Leadscope Enterprise model for mutations in the thymidine kinase (TK) locus in mouse lymphoma cells *in vitro*

1. QSAR identifier

1.1 QSAR identifier (title)

Leadscope Enterprise model for mutations in the thymidine kinase (TK) locus in mouse lymphoma cells *in vitro*, Danish QSAR Group at DTU Food.

1.2 Other related models

MultiCASE CASE Ultra model for mutations in the thymidine kinase (TK) locus in mouse lymphoma cells *in vitro*, Danish QSAR Group at DTU Food.

SciMatics SciQSAR model for mutations in the thymidine kinase (TK) locus in mouse lymphoma cells *in vitro*, Danish QSAR Group at DTU Food.

1.3. Software coding the model

Leadscope Predictive Data Miner, a component of Leadscope Enterprise version 3.1.1-10.

2. General information

2.1 Date of QMRF

January 2015.

2.2 QMRF author(s) and contact details

QSAR Group at DTU Food;

Danish National Food Institute at the Technical University of Denmark;

http://qsar.food.dtu.dk/;

qsar@food.dtu.dk

Eva Bay Wedebye;

National Food Institute at the Technical University of Denmark;

Nikolai Georgiev Nikolov;

National Food Institute at the Technical University of Denmark;

Marianne Dybdahl;

National Food Institute at the Technical University of Denmark;

Sine Abildgaard Rosenberg;

National Food Institute at the Technical University of Denmark;

2.3 Date of QMRF update(s)

2.4 QMRF update(s)

2.5 Model developer(s) and contact details

Jay Russel Niemelä;

National Food Institute at the Technical University of Denmark;

Nikolai Georgiev Nikolov;

National Food Institute at the Technical University of Denmark;

Eva Bay Wedebye;

National Food Institute at the Technical University of Denmark;

Danish QSAR Group at DTU Food;

National Food Institute at the Technical University of Denmark;

http://qsar.food.dtu.dk/;

qsar@food.dtu.dk

2.6 Date of model development and/or publication

January 2014.

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

Roberts, G., Myatt, G. J., Johnson, W. P., Cross, K. P., and Blower, P. E. J. (2000) LeadScope: Software for Exploring Large Sets of Screening Data. *Chem. Inf. Comput. Sci.*, 40, 1302-1314.

Cross, K.P., Myatt, G., Yang, C., Fligner, M.A., Verducci, J.S., and Blower, P.E. Jr. (2003) Finding Discriminating Structural Features by Reassembling Common Building Blocks. *J. Med. Chem.*, *46,* 4770-4775.

Valerio, L. G., Yang, C., Arvidson, K. B., and Kruhlak, N. L. (2010) A structural feature-based computational approach for toxicology predictions. *Expert Opin. Drug Metab. Toxicol.*, 6:4, 505-518.

2.8 Availability of information about the model

The training set is non-proprietary and consists of Gene-Tox data from Grant *et al.* (2000) who derived the data from Mitchell *et al.* (1997). The model algorithm is proprietary from commercial software.

2.9 Availability of another QMRF for exactly the same model

3. Defining the endpoint

3.1 Species

Mouse (lymphoma cells (L5178Y/*tk+/-* - 3.7.2C cells)).

3.2 Endpoint

QMRF 4. Human Health Effects

QMRF 4.10. Mutagenicity

3.3 Comment on endpoint

For training of this model results from the L5178Y/$tk^{+/-}$ - 3.7.2C mouse lymphoma assay (MLA) carried out under the aegis of the US National Toxicology Program (NTP) were used (Mitchell *et al.* 1997). The assay detects chemicals causing mutations (i.e. $tk^{-/-}$) and/or allele loss (i.e. $tk^{o/-}$) affecting the heterozygous thymidine kinase (*tk)* locus in L5178Y/$tk^{+/-}$ - 3.7.2C cells. As the MLA is capable of responding to chemicals acting as clastogens as well as point mutagens mutant frequencies seen in the MLA can be quite high compared with similar assays. Mitchell *et al.* (1997) concluded based on the published US NTP MLA data, that for most chemicals the mouse lymphoma assay is eminently well suited for genotoxicity testing and for predicting the potential for carcinogenicity. Even though many compounds positive in this test are mammalian carcinogens there is not a perfect correlation between this test and carcinogenicity. Correlation is dependent on chemical class and there is increasing evidence that there are carcinogens that are not detected by this test because they appear to act through mechanisms not readily detected in these cells.

