(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: *In vivo* Micronucleus formation v.16.16
Platform version: OASIS TIMES 2.33.1
Name: *In vivo* Micronucleus formation
Author: LMC, University "Prof. As. Zlatarov", Bourgas, Bulgaria
Date: 31 March 2024
E-mail: omekenya@btu.bg
www: http://www.oasis-lmc.org/

Section 1. QSAR identifier

1.1. QSAR identifier (title)

In vivo Micronucleus formation v.16.16

1.2. Other related models

In vivo Liver Clastogenicity

1.3. Software coding of the model

Model version: *In vivo* Micronucleus formation v.16.16
Platform version: OASIS TIMES 2.33.1
Name: *In vivo* Micronucleus formation
Developer: LMC, University "Prof. As. Zlatarov", Bourgas, Bulgaria
Coding language: Delphi 10.2

Section 2. General information

2.0. Abstract

In vivo micronucleus formation model detects chemicals capable to induce chromosomal breakage (clastogenicity) and chromosome lagging due to dysfunction of mitotic apparatus (aneugenicity) in bone marrow or peripheral blood of rats and mice.

If availability of parent chemicals or their metabolites in the target tissue is not rate limiting, then no differences would be expected between the *in vitro* and *in vivo* results, i.e., the toxicodynamic model for *in vitro* should also be valid *in vivo*. Aiming to predict the *in vivo* micronucleus formation in bone marrow, an *in vitro* reactivity component based on interaction of chemicals with DNA and/or proteins has been applied with *in vivo* metabolic

simulator. The result from model prediction indicated that the *in vitro* reactivity overestimates *in vivo* bone marrow genotoxicity. Analysis of the "false positives" model predictions shows that almost all of them are *in vitro* mutagenic in Ames or chromosomal aberrations tests. This analysis indicated that *in vitro* mutagenicity does not always provides prediction of *in vivo* effects. This observation might be explained by neglecting *in vivo* bioavailability in *in vitro* tests. *In vivo*, enzymes are aggregated in multienzyme complexes and the cells could be protected from reactive metabolites via shuttling intermediates between consecutive enzymes. Thus, the product of one enzymatic reaction may become a substrate of the subsequent enzymatic reaction. In this so called channeling effect, some *in vitro* positive metabolites could be "trapped" and thus unavailable to react with the bone marrow. Ultimately, some *in vitro* positive chemicals were predicted to be *in vivo* negative in the bone marrow.

2.1. Date of QMRF

31 March 2024

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)

21 November 2014; 12 June 2015; 11 may 2016; 13 July 2016; 01 September 2016; 30 May 2017; 17 July 2018; 20 August 2019; 8 December, 2021; 2 March 2023; 31 March 2024.

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 2.0** Abstract; **Sections 2.1** Date of QMRF; **Sections 2.3** Date of QMRF update(s); **Sections 2.6** Date of model development and/or publications; **Sections 2.8.** Availability of information about the model; **Sections 3.3** Comment on endpoint; **Section 3.6** Experimental protocol; **Section 3.7.** Endpoint data quality and variability; **Section 4.2.** Explicit algorithm; **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 6.3** data for each descriptor variable for the training set; **Section 6.4** Data for the dependant variable for the training set; **Section 6.5** Other information about the training set

2.5. Model developer(s) and contact details

Name: O. Mekenyan, P. Petkov, S. Kotov, S. Stoeva, V. Kamenska, S. Dimitrov, M. Honma, M. Hayashi, R. Benigni, M. Donner, G. Patlewicz
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; p_petkov@btu.bg; skotov@btu.bg; sstoeva@btu.bg;

2.6. Date of model development and/or publication

Date of the model development: 2014

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

O. Mekenyan, P. Petkov, S. Kotov, S. Stoeva, V. Kamenska, S. Dimitrov, M. Honma, M. Hayashi, R. Benigni, M. Donner, G. Patlewicz, Investigating the Relationship between *in vitro*−*in vivo* Genotoxicity: Derivation of Mechanistic QSAR Models for *in vivo* liver genotoxicity and *in vivo* bone marrow micronucleus formation which encompass metabolism, Chemical Research in Toxicology, Vol. 25, (2012), pp. 277- 296.

