes	7	1	0.750 (0.567 ÷ 0.991)	0.035
83	Bleomycin and Structurally Related Compounds	2	0	0.750 (0.368 ÷ 1.000)	0.249
84	Dichlorophosphine and Dichlorophosphonium Derivatives	2	0	0.750 (0.368 ÷ 1.000)	0.249
85	Fused-Ring Conjugated Lactones	2	0	0.750 (0.368 ÷ 1.000)	0.249
86	Halogenated Oxetanes and Haloepoxides	2	0	0.750 (0.368 ÷ 1.000)	0.249
87	Hypoxanthine Derivatives	2	0	0.750 (0.368 ÷ 1.000)	0.249
88	N-Alkylindolinium and N-Alkylbenzothiazolium Salts	2	0	0.750 (0.368 ÷ 1.000)	0.249
89	Polyethylene Polyamines	2	0	0.750 (0.368 ÷ 1.000)	0.249
90	Quinoline Derivatives	15	5	0.727 (0.544 ÷ 0.899)	0.02
91	Nitroalkanes	4	1	0.714 (0.409 ÷ 0.982)	0.186
92	Anthrones	3	1	0.667 (0.330 ÷ 0.974)	0.311
93	Propyne Derivatives	3	1	0.667 (0.330 ÷ 0.974)	0.311
94	1,4-Diazabutadiene Derivatives	1	0	0.667 (0.224 ÷ 1.000)	0.499
95	Alkyl Xanthate Esters	1	0	0.667 (0.224 ÷ 1.000)	0.499
96	Amidoxime Esters and Amidoximes	1	0	0.667 (0.224 ÷ 1.000)	0.499
97	Antibiotic Aminoglycoside Derivatives	1	0	0.667 (0.224 ÷ 1.000)	0.499
98	Azoalkanes with Activating Elactron-Withdrawing Groups (EWG)	1	0	0.667 (0.224 ÷ 1.000)	0.499
99	Conjugated Benzoylethene Derivatives	1	0	0.667 (0.224 ÷ 1.000)	0.499
100	Dithianes	1	0	0.667 (0.224 ÷ 1.000)	0.499

101	Haloisothiazolinones	1	0	0.667 (0.224 ÷ 1.000)	0.499
102	Heterocyclic urea derevatives	1	0	0.667 (0.800 ÷ 0.951)	0.499
103	N-Hydroxyethyl Lactams	1	0	0.667 (0.800 ÷ 0.951)	0.499
104	Non-Cyclic Alkyl Phosphoramides and Thionophosphoramides	1	0	0.667 (0.800 ÷ 0.951)	0.499
105	Organic Diselenides and Ditellurides	1	0	0.667 (0.800 ÷ 0.951)	0.499
106	PAH Benzylic Alcohol Esters	1	0	0.667 (0.800 ÷ 0.951)	0.499
107	Perfluorinated Hypofluorites	1	0	0.667 (0.800 ÷ 0.951)	0.499
108	Peroxyacyl Nitrates	1	0	0.667 (0.800 ÷ 0.951)	0.499
109	S-Activated Cysteine Derivatives	1	0	0.667 (0.800 ÷ 0.951)	0.499
110	Short-Chain Alkyltin and Alkylgermanium Halides	1	0	0.667 (0.800 ÷ 0.951)	0.499
111	Tertiary aromatic amine	1	0	0.667 (0.800 ÷ 0.951)	0.499
112	Azoxyalkanes	2	1	0.600 (0.228 ÷ 0.956)	0.498

1) Confidence ranges and p-value are calculated at 95% confidence level.

Alerts which are not supported by chemicals from the training set (theoretical alerts) are not included in Table 1. Detailed information for each alert such as structural boundaries, mechanisms, local training sets and references associated with each observed data is provided in ***Appendix 2***.

4.5. Algorithm and descriptor generation

The structural boundaries of the alerts are derived from the chemicals included in the local training sets (see Section 4.3). 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

TIMES Ames mutagenicity model version 18.18

4.7. Chemicals/Descriptors ratio

Provided in Section 4.4.

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 model (+S9) include three sub-domain layers: general parametric requirements, structural features and alerts reliability.

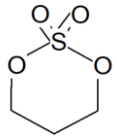
- General properties requirements:

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}	[-18.858; 35.183]	-0.410
MW, Da	[17.996; 2368.505]	138.137

* K_{OW} is calculated by EPI Suite

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 is extracted from 4129 training chemicals contains:
  - 26 164 correct fragments,
  - 1 214 fuzzy fragments (treated as correct fragments),
  - 663 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).

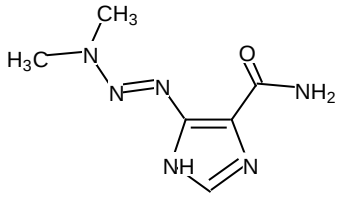
Example chemical belonging to alert with “High reliability”.

