(Q)SAR Model Reporting Format (QMRF)
(version 2.1)

The adequacy of a prediction depends on the following conditions: a) **the (Q)SAR model is scientifically valid**: the scientific validity is established according to the OECD principles for (Q)SAR validation; b) **the (Q)SAR model is applicable to the query chemical**: a (Q)SAR is applicable if the query chemical falls within the defined applicability domain of the model; c) **the (Q)SAR result is reliable**: a valid (Q)SAR that is applied to a chemical falling within its applicability domain provides a reliable result; d) **the (Q)SAR model is relevant for the regulatory purpose**: the predicted endpoint can be used directly or following an extrapolation, possibly in combination with other information, for a particular regulatory purpose.

1. QSAR identifier

1.1 QSAR identifier (title):

ACD/Ames Mutagenicity: QSAR for genotoxicity in Salmonella typhimurium

1.2 Other related models:

ACD/Labs Percepta Impurity Profiling Suite, including probabilistic models for 21 different endpoints related to:

- Genetic toxicity: Mutagenicity (Ames test, Mouse Lymphoma Assay, CHO/CHL all loci composite, and other standard assays), Clastogenicity (Micronucleus test, Chromosomal Aberrations), DNA damage (Unscheduled DNA Synthesis)
- Carcinogenicity (rodent carcinogenicity)
- Reproductive toxicity: Endocrine disruption mechanisms (estrogen receptor binding)

1.3 Software coding the model:

ACD/Labs Percepta v. 2025.1

https://www.acdlabs.com/products/percepta-platform/

2. General information

2.1 Date of QMRF:

26 March 2018

2.2 QMRF author(s) and contact details:

Simona Kovarich[1], Kiril Lanevskij[2,*], Andrius Sazonovas[2,**].

[1] S-IN Soluzioni Informatiche Srl Via G. Ferrari, 14, I-36100 Vicenza, Italy.

[2] Advanced Chemistry Development, Inc. 8 King Street East, Suite 107, Toronto, Ontario, Canada, M5C 1B5.

[*] Contact email: kiril.lanevskij@acdlabs.com

[**] Contact email: andrius.sazonovas@acdlabs.com

2.3 Date of QMRF update(s):

1) 14 August 2024
2) 11 June 2025

2.4 QMRF update(s):

1) Authors of the update: Kiril Lanevskij, Andrius Sazonovas (see section 2.2). Content updates: extensive editing of the entire document, reformatting to a new template, updated metadata. Rewritten sections 4.1-4.7, 5.1-5.2, 6.6, 6.12, 7.9.
2) Authors of the update: Kiril Lanevskij, Andrius Sazonovas (see section 2.2). Updated sections: 1.3, 2.2-2.6, 5.3, 7.9.

2.5 Model developer(s) and contact details:

Advanced Chemistry Development, Inc. 8 King Street East, Suite 107, Toronto, Ontario, Canada, M5C 1B5; info@acdlabs.com

2.6 Date of model development and/or publication:

2011-2025

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

[1] Sazonovas A, Japertas P, Didziapetris R. Estimation of reliability of predictions and model applicability domain evaluation in the analysis of acute toxicity (LD50). SAR QSAR Environ Res. 2010 Jan 1;21(1):127-48.

http://www.ncbi.nlm.nih.gov/pubmed/20373217

[2] Japertas P, Lanevskij K, Juska L, Dapkunas J, Didziapetris R. A comprehensive approach for in silico risk assessment of impurities and degradants in drug products. Toxicol Lett. 2011, 205, S95.

[3] Lanevskij K, Juska L, Dapkunas J, Sazonovas A, Japertas P, Didziapetris R. An In Silico Test Battery for Rapid Evaluation of Genotoxic and Carcinogenic Potential of Chemicals. Poster (Mar 25, 2012, ACS Spring).

2.8 Availability of information about the model:

The model is proprietary, however the algorithm was published in [1], while the details relevant to predicting genotoxicity were provided in [2-3]. The compounds used to derive the model and their experimental data are available within the corresponding software product

2.9 Availability of another QMRF for exactly the same model:

None to date

3. Defining the endpoint - OECD Principle 1: "A DEFINED ENDPOINT"

3.1 Species:

Salmonella typhimurium

3.2 Endpoint:

Mutagenicity: OECD 471 Bacterial Reverse Mutation Test

3.3 Comment on endpoint:

Mutagenicity assessment based on bacterial reverse mutation test using different strains of Salmonella typhimurium.

3.4 Endpoint units:

Dimensionless

3.5 Dependent variable:

Mutagenicity as microbial in vitro Salmonella (composite) gene mutation assay 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 the respective 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:

Experimental dataset was obtained from FDA. Data was collected from EPA GENE-TOX database and scientific literature [4]. For modeling purposes experimental results of microbial in vitro Salmonella (composite) Assay have been transformed into a binary variable, i.e. positive/negative.

