	QMRF identifier (JRC Inventory): **Q15-412-0002**
	QMRF Title: **Lazar model for rodent carcinogenicity**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

Lazar model for rodent carcinogenicity

1.2.Other related models:

lazar models for

- salmonella mutagenicity
- human licer toxicity
- fathead minnow toxicity
- human maximum recommended therapeutic dose

1.3.Software coding the model:

lazar

webinterface for the lazar program (with source code download)

helma@in-silico.ch

lazar.in-silico.ch

2.General information

2.1.Date of QMRF:

20 August 2008

2.2.QMRF author(s) and contact details:

Christoph Helma in silico toxicology gmbh Rastatterstr. 41, CH-4057 Basel helma@in-silico.ch www.in-silico.ch

2.3.Date of QMRF update(s):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

Christoph Helma in silico toxicology gmbh Rastatterstr. 41, CH-4057 Basel helma@in-silico.ch www.in-silico.ch

2.6.Date of model development and/or publication:

2008

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

[1]Helma C (2006) Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and Salmonella mutagenicity. Molecular Diversity, 10,147-158 http://link.springer.com/article/10.1007/s11030-005-9001-5#page-1

[2]Maunz A and Helma C (2008) Prediction of chemical toxicity with local support vector regression and activity-specific kernels. SAR & QSAR in Environmental Research 19 (5-6), 413-431 http://www.tandfonline.com/doi/abs/10.1080/10629360802358430

2.8.Availability of information about the model:

Algorithms have been published, recent modifications are on the lazar website

Source code is available under the GNU License

Data is public

Up to date validation results for all endpoints are on the lazar website

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

lazar models for carcinogenicity and mutagenicity have been submitted with a previous version of the QMRF (Excel based)

3.Defining the endpoint - OECD Principle 1

3.1.Species:

rat

mouse

hamster

3.2.Endpoint:

4.Human Health Effects 4.12.Carcinogenicity

3.3.Comment on endpoint:

sex/species specific models for rat, mouse, hamster

composite models for multiple sex/species/sites and single sex/species/sites

3.4.Endpoint units:

classification carcinogenic/non-carcinogenic

3.5.Dependent variable:

multiple sex/species/sites carcinogenicity: An assignment of carcinogenic categorical activity based on multicell evidence for or against activity:

active: with more than one TD50 or tumor site listed for carcinogenicity experiment sex/species cells (e.g., liver, lung, or Rat Male, Rat Female, etc);

inactive: with no TD50 or tumor site listed AND more than one "no positive results" entry for carcinogenicity experiment sex/species cells (e.g., Rat Male, Rat Female, etc)

single sex/species/sites: An assignment of carcinogenic categorical activity based on minimal evidence for or against activity:

active: with one or more TD50 and tumor site listed for one or more carcinogenicity experiment sex/species cell (e.g., Rat Male, Rat Female, etc);

inactive: with no TD50 or tumor site listed AND one or more "no positive results" entry for one or more carcinogenicity experiment sex/species cell (e.g., Rat Male, Rat Female, etc)

sex/species specific carcinogenicity: Active if at least one target site has been reported, inactive if no positive results have been reported

3.6.Experimental protocol:

Rodent carcinogenicity bioassays from the NTP and the general literature, that meet the inclusion critrea of the Carcingenic Potency Database (details at http://potency.berkeley.edu/methods.html#sources and http://www.epa.gov/nheerl/dsstox/)

3.7.Endpoint data quality and variability:

unclear (lack of repeated experiments and ring trials)

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

modified k-nearest neighbor classification

4.2.Explicit algorithm:

see Helma 2006 and lazar.in-silico.ch, submitted version: https://github.com/helma/lazar-core/commit/7edd0ddb0135f99671b0d7a3e8375cf93ece9860

modified k-nearest neighbor classification with activity specific similarities, weighted voting and exhaustive enumeration of fragments and neighbors

4.3.Descriptors in the model:

linear fragmens (paths) true/false (i.e. present/absent) all statistically relevant paths are used for similarity calculation

4.4.Descriptor selection:

statistical filter (chi-square with Yates correction)

4.5.Algorithm and descriptor generation:

exhaustive breadth first search for paths in chemical graphs (simplified MolFea algorithm)

4.6.Software name and version for descriptor generation:

lazar, submitted version: https://github.com/helma/lazar-core/commit/7edd0ddb0135f99671b0d7a3e8375cf93ece9860

simplified MolFea algorithm

helma@in-silico.ch

lazar.in-silico.ch

4.7.Chemicals/Descriptors ratio:

not applicable (classification based on activities of neighbors, descriptors are used for similarity calculation)

