	QMRF identifier (JRC Inventory): **Q15-42-0004**
	QMRF Title: **ACD/Percepta model for rat acute oral toxicity**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

ACD/Percepta model for rat acute oral toxicity

1.2.Other related models:

1.3.Software coding the model:

ACD labs/Percepta (Release 2014) - Acute Toxicity Prediction Module

The ACD/Labs Acute Toxicity predictor (LD50 module) provides predictions of LD50 values for rats and mice according to various routes of administration

Advanced Chemistry Development, Inc. (ACD/Labs). 8 King Street East, Suite 107, Toronto, Ontario, Canada M5C 1B5

http://www.acdlabs.com/products/percepta/predictors.php

2.General information

2.1.Date of QMRF:

July 2012

2.2.QMRF author(s) and contact details:

Simona Kovarich S-IN Soluzioni Informatiche Via Ferrari 14, I-36100 Vicenza

simona.kovarich@gmail.com http://www.s-in.it/it/

2.3.Date of QMRF update(s):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

[1]A. Sazonovas ACD/Labs, Inc.; Faculty of Chemistry, Vilnius University ACD/Labs, Inc.: A.Mickeviciaus g. 29, LT-08117 Vilnius, Lithuania; Vilnius University: Naugarduko g. 24, LT-03225 Vilnius, Lithuania

[2]P. Japertas ACD/Labs, Inc. A.Mickeviciaus g. 29, LT-08117 Vilnius, Lithuania.

[3]R. Didziapetris ACD/Labs, Inc. A.Mickeviciaus g. 29, LT-08117 Vilnius, Lithuania.

2.6.Date of model development and/or publication:

2010

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

A. Sazonovas A, P. Japertas P & and R. Didziapetris R (2012)., Estimation of reliability of predictions and model applicability domain evaluation in the analysis of acute toxicity (LD50). SAR and QSAR in Environmental Research. 2012, 21 (1),: 127–148.

2.8.Availability of information about the model:

The model is proprietary

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

None to date.

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Rat

3.2.Endpoint:

4.Human Health Effects 4.2.Acute Oral toxicity

3.3.Comment on endpoint:

The median lethal dose (LD50) indicates the dose that kills 50% of the treated animals within 24 hours of administration.

3.4.Endpoint units:

mg/kg

3.5.Dependent variable:

Prior to modeling, the original experimental data were converted to logarithmic form (log LD50) to maintain a linear relationship with the descriptors. The final prediction results returned to the user are converted back to LD50 value (mg/kg).

3.6.Experimental protocol:

LD50 values were collected mainly from the Registry of Toxic Effects of Chemical Substances (RTECS) database. This database was rigorously reviewed and 'cleaned' by removing any non-covalent complexes, salts, compounds with incorrect structures (identified automatically), and unusually high deviations in interspecies correlations. Whenever available, the acute toxicity data from the International Uniform ChemicaL Information Database (IUCLID) Chemical Data Sheets (accessible online via ESIS, the European chemical Substances Information System) were used to validate, correct or exclude entries of RTECS. The IUCLID database provided some new compounds that were not available in RTECS. The final dataset contained 8631 acute oral rat toxicity data.
Additional LD50 values for oral exposure in rats became available after the development of the model (Zhu et al., 2009) and were used for the external validation. This external test set consisted of 2718 data.

3.7.Endpoint data quality and variability:

See section 3.6

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR combined with a similarity-based approach

4.2.Explicit algorithm:

GALAS (Global, Adjusted Locally According to Similarity) modeling methodology
The GALAS model is a combination of two systems: 1) PLS model with multiple bootstrapping, using a predefined set of fragmental descriptors, for the prediction of LD50 ("Global model") and 2) local correction to baseline predictions (LD50 predicted by the Global model) based on the analysis of model performance for similar compounds from the training set.

1)Global Model(linear equation): log *LD50*= *?a ¡f ¡* + *c* + *?* where: *f ¡* = fragmental descriptors, *a ¡*= statistical coefficients of fragmental descripors, *c*= intercept, ? = unexplained variation.

