	QMRF identifier (JRC Inventory): **Q15-410-0003**
	QMRF Title: **ACD/Percepta model for genotoxicity (Ames test)**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

ACD/Percepta model for genotoxicity (Ames test)

1.2.Other related models:

1.3.Software coding the model:

ACD labs/Percepta (2014 Release) - Genotoxicity Predictor Module

The ACD/Labs Genotoxicity predictor (Ames test module) provides predictions of mutagenic potential

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@s-in.it 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:

ACD/Labs Percepta, Advanced Chemistry Development, Inc., 2012. Toronto, On, Canada. www.acdlabs.com

2.6.Date of model development and/or publication:

2012

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

ACD labs/Percepta - Toxicity Suite - Genotoxicity Predictor Module

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

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:

Several strains of S. typhimurium (TA97, TA98, TA100, TA102, TA104, TA1535, TA1537, TA1538) and E. coli strain WP2 uvrA.

3.2.Endpoint:

4.Human Health Effects 4.10.Mutagenicity

3.3.Comment on endpoint:

The Genotoxicity Predictor Module predict the probability of a query compound to be positive in the Ames Test (i.e. mutagenic potential). Genotoxicity qualitative categories: "Ames positive/Genotoxic" (clear positive results were demonstrated in at least one tested strain, with or without metabolic activation), "Ames negative/Safe" (compounds did not increase the frequency of revertants in any of the tested strains), "Weakly positive" (chemicals that consistently exhibited weak mutagenic activity), "Inconclusive" (contradictory results from different studies).

3.4.Endpoint units:

Not applicable since qualitative endpoint

3.5.Dependent variable:

Mutagenicity as microbial *in vitro* Salmonella (composite) gene mutation assay (Ames test) is modelled for study calls, where the positive calls are trained as binary 1 and negative calls as binary 0. The output of the probabilistic QSAR model consists of: the probability that a compound will result in a positive test in this mutagenicity assay ("p-value"); an indication of whether the compound belongs to the model applicability domain according to the calculated RI value; and a "positive" or "negative" call if the compound can be reliably classified on the basis of p and RI values ("Undefined" otherwise)

3.6.Experimental protocol:

Modelling was performed using a standardized Ames genotoxicity dataset containing 8607 compounds compiled from two public databases: Chemical Carcinogenesis Research Information (CCRIS) and Genetic Toxicology Data Bank (GENE-TOX). The results of Ames genotoxicity assays were collected for several strains of S. typhimurium that are most frequently used for testing (TA97, TA98, TA100, TA102, TA104, TA1535, TA1537, TA1538) and E. coli strain WP2 uvrA, with or without metabolic activation.

Compounds in the Ames Genotoxicity database were classified as "Ames positive/Genotoxic", if clear positive results were demonstrated in at least one tested strain, with or without metabolic activation. Compounds that did not increase the frequency of revertants in any of the tested strains were considered safe ("Ames negative"). Some chemicals that consistently exhibited weak mutagenic activity were marked weakly positive while in cases where the results of different studies were contradictory, the corresponding compounds were labelled inconclusive.

3.7.Endpoint data quality and variability:

See section 3.6

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

Classification QSAR combined with an Expert System

4.2.Explicit algorithm:

Probabilistic Model (Binomial PLS model)

The probabilistic model for the prediction of mutagenic potential consists of two parts: 1) Global baseline statistical model employing binomial PLS with multiple bootstrapping, using a predefined

set of fragmental descriptors, and 2) Local correction to baseline predictions based on the analysis of experimental data for similar compounds.The probabilistic model is combined with a knowledge-based expert system (Genotoxicity Hazard Module) that identifies hazardous structural fragments that are known to be hazardous substructures potentially involved in genotoxic activity (27 predefined genotoxicophores).

4.3.Descriptors in the model:

404 fragmental descriptors

4.4.Descriptor selection:

404 fragmental descriptors were used for the development of the GALAS (Global, Adjusted Locally According to Similarity) 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, 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 Michael-type addition), bicyclophosphates, orthocarboxylates, and silatranes (non competitive GABA receptor inhibition). No descriptor selection techniques based on a defined training set were applied, since this is not compatible with the used approach.

4.5.Algorithm and descriptor generation:

No information available (proprietary model)

4.6.Software name and version for descriptor generation:

Algorithm Builder 1.8 software (2006)
(Software used for model development.)
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 confidence of predictions is evaluated via a Reliability Index (RI) calculated for each prediction. The RI is a value ranging from 0 and 1 (0 – unreliable prediction, 1 – idealistic, fully reliable prediction) and is an indicator of how well a particular compound is represented within the training set of the model. Two criteria are applied for

reliability estimation:
1) Similarity of the analyzed molecular structure to compounds in the Self-training Library (prediction is considered unreliable if no similar compounds are found in the training set).
2) Consistency of experimental data for similar compounds (inconsistent data for similar molecules lead to lower RI values).
RI can serve as a valuable tool for interpreting prediction results. If a compound obtains RI lower than a certain cut-off value (here set at 0.3), it means that this compound falls outside the applicability domain of the model and the respective prediction may be less accurate (inconclusive). Any prediction providing an RI value below 0.3 should be considered unreliable.

