	QMRF identifier (JRC Inventory): **Q13-410-0042**
	QMRF Title: **Derek for Windows - Mutagenicity**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

Derek for Windows - Mutagenicity

1.2.Other related models:

1.3.Software coding the model:

Derek for Windows version 13

www.lhasalimited.org/derek

2.General information

2.1.Date of QMRF:

9 June 2009

2.2.QMRF author(s) and contact details:

Kate Langton Lhasa Limited 22-23 Blenheim Terrace, Woodhouse Lane, Leeds, LS2 9HD, UK

kate.langton@lhasalimited.org www.lhasalimited.org

2.3.Date of QMRF update(s):

21st February 2011

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

Lhasa Limited 22-23 Blenheim Terrace, Woodhouse Lane, LS2 9HD, UK

kate.langton@lhasalimited.org www.lhasalimited.org

2.6.Date of model development and/or publication:

Derek for Windows version 13 was released in December 2010 and included updates to the mutagenicity endpoint.

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

[1]Sanderson M & Earnshaw CG (1991). Computer Prediction of Possible Toxic Action from Chemical Structure; The DEREK System. Human and Experimental Toxicology 10, 261-273. http://het.sagepub.com/cgi/content/abstract/10/4/261

[2]Judson PN, Marchant CA & Vessey JD (2003). Using Argumentation for Absolute Reasoning about the Potential Toxicity of Chemicals. Journal of Chemical Information and Computer Sciences 43, 1364-1370. http://pubs.acs.org/doi/abs/10.1021/ci020272g

2.8.Availability of information about the model:

The alerts are available for inspection within the software and representative examples are provided to illustrate a given alert if available. The training set underpinning a given alert is proprietary, though generally based on publicly available data.

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

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Bacterium (primarily Salmonella typhimurium)

3.2.Endpoint:

4.Human Health Effects 4.10.Mutagenicity

3.3.Comment on endpoint:

The model is primarily based on data from Ames test and predicts for the mutagenicity endpoint. Additional data from in vivo lacZ-transgenic assay, in vitro L5178Y +/- assay, in vitro HGPRT gene mutation assay, in vitro Na+/K+ ATPase gene mutation assay has also been considered for the development of a small number of alerts.

3.4.Endpoint units:

Reversion count in Ames test is used to assign activity for mutagenicity

3.5.Dependent variable:

Not applicable

3.6.Experimental protocol:

The model is based primarily on data from the standard strains in the Ames test conducted following standard test protocol. If activity is observed in a non-standard strain this will be mentioned in the comments.

3.7.Endpoint data quality and variability:

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

Expert system

4.2.Explicit algorithm:

Expert system

Expert system based on multiple structure alerts (2D SARs)

None - alert based expert system

4.3.Descriptors in the model:

4.4.Descriptor selection:

4.5.Algorithm and descriptor generation:

4.6.Software name and version for descriptor generation:

4.7.Chemicals/Descriptors ratio:

This is not applicable as the structural alerts are knowledge-based rather than statistically based.

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The scope of the structure-activity relationships describing the mutagenicity endpoint are defined by the developer to be the applicability domain for the model. Therefore, if a chemical matches an alert describing a structure-activity for mutagenicity it can be considered to be within the applicability domain.

5.2.Method used to assess the applicability domain:

The applicability domain of each alert is defined by the alert developer on the basis of the training set data and expert judgement on the chemical and biological factors which affect the mechanism of action for each alert.

5.3.Software name and version for applicability domain assessment:

5.4.Limits of applicability:

Limits for individual alerts are mainly defined by restructions in the scope of the alerts which are available for inspection within the software.

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:

No internal validation has been performed

6.6.Pre-processing of data before modelling:

6.7.Statistics for goodness-of-fit:

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

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

6.10.Robustness - Statistics obtained by Y-scrambling:

6.11.Robustness - Statistics obtained by bootstrap:

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: 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:

Multiple external validation sets exists. Some are publicly available but not attached, others are not available because they are proprietary data.

1. CGX dataset -publicly available but not attached: Kirkland D, Aardema M, Henderson L and Muller L. Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. I. Sensitivity, specificity and relative predictivity. Mutation Research, 2005, 584, 1-256, available at "http://dx.doi.org/10.1016/j.mrgentox.2005.02.004"

2. Vitic 4.0 database National Toxicology Program data – publicly available but not attached

3. Marketed Pharmaceuticals dataset - publicly available but not attached: Snyder RD and Green JW. A review of the genotoxicity of marketed pharmaceuticals. Mutation Research, 2001, 488, 151-169, available at "http://dx.doi.org/10.1016/S1383-5742(01)00055-2"; Snyder RD, Pearl GS, Mandakas G, Choy WN, Goodsaid F and Rosenblum IY. Assessment of the sensitivity of the computational programs DEREK, T O P K A T , a n d M C A S E i n t h e p r e d i c t i o n o f t h e g e n o t o x i c i t y o f pharmaceutical molecules. Environmental and Molecular Mutagenesis, 2004, 43, 143-158, available at "http://dx.doi.org/10.1002/em.20013"; Snyder RD. An update on the genotoxicity and carcinogenicity of marketed pharmaceuticals with reference to in silico predictivity.
Environmental and Molecular Mutagenesis, 2009, 50, 435-450, available at "http://dx.doi.org/10.1002/em.20485".