2)Local Corrections (?) to the LD50 baseline prediction of a query compound

are calculated as a weighted average from the differences between global QSAR predictions and experimental data for the five most similar compounds in the training set. Additional details on the GALAS methodology (e.g. bootstrapping method and calculation of local corrections) are provided in the Supporting Information.

4.3.Descriptors in the model:

4.4.Descriptor selection:

404 fragmental descriptors were used for the development of the GALAS model. The fragmental descriptor set was identified based on general knowledge and considerations regarding all possible chemical structures and include all the fragments, even those that are not detected in the training set molecules at all. The major part of the utilized fragment set was 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. This initial set was expanded with 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. This includes, for example, phosphates, thiophosphates and carbamates (cholinesterase inhibition),
methylene fluorides (Krebs cycle inhibition), mustard derivatives, activated methylene halides, aziridinium and aziridine derivatives (alkylation of macromolecules), activated nitriles (respiratory chain inhibition), activated double bonds (alkylation through the Michael-type addition), bicyclophosphates, orthocarboxylates, and silatranes (non competitive GABA receptor inhibition).

4.5.Algorithm and descriptor generation:

No descriptor selection techniques were applied, since this is not compatible with the used approach.

4.6.Software name and version for descriptor generation:

Algorithm Builder 1.8 software (2006)
Pharma Algorithms, Inc., Toronto ON, Canada
http://www.pharma-algorithms.com

4.7.Chemicals/Descriptors ratio:

Not given.

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The local part of the model provides the basis for estimating reliability of prediction by the means of calculated Reliability Index (RI) values. RI is a number ranging from 0 to 1 (0 – unreliable prediction, 1 – idealistic, fully reliable prediction). Two criteria are applied for reliability estimation:

1) Similarity of the analyzed molecule to compounds in the Self-training

Library (a reliable prediction cannot be made if no similar compounds are found in the Library).

2) Consistency of model predictions with experimental data for similar compounds (highly variable LD50 values for similar molecules lead to lower RI values).

RI can serve as a valuable tool for interpreting prediction results. If a compound has RI lower than a certain cut-off value (here set at 0.3), it means that this compound falls outside of the Model Applicability Domain, and the respective prediction should be considered unreliable from further analysis regardless of calculated LD50 value.

5.2.Method used to assess the applicability domain:

Reliability Index (RI) is given as a product of two indices: RI = SI*DMCI 1) SI (Similarity Index) evaluates how distant the query structure is from the whole training set, and is calculated by obtaining the weighted average of all the individual Similarity Indices SI_i (i.e., calculated from the correlation of two predicted property value vectors) for the test molecule and each of the five most similar compounds from the training set: $SI = ?\,?^{i-1} * SI_i / ??^{i-1}$ 2) DMCI (Data-model consistency index) accounts for the influence of consistency of experimental data with regard to the baseline model for the five most similar compounds on the reliability of the predictions.

DMCI is calculated by comparing the differences between experimental and global model-predicted baseline values for the individual most similar compounds (?*i*) and the suggested correction value (?) for the test compound. The more individual differences are scattered around the calculated average, the more inconsistent are the data for the similar compounds with regards to the global baseline model.

5.3.Software name and version for applicability domain assessment:

5.4.Limits of applicability:

RI < 0.3: unreliable prediction

0.3<RI<0.5: borderline reliability of prediction

0.5<RI<0.75: moderate reliable prediction

RI>0.75: high reliable prediction

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

No

6.2.Available information for the training set:

CAS RN: No

Chemical Name: No

Smiles: No

Formula: No

INChI: No

MOL file: 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:

The dataset consists of 8631 compounds, which were split into a training set of 6464 compounds (only used for QSAR development and AD assessment) and a validation set of 2167 compounds (only used for verifying the validity of the results).