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The applicability domain (AD) of the training set is characterized by the confidence index of a prediction (high confidence index: close to the applicability domain of the training set/reliable prediction, low confidence: far from the applicability domain of the trainingset/unreliable prediction). The confidence index considers (i) the similarity and number of neighbors and (ii) contradictory examples within the neighbors. A formal definition can be found in Helma 2006. The reliability of predictions decreases gradually with increasing distance from the applicability domain (i.e. decreasing confidence index)

5.2.Method used to assess the applicability domain:

see Helma 2006 and Maunz 2008

5.3.Software name and version for applicability domain assessment:

lazar, submitted version: https://github.com/helma/lazar-core/commit/7edd0ddb0135f99671b0d7a3e8375cf93ece9860

integrated into main lazar algorithm

helma@in-silico.ch

lazar.in-silico.ch

5.4.Limits of applicability:

Predictions with low confidence index, unknown substructures and neighbors that might act by different mechanisms

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS RN: Yes

Chemical Name: Yes

Smiles: Yes

Formula: Yes

INChI: Yes

MOL file: Yes

6.3.Data for each descriptor variable for the training set:

No

6.4.Data for the dependent variable for the training set:

All

6.5.Other information about the training set:

DSSTox CPDBAS Version v5c 29 April 2008

6.6.Pre-processing of data before modelling:

none

6.7.Statistics for goodness-of-fit:

not applicable

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

see detailed validation results for each individual endpoint in the attachment

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

not available

6.10.Robustness - Statistics obtained by Y-scrambling:

presently not available, can be delivered on demand

6.11.Robustness - Statistics obtained by bootstrap:

not available

6.12.Robustness - Statistics obtained by other methods:

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: Yes

Chemical Name: Yes

Smiles: Yes

Formula: Yes

INChI: Yes

MOL file: Yes

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:

7.6.Experimental design of test set:

7.7.Predictivity - Statistics obtained by external validation:

7.8.Predictivity - Assessment of the external validation set:

7.9.Comments on the external validation of the model:

No external testsets available, an experimental comparison of leave-one-out vs. external testset validation for lazar mutagenicity models can be found in Benigni 2007 (see attachment). This experiment (and theoretical considerations) indicate that corssvalidation results are a good indicator for real world performance if (i) all test set information has been excluded from the training set and (ii) if the applicability domain is considered.

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

Compounds with similar structures (neighbors) are assumed to have similar activities as the query compound. For the determination of activity specific similarities only statistically relevant subtructures (paths) are used. For this reason there is a priori no bias towards specific mechanistic hypothesis.

8.2.A priori or a posteriori mechanistic interpretation:

a posteriori for individual predictions

8.3.Other information about the mechanistic interpretation:

Hypothesis about biochemical mechanisms can be derived from individual predictions by inspecting neighbors and relevant fragments.
Neighbors are compounds that are similar in respect to a certain endpoint and it is likely that compounds with high similarity act by similar mechanisms as the query compound. Links at the webinterface prove an easy access to additional experimental data and literature citations for the neighbors and the query structure.
Activating and deactivating parts of the query compound are highlighted in red and green on the webinterface. Fragments that are unknown (or too infrequent for statistical evaluation are marked in yellow and additional statistical information about the individual fragments can be retrieved. Please note that lazar predictions are based on neighbors and not on fragments. Fragments and their statistical significance are used for the calculation of activity specific similarities.

9.Miscellaneous information

9.1.Comments:

Please send any open questions, additional requirements etc. to helma@in-silico.ch.

9.2.Bibliography:

[1]Helma C (2006) Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and Salmonella mutagenicity. Molecular Diversity, 10,147-158 http://link.springer.com/article/10.1007/s11030-005-9001-5#page-1

[2]Benigni R, Netzeva TI , Benfenati E, Bossa C and Franke R, Helma C, Hulzebos E, Marchant C, Richard A, Woo Y-T and Yang C (2007) The expanding role of predictive toxicology: an update on the (Q)SAR models for mutagens and carcinogens. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev. 25,53-97 http://www.tandfonline.com/doi/abs/10.1080/10590500701201828

[3]Maunz A and Helma C (2008) Prediction of chemical toxicity with local support vector regression and activity-specific kernels. SAR & QSAR in Environmental Research 19 (5-6), 413-431 http://www.tandfonline.com/doi/abs/10.1080/10629360802358430

9.3.Supporting information:

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

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

Q15-412-0002

10.2.Publication date:

2015-03-05

10.3.Keywords:

Lazar;rat;mouse;carcinogenicity;

10.4.Comments:

old # Q28-43-38-420