Thymidine Kinase (TK) is an enzyme that phosphorylates thymidine to thymidine monophosphate (TMP) in most mammalian cells. If a lethal TPM analogue, such as trifluorothymidine (TFT), the selective agent, is added to the medium it is phosphorylated by TK and cause inhibition of cellular metabolism and halts further cell division (cytotoxicity). Cells deficient in TK due to the mutation TK+/- →$TK^{-/-}$ are resistant to the cytotoxic effects of the lethal TFT, because the TK mediated phosphorylation of TFT does not occur. Thus mutant cells are able to proliferate in the presence of TFT, whereas normal cells, which contain functional TK, are not. The mutant frequency is derived from the number of mutant colonies in selective TFT-containing medium relative to the number of colonies in non-selective medium (no TFT). The colonies are scored using the criteria of normal growth (large) and slow growth (small) colonies. Mutant cells that have suffered the most extensive genetic damage have prolonged doubling times and thus form small colonies. This damage typically ranges in scale from the losses of the entire gene to karyotypically visible chromosome aberrations. The induction of small colony mutants has been associated with chemicals that induce gross chromosome aberrations. Less seriously affected mutant cells grow at rates similar to the parental cells and form large colonies.

3.4 Endpoint units

No units, 1 for positives and 0 for negatives.

3.5 Dependent variable

Mutations in the thymidine kinase (*tk)* locus of mouse lymphoma cells *in vitro*, positive or negative.

3.6 Experimental protocol

The experimental protocol is described in OECD guideline 476 (1997). Briefly, cells (mouse lymphoma cells) in suspension or monolayer culture are exposed to the test substance, both with and without metabolic activation, for a suitable period of time and subcultured to determine cytotoxicity and to allow phenotypic expression prior to mutant selection. Mutant frequency is determined by seeding known numbers of cells in medium containing the selective agent TFT to detect mutant cells, and in medium without selective agent to determine the cloning efficiency (viability). After a suitable incubation time, colonies are counted.

3.7 Endpoint data quality and variability

The US NTP Gene-Tox MLA data were originally gathered by Mitchell *et al.* (1997), who reviewed and evaluated literature containing MLA results published from 1976 through 1993. Mitchell *et al.* (1997) only concluded on MLA data for chemicals that according to a set of criteria were considered adequately tested. As the data were originally compiled from multiple sources some degree of variability in data is expected, especially because the performance of the MLA assay has been regularly improved during the timeframe. Also, Mitchell *et al.* (1997) noted that not all laboratories detected the small colonies. By the use of the set of criteria for inclusion data variability has been minimised.
To balance the training set normal physiological chemicals were included as non-mutagenic by Grant *et al.* (2000). These chemicals have not been tested in the MLA but are assumed negative as they are normal constituent of cells.

4. Defining the algorithm

4.1 Type of model

A categorical (Q)SAR model based on structural features and numeric molecular descriptors.

4.2 Explicit algorithm

This is a categorical (Q)SAR model made by use of partial logistic regression (PLR). The model is a composite model consisting of 2 submodels, using all the negatives (246 chemicals) in each of these and different sub-sets of the positives (see 4.5). The specific implementation is proprietary within the Leadscope software.

4.3 Descriptors in the model

structural features,

aLogP,

polar surface area,

number of hydrogen bond donors,

Lipinski score,

number of rotational bonds,

parent atom count,

parent molecular weight,

number of hydrogen bond acceptors

4.4 Descriptor selection

Leadscope Predictive Data Miner is a software program for systematic sub-structural analysis of a chemical using predefined structural features stored in a template library, training set-dependent generated structural features (scaffolds) and calculated molecular descriptors. The feature library contains approximately 27,000 pre-defined structural features and the structural features chosen for the library are motivated by those typically found in small molecules: aromatics, heterocycles, spacer groups, simple substituents. Leadscope allows for the generation of training set-dependent structural features (scaffold generation), and these features can be added to the pre-defined structural features from the library and be included in the descriptor selection process. It is possible in Leadscope to remove redundant structural features before the descriptor selection process and only use the remaining features in the descriptor

selection process. Besides the structural features Leadscope also calculates eight molecular descriptors for each training set structure: the octanol/water partition coefficient (alogP), hydrogen bond acceptors (HBA), hydrogen bond donors (HBD), Lipinski score, atom count, parent compound molecular weight, polar surface area (PSA) and rotatable bonds. These eight molecular descriptors are also included in the descriptor selection process.

Leadscope has a default automatic descriptor selection procedure. This procedure selects the top 30% of the descriptors (structural features and molecular descriptors) according to X^2-test for a binary variable, or the top and bottom 15% descriptors according to *t*-test for a continuous variable. Leadscope treats numeric property data as ordinal categorical data. If the input data is continuous such as IC_{50} or cLogP data, the user can determine how values are assigned to categories: the number of categories and the cut-off values between categories. (Roberts *et al.*2000).