6.6.Pre-processing of data before modelling:

No information available

6.7.Statistics for goodness-of-fit:

No information available

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

No information available

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

No information available

6.10.Robustness - Statistics obtained by Y-scrambling:

No information available

6.11.Robustness - Statistics obtained by bootstrap:

No information available

6.12.Robustness - Statistics obtained by other methods:

No information available

7.External validation - OECD Principle 4

7.1.Availability of the external validation set:

No

7.2.Available information for the external validation set:

CAS RN: No

Chemical Name: No

Smiles: No

Formula: No

INChI: No

MOL file: 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 external validation set:

Two validation sets were used to externally validate the model: 1) A "Validation Set" of 2167 compounds. Within this set, 1976 chemicals (i.e. 91.2%) have RI>0.3, 1335 chemicals (i.e. 61.6%) have RI>0.5 and 290 chemicals (i.e. 13.4%) have RI>0.75. 2) An "External Validation Set" of 2718 compounds. Within this set, 2501 compounds have RI>0.3 (i.e. 92%), 1804 compounds have RI>0.5 (i.e. 66.4%) and 430 compounds have RI>0.75 (i.e. 15.8%). Only chemicals inside the Applicability Domain of the model (i.e. RI>0.3) were considered for the calculation of statistical

performances.

7.6.Experimental design of test set:

1) The "Validation set" was obtained after the random splitting of the dataset (see section 6.5) into training (70%) and validation (30%) sets. 2) The "External Validation Set" became available after the development of the model (Zhu et al., 2009).

7.7.Predictivity - Statistics obtained by external validation:

1) Statistics for the "Validation set": a) Test set RI>0.3 (N=1976): R2=0.56, RMSE=0.59; b) Test set RI>0.5 (N=1335): R2=0.64, RMSE=0.54; c) Test set RI>0.75 (N=290): R2=0.75, RMSE=0.43; 2) Statistics for the "External Validation Set": a) Test set RI>0.3 (N=2501): R2=0.63, RMSE=0.60, MAE=0.44; b) Test set RI>0.5 (N=1804): R^2=0.70, RMSE=0.55, MAE=0.40; c) Test set RI>0.75 (N=430): R^2=0.81, RMSE=0.44, MAE=0.30

7.8.Predictivity - Assessment of the external validation set:

7.9.Comments on the external validation of the model:

Compounds with unreliable predictions (RI<0.3) were excluded from considerations (approximately 10%), as by definition they fall outside of the model AD and hence provide no meaningful information about the models'performance [1].

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

In addition to fragmental descriptors encoding the general chemical constitution of any compound (e.g. atoms, functional groups, molecular "shape fragments", etc...), predefined fragments derived from existing mechanistic knowledge, called toxicophores (i.e. substructures identified as responsible for the toxic action of the molecules possessing them) were used for model development.

8.2.A priori or a posteriori mechanistic interpretation:

A priori (see section 8.1).

8.3.Other information about the mechanistic interpretation:

No additional information available

9.Miscellaneous information

9.1.Comments:

The ACD/Percepta Acute Toxicity predictor (rat oral) provides, in addition to the LD50 predictions, several other pieces of information including: i) possible “Oral Acute Toxicity Hazard Categories” (defined by OECD) for a compound and displays experimentally assigned categories for similar compounds. ii) a knowledge-based expert system that identifies and visualizes structural fragments potentially involved in hazardous activity, and, for each hazardous fragment, displays plots illustrating the distribution of LD50 values for compounds possessing the same fragment compared to the entire training set. iii) batch calculation.

9.2.Bibliography:

[1]Sazonovas A, Japertas P & Didziapetris R (2012). Estimation of reliability of predictions and model applicability domain evaluation in the analysis of acute toxicity (LD50). SAR and QSAR in Environmental Research 21 (1), 127–148.

[2]Zhu H, Martin TM, Ye L, Sedykh A, Young DM & Tropsha A (2009). Quantitative structure-activity relationship modeling of rat acute toxicity by oral exposure. Chemical Research in Toxicology 22, 1913–1921.

[3]ACD/Labs Percepta, Data sheet. Acute Toxicity Prediction Module.
http://www.acdlabs.com/download/docs/datasheets/datasheet_acute.pdf

[4]ACD/Labs Percepta, Model Performance. Acute Toxicity Prediction Module.
http://www.acdlabs.com/download/docs/model_performance/modelperf_acute_toxicity.pdf

9.3.Supporting information:

Training set(s)Test set(s)Supporting information

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

Q15-42-0004

10.2.Publication date:

2015-03-05

10.3.Keywords:

ACD/Percepta;acute oral toxicity;LD50;rat;

10.4.Comments:

old # Q32-48-43-425