3.7 Endpoint data quality and variability:

No information available

4. Defining the algorithm - OECD Principle 2: "AN UNAMBIGUOUS ALGORITHM"

4.1 Type of model

Hybrid QSAR, combining a linear baseline model utilizing PLS method and the local similarity-based corrections

4.2 Explicit algorithm

Probabilistic GALAS algorithm (Global linear baseline QSAR + local similarity-based corrections)

GALAS (Global, Adjusted Locally According to Similarity) models consist of two parts: (1) a global linear model, and (2) local corrections based on the analysis of global model performance for the most similar compounds from the training set. Experimental values for microbial in vitro Salmonella assay are used during the local part of the modeling to yield final GALAS model. The global QSAR was developed using binomial PLS in combination with bootstrapping technique. This method implies random compound sampling from the initial training set, i.e. generation of new "training sub-sets". Each of the sampled sub-sets is of the same size as the initial training set, however, random manner of their population results in some compounds being selected more than once, others being omitted. This procedure is performed 100 times and an independent PLS model is derived for every sub-set. Each of those PLS models is based on 2D fragmental descriptors:

$$\text{logit}(p) = \Sigma_i a_i \cdot f_i + \text{c}$$

where f_i is the number of occurrences of the i-th fragment in a molecule, a_i – its statistical coefficient, and c – the intercept.

As a result, each global QSAR model actually represents an ensemble of 100 PLS models, providing each compound with a vector of 100 logit-transformed probability predictions ($\ln(p/(1-p))$), each based on a slightly different sub-set of the initial training set. It is defined that two compounds with similar trends in the variation patterns of the 100 value vectors predicted by a global QSAR model are considered similar in terms of the analyzed property, i.e. the differences in the compound sets used to parameterize each of 100 PLS models, constituting a baseline model, affect estimations for the two compounds in a similar way. The correlation coefficient of the two vectors is called an Individual Similarity Index between two compounds (SI_i). An analogous definition of the "property-specific" or dynamic similarity was first used by Tetko and his co-workers [5-9] and this method has been recently used in the analysis of the acute toxicity data [10]. With the available robust similarity measure, it becomes

possible to analyze the performance of the baseline QSAR model in the local chemical environment of a query molecule represented by the most similar compounds in the training set. In case any systematic errors are encountered for sufficiently similar compounds, a local correction (Δ) is calculated. Later on it is possible to train the model quickly and efficiently using new experimental data by just adding it to this second similarity correction calculation procedure, without the time costly baseline model re-training.

4.3 Descriptors in the model

Fragmental descriptors (occurrence counts) – a fixed set of fragmental descriptors, based on the expanded list of Platt's type fragments (see [11]). The fragments included in this list reflect (i) molecular size and topology (linear and branched atom chains, cycles of different sizes, etc.), (ii) aromaticity, (iii) presence of various functional groups, (iv) intramolecular interactions, (v) substructures typical for specific drug classes. 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. Overall, 404 fragmental descriptors were used for the development of the baseline model.

4.4 Descriptor selection

No special descriptor selection techniques had been used to reduce the initial descriptor pool (e.g., excluding statistically insignificant or intercorrelated variables) due to the specifics of employed modeling methodology. Any potential negative influence of insignificant fragments would be remedied by the use of PLS method, but their presence is necessary for providing the so called "dynamic similarity" measure between the molecules. For this purpose, even “blank” fragments (with zero occurrence count) should remain, as these would allow detecting new structural features of a query molecule that were not present in the training set, and would thus decrease its similarity coefficient to training set molecules.

4.5 Algorithm and descriptor generation

The generation of the descriptor matrix following the outlined approach constituted counting the occurrences of any of the pre-defined fragments in the training set molecules. This procedure as well as all the subsequent statistical analysis were performed using Algorithm Builder software.

4.6 Software name and version for descriptor generation

ACD/Labs Algorithm Builder 1.8

ACD/Labs, Inc. 8 King Street East, Suite 107, Toronto, Ontario, M5C 1B5 Canada.

info@acdlabs.com

4.7 Chemicals/Descriptors ratio

Not relevant as PLS method operates on a small number of latent principal components rather than raw descriptors

5. Defining the applicability domain - OECD Principle 3: “A DEFINED DOMAIN OF APPLICABILITY”

5.1 Description of the applicability domain of the model

Applicability domain of the model is defined based on the training set compounds. This procedure takes into account the following two aspects:

- Similarity of the tested compound to the training set. No reliable predictions can be made if there are no similar compounds in the training set

- Consistence of the experimental values with regard to the baseline model for similar compounds. Even if there are similar compounds in the dataset, the quality of prediction could be lower if that data cannot be reproduced by the baseline model. It does not matter what the reason for this inconsistency – experimental variability or sudden change in mechanism of action because of slight structural changes – in any case it indicates possible problems when trying to give accurate predictions.

