*QMRF identifier (JRC Inventory):*To be entered by JRC
*QMRF Title:*Derek Nexus - mutagenicity
*Printing Date:*8 Nov 2023

1.QSAR identifier

1.1.QSAR identifier (title):

Derek Nexus - mutagenicity

1.2.Other related models:

Derek Nexus contains alerts for multiple endpoints, including mutagenicity, chromosome damage, carcinogenicity, hepatotoxicity, teratogenicity and skin irritation

1.3.Software coding the model:

Derek Nexus v6.3.0 contains 156 active alerts for bacterial mutagenicity, together with reasoning rules and secondary functionality that evaluates potentially misclassified and unclassified features in compounds that do not activate bacterial mutagenicity alerts or examples.

2.General information

2.1.Date of QMRF:

2 June 2009

2.2.QMRF author(s) and contact details:

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

2.3.Date of QMRF update(s):

8 November 2023

2.4.QMRF update(s):

Rachael Tennant, Lhasa Limited, Granary Wharf House, 2 Canal Wharf,
Leeds, LS11 5PS, UK
1.3, 2.3, 2.4, 2.6, 4.5, 7.8

2.5.Model developer(s) and contact details:

Lhasa Limited Granary Wharf House, 2 Canal Wharf, Leeds, LS11 5PS

2.6.Date of model development and/or publication:

Derek Nexus 6.3.0 was released on 27 October 2023

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

[1]Sanderson DM & Earnshaw CG (1991). Computer prediction of possible toxic action from chemical structure; The DEREK system. Human and Experimental Toxicology 10, 261-273.

[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.

[3]Marchant CA, Briggs KA & Long A (2003). In silico tools for sharing data and knowledge on toxicity and metabolism: Derek for Windows, Meteor, and Vitic. Toxicology Mechanisms and Methods 18, 177–187.

[4]Judson PN, Stalford SA & Vessey J (2013). Assessing confidence in predictions made by knowledge-based systems. Toxicology Research 2, 70-79.

[5]Williams RV, Amberg A, Brigo A, Coquin L, Giddings A, Glowienke S, Greene N, Jolly R, Kemper R, O'Leary-Steele C, Parenty A, Spirkl HP, Stalford SA, Weiner SK and Wichard J (2016). It's difficult, but important, to make negative predictions. Regulatory Toxicology and Pharmacology 76,

79-86.

2.8.Availability of information about the model:

Derek Nexus is a proprietary, rule-based expert system for the prediction of toxicity. Its knowledge base is composed of alerts, examples and reasoning rules which may each contribute to the predictions made by the system. Each alert in Derek describes a chemical substructure believed to be responsible for inducing a specific toxicological outcome (often referred to as a toxicophore). Alerts are derived by experts, using toxicological data and information regarding the biological mechanism of action. Where relevant, metabolism data may be incorporated into an alert, enabling the prediction of compounds which are not directly toxicity but are metabolised to an active species. The derivation of each alert is described in the alert comments along with supporting references and example compounds where possible. By reporting this information to the user, Derek provides highly transparent predictions. The use of structural alerts for the prediction of toxicity is both widely understood and the subject of many publications.

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

No

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Predictions are made for the domain of bacteria and can be broken down into species (e.g. Salmonella typhimurium and Escherichia coli).

3.2.Endpoint:

TOX 7.6.1. Genetic toxicity in vitro

3.3.Comment on endpoint:

The Derek Nexus model for mutagenicity is developed from Ames test data in both S.typh and E.coli. Supporting data from transgenic rodent mutation assay, in vitro L5178Y TK+/- 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. Additionally, alert writers consider both mechanistic evidence and chemical properties (such as reactivity).

3.4.Endpoint units:

Derek Nexus makes qualitative predictions for and against toxicity through reasoning. For the endpoint of mutagenicity, predictions for toxicity decrease in confidence in the following order: certain> probable>plausible>equivocal. Predictions against toxicity increase in confidence in the following order: inactive (with unclassified and/or misclassified features)<inactive<improbable. Likelihood levels have been shown to correlate with predictivity [Judson et al, 2013]. Multiple data sources (e.g. toxicity data from multiple assays and mechanistic evidence) are synthesised into the structure-activity relationships that underpins Derek Nexus predictions. An appreciation of the assay units

applied by alert writers when building the alert training set. However, predictions are not quantitative and, as a result, do not include units.

