

	QMRM identifier (ECB Inventory): To be entered by JRC	
	QMRM Title: Nonlinear classification ANN QSAR model for Genotoxicity (In Vitro Mammalian Cell Gene Mutation Test - Mouse Lymphoma Assay)	
	Printing Date: Dec 6, 2011	

1. QSAR identifier

1.1. QSAR identifier (title):

Nonlinear classification ANN QSAR model for Genotoxicity (In Vitro Mammalian Cell Gene Mutation Test - Mouse Lymphoma Assay)

1.2. Other related models:

1.3. Software coding the model:

QSARModel 4.0.4; Statistica 7, StatSoft Ltd. Turu 2, Tartu, 51014, Estonia, <http://www.molcode.com>

2. General information

2.1. Date of QMRM:

11.11.2011

2.2. QMRM author(s) and contact details:

Dimitar Dobchev, Tarmo Tamm, Gunnar Karelson, Indrek Tulp, Kaido Tämm, Jaak Jänes, Eneli Härk, Mati Karelson, Molcode model development team Molcode Ltd. Turu 2, Tartu, 51014, Estonia models@molcode.com <http://www.molcode.com>

2.3. Date of QMRM update(s):

2.4. QMRM update(s):

2.5. Model developer(s) and contact details:

Molcode model development team Molcode Ltd Molcode Ltd Turu 2, Tartu, 51014, Estonia models@molcode.com www.molcode.com

2.6. Date of model development and/or publication:

08.11.2011

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

[1]Karelson, M.; Dobchev, D. (2011). Using artificial neural networks to predict cell-penetrating compounds. Expert Opinion on Drug Discovery, 6(8), 783 - 796.

[2]Statistica 7 www.statsoft.com

[3]Karelson M, Dobchev D, Tamm T, Tulp I, Jänes J, Tämm K, Lomaka A, Savchenko D & Karelson G (2008). Correlation of blood-brain penetration and human serum albumin binding with theoretical descriptors. ARKIVOC 16, 38-60.

[4]Karelson M, Karelson G, Tamm T, Tulp I, Jänes J, Tämm K, Lomaka A, Savchenko D & Dobchev D (2009). QSAR study of pharmacological permeabilities. ARKIVOC 2, 218–238.

2.8. Availability of information about the model:

All data and modeling information is available

2.9. Availability of another QMRM for exactly the same model:

3. Defining the endpoint - OECD Principle 1

3.1. Species:

Mouse lymphoma cells

3.2. Endpoint:

4. Human health effects 4.10. Mutagenicity

3.3. Comment on endpoint:

Determination of gene mutations In vitro mammalian cell gene mutation test was carried out using the OECD test guideline 476. The in vitro mammalian cell gene mutation test can be used to detect gene mutations induced by chemical substances. Suitable cell lines include L5178Y mouse lymphoma cells, the CHO, AS52 and V79 lines of Chinese hamster cells, and TK6 human lymphoblastoid cells. In these cell lines the most commonly-used genetic endpoints measure mutation at thymidine kinase (TK) and hypoxanthine-guanine phosphoribosyl transferase (HPRT), and a transgene of xanthine-guanine phosphoribosyl transferase (XPRT). The TK, HPRT and XPRT mutation tests detect different spectra of genetic events. The autosomal location of TK and XPRT may allow the detection of genetic events (e.g. large deletions) not detected at the HPRT locus on X-chromosomes. Many compounds that are positive in this test are mammalian carcinogens; however, there is not a perfect correlation between this test and carcinogenicity [1].

3.4. Endpoint units:

+/- (yes/no)

3.5. Dependent variable:

Genotoxic index (-1,+1)

3.6. Experimental protocol:

The L5178Y TK+/- mouse lymphoma cells are used throughout the study. The cells used in the mutagenesis assay should have a high cloning efficiency and low spontaneous mutation frequency. The cells are maintained as suspension culture in RPMI 1640 media in culture flasks and incubated at 37°C and 5% CO₂. The cells have a doubling time as 10-11 hours. Each week the cells will be grown in the RPMI 1640 media containing THMG (thymidine, hypoxanthine, methotrexate and glycine) to select against newly arising TK-/- mutants, and then placed in the RPMI 1640 media containing THG (thymidine, hypoxanthine, and glycine) for 1-3 days prior to use in mutagenesis study.

Cells will be exposed to the test agent both in the presence and absence of an appropriate metabolic activation system. Cofactor-supplemented liver S9 from Aroclor-induced rats will be used in each assay.

The test agent will be freshly dissolved with distilled water prior to each use. A preliminary range finding experiment will be conducted using 10 doses over a 3-4 log range with 5000 µg/mL as the top concentration. The procedures for range finding are identical to that used for mutagenesis except that the cultures are terminated after 24-48 hours without further cloning. The toxicity is indicated by the decrease of cell number in the suspension culture compared with that in untreated control. Four to five concentrations will be selected based on the result and used in the mutagenesis assay. The highest dose should produce a low level of survival (approximately 10-15%), and the survival in the lowest dose should be the same as the negative control. Negative control without treatment and positive control with known mutagens should be included in each assay.