2.8. Availability of information about the model

In vivo MNT model 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: omekenya@btu.bg.

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/micronucleus-formation-in-vivo.aspx

2.9. Availability of another QMRF for exactly the same model

Section 3. Defining the endpoint – OECD Principle 1

3.1. Species

Mammalian erythrocytes and/or peripheral blood

3.2. Endpoint

Chromosome aberration

3.3. Comment on endpoint

The mammalian *in vivo* micronucleus test is used for the detection of damage induced by the test chemical to the chromosomes or the mitotic apparatus of erythroblasts. The test

evaluates micronucleus formation in erythrocytes sampled either in the bone marrow or peripheral blood cells of animals, usually rodents.

The purpose of the micronucleus test is to identify substances that cause cytogenetic damage which results in the formation of micronuclei containing either lagging chromosome fragments or whole chromosomes.

3.4. Endpoint units

Qualitative – positive/ negative

3.5. Dependent variable

Observed MNT

3.6. Experimental protocol

Standard micronuclei tests performed in mammalian erythrocytes and/or peripheral blood.

Test No. 474: Mammalian Erythrocyte Micronucleus Test

http://www.oecd-ilibrary.org/environment/test-no-474-mammalian-erythrocyte-micronucleus-test_9789264224292-en

3.7. Endpoint data quality and variability

Quality: High

Data compilation: Structurally diverse chemicals collected from scientific papers; ECVAM DB.

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

SAR

4.2. Explicit algorithm

Prediction of MNT formation

Alerting group approach with a pattern recognition type of models have been used in order to delineate reactivity of chemicals toward DNA and/or proteins within a given interaction mechanism.

The explicit generation of metabolites allowed the reactivity model to be applied not only to parent chemicals but also on their stable metabolites.

This model allows application of detoxification pathways.

Detoxification is applied to justify *in vivo* bone marrow negative result of chemical which are *in vitro* positive. It is assumed that such chemicals are involved in enzyme channeling effect.

4.3. Descriptors in the model

Descriptors in the model are structural boundaries associated with alerting groups related to interactions with macromolecules (DNA and/ or proteins). Alerts in the TIMES MNT model constitute expertly-derived sets of structural fragments incorporating knowledge for the interactions of chemicals (parents and metabolites) with DNA/proteins. 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/proteins. 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

The main characteristics of each alerting group are:

- 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.

Table 1. Main characteristics of the alerts in the TIMES MNT model.