3.5.Dependent variable:

Data from the Ames test and mechanistic studies (e.g. measures of electrophilicity) are synthesised to arrive at an expert conclusion of whether compounds within the model training set is likely to be a mutagen.

3.6.Experimental protocol:

The model is based primarily on data from the Ames test conducted following standard test protocol (OECD TG471). If activity is observed in a non-standard assay or protocol this will be mentioned in the comments.

3.7.Endpoint data quality and variability:

Alert writers use all available and relevant information in the public domain (and proprietary data, where available) for alert development. Wherever possible, primary references are used as data sources: (i) the data are subject to expert assessment prior to inclusion in the alert training set using, amongst other criteria, OECD test guidelines and (ii) the references themselves are cited in the alert comments enabling users to conduct their own expert assessments on data quality.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

Expert derived structural alerts for mutagenicity (2D SARs), physicochemical properties and associated reasoning. Following alert evaluation, Derek evaluates whether non-alerting query compounds contain any features that are either (i) also present in non-alerting mutagens in a large Ames test reference set (misclassified features) or (ii) not present in a large Ames test reference set (unclassified features).

4.2.Explicit algorithm:

logic of argumentation

structural alerts

feature-based database search

4.3.Descriptors in the model:

[1]Markush structures encoding activating and deactivating features (known as patterns in the Derek Nexus knowledge base)

[2]count of non-hydrogen atoms

[3]ClogP

[4]2D structural fragments

4.4.Descriptor selection:

There is an a priori assumption that patterns and associated reasoning will be used to model toxicity within Derek Nexus. Further, experts identified that misclassified and unclassified features were useful descriptors for determining the reliability of negative predictions for non-alerting compounds.

4.5.Algorithm and descriptor generation:

Alert writers design the patterns to describe the activating and deactivating features found during expert assessment of the alert training set. Misclassified and unclassified features are generated by processing a large Ames test reference set (comprising 5858 mutagens and 6532 non-mutagens) against Derek Nexus (v6.3.0) and fragmenting.

4.6.Software name and version for descriptor generation:

Alert writers use the Derek Knowledge Editor (v2.0) for the implementation of patterns. ClogP predictions generated using the BioByte model (v5.9). Fragmentation is completed using an in-house algorithm.

4.7.Chemicals/Descriptors ratio:

This is not applicable to structural alerts as these 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 scopes 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 activates an alert describing a structure-activity for mutagenicity it can be considered to be within the applicability domain. If a compound does not activate an alert or reasoning rule then Derek makes a negative prediction. The applicability of the negative prediction to the query compounds can be determined by an expert, if required, by investigating the presence (or absence) of misclassified and/or unclassified features.

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. For non-alerting compounds, users should determine the applicability of negative predictions by evaluating the information supplied by Derek (i.e. the presence or absence of misclassified and/or unclassified features).

5.3.Software name and version for applicability domain assessment:

This is not applicable.

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

NanoMaterial: 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:

Non-proprietary elements of the training set are available through the references, and illustrated by the examples, within Derek Nexus. The illustrative examples are not available, due to the proprietary nature of Derek Nexus.

6.6.Pre-processing of data before modelling:

This is not applicable.

6.7.Statistics for goodness-of-fit:

This is not applicable.

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

This is not applicable.

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

This is not applicable.

6.10.Robustness - Statistics obtained by Y-scrambling:

This is not applicable.

6.11.Robustness - Statistics obtained by bootstrap:

This is not applicable.

6.12.Robustness - Statistics obtained by other methods:

This is not applicable.

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

NanoMaterial: 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:

External validation is carried out on each knowledge base release. The data sets used for validation are available in the public domain, but the curated versions used at Lhasa are proprietary, so are not made available. Three proprietary data sets have been used for alert validation.

7.6.Experimental design of test set:

Proprietary data sets were sought.

