QMRF CHAPTER TITLES	SECTION TITLES	MODEL DESCRIPTION *(Fill in this column)*
1. QSAR identifier		
	1.1. QSAR identifier (title):	Derek Nexus - chromosome damage
	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 v5.0 contains 98 alerts for chromosome damage, together with reasoning rules. These describe numerical (e.g. aneugenic) and structural (i.e. clastogenic) chromosomal aberrations in vitro and in vivo.
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
	2.3. Date of QMRF update(s):	08 January 2016
	2.4. QMRF update(s):	Lilia Fisk, Lhasa Limited, Granary Wharf House, 2 Canal Wharf, Leeds, LS11 5PS, UK 1.3, 2.3, 2.4, 2.6
	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 5.0 was released on 18 January 2016
	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. hamster, mouse, human) and prediction type (in vitro or in vivo).
	3.2. Endpoint:	6.Other. 6.6.Chromosome damage
	3.3. Comment on endpoint:	The Derek Nexus model for chromosome damage is developed from several sources of data. Sources of primary data used for alert development include [1] in vitro and in vivo chromosome aberration test, [2] in vitro and in vivo micronucleus test, [3] in vitro L5178Y TK+/- assay. Alert writers consider both mechanistic evidence and chemical properties (such as reactivity). Depending on evidence in vitro and/or in vivo prediction can be made.
	3.4. Endpoint units:	Derek Nexus makes predictions for and against toxicity through reasoning. For the endpoint of chromosome damage, 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. An appreciation of the assay units applied by alert writers when building the alert training set. However, predictions are not restricted to a specific assay and, as a result, do not include units.
	3.5. Dependent variable:	Toxicological data and mechanistic studies (e.g. tubulin binding studies) are synthesised to arrive at an expert conclusion of whether compounds within the model training set is likely to be an aneugen or a clastogen.
	3.6. Experimental protocol:	The model is based primarily on data from in vitro and in vivo chromosome aberration and micronucleus tests. Whilst standard test protocols exist for such studies (CA: OECD TG473, 475; MN: OECD TG474, 487) much of the historical data predates such guidance. Data for such assays may be used for model development, where alert writers have deemed the quality to be acceptable.
	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 chromosome damage (2D SARs), physicochemical properties and associated reasoning.
	4.2. Explicit algorithm:	[1] Structural alerts, [2] 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] ClogP and [3] count of non-hydrogen atoms.
	4.4. Descriptor selection:	There is an a priori assumption that patterns, physicochemical properties 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 (v5.3).
	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 scopes of the structure-activity relationships describing the chromosome damage 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 chromosome damage 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:	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.
	6.2. 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:	This is not applicable.
	6.4. Data for the dependent variable for the training set:	This is not applicable.
	6.5. Other information about the training set:	This is not applicable.
	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:	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.
	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:	This is not applicable.
	7.4. Data for the dependent variable for the external validation set:	This is not applicable.
	7.5. Other information about the external validation set:	One proprietary and two published data sets have been used for validation of in vitro chromosome damage alerts: [1] Sofuni, [2] Kirkland et al. Three proprietary data sets have been used of validation for in vivo chromosome damage alerts. Further, the relationship between likelihood levels and prediction accuracy has been assessed [Judson et al 2013].
	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:	Data derived from in vitro and in vivo chromosome damage tests covering 3361 and 1802 unique compounds correspondingly. The compounds in the data sets 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:	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 chromosome damage 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. Chromosome damage predictions from Derek have been used in the assessment food/flavour chemicals [Cotterill et al]. Additionally, Derek mutagenicity predictions are submitted as part of the regulatory requirements on genotoxic impurities in pharmaceuticals [Sutter et al].
	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] T. Sofuni; Revised Edition 1998 Data book of chromosomal aberration test in vitro: Sofuni T (editor), Life-Science Information Center, Tokyo. [6] D. 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, 1-256. [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] 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.
	9.3. Supporting information:	No information is available.
10. Summary (JRC Inventory)		
	10.1. QMRF number:	tbc by JHRC
	10.2. Publication date:	tbc by JHRC
	10.3. Keywords:	tbc by JHRC
	10.4. Comments:	tbc by JHRC