No.	Alert name	Correct	Incorrect	Performance	p-value
1	Haloalkane Derivatives Containing Chain Heteroatom	15	0	0.941 (0.829 ÷ 1.000)	7.7E-6
2	Nitrogen Mustards	12	0	0.929 (0.794 ÷ 1.000)	0.0001
3	C-Nitroso Compounds	11	0	0.923 (0.779 ÷ 1.000)	0.0002
4	C-Nitroso compounds protein binding	11	0	0.923 (0.779 ÷ 1.000)	0.0002
5	Nitrogen and Sulfur Mustards	8	0	0.900 (0.717 ÷ 1.000)	0.0018
6	Sulfonates and Sulfates DNA binding	8	0	0.900 (0.717 ÷ 1.000)	0.0018
7	Carbamates	10	0	0.917 (0.762 ÷ 1.000)	0.0004
8	Sufonates and Sulfates DNA binding	8	0	0.900 (0.717 ÷ 1.000)	0.0018
9	N-Hydroxylamines	16	1	0.895 (0.761 ÷ 1.000)	3.4E-5
10	Hydrazine Derivatives	15	1	0.889 (0.748 ÷ 0.997)	0.0001
11	alpha-Activated Haloalkanes	14	1	0.882 (0.734 ÷ 0.996)	0.0001
12	Alkylated nitrosoureas and nitrosoguanidines	6	0	0.875 (0.652 ÷ 1.000)	0.0084
13	Arenediazonium and Diazonium Salts	6	0	0.875 (0.652 ÷ 1.000)	0.0084
14	Epoxides, Aziridines, Thiiranes and Oxetanes	27	3	0.875 (0.762 ÷ 0.974)	5,8E-7
15	p-Aminobiphenyl Analogs	6	0	0.875 (0.652 ÷ 1.000)	0.0084
16	Sulfonates and Sulfates protein binding	6	0	0.875 (0.652 ÷ 1.000)	0.0084
17	Alkylphosphates, Alkylthiophosphates and Alkylphosphonates	6	0	0.875 (0.652 ÷ 1,000)	0.0084
18	(Thio)Phosphates	6	0	0.875 (0.652 ÷ 1.000)	0.0084
19	Haloalkane Derivatives with Labile Halogen	13	1	0.875 (0.719 ÷ 0.996)	0.0003
20	Conjugated Nitroalkenes and Five-Membered Nitro- and Amino Heterocycles	6	0	0.875 (0.652 ÷ 1.000)	0.0084
21	Heterocyclic nitro compounds	5	0	0.857 (0.607 ÷ 1.000)	0.019
22	Hydroxamic acid	5	0	0.857 (0.607 ÷ 1.000)	0.019
23	Epoxides, Aziridines and Thiiranes	28	4	0.853 (0.735 ÷ 0.959)	1,3E-6

24	N-Nitroso Compounds	10	1	0.846 (0.659 ÷ 0.994)	0.0024
25	Single-ring Substituted Primary Aromatic Amines	20	3	0.840 (0.698 ÷ 0.965)	0.0001
26	Acyclic Triazenes	4	0	0.833 (0.549 ÷ 1.000)	0.041
27	Heterocyclic Aromatic Amines	4	0	0.833 (0.549 ÷ 1.000)	0.041
28	Heterocyclic N-Hydroxylamines	4	0	0.833 (0.549 ÷ 1.000)	0.041
29	Polycyclic Aromatic Hydrocarbon, Naphthaleneimide and Carbazole Derivatives	4	0	0.833 (0.549 ÷ 1.000)	0.041
30	Pyrrolizidine derivatives	4	0	0.833 (0.549 ÷ 1.000)	0.041
31	Benzodiazepine derivatives	3	0	0.800 (0.473 ÷ 1.000)	0.091
32	Phenyl-azo-naphthole derivatives	3	0	0.800 (0.473 ÷ 1.000)	0.091
33	Quinoxaline-Type 1,4-Dioxides	3	0	0.800 (0.473 ÷ 1.000)	0.091
34	Vinylcarbamate epoxides	3	0	0.800 (0.473 ÷ 1.000)	0.091
35	Polarized Haloalkene Derivatives	3	0	0.800 (0.473÷ 1.000)	0.091
36	Quinoline Derivatives	3	0	0.800 (0.473÷ 1.000)	0.091
37	Hydroxylated phenols	11	2	0.800 (0.605 ÷ 0.971)	0.0044
38	Substituted Anilines	13	3	0.778 (0.590 ÷ 0.947)	0.0038
39	Aminoacridine DNA Intercalators	5	1	0.750 (0.473 ÷ 0.987)	0.069
40	Benzoquinolines and Acridines derivatives	5	1	0.750 (0.473 ÷ 0.987)	0.069
41	Amphetamine derivatives	2	0	0.750 (0.368 ÷ 1.000)	0.20
42	Azoxyalkanes	2	0	0.750 (0.368 ÷ 1.000)	0.20
43	Hydrazine derivatives specific	2	0	0.750 (0.368 ÷ 1.000)	0.20
44	Pyrimidines and Purines	2	0	0.750 (0.368 ÷ 1.000)	0.202
45	Haloalcohols	5	1	0.750 (0.473 ÷ 0.987)	0.069
46	Vicinal Dihaloalkanes	2	0	0.750 (0.368 ÷ 1.000)	0.202
47	Halogenated Vicinal Hydrocarbons	2	0	0.750 (0.368 ÷ 1.000)	0.202