5.2.Method used to assess the applicability domain:

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 8607 compounds, which were split into a training set of 6895 compounds (only used for QSAR development and AD assessment) and a validation set of 1712 compounds (only used for verifying the validity of the results).

6.6.Pre-processing of data before modelling:

A standardized Ames genotoxicity data set containing more than 8500 compounds was compiled from public databases (e.g., Chemical Carcinogenesis Research Information (CCRIS), Genetic Toxicology Data Bank (GENE-TOX)). The results of Ames genotoxicity assays were collected for several strains of S. typhimurium that are most frequently used for testing (TA97, TA98, TA100, TA102, TA104, TA1535, TA1537, TA1538 and

also E. coli strain WP2 uvrA), with or without metabolic activation.

Compounds were classified as “Ames positive” if it demonstrated clear positive results in at least one tested strain with or without metabolic activation. Compounds that did not increase the frequency of revertants in all tested strains were considered safe (“Ames negative”). Some chemicals that consistently exhibited weak mutagenic activity were marked “Weakly positive” while in those cases when the results of different studies were contradictory the corresponding compounds were labeled “Inconclusive”.

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:

The validation sets consists of 1712 compounds. Within this set, 1483 chemicals (i.e. 86.6%) have RI>0.3 and 1117 chemicals (i.e. 65.2%) have RI>0.5. 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 splitting of the dataset (see section 6.5) into training (80%) and validation (20%) sets.

7.7.Predictivity - Statistics obtained by external validation:

1) RI>0.3 (N=1483): Accuracy = 89.0% (1320/1483 compounds correctly classified), Sensitivity = 93.3% (928/995 positive compounds correctly classified); Specificity = 80.3% (392/488 negative compounds correctly classified). 2) RI>0.5 (N=1117): Accuracy = 93.4% (1043/1117 compounds correctly classified), Sensitivity = 97.2% (786/809 positive compounds correctly classified); Specificity = 83.4% (257/308 negative compounds correctly classified). Additional information on the external validation of the model is provided as Supporting Information.

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 13%), as by definition they fall outside of the model AD and hence provide no meaningful information about the model performance.

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

Predictions obtained by the probabilistic model are combined with and supported by the Genotoxicity Hazard module, which is a knowledge-based expert system that identifies structural fragments that may be responsible for the mutagenic activity of the analyzed molecules. The Genotoxicity Hazard system searches through a list of 27 predefined genotoxicophores derived from existing mechanistic knowledge (fragments collected from toxicological literature). Most genotoxicophores on the list are not defined by a single hazardous fragment, but represent groups of similar substructures sharing common properties with respect to causing DNA damage. Each identified hazard is accompanied by information describing its mode of action.

8.2.A priori or a posteriori mechanistic interpretation:

A priori (see section 8.1).

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

Additional features provided by the ACD/Percepta Genotoxicity Predictor module: i) genotoxic potential of different parts of the molecule are visualized by color-mapping the contributions of different atoms (or fragments) onto the structure (red: associated with genotoxicity; green: not associated with the genotoxic effect); ii) experimental Ames test results for up to 5 similar structures, taken from the training set, are displayed with each prediction; iii) the Genotoxic Hazard module provides experimental data for 5 similar compounds that posses the same hazardous substructure as the analyzed compound. Bar charts showing the distribution of experimental data in various bacterial srains are also presented; iv) a searchable database of over 5500 experimental Ames test data provides information about individual studies conducted with each compound

corresponding to various bacterial strains tested, the presence or absence of metabolic activation, and other experimental conditions; v) model predictions are trainable, i.e. the accuracy and reliability of predictions can be improved using experimental data provided by the user; vi) batch calculation.

9.2.Bibliography:

[1]ACD/Labs Percepta, Data sheet. Genotoxicity Predictor Module. http://www.acdlabs.com/download/docs/datasheets/datasheet_genotox.pdf

[2]ACD/Labs Percepta, Model Performance. Genotoxicity Predictor Module. http://www.acdlabs.com/download/docs/model_performance/modelperf_genotox.pdf

9.3.Supporting information:

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

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

Q15-410-0003

10.2.Publication date:

2015-03-05

10.3.Keywords:

ACD/Percepta;genotoxicity;mutagenicity;Ames;S. Typhimurium;E. coli;

10.4.Comments:

old # Q31-47-42-424