Methylmethanesulphonate (MMS, without S9 mixture) and Cyclophosphamide or Benzo(a)pyrene (with S9 mixture) will be used as the positive controls. Both mutagens are dissolved in DMSO, and corresponding solvent control will also be included.

Cells (6 x 10⁶ cells in 10 ml medium for each culture) are treated with test agents with and without S9 mixture, and incubated at 37°C for 3-4 hours. Chemicals are removed and cells are washed twice by centrifugation then resuspended in non-selective medium at a density of 3 x 10⁵ cells/ml, and maintained in the incubator for 2 days at 37°C. The 2 day maintenance after exposure is the expression period for mutation. During this period, cell density is checked daily and adjusted to 3 x 10⁵ cells/ml. On the second day of expression, 1.6 cells are seeded on 96-well plates to determine the survival and the mutation frequency. For each dose group, two 96-well plates containing 1.6 cells/well in non-selective medium are set up for viability measurement, another two 96-well plates with 2000 cells/well in selective medium containing TFT (4 µg/ml) are used for mutant counting. Plates are incubated at 37°C in an atmosphere of 5% CO₂ - 95% air. Colonies are counted and sized 11-14 days after cloning using. The mutant frequency is calculated and adjusted based on the survival percentage.

The toxicity of the test agent will be indicated by a decrease in colony forming efficiency (CFE: number of colonies/number of cells plated [%]). The mutagenicity of the test agent will be evident by the increase in mutation frequency based on the number of mutants and adjusted by the survival fraction of cells.

A test agent will be considered to be positive in the mouse lymphoma cell mutagenesis assay if it induces a statistically significant dose-related increase in the mutant frequency, or generates a reproducible and statistically significant increase in the mutant frequency for at least one concentration.

A positive result in mouse lymphoma cell mutagenesis assay indicates that under the experimental conditions, the test compound induces gene mutation in the cells used. A high amount of small colonies is an indication of a compound's clastogenicity. A test agent which does not produce either a statistically significant dose-related increase or a reproducible and statistically significant increase of the mutant frequency in any one of the concentrations tested will be considered nonmutagenic in this system. A negative result indicates that under the experimental conditions, the test compound does not produce gene mutation in the cells used [2-5].

3.7. Endpoint data quality and variability:

Experimental data from a number of different publications was used, as assembled in publication listed in 9.2

4. Defining the algorithm - OECD Principle 2

4.1. Type of model:

Nonlinear QSAR: Backpropagation Neural Network (Multilayer Perceptron) classification

4.2. Explicit algorithm:

The algorithm is based on neural network predictor with structure 7-5-5-1 Standard Backpropagation Neural Network (Multilayer Perceptron) using Levenberg-Marquardt optimization algorithm

4.3. Descriptors in the model:

- [1]No. of occupied electronic levels (AM1)
- [2]Bonding Information content (order 1) unitless
- [3]Randic index (order 0) unitless
- [4]Molecular volume (AM1) A^3
- [5]HA dependent HDCA-2 (AM1) $C.A^2$
- [6]Gravitation index (all bonds) (AM1) amu/A^2
- [7]HOMO - LUMO energy gap (AM1) eV

4.4.Descriptor selection:

Initial pool of ~1000 descriptors. Reduction of the initial descriptor pool - stepwise descriptor selection based on the highest F statistics (low p values) followed with variance cleaning (small variance of the descriptor 10^{-6}). Further, the best 10 descriptors were selected and used for building different (topology) ANN models. 29 networks with different structures and descriptors were tested in order to find the best ANN with lowest RMS (root-mean-squared error) for training, selection and test sets. Then 1002 epochs were used to train the final network with architecture depicted in 4.2. Optimization of the weights was performed by Levenberg-Marquardt algorithm using linear(inputs) and hyperbolic(hidden) and logistic(output) activation functions.

4.5.Algorithm and descriptor generation:

All descriptors were generated using QSARModel on structure optimized by AM1 semiempirical quantum mechanical model.

4.6.Software name and version for descriptor generation:

QSARModel 4.0.4; Statistica 7, StatSoft Ltd.