48	Quinoneimines protein binding	16	5	0.739 (0.562 ÷ 0.904)	0.0042
49	Benzidine-based azo dyes	4	1	0.714 (0.409 ÷ 0.982)	0.131
50	Quinoid compounds	15	6	0.696 (0.511 ÷ 0.845)	0.014
51	Fused-Ring Primary Aromatic Amines	8	3	0.692 (0.452 ÷ 0.917)	0.062
52	alpha,beta-Unsaturated Carbonyls and Related Compounds	8	3	0.692 (0.452 ÷ 0.917)	0.062
53	Diazenes	3	1	0.667 (0.330 ÷ 0.974)	0.24
54	DNA Intercalators with Carboxamide and Aminoalkylamine Side Chain	3	1	0.667 (0.330 ÷ 0.974)	0.24
55	Alkylnitrites	1	0	0.667 (0.224 ÷ 1.000)	0.448
56	Alpha-Haloethers	1	0	0.667 (0.224 ÷ 1.000)	0.448
57	Atrazine derivatives	1	0	0.667 (0.224 ÷ 1.000)	0.448
58	Benzanthrone Derivatives	1	0	0.667 (0.224 ÷ 1.000)	0.448
59	Bleomycin and Structurally Related Chemicals	1	0	0.667 (0.224 ÷ 1.000)	0.448
60	Coumarins and Thiocoumarins	1	0	0.667 (0.224 ÷ 1.000)	0.448
61	Dialkyl Alkylphosphonates	1	0	0.667 (0.224 ÷ 1.000)	0.448
62	Diazoalkanes	1	0	0.667 (0.224 ÷ 1.000)	0.448
63	Ethenyl Pyridines	1	0	0.667 (0.224 ÷ 1.000)	0.448
64	N-Haloacylamides	1	0	0.667 (0.224 ÷ 1.000)	0.448
65	Organic Azides	1	0	0.667 (0.224 ÷ 1.000)	0.448
66	Polyhaloethane derivatives	1	0	0.667 (0.224 ÷ 1.000)	0.448
67	Primary thioamides	1	0	0.667 (0.224 ÷ 1.000)	0.448
68	Quinolone Derivatives	1	0	0.667 (0.224 ÷ 1.000)	0.448
69	Short-Chain Alkyltin and Alkylgermanium Halides	1	0	0.667 (0.224 ÷ 1.000)	0.448
70	Sulfapyridine Azo derivatives and Sulfapyridines	1	0	0.667 (0.224 ÷ 1.000)	0.448
71	Sultones	1	0	0.667 (0.224 ÷ 1.000)	0.448

