Leadscope Enterprise commercial model for Rodent carcinogenicity in vivo, Danish QSAR Group at DTU Food

1. QSAR identifier

1.1 QSAR identifier (title)

Leadscope Enterprise commercial model for Rodent carcinogenicity *in vivo*, Danish QSAR Group at DTU Food .

1.2 Other related models

MultiCASE CASE Ultra commercial model RC_AFU for Rodent carcinogenicity *in vivo*, Danish QSAR Group at DTU Food.

Leadscope Enterprise commercial model for Rodent carcinogenicity in male rat *in vivo*, Danish QSAR Group at DTU Food.

Leadscope Enterprise commercial model for Rodent carcinogenicity in female rat *in vivo*, Danish QSAR Group at DTU Food.

Leadscope Enterprise commercial model for Rodent carcinogenicity in male mouse *in vivo*, Danish QSAR Group at DTU Food.

Leadscope Enterprise commercial model for Rodent carcinogenicity in female mouse *in vivo*, Danish QSAR Group at DTU Food.

Leadscope Enterprise commercial model for Rodent carcinogenicity in rat *in vivo*, Danish QSAR Group at DTU Food.

Leadscope Enterprise commercial model for Rodent carcinogenicity in mouse *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

Trine Klein Reffstrup;

National Food Institute at the Technical University of Denmark;

Eva Bay Wedebye;

National Food Institute at the Technical University of Denmark;

Sine Abildgaard Rosenberg;

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;

2.3 Date of QMRF update(s)

2.4 QMRF update(s)

2.5 Model developer(s) and contact details

Leadscope Inc.;

1393 Dublin Road, Columbus, Ohio, 43215, USA;

www.leadscope.com

2.6 Date of model development and/or publication

Commercial model updated by Leadscope/FDA (U.S. Food and Drug Administration) regularly. For the Danish QSAR predictions database the 2013 version was used.

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 consists of non-proprietary studies and is composed of data harvested from FDA approval packages and the published literature (for more details see 6.5). The model algorithm is proprietary from commercial software. The model has been created by FDA and Leadscope based on FDA data as part of a Research Cooperation Agreement (RCA).

The training set was constructed using the NTP (U.S. National Toxicology Program) Rodent carcinogenicity database, the Lois Gold Carcinogen Potency Database, FDA/CDER (U.S. Food and Drug Administration / Center for Drug Evaluation and Research) archives, and the scientific literature. The model including the training data set is commercially available from Leadscope.

2.9 Availability of another QMRF for exactly the same model

Leadscope Model Applier – Rodent Carcinogenicity Suite – carc rodent (v2). Date: August 6, 2012.

3. Defining the endpoint

3.1 Species

Rodent.

3.2 Endpoint

4.Human health effects

4.12.Carcinogenicity

3.3 Comment on endpoint

The two-year rodent bioassay is considered the regulatory standard for evaluating the carcinogenic potential of a chemical and provides information on the possible health hazards likely to arise from repeated exposure for a period lasting up to the entire lifespan of the species used. The assay is usually performed in both sexes of rats or mice for a period of two years, with the chemical administered at high doses, primarily by the oral route. Extrapolation from the rodent assay to humans is based on two principal assumptions. The first assumption is that a carcinogenic response in rats/mice predicts possible carcinogenicity in humans (interspecies extrapolation). The second assumption is that carcinogenicity detected at a high dose implies carcinogenicity at low doses, although at a lower rate (dose extrapolation).

3.4 Endpoint units

No units, 1 for positives and 0 for negatives.

3.5 Dependent variable

Carcinogenicity in rodents *in vivo*, positive or negative.

3.6 Experimental protocol

Male and female rats and mice (50-70 animals/group) are divided randomly into one or two control groups and three treatment groups. Historically, the highest dose in the studies generally approximates the maximum tolerated dose (MTD) in the test specie. The test substance is normally administered in the feed or by oral gavage for two years. In NTP studies the most often used rodent strains are the inbred Fisher 344 rat and the hybrid B6C3F1 (C3HxC57B16) mouse. The Sprague-Dawley derived CD rat and the CD-1 Swiss-Webster derived mouse are the predominant strains in pharmaceutical studies submitted to the FDA. Tumor findings are classified as positive if either benign and/or malignant findings are statistically significant in pair-wise comparison to concurrent controls ($p \leq 0.01$) by Fisher's Exact Test or equivalent statistical analysis. The tumor findings are adjusted for rare (with a spontaneous background incidence rate equal to or less than 1 %) and common events (Contrera *et al*. 2005).

3.7 Endpoint data quality and variability

Data in the training set originates from multiple sources and therefore some variability in the experimental procedures (e.g. strains, concentration ranges) and experimental results is expected.

In an external validation exercise (see 7.) performed, the activity scores for 1028 duplicate compounds between the training set and the external test set were compared. Depending on the endpoint, concordance in activity scores ranged from 86.7% to 90.7% (Stavitskaya *et al.*, 2013).

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 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,

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 / dynamic features) 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).

For this model scaffolds were generated by Leadscope for the training set structures and added to the Leadscope library of structural features. No information is available to the authors of this QMRF about possible automatic or manual reduction of scaffolds.

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 sub-models and they therefore have a tendency to be over-fitted 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. For this model, 430 predictors including scaffolds were selected to build the model. These include 7 Leadscope calculated molecular descriptors and 423 structural features. The 430 predictors in total were distributed on 3 PLS factors.

