Commercial ACD/Labs model for acute toxicity (LD50) in mice, oral administration

1. QSAR identifier

1.1 QSAR identifier (title)

Commercial ACD/Labs model for acute toxicity (LD50) in mice, oral administration, Danish QSAR Group at DTU Food.

1.2 Other related models

Commercial ACD/Labs model for acute toxicity (LD50) in rats, oral administration, Danish QSAR Group at DTU Food.

Commercial ACD/Labs model for acute toxicity (LD50) in rats, intraperitoneal administration, Danish QSAR Group at DTU Food.

Commercial ACD/Labs model for acute toxicity (LD50) in mice, intraperitoneal administration, Danish QSAR Group at DTU Food.

Commercial ACD/Labs model for acute toxicity (LD50) in mice, intravenous administration, Danish QSAR Group at DTU Food.

Commercial ACD/Labs model for acute toxicity (LD50) in mice, subcutaneous administration, Danish QSAR Group at DTU Food.

1.3. Software coding the model

ACD/Labs Tox Suite version 2.95.1.

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

Advanced Chemistry Development Inc.;

8 King Street East, Suite 107, Toronto, Ontario, Canada M5C 1B5;

http://www.acdlabs.com/home/

2.6 Date of model development and/or publication

The model was developed in 2010 and has been published in Sazonovas *et al.* (2010).

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

ACD/Percepta: Mouse Oral Acute Toxicity (2012). QMRF Editor pdf file kindly provided by ACD Inc. QMRF author Simona Kovarich (S-IN Soluzioni Informatiche, Via Ferrari 14, I-36100, Vicenza, simona.kovarich@gmail.com, http://www.s-in.it/it/).

Data sheet (2014) Acute Toxicity Prediction Module. ACD/Labs Percepta. Downloaded from http://www.acdlabs.com/products/percepta/predictors.php on the 13th of August 2014.

Japertas, P., Didziapetris, R., Sazonovas, A. and Petrauskas, A. (2007) Acute Toxicity (LD50) Modeling Utilizing Fragmental QSAR, Similarity Analysis and Reliability of Prediction. Poster 47 presented at the 234th ACS National Meeting, Boston, MA, August 19-23, 2007. Abstract available at http://oasys2.confex.com/acs/234nm/techprogram/S24347.HTM.

Model Performance (2014) ACD/Labs Acute Toxicity Module. ACD/Labs Tox Suite. Downloaded from http://www.acdlabs.com/products/percepta/predictors.php on the 13th of August 2014.

Sazonovas, A., Japertas, P., and Didziapetris, R. (2010) Estimation of reliability of predictions and model applicability domain evaluation in the analysis of acute toxicity (LD_{50}). *SAR and QSAR in Environmental Research*, 21:1-2, 127-148.

2.8 Availability of information about the model

The training and validation sets are proprietary and were compiled and prepared by ACD/Labs. The model algorithm is proprietary from commercial software. The model has been published in the paper by Sazonovas *et al.* (2010).

2.9 Availability of another QMRF for exactly the same model

ACD/Percepta: Mouse Oral Acute Toxicity (2012).

3. Defining the endpoint

3.1 Species

Mouse (oral).

3.2 Endpoint

QMRF 4. 2. Acute Oral toxicity

OECD 401 Acute Oral Toxicity DELETED

3.3 Comment on endpoint

In drug development a drug candidate has to go through various stages in order to produce a drug that is efficacious and safe. Also, the drug must pass many regulatory requirements before being approved.
In the preclinical stage the drug is tested in animals to analyse the bioactivity, safety, and efficacy of the formulated drug product. This stage is critical to the drug's eventual success and, as such, is scrutinized by many regulatory entities. One of the many required *in vivo* tests in the preclinical stage is the Acute toxicity test, which goal is to determine toxic dose levels and observe clinical indications of toxicity. Data obtained from the Acute toxicity test helps determine doses for the Repeated dose preclinical *in vivo* test, that test a drug for potential chronic toxicity, as well as for the Phase I clinical studies, that evaluate pharmacokinetic parameters and tolerance in humans, generally in healthy volunteers.

