	QMRF identifier (JRC Inventory): **Q13-412-0043**
	QMRF Title: **Derek for Windows - Carcinogenicity**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

Derek for Windows - Carcinogenicity

1.2.Other related models:

The Derek for Windows model for Carcinogenicity is the only model and invokes no submodels.

1.3.Software coding the model:

Derek for Windows version 13
https://www.lhasalimited.org/derek_nexus/

2.General information

2.1.Date of QMRF:

26 July 2010

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):

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 carcinogenicity endpoint.

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

Sanderson DM & 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

2.8.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. Alerts are developed against a set of compounds which are generally gleaned from publicly available data, though a formal training set is not recorded for all alerts.

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

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Mammal (principally rat and mouse)

3.2.Endpoint:

4.Human Health Effects 4.12.Carcinogenicity

3.3.Comment on endpoint:

The model is based primarily on data from chronic carcinogenicity studies in the rat or mouse. If available, human data (cohort studies) is also used.

3.4.Endpoint units:

Tumour incidence in the chronic carcinogenicity study is used to assign activity for carcinogenicity. Site of action and tumour type are also considered as evidence of a common mechanism for all compounds covered by an alert.

3.5.Dependent variable:

Not applicable

3.6.Experimental protocol:

Where available, alerts are based on data generated from studies with a protocol within the guidelines outlined by the NTP (http://ntp.niehs.nih.gov/?objectid=36305D16-F1F6-975E-79776DAD38EC101E). Genotoxicity data from Ames studies, run in the standard strains to a standard protocol, are also considered in the development of alerts for genotoxic carcinogenicity.

3.7.Endpoint data quality and variability:

Where available, alerts are based on data generated from studies with a protocol within the guidelines outlined by the NTP (http://ntp.niehs.nih.gov/?objectid=36305D16-F1F6-975E-79776DAD38EC101E), the data forming the basis of each alert is fully referenced within the alert.

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:

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 carcinogenicity 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 carcinogenicity, 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 restrictions 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:

The training set for each alert is not available. The number of compounds upon which an alert is built can vary substantially from alert to alert depending on the amount of information available in the public domain or donated proprietary data. However an alert will not be built against a single compound. No internal validation has been performed.

6.6.Pre-processing of data before modelling:

Data is not processed before an alert is developed, though an alert will not be built against a single compound, and data generated from standard test protocols is always used unless the alert comments specifically mention otherwise.

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:

Yes

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:

Three external validation sets exist which have been used for the external validation of the alerts within the carcinogenicity endpoint. All datasets are publicly available but not attached, though they are free to download from the given addresses.

1) Carcinogenic Potency Database (CPDB) downloaded from DSSTox (v e r s i o n 5 d , r e v i s e d 2 0 N o v e m b e r 2 0 0 8) . A v a i l a b l e a t " h t t p : / / w w w . e p a . g o v / N C C T / d s s t o x / s d f _ c p d b a s . h t m l " .

2) Toxicity Reference Database (ToxRefDB) Chronic & Cancer Endpoints d a t a (r e v i s e d 0 6 A p r i l 2 0 0 9) . A v a i l a b l e a t " h t t p : / / w w w . e p a . g o v / N C C T / t o x r e f d b / " .

3) Istituto Superiore di Sanita, Chemical Carcinogens: Structures and Experimental Data (ISSCAN) downloaded from DSSTox (version 3a, r e v i s e d 1 9 S e p t e m b e r 2 0 0 8) . A v a i l a b l e a t "http://www.epa.gov/ncct/dsstox/sdf_isscan_external.html".

7.6.Experimental design of test set:

In addition to the publicly available dataset, proprietary datasets were sought for robust external evaluations.

7.7.Predictivity - Statistics obtained by external validation:

The positive predictivity for each alert for the three 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 1818, which is sufficiently large to validate the model (although there may be some compounds which are common to both 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 three datasets used for external validation were tested against each alert in the carcinogenicity 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 carcinogenicity 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. The model may be used to consider the likelihood of carcinogenicity when preparing a regulatory submission (e.g. for REACH).

9.2.Bibliography:

[1]Carcinogenic Potency Database (CPDB) downloaded from DSSTox (version 5d, revised 20 November 2008) http://www.epa.gov/NCCT/dsstox/sdf_cpdbas.html

[2]Toxicity Reference Database (ToxRefDB) Chronic & Cancer Endpoints data (revised 06 April 2009) http://www.epa.gov/NCCT/toxrefdb/

[3]Istituto Superiore di Sanita, Chemical Carcinogens: Structures and Experimental Data (ISSCAN) downloaded from DSSTox (version 3a, revised 19 September 2008) http://www.epa.gov/ncct/dsstox/sdf_isscan_external.html

9.3.Supporting information:

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

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

Q13-412-0043

10.2.Publication date:

2013-06-27

10.3.Keywords:

Lhasa Limited;carcinogenicity;Derek for Windows;mammal;rat;mouse;

10.4.Comments:

former Q13-33-36-314