Leadscope Enterprise model for sex-linked recessive lethal test in *Drosophila melanogaster in vivo*

1. QSAR identifier

1.1 QSAR identifier (title)

Leadscope Enterprise model for sex-linked recessive lethal test in *Drosophila melanogaster in vivo*, Danish QSAR Group at DTU Food.

1.2 Other related models

MultiCASE CASE Ultra model for sex-linked recessive lethal test in *Drosophila melanogaster in vivo*, Danish QSAR Group at DTU Food.

SciMatics SciQSAR model for sex-linked recessive lethal test in *Drosophila melanogaster in vivo*, 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

Eva Bay Wedebye;

National Food Institute at the Technical University of Denmark;

Jay Russel Niemelä;

National Food Institute at the Technical University of Denmark;

Nikolai Georgiev Nikolov;

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 compiled by Lee and co-workers (1983) plus data from EMIC (US Environmental Mutagen Information Center), IARC (International Agency for Research on Cancer) and NTP (US National Toxicology Program) etc. 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

Drosophila melanogaster (germ cells).

3.2 Endpoint

QMRF 4.10. Mutagenicity

EC B.20. Sex-Linked recessive Lethal Test in Drosophila Melanogaster

3.3 Comment on endpoint

Drosophila melanogaster, generally known as the common fruit fly, is the test organism most often used to detect transmissible mutations in germ cells of eukaryotes. The short generation times of 10 days, low cost of culture media, and a large number of well-defined genetic tests for mutations are the principal advantages of using *D. melanogaster* as compared with using the mouse or rat, which are the only other well-developed systems for testing mutagenesis in germ cells. The sex-linked recessive lethal (SLRL) test using *D. melanogaster* detects the occurrence of chromosome aberrations and mutations, both point mutations and small deletions, in different stages of germ cell development of the insect. It is capable of detecting both direct-acting mutagens and promutagens, i.e. compounds that require activation to become mutagenic. The test is therefore not specific for any one class of chemicals.

This SLRL test is a forward mutation assay capable of screening for mutations in around 800 loci on the X-chromosome. This represents about 80 % of all X-chromosomal loci. The X-chromosome represents approximately one-fifth of the entire genome. Therefore the test gives a good estimate of mutation frequency in the entire genome. A lethal mutation is a change in the genome which, when expressed, causes death to the carrier. A recessive mutation is a change in the genome which is expressed in the homozygous or hemizygous condition. Sex-linked genes are present on the sex (X or Y) chromosomes. Mutations in the X-chromosome of *D. melanogaster* are phenotypically expressed in males carrying the mutant gene. When the mutation is lethal in the hemizygous condition, its presence is inferred from the absence of one class of male offspring out of the two that are normally produced by a heterozygous female.
Positive results from the SLRL-test in *D. melanogaster* indicate that a substance induces mutations in the germ line of the insect. Negative results indicate that, under the test conditions, the test substance does not induce mutations in the germ line of the insect. A high correlation between mutagenesis in the SLRL test and carcinogenesis has been found. (Lee *et al.*1983)

3.4 Endpoint units

No units, 1 for positives and 0 for negatives.

3.5 Dependent variable

Sex-linked recessive lethal (SLRL) test in *D. melanogaster in vivo,* positive or negative.

3.6 Experimental protocol

Data has been generated using similar experimental protocols similar to that described in OECD guideline 477 (1984). Briefly, 3 to 5 days old wild-type males are treated with the test substance and mated individually to an excess of virgin females. The females are replaced with fresh virgins every 2 to 3 days to cover the entire germ cell cycle. The offspring of these females are scored for lethal effects corresponding to the effects on mature sperm, mid or late-stage spermatids, early spermatids, spermatocytes and spermatogonia at the time of treatment.
Heterozygous F1 females from the above crosses are mated individually with their brothers. In the F2 generation each separate cross is scored for the absence of phenotypically wild-type males. If a culture appears to have arisen from a F1 female carrying a lethal mutation in the parental X-chromosome (i.e. no males with the treated chromosome are observed), a daughter of that female with the same genotype should be tested to ascertain whether the lethality is repeated at the next generation.
The assay has a low sensitivity for genotoxins other than direct-acting agents and simple promutagens, but a very high specificity.
From Lee *et al.* (1983): A positive mutagenic response was the demonstration of a difference between the mutation frequencies in a treated and a concurrent control group that was statistically significant at the 5% level. If the investigation of a compound did not have a concurrent control, but the mutation frequency was significantly higher than 0.5%, the compound was accepted as a mutagen. The frequency of 0.5% was selected because the spontaneous frequencies for the standard strains range from 0.1 to 0.3%.
A test was considered negative if both of the following criteria were met: (1) The observed increase in the treated group over control is less than 0.2% and sample size is large enough so that an observed increase of 0.2% would be statistically significant. (2) The second criterion takes into account the possible differential response of different post-meiotic germ cell stages to direct and indirect mutagens as revealed by a mating pattern analysis. If none of the mating's analysed gives a positive result, the data must indicate a statistically negative response in at least 2 mating's, preferably representing mature sperm and early spermatids.

3.7 Endpoint data quality and variability

The data set from Lee *et al.* (1983) consists of data compiled from publications in the EMIC (Environmental Mutagen Information Center) file for the period 1968 to 1978. The publications were reviewed thoroughly and only data that meet the criteria defined by the Working Group were included. As training set data originates from different sources a certain degree of variability in the experimental protocols (strain, mating protocols, route of administration etc.) and data is expected although this variability has been diminished by the criteria for inclusion.

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 (184 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 5. was chosen.

For this model scaffolds were generated by Leadscope for the training set structures and added to the Leadscope library of structural features. Descriptors were then automatically selected among the 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 lightly unbalanced training set (i.e. 185 positives vs. 184 negatives) 2 submodels for smaller individual training sets consisting of the 184 negatives and an equal number of positives selected by random among the 186 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

370 compounds are in the training set: 186 positives and 184 negatives.

6.6 Pre-processing of data before modelling

From the original data set from Lee *et al.* (1983) only compounds for which SMILES codes could be found and that had a structure acceptable for the commercial software 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.

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 57.0% (1055/(5*370)) of the predictions which were within the applicability domains of the respective submodels:

- Sensitivity (true positives / (true positives + false negatives)): 79.1%
- Specificity (true negatives / (true negatives + false positives)): 80.3%
- Concordance ((true positives + true negatives) / (true positives + true negatives + false positives + false negatives)): 79.6%

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 sex-linked recessive lethal (SLRL) *in vivo* test in *Drosophila melanogaster*.

9.2 Bibliography

Lee, W.R., Abrahamson, S., Valencia, R., von Halle, E.S., Würgler, F.E., and Zimmering, S. (1983) The sex-linked recessive lethal test for mutagenesis in *Drosophila melanogaster*. A report of the U.S. Environmental Protection Agency (EPA) Gene-Tox Program. *Mutation research,* 123, 183-279.

OECD guideline 477 (1984) Genetic Toxicology: Sex-linked Recessive Lethal Test in *Drosophila melanogaster*. OECD guidelines for testing of chemicals. 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