The acute toxicity test may also serve as a basis for classification and labelling of new chemicals and it is used as an initial step in the establishment of a dosage regimen in subchronic and other studies of the chemical. Results from the test may also provide initial information on the mode of toxic action of the chemical.

Data used to train and validate this model originate from Acute toxicity tests in mice administered the test compound orally. The test endpoint is the statistically derived mean lethal dose or lethal dose 50 (LD50), which is the dose expected to cause death in 50% of the members of a tested population within a specified time interval after administration. The data were compiled by ACD/Labs from the Registry of Toxic Effects of Chemical Substances (RTECS®), a compendium of toxicological data, and from the International Uniform ChemicaL Information Database (IUCLID) (for a few compounds not available in RTECS®).

Acute toxicity (LD50) is a complex endpoint as it includes a wide variety of biological mechanisms, of which many either still left undiscovered or fully understood. Results from Acute toxicity tests are in general subjected to high variability, and this is partly due to the complexity of the endpoint but also factors related to the experimental protocol may play a role in the data variability.

3.4 Endpoint units

mg/kg.

3.5 Dependent variable

logLD50.
The experimental LD50 values were converted to logarithmic form (logLD50) in order to maintain a linear relationship with the fragmental descriptors in the model. The final prediction results returned to the user are converted back to an LD50 value in mg/kg (Sazonovas *et al.* 2010).

3.6 Experimental protocol

The experimental LD50 data in the training and validation sets were obtained using an experimental protocol similar to that described in OEDC guideline 401 (1987). Briefly, the test substance is administered in graduated doses to several groups of experimental animals, one dose being used per group. Subsequently observations of effects and deaths are made. Animals which die during the test are necropsied, and at the conclusion of the test the surviving animals are sacrificed and necropsied. The dose (in mg/kg) at which 50% of the animals are expected to die within a specified time interval after administration is calculated and this is the LD50.

OECD deleted the OECD test guideline 401 back in 2002. It has been replaced with other more ethically correct *in vivo* testing methods for acute toxicity.

3.7 Endpoint data quality and variability

As data in the training and validation set originate from various sources and as the Acute toxicity test in general is somewhat unreliable (see 3.3) a certain degree of variability in data is expected. This variability have been reduced during the compilation of data as whenever available, similar data from the IUCLID Chemical Data Sheets were used to validate, correct or exclude entries of RTECS® (Sazonovas *et al.* 2010).

4. Defining the algorithm

4.1 Type of model

A continuous (Q)SAR model based on 404 pre-defined fragmental descriptors and combined to a similarity-based approach.

4.2 Explicit algorithm

This is a continuous (Q)SAR model made by use of the GALAS (Global, Adjusted Locally According to Similarity) modelling methodology (see 4.5). The explicit algorithm and implementation is proprietary within the ACD/Labs software.

4.3 Descriptors in the model

Conventional fragmental descriptors (atoms, functional groups, molecular 'shape' fragments, etc.).

Complex fragments known to be responsible for toxic effects (i.e. toxicophores).

4.4 Descriptor selection

The fragmental descriptors in the model were identified based on general knowledge and considerations regarding all possible chemical structures and not on the basis of fragmentation of the training set molecules. Therefore, even fragments not present in the training set molecules are included as descriptors in the model. No descriptor selection procedures have been used and in total 404 pre-defined fragmental descriptors were used in the model. The major part of these were intended for the description of the general chemical constitution of any compound and comprised conventional fragmental descriptors, such as atoms, functional groups, molecular 'shape fragments', etc. In addition to these a group of more complex fragments, generally called toxicophores, i.e. substructures identified to be responsible for the toxic action of the molecules possessing them, where included in the fragment set. This part of the fragment set was added in order to account for, at least, the most widely-known specific mechanisms and interactions leading to the manifestation of acute toxicity. (Sazonovas *et al.* 2010)

4.5 Algorithm and descriptor generation

The GALAS modelling methodology was used to build this model. The GALAS model is described in details in Sazonovas *et al.* (2010) and only a brief description of the model is given here.
The GALAS model is a combination of the following two systems:

1) A global baseline QSAR model: a fragment-based QSAR model to predict LD50 values was constructed and constitutes the baseline model. It is based on the Partial Least Squares (PLS) regression method, a linear additive method that learns trends, in combination with a bootstrapping technique. The baseline

QSAR model is actually an ensemble of 100 PLS-bootstrapping QSAR models, and the final LD50 baseline prediction is derived as a mean average of the 100 LD50 predictions from these models.