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

In this model, 430 predictors in total were used and distributed on 3 PLS factors. The training set consists of 1531 compounds. The descriptor/chemical ratio is 1:3.56 (430:1531).

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 is 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 the applicability domain and negatives out of the applicability domain, respectively. When these predictions are weeded 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 for which the molecular descriptors could not be calculated.

Only predictions within the Leadscope model dependant structural domain, and positive predictions with probability equal to or greater than 0.7 or 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 Leadscope. 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

The training set is commercially available embedded in software from Leadscope.

6.2 Available information for the training set

No

6.3 Data for each descriptor variable for the training set

No

6.4 Data for the dependent variable for the training set

No

6.5 Other information about the training set

1531 compounds are in the training set: 711 positives and 820 negatives.

6.6 Pre-processing of data before modelling

"Rodent carcinogenicity studies of compounds that have been tested in at least the male and female animals of one rodent species (i.e., 2 cells) were included. Although the MCASE database modules contained compounds tested by nonoral routes of administration (inhalation, intravenous, intramuscular, dermal, intraperitoneal, and subcutaneous), the majority of acceptable studies used an oral route of exposure (feed, gavage, or drinking water). The duration of acceptable carcinogenicity studies was limited to ≥18 months for negative (inactive) compounds. All studies with compound-related tumor findings (positive studies) were acceptable regardless of duration of treatment, with one exception. Positive (active) nonoral studies were included if tumors were induced at other than the site of application." (Matthews and Contrera, 1998). Further information can be found in (Matthews and Contrera, 1998), (Contrera *et al.* 2003) and (Contrera *et al.* 2005).

6.7 Statistics for goodness-of-fit

Concordance: 77.3, sensitivity: 71.4, specificity: 82.5.

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 ten times 50% cross-validation was performed. This was done by randomly removing 50% of the full training set used to make the "mother model". The validation sub-model was applied to predict the removed 50%. This procedure was repeated ten times.

Predictions for the ten validation sub-models were pooled and Cooper's statistics calculated. This gave the following results for the 94.8% of the predictions which were within the applicability domain:

- Sensitivity (true positives / (true positives + false negatives)): 65.9%
- Specificity (true negatives / (true negatives + false positives)): 76.2%
- Concordance ((true positives + true negatives) / (true positives + true negatives + false positives + false negatives)): 71.3%

NB This cross-validation was performed using Leadscope's own functionality. This functionality transfers information about descriptors selected from the "mother model", which in some cases can give optimistic cross-validation results. Furthermore, only the structural domain, and not the statistical domain, could be applied, which will give pessimistic cross-validation results compared to predictions which are within the applicability domain applied in predictions generation for the Danish online (Q)SAR database.

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

No

7.2 Available information for the external validation set

No

7.3 Data for each descriptor variable for the external validation set

No

7.4 Data for the dependent variable for the external validation set

No

7.5 Other information about the training set

The external data set was comprised of non-proprietary data was harvested by Leadscope from FDA approval packages and the published literature. The entire set contained 2115 compounds, but 1393 of the compounds were removed as they were already part of the training set or were stereo or geometric isomers of structures already in the training set or were duplicates or perceived duplicates within the set. Therefore the final external test set consisted of 722 compounds (34-52 % active and 48-66 % inactive) (Stavitskaya *et al*., 2013).

In the external validation the definition of the applicability domain in Leadscope used was point 1 in 5.1 combined with the probability domain where positive predictions with a probability equal to or greater than 0.6 and negative predictions with probability equal to or less than 0.4 are accepted (i.e. different cut-offs than used for the Danish online QSAR predictions database, point 2 in 5.1).

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

See section 3.7 for information on concordance between experimental tests in training sets and validation sets.

7.9 Comments on the external validation of the model

An external validation is not available for the present model on carcinogenicity in rodents in which data from rat and mice studies are comprised. However, external validations are available for the separate

models on carcinogenicity in male rats, female rats, male mice, female mice, combined rats and combined mice, see Stavitskaya et al. 2013 and the QMRFs for Leadscope models listed section 1.2.

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 if a chemical has the potential to cause cancer in rodents.

9.2 Bibliography

Contrera, J.F., Matthews, J. and Benz, R.D. (2003). Predicting the carcinogenic potential of pharmaceuticals in rodents using molecular structural similarity and E-state indices. *Regulatory Toxicology and Pharmacology*, 38, 243-259.

Contrera, J.F., MacLaughlin, P., Hall, L.H., Kier, L.B. (2005) QSAR Modeling of Carcinogenic Risk Using Discriminant Analysis and Topological Molecular Descriptors. *Current Drug Discovery Technologies*, 2, 55-67.

Matthews, J. and Contrera, J.F. (1998) A new highly specific method for predicting the carcinogenic potential of pharmaceuticals in rodent using enhanced MCASE (Q)SAR-ES software. *Regulatory Toxicology and Pharmacology*, 28, 242-264.

Stavitskaya, L., Kruhlak, N.L., Cross, K.P., Minnier, B.L., Bower, D.A., Chakravarti, S., Saiakhov R.D., Benz, R.D.. (2013) P217 Development of Improved In Silico Models for Predicting Rodent Carcinogenicity. Poster at the American College of Toxicology Annual Meeting 2012. Poster abstract published in *International Journal of Toxicology*, 32 (1), 72.

9.3 Supporting information