Turu 2, Tartu, 51014, Estonia

<http://www.molcode.com>

4.7.Chemicals/Descriptors ratio:

according to the training set = 42

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Applicability domain based on training set:

By descriptor value range (between min and max values): The model is suitable for compounds (small organic molecules with functional groups as halogens, nitro, hetero benzens, alcohols, acids etc) that have descriptors in the following range augmented with the confidence in 5.2:

(the following is in table format - first row Descs IDs, second row- min desc values, third row - max desc values)

Desc 1 2 3 4 5 6 7

min 6.0 0.3 1.0 23.7 0.0 151.3 3.2

max 204.0 1.0 53.4 786.6 13.8 9365.2 14.4

5.2.Method used to assess the applicability domain:

Quantitative approach - range of descriptor values in training set with augmented with $\pm 30\%$ confidence

Descriptor values must fall between maximal and minimal descriptor values (see 5.1) of training set augmented by $\pm 30\%$.

5.3. Software name and version for applicability domain assessment:

QSARModel 4.0.4; Statistica 7, StatSoft Ltd.

Turu 2, Tartu, 51014, Estonia,

<http://www.molcode.com>

5.4. Limits of applicability:

See 5.2

6. Internal validation - OECD Principle 4

6.1. Availability of the training set:

Yes

6.2. Available information for the training set:

CAS RN: Yes

Chemical Name: Yes

Smiles: No

Formula: No

INChI: No

MOL file: Yes

6.3. Data for each descriptor variable for the training set:

All

6.4. Data for the dependent variable for the training set:

All

6.5. Other information about the training set:

data points: 299

6.6. Pre-processing of data before modelling:

6.7. Statistics for goodness-of-fit:

NB: This is classification ANN model and its statistical parameters are different than the usual multilinear models.

The bellow information is in table format (we are well aware that this reporting tool doesnt format properly the data):

Train. Genotoxic Index.-1	Train. Genotoxic Index.1	Sel. Genotoxic Index.-1	Sel. Genotoxic Index.1
---------------------------	--------------------------	-------------------------	------------------------

Total	146	153	12	28	24	16
-------	-----	-----	----	----	----	----

Correct	117	125	10	20	16	13
---------	-----	-----	----	----	----	----

Wrong	29	28	2	8	8	3
-------	----	----	---	---	---	---

Correct(%)	80	82	83	71	67	81
------------	----	----	----	----	----	----

Wrong(%)	20	18	17	29	33	19
----------	----	----	----	----	----	----

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

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

6.10. Robustness - Statistics obtained by Y-scrambling:

6.11. Robustness - Statistics obtained by bootstrap:

6.12. Robustness - Statistics obtained by other methods:

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

Yes

7.2.Available information for the external validation set:

CAS RN:Yes

Chemical Name:Yes

Smiles:No

Formula:No

INChI:No

MOL file:Yes

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

All

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

All

7.5.Other information about the external validation set:

The method used two randomly selected validation sets – selection (40) and test(40)

7.6.Experimental design of test set:

Randomly selected 40 and 40 data points

7.7.Predictivity - Statistics obtained by external validation:

see 6.7 and 6.12

7.8.Predictivity - Assessment of the external validation set:

see 6.7 and 6.12

7.9.Comments on the external validation of the model:

This is a classification method which stresses on the correct prediction of the positive and negative Genotoxic index. Overall predictions for the selection set (used to stop the ANN training and not to over fit it) and the test set (used to test the external prediction of the net after training) are significant according to 6.7.

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

Since this model is nonlinear (ANN) the mechanistic interpretation related to the descriptors in the model is not so straightforward as compared to simple multilinear models. However, according to the descriptor No. of occupied electronic levels (AM1) the genotoxic index is negative for larger values. It indicates that the compounds with open shells and thus unpaired electrons (resulting in residual charges and lone pairs) lead to genotoxic index +1.

8.2.A priori or a posteriori mechanistic interpretation:

a posteriori mechanistic interpretation, consistent with published scientific interpretations of experiments

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

Supporting information for :Training set(s)
Selection set(s)
Test set(s)

9.2.Bibliography:

- [1]In Vitro Mammalian Cell Gene Mutation Test OECD TG 476, 1997.
[2]<http://www.genpharmtox.de/downloads/AssaySheetMOLY.pdf>
[3]Oberly T., Yount D., Garriott M. A comparison of the soft agar and microtitre methodologies for the L5178Y TK+/- mouse lymphoma assay, Mutation Research, 1997, 388, 59-66
[4]Kirkland D., Aardema M., Henderson L. and Müller L. Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity, Mutation Research, 2005, 584, 1–256.
[5]Ronald D. Snyder An Update on the Genotoxicity and Carcinogenicity of Marketed Pharmaceuticals with Reference to In Silico Predictivity, Environ. Mol. Mutagen., 2009, 50, 435–450

9.3.Supporting information:

Training set(s)

MLA_train.sdf	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf329_MLA_train.sdf
---------------	---

Test set(s)

MLA_test.sdf	http://qsardb.jrc.it:80/qmrf/download_attachment.jsp?name=qmrf329_MLA_test.sdf
--------------	---

10.Summary (JRC Inventory)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

To be entered by JRC

10.4.Comments:

To be entered by JRC