2) Local corrections: any baseline LD50 prediction is subjected to the local similarity correction procedure, in order to capture the deviations from linear trend of the responses that may occur in specific chemical spaces. The correction procedure is based on the analysis of the performance of the global baseline model in the local environment of the query compound. This means a comparison of the experimental data and the baseline LD50 predictions for the five most similar compounds in the training set. These five training set structures were identified using the individual Similarity Index (SI_i) (see Sazonovas *et al.* 2010). If baseline LD50 predictions for the five training set compounds show any systematic deviations from their experimental values, a local correction is applied to the LD50 baseline prediction of the query compound. The required correction is calculated as a weighted average of the differences between the baseline LD50 predictions and experimental data for the five training set compounds.

Because the 404 fragmental descriptors in the model were all predefined and not generated from the training set compounds, no descriptor generation was performed. The fragmentation of the training set compounds and subsequent search for occurrence of the 404 pre-defined fragmental descriptors were performed using Algorithm Builder version 1.8 from Pharma Algorithms Inc.

4.6 Software name and version for descriptor generation

ACD/Labs Tox Suite version 2.95.1.

4.7 Descriptors/chemicals ratio

This model was build using a training set consisting of 14,678 compounds. This gives a descriptor/chemicals ratio of 1:36 (404:14,678).

5. Defining Applicability Domain

5.1 Description of the applicability domain of the model

In addition to the predicted LD50 value for a query compound ACD/Labs also provide a Reliability Index (RI). The RI gives a quantitative assessment (number between 0 and 1) of the reliability of the LD50 prediction for a query compound and by using a RI range in which a prediction is accepted, e.g. RI greater than 0.3, an applicability domain can be defined for the model. The RI takes into account the similarity of the query compound to the training set compounds, the difference between predicted LD50 and experimental values for the five most similar training set compounds, and the consistence of experimental values for the five similar compounds. The RI is calculated using the equation:

$$RI = SI * DMCI$$

- Similarity Index (SI): this index is a measure of how distant the query structure is from the whole training set in terms of structural similarity. The SI is calculated as a weighted average of all the individual Similarity Indices (SI_i), i.e. the similarity of the query structure with each of the training set compounds. SI is a value between 0 and 1, zero meaning no similarity to any training set structure.
- Data-model consistency index (DMCI): this index accounts for the influence of consistency of experimental data with regard to the baseline model for the five most similar training set compounds on the reliability of a prediction. It is calculated as the differences between the experimental LD50 and the baseline LD50 prediction for each of the five training set compounds and the calculated correction value used in the correction of the predicted baseline LD50 for the query structure. DMCI is a value between 0 and 1, zero meaning no consistency between experimental values and baseline model predictions.

More detailed descriptions and equations for the indices can be found in Sazonovas *et al.* (2010).

If the RI approaches zero, then a given compound is far from the model applicability domain and the respective prediction is unreliable. This is observed when either the SI or DMCI approaches zero, i.e. when either no similar structures are present in the training set, or such structures have inconsistent experimental data (respectively to baseline model). If the RI approaches 1, then a given compound is within the model applicability domain and the respective prediction is reliable. This is only observed if both the SI and DMCI approaches 1, i.e. similar compounds are found in the training set and their baseline LD50 predictions are consistent with their experimental values.

Correlations between the RI and the Root Mean Square Error (RMSE) and R2 from external validations have been shown (Japertas *et al.* 2007, Sazonovas *et al.* 2010). That is RMSE decreases and R2 increases as RI increases. Thus by increasing the RI cut off for the applicability domain the performance of the model will increase correspondingly. The cost of increasing RI and performance is a smaller applicability domain, i.e. a smaller number of query compounds that the model can predict.

