(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* Liver TGR mutagenicity v.11.11
Platform version: OASIS TIMES 2.33.1
Name: *In vivo* Liver TGR mutagenicity
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 Liver TGR mutagenicity v.11.11

1.2. Other related models

In vivo Comet Genotoxicity v.15.15

1.3.Software coding of the model

Model version: *In vivo* Liver TGR mutagenicity v.11.11
Platform version: OASIS TIMES 2.33.1
Name: *In vivo* Liver TGR mutagenicity
Developer: LMC, University "Prof. As. Zlatarov", Bourgas, Bulgaria
Coding language: Delphi 10.2

Section 2. General information

2.0. Abstract

Transgenic Rodent mutation assay (TGR) is practical and widely available in vivo tests for gene mutations. The TGR assay provides quick and statistically reliable data for mutations in the DNA from any tissue.

The model consists of two components – a reactivity and metabolism component.
The reactivity component is based on in vitro structural alerts for DNA binding and active structural alerts accounting for in vivo generated reactive oxygen species. The metabolism component is accounting for metabolic activation and detoxification of

chemical [1]. The metabolism simulator was developed comprising a set of structurally generalized molecular transformations. A database of 701 in vivo metabolic pathways of chemicals was compiled and formed the training set used to derive the rat in vivo metabolic simulator.

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 macromolecules in liver. In vitro positive chemicals which are not involved in “trapping” detoxification pathways are considered capable of causing DNA damage and hence in vivo liver mutagenic effect. In vitro negative chemicals are also expected to be in vivo negative in liver.

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; 15 June 2015; 11 May 2016; 13 July 2016; 02 September, 2016; 30 May 2017; 18 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 2.7** Reference(s) to the main scientific and/or software package; **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 6.1** Availability of the training set; **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: O. Mekenyan, P. Petkov, S. Kotov, S. Dimitrov, M. Honma
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;

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, 2012, 25, 277- 296.

2.8. Availability of information about the model

In vivo TGR 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/in-vivo-liver-genotoxicity-(1).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

Rodents (mainly rats)

3.2.Endpoint

Gene mutation

3.3. Comment on endpoint

Transgenic Rodent Mutation assay (TGR) is an *in vivo* test for gene mutations of regulatory concern. The TGR is a simple and statistically reliable test providing mutagenicity data from any tissue.

3.4. Endpoint units

Qualitative – positive/ negative

3.5. Dependent variable

Observed Liver TGR

3.6. Experimental protocol

Transgenic Rodent Somatic and Germ Cell Gene Mutation Assay.
http://www.oecd-ilibrary.org/environment/test-no-488-transgenic-rodent-somatic-and-germ-cell-gene-mutation-assays_9789264122819-en

3.7. Endpoint data quality and variability

Quality: High

Data compilation: Structurally diverse chemicals collected from NIHS Japan; 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 liver TGR mutagenicity

Alerting group approach has been used to investigate reactivity of chemicals toward DNA in liver of mammals.

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

This model allows application of so called “trapping” detoxification pathways. This type of detoxification is applicable for chemicals which are *in vitro* positive but *in vivo* liver

(TGR) negative. It is assumed that such chemicals are involved in enzyme channeling effect.

4.3. Descriptors in the model

Descriptors in the model are structural alerts related to interactions with DNA. Alerts in the TIMES *in vivo* TGR 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 in vivo 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 selection

Table 1 summarizes the main characteristics of each DNA alert in TIMES *in vivo* TGR:

- 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 DNA alerts in the TIMES *in vivo* TGR.

No.	Alert name	Correct	Incorrect	Performance	p-value
1	N-Hydroxylamines	10	0	0.917 (0.762 ÷ 1.000)	0.0047
2	N-Nitrosamines	7	0	0.889	0.022