5.2 Method used to assess the applicability domain:

The two aspects mentioned in Section 5.1 receive their quantitative assessment in terms of Similarity Index (*SI*) and Data-Model Consistency Index (*DMCI*). The SI, evaluating how distant the query structure is from the whole training set, is calculated by weighted averaging of all the individual Similarity Indices (SI_i) for the test molecule and each of the 5 most similar compounds from the training set. *DMCI* is calculated by comparing the differences between experimental and global QSAR predicted values for the 5 most similar compounds and the suggested similarity correction value (Δ) for the test compound, calculated by averaging these differences. The more individual differences are scattered around the calculated average (Δ), the more inconsistent are the data for the similar compounds with regards to the global QSAR. The final prediction Reliability Index is calculated as a product of the aforementioned two indices:

$$RI = SI \cdot DMCI$$

Both *SI* and *DMCI* are scaled to vary from 0 to 1, so the resulting *RI* also varies in this range. Lower values suggest a compound being further from the Model Applicability Domain and the prediction less reliable (low *SI* or low *DMCI* either alone or in combination can be the reason). On the other hand, high *RI* values indicate an increasing confidence about the quality of the prediction (both *SI* and *DMCI* must be high to yield such a result).

5.3 Software name and version for applicability domain assessment:

ACD/Labs Percepta v. 2025.1

5.4 Limits of applicability:

For the purpose of applicability domain assessment, Reliability Index (*RI*) values of predictions are categorized as follows:

- $RI < 0.3$: unreliable predictions
- $0.3 \leq RI < 0.5$: borderline reliability of predictions
- $0.5 \leq RI < 0.75$: moderate reliability of predictions
- $RI \geq 0.75$: high reliability of predictions

6. Defining goodness-of-fit and robustness (internal validation) – OECD Principle 4: "APPROPRIATE MEASURES OF GOODNESS-OF-FIT, ROBUSTENESS AND PREDICTIVITY"

6.1 Availability of the training set

The training data is available through ACD/Labs Percepta software (Genotoxicity module).

6.2 Available information for the training set

The training set is not attached to the QMRF itself

6.3 Data for each descriptor variable for the training set

Not available

6.4 Data for the dependent variable for the training set

Not available

6.5 Other information about the training set

The entire dataset used in model development and validation consists of 7826 compounds, including 3875 positive compounds (i.e. 49.5%).

6.6 Pre-processing of data before modelling

Inorganic compounds have been excluded. All provided statistical information corresponds to the final filtered data set

6.7 Statistics for goodness-of-fit

Sensitivity = 85.6%; Specificity = 81.5%; Concordance = 83.6%

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

Not available

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

Not available

6.10 Robustness - Statistics obtained by Y-scrambling

Not available

6.11 Robustness - Statistics obtained by bootstrap

Not available

6.12 Robustness - Statistics obtained by other methods

Cross-validation or any other robustness determination procedures were not applied to the reported model since its performance was evaluated using an external validation set. External predictivity always prevails over any internal validation results, whereas consistency between prediction statistics obtained on training and test sets provides sufficient evidence for the robust performance of the model.

7. Defining predictivity (external validation) – OECD Principle 4: "APPROPRIATE MEASURES OF GOODNESS-OF-FIT, ROBUSTENESS AND PREDICTIVITY"

7.1 Availability of the external validation set

The external validation data is available through ACD/Labs Percepta software (Genotoxicity module).

7.2 Available information for the external validation set

The external validation set is not attached to the QMRF itself

7.3 Data for each descriptor variable for the external validation set

Not available

7.4 Data for the dependent variable for the external validation set

Not available

7.5 Other information about the external validation set

The part of the dataset used for model validation consists of 1577 compounds, including 794 positive compounds (i.e. 50.3%).

7.6 Experimental design of test set

Random splitting of the initial dataset into the training and validation sets at ~80%:20% ratio.

7.7 Predictivity - Statistics obtained by external validation

Sensitivity = 87.1%; Specificity = 81.7%; Concordance = 84.6

Only chemicals inside the Applicability Domain (i.e., $RI \geq 0.3$) were considered for the calculation of statistical performances (1332 compounds, i.e., 84.5% of the entire test set).