72	Sultones protein binding	1	0	0.667 (0.224 ÷ 1.000)	0.448
73	Thiols	1	0	0.667 (0.224 ÷ 1.000)	0.448
74	Four-and Five-Membered Lactones	1	0	0.667 (0.224 ÷ 1.000)	0.448
75	Quinones and Trihydroxybenzenes	5	2	0.667 (0.379 ÷ 0.935)	0.153
76	Carboxylic Acid Amides	1	0	0.667 (0.224 ÷ 1.000)	0.448
77	Haloalkene Derivatives with Electron-Withdrawing Groups	1	0	0.667 (0.224 ÷ 1.000)	0.448
78	N-Nitrosamines	3	1	0.667 (0.330 ÷ 0.947)	0.240
79	Specific Imine and Thione Derivatives	10	5	0.647 (0.426 ÷ 0.858)	0.078
80	Quinoneimine, Thionine and Phenoxazinium Derivatives	9	5	0.625 (0.395 ÷ 0.846)	0.120
81	Dicarbonyl Compounds	4	2	0.625 (0.315 ÷ 0.919)	0.254
82	N-Substituted Aromatic Amines	4	2	0.625 (0.315 ÷ 0.919)	0.254
83	Acridone, Thioxanthone, Xanthone, Phenazine and Other Fused-Ring Heterocyclic DNA Intercalators	2	1	0.600 (0.228 ÷ 0.956)	0.423
84	N,N-Dialkyldithiocarbamate Derivatives and Azaarene Dithiocarbamates	2	1	0.600 (0.228 ÷ 0.956)	0.423
85	Fused-Ring Nitroaromatics	2	1	0.600 (0.228 ÷ 0.956)	0.423
86	Alpha,Beta-Unsaturated Aldehydes	3	2	0.571 (0.239 ÷ 0.895)	0.404
87	Substituted Phenols	4	3	0.556 (0.254 ÷ 0.851)	0.389
88	(Thio)Acyl and (thio)carbamoyl halides, cyanides, azides, etc.	1	1	0.500 (0.094 ÷ 0.906)	0.695
89	N-Nitrosoamine Derivatives	1	1	0.500 (0.094 ÷ 0.906)	0.695
90	alpha,beta-Unsaturated Carboxylic Acids and Esters	2	2	0.500 (0.147 ÷ 0.853)	0.606
91	Monohaloalkanes	0	1	0.333 (0.000 ÷ 0.776)	1.000
92	Nitroazoarenes and p-Monosubstituted Azobenzene Derivatives	0	1	0.333 (0.000 ÷ 0.776)	1.000
93	Isothiocyanates	0	1	0.333 (0.000 ÷ 0.776)	1.00
94	N-methylol derivatives	0	1	0.333 (0.000 ÷ 0.776)	1.00

95	Nitroalkanes	0	1	0.333 (0.000 ÷ 0.776)	1.00
96	Propargyl Derivatives	0	1	0.333 (0.000 ÷ 0.776)	1.00
97	Propyne Derivatives	0	1	0.333 (0.000 ÷ 0.776)	1.00
98	Geminal Polyhaloalkane Derivatives	3	8	0.308 (0.083 ÷ 0.548)	0.931
99	Nitroaniline Derivatives	0	1	0.333 (0.000 ÷ 0.776)	1.00
100	Pyrazolone and Pyrazolidine Derivatives	0	2	0.250 (0.000 ÷ 0.632)	1.00
101	Arenesulphonamides	0	0	N/A	N/A
102	Polynitroarenes	0	0	N/A	N/A
103	Nitrophenols, Nitrophenyl Ethers and Nitrobenzoic Acids	0	0	N/A	N/A
104	1,4-Diazabutadiene Derivatives	0	0	N/A	N/A
105	4,4’-Bipyridinium Salts and N-Oxides	0	0	N/A	N/A
106	Acyl Halides	0	0	N/A	N/A
107	alpha,omega-Dihaloalkanes	0	0		
108	Alkyl Xanthate Esters	0	0	N/A	N/A
109	Alpha-Beta Conjugated Alkene Derivatives with Geminal Electron-Withdrawing Groups	0	0	N/A	N/A
110	Amidoxime Esters and Amidoximes	0	0	N/A	N/A
111	Aminophenoxazinone derivative	0	0	N/A	N/A
112	Anthrones	0	0	N/A	N/A
113	Antibiotic Aminoglycoside Derivatives	0	0	N/A	N/A
114	Aromatic ester hydroxylamine	0	0	N/A	N/A
115	Azoalkanes with Activating EWG	0	0	N/A	N/A
116	Benzofuranyl carbamate derivatives	0	0	N/A	N/A
117	Arenecarbonyl Compounds	0	0	N/A	N/A
118	Arenecarboxylic Acid Esters	0	0	N/A	N/A
119	Flavonoids	0	0	N/A	N/A
120	Gallic Acid Esters	0	0	N/A	N/A
121	Halofuranones	0	0	N/A	N/A
122	Nitroarenes with Other Active Groups	0	0	N/A	N/A
123	alpha-Activated benzyls	0	0	N/A	N/A
124	Amino Anthraquinones	0	0	N/A	N/A