When developing this model, intermediate models with application of different modelling approaches in Leadscope were tried:

1. 'Single model' using only the Leadscope pre-defined structural features, i.e. no scaffolds, and calculated molecular descriptors for descriptor selection.
2. 'Single model' using both the Leadscope pre-defined structural features and the training set dependent features (scaffolds generation) as well as the calculated molecular descriptors in the descriptor selection.
3. 'Single model' using both Leadscope pre-defined structural features and the training set dependent features (scaffolds generation), with subsequent removal of redundant structural features, and calculated molecular descriptors for descriptor selection.
4. 'Composite model' using only the Leadscope pre-defined structural features, i.e. no scaffolds, and calculated molecular descriptors in the descriptor selection.
5. 'Composite model' using both Leadscope pre-defined structural features and the training set dependent features (scaffolds generation) as well as the calculated molecular descriptors in the descriptor selection.

Based on model performance as measured by a preliminary cross-validation the model developed using approach number 4. was chosen.

In this model the descriptors were automatically selected among the pre-defined structural features and the eight molecular descriptors.

4.5 Algorithm and descriptor generation

For descriptor generation see 4.4.

After selection of descriptors the Leadscope Predictive Data Miner program performs partial least squares (PLS) regression for a continuous response variable, or partial logistic regression (PLR) for a binary response variable, to build a predictive model. By default the Predictive Data Miner performs leave-one-out or leave-groups-out (in the latter case, the user can specify any number of repetitions and percentage of structures left out) cross-validation on the training set depending on the size of the training set. In the cross-validation made by Leadscope the descriptors selected for the 'mother model' are used when building the validation submodels and they therefore have a tendency to be overfittet and give overoptimistic validation results.

In this model because of the categorical outcome in the response variable PLR was used to build the predictive model. Because of the unbalanced training set (i.e. 282 positives vs. 246 negatives) 2 submodels for smaller individual training sets consisting of the 246 negatives and an equal number of positives selected by random among the 282 positives were made. The descriptors for each of the 2 submodels were automatically selected from the Leadscope feature library based solely on the training set compounds used to build the individual submodel and was not affected by the training set chemicals in the composite model. Therefore, a different number of descriptors (structural features and molecular descriptors) were selected and distributed on varying number of PLS factors for each submodel.

4.6 Software name and version for descriptor generation

Leadscope Predictive Data Miner, a component of Leadscope Enterprise version 3.1.1-10.

4.7 Descriptors/chemicals ratio

As this model is a composite model consisting of 2 submodels with varying training set size and using different descriptors and number of PLS factors (see 4.5), an overall descriptor/chemical ratio for this model cannot be calculated.

5. Defining Applicability Domain

5.1 Description of the applicability domain of the model

The definition of the applicability domain consists of two components; the definition of a structural domain in Leadscope and the in-house further probability refinement algorithm on the output from Leadscope to reach the final applicability domain call.

1. Leadscope

For assessing if a test compound is within the structural applicability domain of a given model Leadscope examines whether the test compound bears enough structural resemblance to the training set compounds used for building the model (i.e. a structural domain analysis). This is done by calculating the distance between the test compound and all compounds in the training set (distance = 1 - similarity). The similarity score is based on the Tanimoto method. The number of neighbours is defined as the number of compounds in the training set that have a distance equal to or smaller than 0.7 with respect to the test compound. The higher the number of neighbours, the more reliable the prediction for the test compound. Statistics of the distances are also calculated. Effectively no predictions are made for test compounds which are not within the structural domain of the model or for which the molecular descriptors could not be calculated in Leadscope.

2. The Danish QSAR group

In addition to the general Leadscope structural applicability domain definition the Danish QSAR group has applied a further requirement to the applicability domain of the model. That is only positive predictions with a probability equal to or greater than 0.7 and negative predictions with probability equal to or less than 0.3 are accepted. Predictions within the structural applicability domain but with probability between 0.5 to 0.7 or 0.3 to 0.5 are defined as positives out of applicability domain and negatives out of applicability domain, respectively. When these predictions are wed out the performance of the model in general increases at the expense of reduced model coverage.

5.2 Method used to assess the applicability domain

Leadscope does not generate predictions for test compounds which are not within the structural domain of the model or for which the molecular descriptors could not be calculated.

Only positive predictions with probability equal to or greater than 0.7 and negative predictions with probability equal to or less than 0.3 are accepted.

5.3 Software name and version for applicability domain assessment

Leadscope Predictive Data Miner, a component of Leadscope Enterprise version 3.1.1-10.