The Danish QSAR group used a cut-off value of RI greater than 0.5 to define the model's applicability domain.

5.2 Method used to assess the applicability domain

Only predictions within the applicability domain of the model, i.e. with a related RI greater than 0.5, where accepted.

5.3 Software name and version for applicability domain assessment

ACD/Labs Tox Suite version 2.95.1.

5.4 Limits of applicability

Predictions with a related RI value equal to or less than 0.5 are not within the model's applicability domain and therefore out of domain.

6. Internal validation

6.1 Availability of the training set

No

6.2 Available information for the training set

None

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

The data set of 19,571 compounds was divided into two unequal parts, a training set and a validation set (se 7.) The training set consists of 14,678 compounds, that were used in the model building process and the in the assessment of the model's applicability domain. The remaining 4893 compounds were used as a validation set (see 7.7) (Sazonovas *et al.* 2010).

6.6 Pre-processing of data before modelling

Comparisons of RTECS® LD50 data with similar data from IUCLID were done whenever available (see 3.7). The data were furthermore rewieved and 'cleaned' by removing any non-covalent complexes, salts, compounds with incorrect structures (identified automatically), and unusually high deviations in interspecies correlations. The data set was then randomly divided into a training (70%) and a validation (30%) set (the training set was only used for QSAR development and applicability domain assessment, whereas the validation set was only used for assessing the performance of the model) (Sazonovas *et al.* 2010).

6.7 Statistics for goodness-of-fit

Not performed.

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.

7. External validation

7.1 Availability of the external validation set

No

7.2 Available information for the external validation set

None

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 external validation set

The validation set consists of 4893 compounds from the data set of 19,571 compounds (of which the remaining 14,678 were used in the training set of the model (see 6.6)). The validation set was only used for estimating the model's performance (Sazonovas *et al.* 2010).

7.6 Experimental design of test set

See 3.6.

7.7 Predictivity – Statistics obtained by external validation

Results for the 69.6% (3405/4893) of the validation set compounds within the applicability domain (i.e. with an RI greater than 0.5) of the model:

RMSE = 0.35

R2 = 0.55

7.8 Predictivity – Assessment of the external validation set

The validation set was compiled and evaluated in the same way as the training set and it is therefore assessed as very useful for external validation of this model.

7.9 Comments on the external validation of the model

The external validation of the model was made by ACD/Labs, who reported statistics from external validations for different RI cut-offs (i.e. model applicability domains) - the results can be seen in the Model Performance (2014) and the paper by Sazonovas *et al*. (2010).

The Danish QSAR group uses a cut-off of RI greater than 0.5 for the applicability domain and for this reason only validation statistic results for this cut-off is shown here.

8. Mechanistic interpretation

8.1 Mechanistic basis of the model

The pre-defined complex fragmental descriptors, toxicophores, are substructures identified to be responsible for the toxic action of the molecules possessing them, and were derived from existing mechanistic knowledge (Sazonovas *et al*. 2010).

8.2 A priori or posteriori mechanistic interpretation

A priori.

8.3 Other information about the mechanistic interpretation

As mentioned under 3.3 acute toxicity is a complex endpoint involving many biological mechanisms.

9. Miscellaneous information

9.1 Comments

The model can be used to predict the LD50 value, in mg/kg, of a compound when given to mice given by the oral route. It is important to notice that extrapolation of LD50 values in animals to man is valid only to a very limited degree (OECD guideline 401 1987).

9.2 Bibliography

OECD guideline 401 (1987) OECD Guidelines for the Testing of Chemicals No. 401: Acute Oral Toxicity. Organisation for Economic Cooperation and Development; Paris, France. DELETED on the 17th of December 2002 by the OECD Council. Available online at: http://ntp.niehs.nih.gov/iccvam/docs/acutetox_docs/udpproc/udpfin01/append/appi.pdf.

9.3 Supporting information