125	Benzoyl cyclohexanedione derivatives	0	0	N/A	N/A
126	Bipyridilium Herbicides	0	0	N/A	N/A
127	Carboxylic acid Anhydrides	0	0	N/A	N/A
128	Chlorinated Diphenylmethane and Benzophenone Derivatives	0	0	N/A	N/A
129	Conjugated Benzoylethylene Derivatives with EWG	0	0	N/A	N/A
130	Cyanohydrins	0	0	N/A	N/A
131	Dichlorophosphine and Dichlorophosphonium Derivatives	0	0	N/A	N/A
132	Dithianes	0	0	N/A	N/A
133	Fluoro bis-benzothiazole derivative	0	0	N/A	N/A
134	Formaldehyde Releasers	0	0	N/A	N/A
135	Fused-Ring Conjugated Lactones	0	0	N/A	N/A
136	Haloalkene Cysteine S-Conjugates	0	0	N/A	N/A
137	Haloazaarene and Fused-Ring Haloquinoline Derivatives	0	0	N/A	N/A
138	Halogenated Oxetanes and Haloepoxides	0	0	N/A	N/A
139	Haloisothiazolinones	0	0	N/A	N/A
140	Heterocyclic Nitroso compounds	0	0	N/A	N/A
141	Heterocyclic urea derevatives	0	0	N/A	N/A
142	Hexahydrotriazine Derivatives	0	0	N/A	N/A
143	Hydroxybenzophenone Derivatives	0	0	N/A	N/A
144	Hypoxanthine Derivatives	0	0	N/A	N/A
145	Isocyanates and Diisocyanates	0	0	N/A	N/A
146	N-Acetoxyamines	0	0	N/A	N/A
147	N-Acyloxy(Alkoxy) Arenamides	0	0	N/A	N/A
148	N-Alkyl-N-nitrosocarbamates	0	0	N/A	N/A
149	N-Alkylindolinium and N-Alkylbenzothiazolium Salts	0	0	N/A	N/A
150	N-Aryl-N-Acetoxy(Benzoyloxy) Acetamides				
151	Nitrobiphenyls and Bridged Nitrobiphenyls	0	0	N/A	N/A
152	Specific 5-Substituted Uracil Derivatives	0	0	N/A	N/A
153	N-Hydroxyethyl Lactams	0	0	N/A	N/A
154	N-Oxicarbonyl amides, N-Acyloxy-N-alkoxyamides	0	0	N/A	N/A
155	N-Trihalomethyl Imides (Theoretical)	0	0	N/A	N/A
156	Non-Aromatic Hydroxylamine Derivatives	0	0	N/A	N/A

157	Non-Cyclic Alkyl Phosphoramides and Thionophosphoramides	0	0	N/A	N/A
158	Organic Diselenides and Ditellurides	0	0	N/A	N/A
159	Organic Peroxy Compounds	0	0	N/A	N/A
160	p-Substituted Mononitrobenzenes	0	0	N/A	N/A
161	PAH Benzylic Alcohol Esters	0	0	N/A	N/A
162	Perfluorinated Hypofluorites	0	0	N/A	N/A
163	Peroxyacyl Nitrates	0	0	N/A	N/A
164	Polyethylene Polyamines	0	0	N/A	N/A
165	Quinone Methides	0	0	N/A	N/A
166	S-Activated Cysteine Derivatives	0	0	N/A	N/A
167	Specific Acetate Esters	0	0	N/A	N/A
168	Sterically Hindered Piperidine Derivatives	0	0	N/A	N/A
169	Substituted Benzoindoline and Indole Derivatives	0	0	N/A	N/A
171	Substituted Chlorophenylalkylurea Derivatives	0	0	N/A	N/A
170	Substituted Nitropyridines, Aminopyridines and N-Oxides	0	0	N/A	N/A
172	Sulfonyl Halides	0	0	N/A	N/A
173	Tertiary Heterocyclic Amines	0	0	N/A	N/A
174	Thiadiazole-dioxide derivatives	0	0	N/A	N/A
175	Thiazolidinediones	0	0	N/A	N/A
176	Tri-Methylindole derivatives	0	0	N/A	N/A
177	Triarylimidazole and Structurally Related DNA Intercalators	0	0	N/A	N/A
178	Triazinone derivative	0	0	N/A	N/A
179	Trifluorometyl benzamide derivative	0	0	N/A	N/A
180	Trifluoromethyl pyridinone derivatives	0	0	N/A	N/A

