	QMRF identifier (JRC Inventory): **To be entered by JRC**
	QMRF Title: **Derek Nexus - teratogenicity**
	Printing Date: **22 May 2020**

1.QSAR identifier

1.1.QSAR identifier (title):

Derek Nexus - teratogenicity

1.2.Other related models:

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

1.3.Software coding the model:

Derek Nexus v6.1 contains 51 alerts for Teratogenicity together with reasoning rules.

2.General information

2.1.Date of QMRF:

9 January 2017

2.2.QMRF author(s) and contact details:

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

2.3.Date of QMRF update(s):

22 May 2020

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, 3.4, 4.6

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 6.1 was released on 11 June 2020

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.

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 class of mammals and can be broken down into species (e.g. mouse, human, guinea pig).

3.2.Endpoint:

TOX 7.8.2. Developmental toxicity / teratogenicity

3.3.Comment on endpoint:

The Derek Nexus model for teratogenicity is developed from several sources of data. Sources of primary data used for alert development include, teratogenicity studies in animals, in vitro data (embryo culture assays) and human case reports on teratogenicity including FDA pregnancy categories. Additionally, alert writers consider both mechanistic evidence and chemical properties (such as reactivity). The endpoint of teratogenicity is linked to four other endpoints, 5alpha-reductase inhibition, androgen receptor modulation, glucocorticoid receptor agonism and oestrogen receptor modulation, so that when these endpoints are fired, they extrapolate to the endpoint of teratogenicity (at a lower reasoning level).

Structural alerts for 5alpha-reductuase inhibition are developed from several sources of in vitro and in vivo data. In vitro data consisted of 5alpha-reductase type-II inhibition studies using enzymes isolated from a variety of species. Positive calls for in vitro inhibition were based on potency. In vivo studies consisted of studies that reported observations indicative of 5alpha-reductase inhibition and positive calls were based on the observed effect of the compound. Structures were standardised and an expert conclusion of whether a compound can be classified as a 5alpha-reductase inhibitor was determined given the reported experimental results.

Structural alerts for androgen receptor modulation are developed from in vitro and in vivo androgen receptor bioactivity data extracted from ChEMBL20. In vitro assays were capable of measuring compound agonism

and/or antagonism towards the androgen receptor. In vitro assays were developed from a range of cell types and from various species. Positive calls for a compounds activity in in vitro assays were based on potency. In vivo assays consisted of studies that reported observations indicative of disruption of the androgen receptor pathway. Positive calls for in vivo assays were made based on the observed effect of the compound. Structures were standardised and an expert conclusion of whether a compound can be classified as an androgen receptor modulator was determined given the reported experimental results.

Structural alerts for glucocorticoid receptor agonism are developed from in vitro and in vivo glucocorticoid receptor bioactivity data extracted from ChEMBL20. In vitro assays were capable of measuring compound agonism and/or antagonism towards the glucocorticoid receptor. In vitro assays were developed from a range of cell types and from various species. Positive calls for a compounds activity in in vitro assays were based on potency. In vivo assays consisted of studies that reported observations indicative of disruption of the glucocorticoid receptor pathway. Positive calls for in vivo assays were made based on the observed effect of the compound. Structures were standardised and an expert conclusion of whether a compound can be classified as a glucocorticoid receptor modulator was determined given the reported experimental results.

Structural alerts for oestrogen receptor modulation were developed from in vitro and in vivo oestrogen receptor bioactivity data extracted from ChEMBL17. In vitro assays were capable of measuring compound inhibition, agonism or antagonism towards the oestrogen receptor. In vitro assays were developed from a range of cell types and from various species. Positive calls for a compounds activity in in vitro assays were based on potency. In vivo assays consisted of studies that reported observations indicative of disruption of the oestrogen receptor pathway. Positive calls for in vivo assays were made based on the observed effect of the compound. Structures were standardised and an expert conclusion of whether a compound can be classified as an oestrogen receptor modulator was determined given the reported experimental results.

3.4.Endpoint units:

Derek Nexus makes predictions for and against toxicity through reasoning. For the endpoint of teratogenicity, predictions for toxicity decrease in confidence in the following order: certain> probable>plausible>equivocal. Predictions against toxicity increase in confidence in the following order: doubted<improbable. These 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. Units are considered by the alert writers when building the alert training set, however, as predictions are made using data from multiple assays these

do not include units as default.

3.5.Dependent variable:

Toxicological data and mechanistic studies are reviewed to arrive at an expert conclusion of whether compounds within the model training set are likely to be teratogenic.

3.6.Experimental protocol:

The models for teratogenicity, 5alpha-reductase inhibition, androgen receptor modulation, glucocorticoid receptor agonism and oestrogen receptor modulation are based on a variety of data sources outlined in section 3.3. In some cases, the data is derived from OECD compliant protocols (such as 483, 455, 457 and 458), whilst in others the data stems from modified test protocols.

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:

Predictions are based on expert derived structural alerts for teratogenicity, 5alpha-reductase inhibition, androgen receptor modulation, glucocorticoid receptor agonism and oestrogen receptor modulation (2D SARs), that take into account toxicological and mechanistic evidence, and where appropriate stereochemistry metabolism and physicochemical properties of compounds.

4.2.Explicit algorithm:

Structural alerts

logic of argumentation

4.3.Descriptors in the model:

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

[2]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.

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.

4.6.Software name and version for descriptor generation:

Alert writers use the Derek Knowledge Editor (v2.0) for the implementation of patterns and prediction of ClogP is made using the BioByte model (v5.9)

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 teratogenicity, 5alpha-reductase inhibition, androgen receptor modulation, glucocorticoid receptor agonism and oestrogen receptor modulation endpoints 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 teratogenicity/ 5alpha-reductase inhibition/androgen receptor modulation/glucocorticoid receptor agonism/oestrogen receptor modulation it can be considered to be within the applicability domain. If a compound does not activate an alert or reasoning rule in Derek, a result of 'nothing to report' is presented to the user. This can be interpreted as a negative prediction or that the query compound is outside the domain of the model. Which of these is more appropriate may depend on the endpoint of interest.

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:

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:

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:

8.Providing a mechanistic interpretation - OECD Principle 5

8.1.Mechanistic basis of the model:

All alerts describing structure-activity relationships for the teratogenicity, 5alpha-reductase inhibition, androgen receptor modulation, glucocorticoid receptor agonism and oestroen receptor modulation endpoints have a mechanistic basis wherever possible. Mechanistic information can be 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.

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]Langton K, Patlewicz GY, Long A, Marchant CA, Basketter DA (2006).

9.3.Supporting information:

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

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

To be entered by JRC

10.2.Publication date:

To be entered by JRC

10.3.Keywords:

To be entered by JRC

10.4.Comments:

To be entered by JRC