7.7.Predictivity - Statistics obtained by external validation:

The software reports the number of positive and negative compounds from the validation data sets that activate each alert and calculates positive predictivity using this data.

7.8.Predictivity - Assessment of the external validation set:

Ames test data covering 10,480 unique compounds. The compounds in the dataset are primarily small and medium-sized chemicals and so are representative of the structures used to build the model.

7.9.Comments on the external validation of the model:

No information is available.

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 toxicological and mechanistic evidence 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 Nexus may be used to assess the toxicity of a wide range of chemical classes, including food, drug, cosmetic, and industrial chemicals, and the system provides predictions for over 50 toxicological endpoints, including mutagenicity, chromosome damage, carcinogenicity, skin sensitisation and reproductive toxicity. Mutagenicity predictions from Derek have been used in the assessment of pharmaceuticals [Hillebrecht et al], food/flavour chemicals [Cotterill et al, Ono et al] and industrial/environmental chemicals [Hayashi et al]. Derek mutagenicity predictions are also submitted as part of the regulatory requirements on genotoxic impurities in pharmaceuticals [Sutter et al, Dobo et al, ICH].

9.2.Bibliography:

[1]Sanderson DM & Earnshaw CG (1991). Computer prediction of possible toxic action from chemical structure; The DEREK system. Human and Experimental Toxicology 10, 261-273.

[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.

[3]Marchant CA, Briggs KA & Long A (2003). In silico tools for sharing data and knowledge on toxicity and metabolism: Derek for Windows, Meteor, and Vitic. Toxicology Mechanisms and Methods 18, 177–187.

[4]Judson PN, Stalford SA & Vessey J (2013). Assessing confidence in predictions made by knowledge-based systems. Toxicology Research 2, 70-79.

[5]Williams RV, Amberg A, Brigo A, Coquin L, Giddings A, Glowienke S, Greene N, Jolly R, Kemper R, O'Leary-Steele C, Parenty A, Spirkl HP, Stalford SA, Weiner SK and Wichard J (2016). It's difficult, but important, to make negative predictions. Regulatory Toxicology and Pharmacology 76, 79-86.

[6]Hillebrecht A, Muster W, Brigo A, Kansy M, Weiser T & Singer T (2011). Comparative evaluation of in silico systems for Ames test mutagenicity prediction: scope and limitations. Chemical Research in Toxicology 24, 843-854.

[7]Cotterill JV, Chaudhry MQ, Matthews W & Watkins RW (2008). In silico assessment of toxicity of heat-generated food contaminants. Food and Chemical Toxicology 46, 1905-1918.

[8]Ono A, Takahashi M, Hirose A, Kamata E, Kawamura T, Yamazaki T, Sato K, Yamada M, Fukumoto T, Okamura H, Mirokuji Y & Honma M (2012). Validation of the (Q)SAR combination approach for mutagenicity prediction of flavor chemicals. Food and Chemical Toxicology 50, 1538-1546.

[9]Hayashi M, Kamata E, Hirose A, Takahashi M, Morita T & Ema M (2005). In silico assessment of chemical mutagenesis in comparison with results of Salmonella microsome assay on 909 chemicals. Mutation Research 588, 129-135.

[10]Sutter A, Amberg A, Boyer S, Brigo A, Contrera JF, Custer LL, Dobo KL, Gervais V, Glowienke S, van Gompel J, Greene N, Muster W, Nicolette J, Reddy MV, Thybaud V, Vock E, White AT & Müller L (2013). Use of in silico systems and expert knowledge for structure-based assessment of potentially mutagenic impurities. Regulatory Toxicology and Pharmacology 67, 39-52.

[11]Dobo KL, Greene N, Fred C, Glowienke S, Harvey JS, Hasselgren C, Jolly R, Kenyon MO, Munzner JB, Muster W, Neft R, Reddy MV, White AT & Weiner S (2012). In silico methods combined with expert knowledge rule out mutagenic potential of pharmaceutical impurities: an industry survey. Regulatory Toxicology and Pharmacology 62, 449-455.

[12]International Conference on Harmonisation (2013). Assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk. M7 Draft Guideline, Step 2.

9.3.Supporting information:

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