5.4 Limits of applicability

The Danish QSAR group applies an overall definition of structures acceptable for QSAR processing which is applicable for all the in-house QSAR software, i.e. not only CASE Ultra. According to this definition accepted

structures are organic substances with an unambiguous structure, i.e. so-called discrete organics defined as: organic compounds with a defined two dimensional (2D) structure containing at least two carbon atoms, only certain atoms (H, Li, B, C, N, O, F, Na, Mg, Si, P, S, Cl, K, Ca, Br, and I), and not mixtures with two or more 'big components' when analyzed for ionic bonds (for a number of small known organic ions assumed not to affect toxicity the 'parent molecule' is accepted). Calculation 2D structures (SMILES and/or SDF) are generated by stripping off ions (of the accepted list given above). Thus, all the training set and prediction set chemicals are used in their non-ionized form. See 5.1 for further applicability domain definition.

6. Internal validation

6.1 Availability of the training set

Yes

6.2 Available information for the training set

CAS

SMILES

6.3 Data for each descriptor variable for the training set

No

6.4 Data for the dependent variable for the training set

All

6.5 Other information about the training set

528 compounds are in the training set: 282 positives and 246 negatives.

This model is based on more or less on the same training set used for the QSAR model published in Grant *et al.* (2000). The training set used for the published MultiCASE model contained 570 compounds, and results from 10 times 10%-out cross-validation of this model a sensitivity of 70%, specificity of 81% and concordance of 75% (Grant *et al.* 2000). Of the initial training set of 570 compounds only compounds for which SMILES codes could be generated and that made it through the pre-processing procedure described in 6.6 were used in the final training set of this model.

6.6 Pre-processing of data before modelling

Only structures acceptable for Leadscope were used in the final training set. That is only discrete organic chemicals as described in 5.4 were used. In case of replicate structures, one of the replicates was kept if all the compounds had the same activity and all were removed if they had different activity. No further structures accepted by the software were eliminated (i.e. outliers).

6.7 Statistics for goodness-of-fit

Not performed.

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

Not performed. (It is not a preferred measurement for evaluating large models).

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

A five times two-fold 50 % cross-validation was performed. This was done by randomly removing 50% of the full training set used to make the "mother model", where the 50% contains the same ratio of positive and negatives as the full training set. A new model (validation submodel) was created on the remaining 50% using the same settings in Leadscope but with no information from the "mother model" regarding descriptor selection etc. The validation submodel was applied to predict the removed 50% (within the defined applicability domain for the submodel). Likewise, a validation submodel was made on the removed 50% of the training set and this model was used to predict the other 50% (within the defined applicability domain for this submodel). This procedure was repeated five times.

Predictions within the defined applicability domain of the ten validation submodels were pooled and Cooper's statistics calculated. This gave the following results for the 64% (1689/(5*528)) of the predictions which were within the applicability domains of the respective submodels:

- Sensitivity (true positives / (true positives + false negatives)): 85.1%
- Specificity (true negatives / (true negatives + false positives)): 83.8%
- Concordance ((true positives + true negatives) / (true positives + true negatives + false positives + false negatives)): 84.4%

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.

7. External validation

7.1 Availability of the external validation set

7.2 Available information for the external validation set

7.3 Data for each descriptor variable for the external validation set

7.4 Data for the dependent variable for the external validation set

7.5 Other information about the training set

7.6 Experimental design of test set

7.7 Predictivity – Statistics obtained by external validation

7.8 Predictivity – Assessment of the external validation set

7.9 Comments on the external validation of the model

External validation not performed.

8. Mechanistic interpretation

8.1 Mechanistic basis of the model

The global model identifies structural features and molecular descriptors which in the model development was found to be statistically significant associated with effect. Many predictions may indicate modes of action that are obvious for persons with expert knowledge for the endpoint.

8.2 A priori or posteriori mechanistic interpretation

A posteriori mechanistic interpretation. The identified structural features and molecular descriptors may provide basis for mechanistic interpretation.

8.3 Other information about the mechanistic interpretation

9. Miscellaneous information

9.1 Comments

The model can be used to predict results for the L5178Y/*tk+/*--3.7.2C mouse lymphoma *in vitro* assay (MLA).

9.2 Bibliography

Grant, S.G., Zhang, Y.P., Klopman, G., and Rosenkranz, H.S. (2000) Modeling the mouse lymphoma forward mutational assay: the Gene-Tox program database. *Mutation Research*, 465, 201–229.

Mitchell, A.D., Auletta, A.E., Clive, D., Kirby, P.E., Moore, M.M., and Myhr, B.C. (1997) The L5178Y/*tk+/-* mouse lymphoma specific gene and chromosomal mutation assay - A phase III report of the U.S. Environmental Protection Agency Gene-Tox Program. *Mutation Research*, 394, 177–303.

OECD guideline 476 (1997) OECD Guidelines for the Testing of Chemicals No. 476: *In Vitro* Mammalian Cell Gene Mutation Test. Organisation for Economic Cooperation and Development; Paris, France. Available online at: http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788.

9.3 Supporting information