				(0.688 ÷ 1.000)	
3	N-Nitroso Compounds	7	0	0.889 (0.688 ÷ 1.000)	0.022
4	Polycyclic Aromatic Hydrocarbon, Naphthaleneimide and Carbazole Derivatives	6	0	0.875 (0.652 ÷ 1.000)	0.038
5	Diazenes	5	0	0.857 (0.607 ÷ 1.000)	0.064
6	Fused-Ring Primary Aromatic Amines	5	0	0.857 (0.607 ÷ 1.000)	0.064
7	Fused-Ring Nitroaromatics	4	0	0.833 (0.549 ÷ 1.000)	0.110
8	Single-ring Substituted Primary Aromatic Amines	7	1	0.800 (0.567 ÷ 0.991)	0.085
9	Quinoline Derivatives	3	0	0.800 (0.473 ÷ 1.000)	0.19
10	Epoxides, Aziridines, Thiiranes and Oxetanes	17	4	0.783 (0.617 ÷ 0.935)	0.028
11	Hydrazine Derivatives	6	1	0.778 (0.524 ÷ 0.989)	0130
12	Cyclopropylpyrroloindole derivatives	2	0	0.750 (0.368 ÷ 1.000)	0.33
13	Pyrrolizidine derivatives	2	0	0.750 (0.368 ÷ 1.000)	0.33
14	Conjugated Nitroalkenes and Five-Membered Nitro- and Amino Heterocycles	2	0	0.750 (0.368 ÷ 1.000)	0.328
15	C-Nitroso Compounds	2	0	0.750 (0.368 ÷ 1.000)	0.328
16	Specific Imine and Thione Derivatives	2	0	0.750 (0.368 ÷ 1.000)	0.328
17	Arenediazonium and Diazonium Salts	4	1	0.714 (0.409 ÷ 0.982)	0.293
18	Substituted Allyl Alcohols	4	1	0.714 (0.409 ÷ 0.982)	0.293
19	Alkylphosphates, Alkylthiophosphates and Alkylphosphonates	1	0	0.667 (0.224 ÷ 1.000)	0.57
20	Benzanthrone Derivatives	1	0	0.667 (0.224 ÷ 1.000)	0.57
21	Benzodiazepine derivatives	1	0	0.667 (0.224 ÷ 1.000)	0.57
22	Coumarins and Thiocoumarins	1	0	0.667 (0.224 ÷ 1.000)	0.57
23	Four-and Five-Membered Lactones	1	0	0.667 (0.224 ÷ 1.000)	0.57
24	Haloalkane Derivatives Containing Chain	1	0	0.667 (0.224 ÷ 1.000)	0.57

	Heteroatom				
25	Heterocyclic N-Hydroxylamines	1	0	0.667 (0.224 ÷ 1.000)	0.57
26	Heterocyclic nitro compounds	1	0	0.667 (0.224 ÷ 1.000)	0.57
27	Hydroxamic acid	1	0	0.667 (0.224 ÷ 1.000)	0.57
28	Nitroaniline Derivatives	1	0	0.667 (0.224 ÷ 1.000)	0.57
29	Quinones and Trihydroxybenzenes	1	0	0.667 (0.224 ÷ 1.000)	0.57
30	Haloalkane Derivatives with Labile Halogen	1	0	0.667 (0.224 ÷ 1.000)	0.57
31	Dicarbonyl Compounds	1	0	0.667 (0.224 ÷ 1.000)	0.57
32	Haloalkene Derivatives with Electron-Withdrawing Groups	1	1	0.500 (0.094 ÷ 0.906)	0.81
33	Geminal Polyhaloalkane Derivatives	1	1	0.500 (0.094 ÷ 0.906)	0.81
34	Azoxyalkanes	0	1	0.333 (0.000 ÷ 0.776)	1.00
35	Thiols	0	1	0.333 (0.000 ÷ 0.776)	1.00
36	Quinone Methides	0	1	0.333 (0.000 ÷ 0.776)	1.00
37	Halofuranones	0	0	N/A	N/A
38	Polarized Haloalkene Derivatives	0	0	N/A	N/A
39	Quinolone Derivatives	0	0	N/A	N/A
40	Sulfonates and Sulfates	0	0	N/A	N/A
41	Vicinal Dihaloalkanes	0	0	N/A	N/A
42	Haloalcohols	0	0	N/A	N/A
43	1,4-Diazabutadiene Derivatives	0	0	N/A	N/A
44	4,4’-Bipyridinium Salts and N-Oxides	0	0	N/A	N/A
45	Acridone, Thioxanthone, Xanthone, Phenazine and Other Fused-Ring Heterocyclic DNA Intercalators	0	0	N/A	N/A
46	Acyclic Triazenes	0	0	N/A	N/A
47	Acyl Halides	0	0	N/A	N/A
48	Alkyl Xanthate Esters	0	0	N/A	N/A
49	Alkylnitrites	0	0	N/A	N/A