Chemical ID:

CAS: 4342-03-4

Name: Dacarbazine

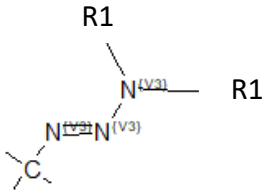
Depiction:



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=19 and M.

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

Section 6. Internal validation – OECD Principle 4

6.1. Availability of the training set

Training set of the TIMES Ames model (+S9) includes 4129 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

Descriptors in the models are structural alerts. The main characteristics of each alert are provided in Table 1 (Section 4.4).

6.4. Data for the dependent variable for the training set

The training set of 4129 chemicals include:

- 2060 chemicals have positive observed Ames data
- 2069 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 0.499 (0.484 ÷ 0.514)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 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

During the internal validation the original training set is separated many times randomly into two parts – one becomes a training set and the other becomes a test set. The model is re-derived many times using each new training set. Then, performance is estimated for the training sets and test sets. The averaged value of all training set performances is compared to the averaged value of all test set performances in order to assess the amount of optimism in the goodness-of-fit (GOF optimism) in the original model. GOF optimism is calculated

as average performance over training sets minus average performance over test sets. Results are provided in Table 2.

Table 2. Performance of the original model over its training set (goodness-of-fit, GOF) vs. expected performance over set different from the training set (GOF – GOF optimism).

	Performance $_{est.}$[1)] model	*p-value*[1)]	Performance[2) 3)] different set
All predictions (accuracy)	0.901 (0.892 ÷ 0.910)	$< 10^{-10}$	0.896
Positive chemicals (sensitivity)	0.890 (0.876 ÷ 0.903)	$< 10^{-10}$	0.879
Negative chemicals (specificity)	0.912 (0.900 ÷ 0.924)	$< 10^{-10}$	0.911

1) Confidence ranges and *p-value* are calculated at 95% confidence level.
2) Estimated performance for training set minus GOF optimism calculated from internal validation.
3) Estimation of expected performance over sets different from training sets.

Addition information including mathematical formalism underlying the above statistical measures are provided in ***Appendix 3***.

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

Not performed

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

Method 1. *k-fold* cross-validation

In ***k-fold cross-validation*** the original training set is partitioned into *k* equally sized subsets. Each time a single subset is used as a test set and the remaining *k-1* subsets are used as training set. In this manner the process is repeated *k* times and each data from the original training set is used once as a test data and *k-1* times as a training data. The advantage of this method is that any data is used for both training and validation and each data is used exactly once as a test data. Commonly the 10-fold cross-validation is used (90% training data, 10% test data). In addition, 4-fold cross validation (75% training data, 25% test data) is also performed and the results from both procedures are provided in Table 3.

Table 3. Results from *k-fold* (*10-fold* and *4-fold*) cross-validation.

	10-fold		4-fold	
	Training sets	Test sets	Training sets	Test sets
Unique chemicals, %	90.0 (90.0 ÷ 90.0)	10.0 (10.0 ÷ 10.0)	75.0 (75.0 ÷ 75.0)	25.0 (25.0 ÷ 25.0)
Performance $_{est.}$, all predictions (accuracy)	0.901 (0.898 ÷ 0.904)	0.896 (0.866 ÷0.926)	0.901 (0.898 ÷ 0.905)	0.896 (0.886 ÷ 0.906)
p-value, accuracy	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$
Performance $_{est.}$, positive chemicals (Sensitivity)	0.890 (0.884 ÷ 0.896)	0.879 (0.824 ÷ 0.934)	0.890 (0.885 ÷ 0.894)	0.880 (0.873 ÷ 0.886)
p-value, Sensitivity	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$
Performance $_{est.}$, negative chemicals (Specificity)	0.9128 (0.908 ÷ 0.916)	0.908 (0.869 ÷ 0.947)	0.912 (0.907 ÷ 0.917)	0.911 (0.897 ÷ 0.925)
p-value, Specificity	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$

1) Confidence ranges and p-value are calculated at 95% confidence level

Method 2. *Monte Carlo* cross-validation

In ***Monte Carlo cross-validation*** the original training set is split randomly into training and test set. The advantage of this method (compared to *k-fold* cross validation) is that the proportion between training and test sets does not depend on the number of repetitions in the internal validation procedure. The ***Monte Carlo cross-validation*** (similarly to the *bootstrapping*) suppose creating a large number of new training/test sets (1000 – 10000). Results from application of this statistical method are provided in Table 5.

Table 4. Results from Monte Carlo cross-validation (1000 repetitions).