7.8 Predictivity - Assessment of the external validation set

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

7.9 Comments on the external validation of the model

The training and validation statistics reported in Sections 6-7 apply to the original model developed using the data set characterized in Section 6.5. The current version of the software uses the same baseline statistical model operating on a significantly expanded built-in self-training library for similarity corrections ("AMES Test v. 1.5", containing 10967 molecules).

8. Providing a mechanistic interpretation - OECD Principle 5: "A MECHANISTIC INTERPRETATION, IF POSSIBLE"

8.1 Mechanistic basis of the model

The model is based on both fragmental structural descriptors and toxicophores, i.e. substructures identified to be responsible for the toxic action of the molecules possessing them. To enhance a mechanistic understanding, predictions obtained by the probabilistic model can be combined with and supported by the Genotoxicity Hazard System, which is a knowledge-based expert system that identifies structural fragments that may be responsible for the mutagenic activity of the analyzed molecules.

8.2 A priori or a posteriori mechanistic interpretation

A priori (see section 8.1)

8.3 Other information about the mechanistic interpretation

The software displays up to 5 most similar structures (included in the training set of the model) to the analyzed molecule with experimental results (positive/negative). The analysis of similar structures provides additional information to gain insight into the possible mechanisms of action and support the in silico prediction for the query compound.

9. Miscellaneous information

9.1 Comments

ACD/Labs Impurity Profiling Suite provides a battery of in silico tests to accurately assess the genotoxic and carcinogenic potential of impurities and degradants. The impurities package offers probabilistic predictive models for 21 different endpoints that cover various mechanisms of hazardous activity (including Mutagenicity, Clastogenicity, DNA damage mechanisms, Carcinogenicity and Endocrine Disruption mechanisms). These predictors are supplemented with a knowledge-based expert system that identifies potentially hazardous structural fragments that could be responsible for genotoxic and/or carcinogenic activity of the compound of interest. The expert system was able to recognize >94% of mutagens in ACD/Ames test database, and >90% of compounds marked as potent carcinogens in the FDA's OFAS Food-Additive Knowledgebase [3].

9.2 Bibliography

[1] Sazonovas A, Japertas P, Didziapetris R. Estimation of reliability of predictions and model applicability domain evaluation in the analysis of acute toxicity (LD50). SAR QSAR Environ Res. 2010 Jan 1;21(1):127-48.

[2] Japertas P, Lanevskij K, Juska L, Dapkunas J, Didziapetris R. A comprehensive approach for in silico risk assessment of impurities and degradants in drug products. Toxicol Lett. 2011, 205, S95.

[3] Lanevskij K, Juska L, Dapkunas J, Sazonovas A, Japertas P, Didziapetris R. An In Silico Test Battery for Rapid Evaluation of Genotoxic and Carcinogenic Potential of Chemicals. Poster (Mar 25, 2012, ACS Spring).

[4] Matthews EJ, Kruhlak NL, Benz RD, Contrera JF. A comprehensive model for reproductive and developmental toxicity hazard identification: I. Development of a weight of evidence QSAR database. Regul Toxicol Pharmacol. 2007, 47, 115.

[5] Zhu H, Tropsha A, Fourches D, Varnek A, Papa E, Gramatica P, Oberg T, Dao P, Cherkasov A, Tetko IV. Combinatorial QSAR modeling of chemical toxicants tested against Tetrahymena pyriformis. *J. Chem. Inf. Model.* 2008, 48, 4, 766–784.

[6] Tetko IV, Sushko I, Pandey AK, Zhu H, Tropsha A, Papa E, Oberg T, Todeschini R, Fourches D, Varnek A. Critical assessment of QSAR models of environmental toxicity against Tetrahymena pyriformis: focusing on applicability domain and overfitting by variable selection. *J. Chem. Inf. Model.* 2008, 48, 9, 1733–1746.

[7] Tetko IV, Neural network studies. 4. Introduction to associative neural networks, *J. Chem. Inf. Comput. Sci.* 2002, 42, 717-728.

[8] Tetko IV, Bruneau P, Application of ALOGPS to predict 1-octanol/water distribution coefficients, logP, and logD, of AstraZeneca in-house database. *J. Pharm. Sci.* 2004, 93, 3103-3110.

[9] Tetko IV, Tanchuk VY. Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. *J. Chem. Inf. Comput. Sci.* 2002, 42, 1136-1145.

[10] Sazonovas A, Japertas P, Didziapetris, R. Estimation of reliability of predictions and model applicability domain evaluation in the analysis of acute toxicity (LD50). *SAR QSAR Environ. Res.* 2010, 21, 127-148.

[11] Platts JA, Butina D, Abraham MH, Hersey A. Estimation of molecular linear free energy relation descriptors using a group contribution approach. *J. Chem. Inf. Comput. Sci.* 1999, 39, 835-845.