50	Alpha,Beta-Unsaturated Aldehydes	0	0	N/A	N/A
51	Alpha-Beta Conjugated Alkene Derivatives with Geminal Electron-Withdrawing Groups	0	0	N/A	N/A
52	Alpha-Haloethers	0	0	N/A	N/A
53	Amidoxime Esters and Amidoximes	0	0	N/A	N/A
54	Amino Anthraquinones	0	0	N/A	N/A
55	Aminoacridine DNA Intercalators	0	0	N/A	N/A
56	Aminophenoxazinone derivative	0	0	N/A	N/A
57	Anthrones	0	0	N/A	N/A
58	Antibiotic Aminoglycoside Derivatives	0	0	N/A	N/A
59	Aromatic ester hydroxylamine	0	0	N/A	N/A
60	Azoalkanes with Activating EWG	0	0	N/A	N/A
61	Benzofuranyl carbamate derivatives	0	0	N/A	N/A
62	Benzoyl cyclohexanedione derivatives	0	0	N/A	N/A
63	Bleomycin and Structurally Related Chemicals	0	0	N/A	N/A
64	Chlorinated Diphenylmethane and Benzophenone Derivatives	0	0	N/A	N/A
65	Conjugated Benzoylethylene Derivatives with EWG	0	0	N/A	N/A
66	Diazoalkanes	0	0	N/A	N/A
67	Dichlorophosphine and Dichlorophosphonium Derivatives	0	0	N/A	N/A
68	Dithianes	0	0	N/A	N/A
69	DNA Intercalators with Carboxamide and Aminoalkylamine Side Chain	0	0	N/A	N/A
70	Flavonoids	0	0	N/A	N/A
71	Fluoro bis-benzothiazole derivative	0	0	N/A	N/A
72	Formaldehyde Releasers	0	0	N/A	N/A
73	Fused-Ring Conjugated Lactones	0	0	N/A	N/A

74	Haloalkene Cysteine S-Conjugates	0	0	N/A	N/A
75	Haloazaarene and Fused-Ring Haloquinoline Derivatives	0	0	N/A	N/A
76	Halogenated Oxetanes and Haloepoxides	0	0	N/A	N/A
77	Haloisothiazolinones	0	0	N/A	N/A
78	Heterocyclic Nitroso compounds	0	0	N/A	N/A
79	Heterocyclic urea derevatives	0	0	N/A	N/A
80	Hydroxybenzophenone Derivatives	0	0	N/A	N/A
81	Hypoxanthine Derivatives	0	0	N/A	N/A
82	Monohaloalkanes	0	0	N/A	N/A
83	N,N-Dialkyldithiocarbamate Derivatives and Azaarene Dithiocarbamates	0	0	N/A	N/A
84	N-Acetoxyamines	0	0	N/A	N/A
85	N-Acyloxy(Alkoxy) Arenamides	0	0	N/A	N/A
86	N-Alkylindolinium and N-Alkylbenzothiazolium Salts	0	0	N/A	N/A
87	N-Aryl-N-Acetoxy(Benzoyloxy) Acetamides	0	0	N/A	N/A
88	N-Hydroxyethyl Lactams	0	0	N/A	N/A
89	N-methylol derivatives	0	0	N/A	N/A
90	N-Trihalomethyl Imides	0	0	N/A	N/A
91	Nitroalkanes	0	0	N/A	N/A
92	Nitroarenes with Other Active Groups	0	0	N/A	N/A
93	Nitroazoarenes and p-Monosubstituted Azobenzene Derivatives	0	0	N/A	N/A
94	Nitrobiphenyls and Bridged Nitrobiphenyls	0	0	N/A	N/A
95	Nitrogen and Sulfur Mustards	0	0	N/A	N/A
96	Nitrophenols, Nitrophenyl Ethers and Nitrobenzoic Acids	0	0	N/A	N/A
97	Non-Aromatic Hydroxylamine Derivatives	0	0	N/A	N/A
98	Non-Cyclic Alkyl	0	0	N/A	N/A