4. Proprietary dataset 1 - not available: A proprietary collection of Ames test data for 575 chemicals

5. Proprietary dataset 2 - not available: A proprietary collection of Ames test data for 454 chemicals contributed by Bayer Schering Pharma AG

6. A collection of Ames test data for 6920 compounds compiled from US Food and Drug Administration (FDA) SAR Genetox Database (extracted 7.External validation - OECD Principle 4 2 September 2010) - proprietary dataset. 7) Hansen K, Mika S, Schroeter T, Sutter A, Ter Laak A, Steger-Hartmann T, Heinrich N and Muller KR. Benchmark data set for in silico prediction of Ames mutagenicity. Journal of Chemical Information and M o d e l i n g , 2 0 0 9 , 4 9 , 2 0 7 7 - 2 0 8 1 , a v a i l a b l e a t "http://dx.doi.org/10.1021/ci900161g" - publicly available but not attached.

7.6.Experimental design of test set:

Proprietary datasets were sought.

7.7.Predictivity - Statistics obtained by external validation:

The positive predictivity for each alert for the seven datasets is available within the software.

7.8.Predictivity - Assessment of the external validation set:

The total number of compounds in the validation datasets is 3703 and so is sufficiently large to validate the model (although there may be some

compounds which are common to multiple datasets).

The compounds in the datasets are primarily small chemicals and so are representative of the structures used to build the model.

7.9.Comments on the external validation of the model:

The seven datasets used for external validation were tested against each alert in the mutagenicity model and the positive predictivity calculated for each alert and each dataset.

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

All alerts describing structure-activity relationships for the mutagenicity endpoint have a mechanistic basis wherever possible. Mechanistic information is detailed in the comments associated with an alert and can include information on both the mechanism of action and biological target.

8.2.A priori or a posteriori mechanistic interpretation:

The mechanistic basis of the model was developed a priori by examining the active and inactive structures before developing the structure-activity relationship.

8.3.Other information about the mechanistic interpretation:

All references supporting the mechanistic basis of an alert are detailed and available for inspection within the software.

9.Miscellaneous information

9.1.Comments:

Derek for Windows is an knowledge-based expert system containing mechanistically-based rules which are built using all the underlying evidence available to the SAR developer. Therefore, there is no defined training or test set, and therefore there are no internal validation statistics to report.

9.2.Bibliography:

[1]Kirkland D, Aardema M, Henderson L & Muller L (2005). Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens. I. Sensitivity, specificity and relative predictivity. Mutation Research 584, 1-25. http://dx.doi.org/10.1016/j.mrgentox.2005.02.004

[2]Snyder RD & Green JW (2001). A review of the genotoxicity of marketed pharmaceuticals. Mutation Research 488, 151-169. http://dx.doi.org/10.1016/S1383-5742(01)00055-2

[3]Snyder RD, Pearl GS, Mandakas G, Choy WN, Goodsaid F & Rosenblum IY (2004). Assessment of the sensitivity of the computational programs DEREK, TOPKAT, and MCASE in the prediction of the genotoxicity of pharmaceutical molecules. Environmental and Molecular Mutagenesis 43, 143-158. http://dx.doi.org/10.1002/em.20013

[4]Snyder RD (2009)An update on the genotoxicity and carcinogenicity of marketed pharmaceuticals with reference to in silico predictivity. Environmental and Molecular Mutagenesis, 50. 435-450 http://dx.doi.org./10.1002/em.20485

[5]Hansen K, Mika S, Schroeter T, Sutter A, Ter Laak A, Steger-Hartmann T, Heinrich N and Muller

KR (2009) Benchmark data set for in silico prediction of Ames mutagenicity. Journal of Chemical Information and Modeling, 49, 2077-2081 http://dx.doi.org/10.1021/ci900161g

9.3.Supporting information:

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

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

Q13-410-0042

10.2.Publication date:

2013-06-27

10.3.Keywords:

Lhasa Limited;Derek for Windows;Ames;mutagenicity;

10.4.Comments:

former Q13-33-36-312