(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 Clastogenicity v.12.12
Platform version: OASIS TIMES 2.33.1
Name: *In vivo* Liver Clastogenicity
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 Clastogenicity

1.2. Other related models

In vivo Micronucleus formation

1.3. Software coding of the model

Model version: *In vivo* Liver Clastogenicity v.12.12
Platform version: OASIS TIMES 2.33.1
Name: *In vivo* Liver Clastogenicity
Developer: LMC, University "Prof. As. Zlatarov", Bourgas, Bulgaria
Coding language: Delphi 10.2

Section 2. General information

2.0. Abstract

In vivo liver clastogenicity model identifies chemicals that cause DNA and/or protein damage in liver of rats or mice, taking into account in vivo detoxification of chemicals. The training set consists of 115 chemicals [1, 2] with integrated liver clastogenicity outcome based on two mutagenicity assays: chromosomal aberration assay (CA) and micronucleus test in liver (MNT). If one of the assays shows positive result, the ultimate liver clastogenicity assignation is considered to be positive.

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; 25 June 2015; 10 May 2016; 13 July 2016; 01 September 2016; 30 May 2017; 16 July 2018; 20 August 2019; 8 December, 2021; 15 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;

2.5.Model developer(s) and contact details

Name: O. Mekenyan, P. Petkov, S. Kotov, S. Stoeva, 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; 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

1. Suzuki, H., Takazawa H., Kobayashi K. et al Mutagenesis, 2009, 24, 9-16

2. Takazawa H., Suzuki H., Ogawa I. et al Mutat. Res., 2010, 698, 30-37

3. 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 Clastogenicity 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)-(1).aspx

2.9. Availability of another QMRF for exactly the same model

Not applicable

Section 3. Defining the endpoint – OECD Principle 1

3.1. Species

Rodents (mainly rats)

3.2. Endpoint

DNA and protein damage

3.3. Comment on endpoint

In vivo liver clastogenicity is assessed by two mutagenicity tests – *in vivo* liver CA and MNT. These tests indicate for ability of environmental and/or synthetic chemicals to elicit DNA and/or protein damages in liver of rodents.

3.4. Endpoint units

Qualitative – positive/ negative

3.5. Dependent variable

Observed Liver Muta

3.6. Experimental protocol

In vivo liver CA and MN tests performed in liver of rodents.

3.7. Endpoint data quality and variability

Quality: High

Data compilation: Structurally diverse chemicals collected from Suzuki et al., Mutat. Res., (2005).

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 clastogenicity

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

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 with are *in vitro* positive but *in vivo* liver (CA, MNT) 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 and proteins. Alerts in the TIMES in vivo Clastogenicity model constitute expertly-derived sets of structural fragments incorporating knowledge for the interactions of chemicals (parents and metabolites) with DNA and 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 and 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 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).

4.4. Descriptor section

The main characteristics of each alert in TIMES in vivo Clastogenicity model 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.

4.5. Algorithm and descriptor generation

Not applicable

The structural boundaries of the alerts are derived from the chemicals included in the local training sets. For derivation of each alert mechanistically justifiable structural fragments for interaction with DNA and proteins 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 Clastogenicity v.12.12

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}	[-3.012; 13.284]	2.175
MW, Da	[58.00; 698.337]	182.127

* K_{OW} is calculated by EPI Suite

- Structural domain extracted from 115 training chemicals contains:
  - 487 correct fragments,
  - 45 fuzzy fragments (treated as correct fragments),
  - 130 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 115 chemicals and is embedded 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

Not applicable

6.4. Data for the dependent variable for the training set

The training set of 115 chemicals includes:

- 78 chemicals with positive in vivo Clastogenicity data

- 37 chemicals with negative in vivo Clastogenicity data

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) = 81%
- Specificity = (predicted negative/observed negative) = 84%
- Concordance = (correct predicted positive and negative chemicals in respect to all training set chemicals) = 82%

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. 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 hence 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 and/or proteins should be suitable as part of the *in vivo* model for liver clastogenicity. Only those toxicophores having clear interpretation for the molecular mechanism causing the ultimate effect were included in the model. Some of the specified alerts interact directly with DNA or nuclear proteins, whereas others are applied in a combination of two-dimensional SAR models assessing the degree of activation of the alerts from the rest of the molecules. In the *in vivo* liver clastogenicity model *in vitro* reactivity component (from the CA model) is combined with *in vivo* metabolism simulator. 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 and/or protein 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.