	Phosphoramides and Thionophosphoramides				
99	Organic Azides	0	0	N/A	N/A
100	Organic Diselenides and Ditellurides	0	0	N/A	N/A
101	Organic Peroxy Compounds	0	0	N/A	N/A
102	p-Aminobiphenyl Analogs	0	0	N/A	N/A
103	p-Substituted Mononitrobenzenes	0	0	N/A	N/A
104	PAH Benzylic Alcohol Esters	0	0	N/A	N/A
105	Perfluorinated Hypofluorites	0	0	N/A	N/A
106	Peroxyacyl Nitrates	0	0	N/A	N/A
107	Polyethylene Polyamines	0	0	N/A	N/A
108	Polynitroarenes	0	0	N/A	N/A
109	Propyne Derivatives	0	0	N/A	N/A
110	Quinoneimine, Thionine and Phenoxazinium Derivatives	0	0	N/A	N/A
111	Quinoxaline-Type 1,4-Dioxides	0	0	N/A	N/A
112	S-Activated Cysteine Derivatives	0	0	N/A	N/A
113	Short-Chain Alkyltin and Alkylgermanium Halides	0	0	N/A	N/A
114	Specific 5-Substituted Uracil Derivatives	0	0	N/A	N/A
115	Specific Acetate Esters	0	0	N/A	N/A
116	Substituted Benzoindoline and Indole Derivatives	0	0	N/A	N/A
117	Substituted Chlorophenylalkylurea Derivatives	0	0	N/A	N/A
118	Substituted Nitropyridines, Aminopyridines and N-Oxides	0	0	N/A	N/A
119	Sulfonyl Halides	0	0	N/A	N/A
120	Sultones	0	0	N/A	N/A
121	Tertiary Heterocyclic Amines	0	0	N/A	N/A
122	Thiadiazole-dioxide derivatives	0	0	N/A	N/A
123	Thiazolidinediones	0	0	N/A	N/A
124	Tri-Methylindole derivatives	0	0	N/A	N/A
125	Triarylimidazole and Structurally Related DNA	0	0	N/A	N/A

	Intercalators				
126	Triazinone derivative	0	0	N/A	N/A
127	Triflorometyl benzamide derivative	0	0	N/A	N/A
128	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 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, *In vivo* Liver TGR mutagenicity v.11.11

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

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}	[-8.369; 12.193]	2.175
MW, Da	[54.03; 1018.63]	182.127

* K_{OW} is calculated by EPI Suite

- Structural domain extracted from 137 training chemicals contains:
  - 681 correct fragments,
  - 55 fuzzy fragments (treated as correct fragments),
  - 73 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 137 chemicals is included in the software implementation of the model.

6.2. Available information for the training set

Chemical names, CAS numbers, SMILES, data source 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

The training set of 137 chemicals includes:

- 79 chemicals with positive in vivo TGR data
- 58 chemicals with negative in vivo TGR 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 0.576 (0.494 ÷ 0.657)[1)].

1) Confidence range is calculated at 95% confidence level

6.5. Other information about the training set

Not available

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

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. Comments on the internal validation of the model

Not performed

Section 7. 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*. In this respect, *in vitro* reactivity taking into account interactions of chemicals with DNA should be suitable as part of the *in vivo* model for liver TGR mutagenicity. Only those toxicophores having clear interpretation for the molecular mechanism causing the ultimate effect were included in the model. Some of the alerts are known to interact directly with DNA, whereas others are applied in a combination of specific structural requirements assessing the degree of activation of the alerts from the rest of the molecules. In the *in vivo* liver TGR mutagenicity model *in vitro* reactivity component (from *in vitro* Ames model) is combined with *in vivo* metabolism simulator and *in vivo*-only active structural alerts, accounting for generated *in vivo* reactive oxygen species (ROS). The simulator was developed comprising a set of structurally generalized molecular transformations. A database of 647 *in vivo* metabolic pathways of chemicals was compiled and formed the training set used to derive the rat *in vivo* metabolic simulator. *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 macromolecules in liver. *In vitro* positive chemicals which are not involved in "trapping" detoxification pathways are considered capable of causing DNA damage and hence *in vivo* liver mutagenic effects. *In vitro* negative chemicals are also expected to be *in vivo* negative in liver.

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.