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 macromolecules 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, *In vivo* Micronucleus formation v.16.16

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

In order to belong to the model domain a target structure must meet the requirements of all layers of the domain.

- General properties requirements:

Property	Domain	Target chemical
log K_{OW}	[-9.89; 14.76]	2.175
MW, Da	[43.022; 1416.92]	182.127

* K_{OW} is calculated by EPI Suite

- Structural domain extracted from 493 training chemicals contains:
  - 1808 correct fragments,
  - 187 fuzzy fragments (treated as correct fragments),
  - 257 incorrect fragments.

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

6.1. Availability of the training set

The training set consisting of 493 chemicals and is embedded in the software implementation of the model.

6.2. Available information for the training set

Chemical names, CAS numbers, SMILES, references are available.

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

Data for the dependent variable of the training set are embedded in the software implementation of the model.

The training set consists of:

- 221 chemicals with positive in vivo micronucleus data
- 271 chemicals with negative in vivo micronucleus data

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 0.448 (0.405 ÷ 0.492) [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 TG474.

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) = 87%
- Specificity = (predicted negative/observed negative) = 85%
- Concordance = (correct predicted positive and negative chemicals in respect to all training set chemicals) = 86%

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

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

7.1. Availability of the external validation set

Not available

7.2. Available information for the external validation set

Not available

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

Not available

7.6. Experimental design of test set

Not available

7.7. Predictivity – Statistics obtained by external validation

Not available

7.8. Predictivity – Assessment of the external validation set

Not available

7.9. Comment on the external validation of the model

Not available

Section 8. Providing a mechanistic interpretation – OECD Principle 5

8.1. Mechanistic basis of the model

If availability of parent chemicals or their metabolites in the target tissue is not rate limiting, then no differences would be expected between the *in vitro* and *in vivo* results, i.e., the toxicodynamic model for *in vitro* should also be valid *in vivo*. Aiming to predict the *in vivo* micronucleus formation in bone marrow, an *in vitro* reactivity component based on interaction of chemicals with DNA and/or proteins has been applied with *in vivo* metabolic

simulator. The result from model prediction indicated that the *in vitro* reactivity overestimates *in vivo* bone marrow genotoxicity. Analysis of the "false positives" model predictions shows that almost all of them are *in vitro* mutagenic in Ames or chromosomal aberrations tests. This analysis indicated that *in vitro* mutagenicity does not always provides prediction of *in vivo* effects. This observation might be explained by neglecting *in vivo* bioavailability in *in vitro* tests. *In vivo*, enzymes are aggregated in multienzyme complexes and the cells could be protected from reactive metabolites via shuttling intermediates between consecutive enzymes. Thus, the product of one enzymatic reaction may become a substrate of the subsequent enzymatic reaction. In this so called channelling effect, some *in vitro* positive metabolites could be "trapped" and thus unavailable to react with the bone marrow. *In vitro* positive chemicals which are not involved in "trapping" detoxification pathways are considered capable of causing *in vivo* micronucleus effect in the bone marrow. Ultimately, *in vitro* negative chemicals are also expected to be *in vivo* negative in the bone marrow, since bio activation in remote tissues is considered rather as exception than a rule.

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

Not available

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.