	75% training set		63% training set	
	Training sets	Test sets	Training sets	Test sets
Unique chemicals, %	75.0 (75.0 ÷ 75.0)	25.0 (25.0 ÷ 25.0)	63.0 (63.0 ÷ 63.0)	37.0 (37.0 ÷ 37.0)
Performance $_{est.}$, all predictions (Accuracy)	0.901 (0.896 ÷ 0.907)	0.896 (0.879 ÷ 0.912)	0.901 (0.894 ÷ 0.908)	0.895 (0.882 ÷ 0.908)
p-value[1)], Accuracy	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$
Performance $_{est.}$, positive chemicals (Sensitivity)	0.890 (0.882 ÷ 0.898)	0.879 (0.854 ÷ 0.904)	0.890 (0.879 ÷ 0.900)	0.878 (0.858 ÷ 0.897)

p-value, Sensitivity	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$
Performance $_{est.}$, negative chemicals (Specificity)	0.912 (0.905 ÷ 0.919)	0.911 (0.890 ÷ 0.932)	0.912 (0.902 ÷ 0.922)	0.911 (0.895 ÷ 0.927)
p-value, Specificity	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$

1) Confidence ranges and *p-value* are calculated at 95% confidence level.

6.10. Robustness - Statistics obtained by Y-scrambling

Not performed

6.11. Robustness - Statistics obtained by bootstrap

In bootstrapping a newly derived training sets is populated from the original training set of the model by random sampling with replacement until the size of the new training set reaches the size of the original training set. The data not selected for the new training set becomes the new test set. On average, about 63% of original training set data goes into the new training set (some data appear more than once) and 37% remains in the new test set. One of the advantages of this method is that the new training sets and the original training set are equally sized. The process is repeated many times and the average results are provided in Table 5.

Table 5. Results from bootstrapping (1000 repetitions).

	Training sets	**Test sets**
Unique chemicals, %	63.2 (62.2 ÷ 64.3)	36.8 (35.7 ÷ 37.8)
Performance $_{est.}$, all predictions (Accuracy)	0.901 (0.892 ÷ 0.910)	0.895 (0.883 ÷ 0.908)
p-value [1)], accuracy	$< 10^{-10}$	$< 10^{-10}$
Performance $_{est.}$, positive chemicals (Sensitivity)	0.890 (0.877 ÷ 0.903)	0.878 (0.859 ÷ 0.897)
p-value [1)], Sensitivity	$< 10^{-10}$	$< 10^{-10}$
Performance $_{est.}$, negative chemicals (Specificity)	0.912 (0.900 ÷ 0.925)	0.911 (0.895 ÷ 0.928)
p-value [1)], Specificity	$< 10^{-10}$	$< 10^{-10}$

1) Confidence ranges and *p-value* are calculated at 95% confidence level.

6.12. Robustness - Statistics obtained by other methods

Not performed

6.13. Comment on the internal validation of the model

The first evident observation from above results is that averaged estimations are practically unchangeable, no matter what kind of sampling we use for the internal validation. The variability of averaged performances is 0.0033 and below even for test sets. Also their p-values are extremely low, which show very high reliability of these estimations.
The difference between performances of training and test sets - which is a measure for optimism in goodness-of-fit, - is around 0.006 for all predictions, 0.011 for positive chemicals (sensitivity) and 0.002 for negative chemicals (specificity). These values are very low and show that the model is very well balanced and provides high quality for both positive and negative chemicals. A similar quality is also expected for predictions of external chemicals.

Section 7. External validation – OECD Principle 4

7.1. Availability of the external validation set

12140 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 12140 chemicals 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 substances from three categories:

- Class A: Strong positive (672 chemicals)
- Class B: Positive (1085 chemicals)
- Class C: Negative (10383 chemicals)

7.7. Predictivity – Statistics obtained by external validation

Performance for the entire list with chemicals which belong and do not belong to model domain (*In Domain* and *Out of domain)*:

- Sensitivity of the Class A: 61%
- Sensitivity of Class B: 49%
- Specificity of Class C: 80%

Performance for the chemicals which belong to model domain, i.e. "*In Domain":*

- Sensitivity of the Class A: 82%
- Sensitivity of Class B: 50%
- Specificity of Class C: 90%

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 model (+S9) with respect to the strong mutagens from Class A and non-mutagenic chemicals from Class C (belonging to the model domain) is consistent with accuracy of the experimental Ames data (~85%). In turn, performance of the model with respect to Class B mutagens is lower (about 50%) as compared with Class A mutagens probably due to dose-dependency of the obtained experimental data.

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. Additional information about mechanisms of alerts is provided in **Section 4.4**.

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. Fitting of model variable to the observed data,
d. Verification of model quality,
e. Depending on the results found in step *d* model building could continue with step *a*, *b* or *f*,
f. 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.