	QMRF identifier (JRC Inventory): **Q13-412-0039**
	QMRF Title: **TOPKAT NTP Rodent Carcinogenicity Model (Female Mouse)**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

TOPKAT NTP Rodent Carcinogenicity Model (Female Mouse)

1.2.Other related models:

TOPKAT NTP Rodent Carcinogenicity Model (Male Mouse)

TOPKAT NTP Rodent Carcinogenicity Model (Female Rat)

TOPKAT NTP Rodent Carcinogenicity Model (Male Rat)

1.3.Software coding the model:

DS TOPKAT v2.1

See section 9.1

Accelrys Software Inc., 10188 Telesis Court, Suite 100, San Diego, CA 92121, USA

http://accelrys.com/products/discovery-studio/toxicology/

2.General information

2.1.Date of QMRF:

14 October 2008

2.2.QMRF author(s) and contact details:

Deqiang Zhang Accelrys Software Inc. 10188 Telesis Ct. Suite 100, San Diego, CA 92121, USA

dzhang@accelrys.com http://www.accelrys.com

2.3.Date of QMRF update(s):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

Accelrys Accelrys Software Inc. 10188 Telesis Ct. Suite 100, San Diego, CA 92121, USA

dzhang@accelrys.com http://www.accelrys.com

2.6.Date of model development and/or publication:

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

TOPKAT software http://accelrys.com/products/discovery-studio/toxicology/

2.8.Availability of information about the model:

The model is proprietary. The algorthim is confidential, although the

training set can be recovered from the software by similarity searching.

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

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Female mouse

3.2.Endpoint:

4.Human Health Effects 4.12.Carcinogenicity

3.3.Comment on endpoint:

3.4.Endpoint units:

None (classification model)

3.5.Dependent variable:

3.6.Experimental protocol:

US National Toxicology Program protocol: a chemical's carcinogenicity is determined by testing it in both sexes of rat and mouse. The rodents are exposed to the chemical, generally for their lifetime. The pathological and other biological data are evaluated by a committee of experts, and on the basis of a majority rule, the chemical is assigned to one of the four categories, namely: CE, SE, EE, or NE, which respectively, indicate Clear Evidence, Some Evidence, Equivoval Evidence, and No Evidence of carcinogenicity.

3.7.Endpoint data quality and variability:

Endpoint data were evaluated by a committee of experts in US NCI and assigned to four different categories by majority rule.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

Linear Discriminant Analysis Model

4.2.Explicit algorithm:

Linear discriminant analysis model

Algorithm is confidential

4.3.Descriptors in the model:

4.4.Descriptor selection:

All the E-state keys, symmetry indices, shape indices, molecular weight and VlogP were calculated for the training set. These descriptors underwent a principal component analysis and the ones corresponding to the largest variances were selected to do a linear discriminant analysis.

4.5.Algorithm and descriptor generation:

Standard LDA procedures using SAS statistics software were used. The descriptors were calculated according to methods described in: Hall et. al. (1991), Kier (1986), and Gombar & Jain (1987).

4.6.Software name and version for descriptor generation:

TOPKAT

The descriptors are calculated internally by TOPKAT

Accelrys Software Inc., 10188 Telesis Court, Suite 100, San Diego, CA 92121, USA

http://accelrys.com/products/discovery-studio/toxicology/

4.7.Chemicals/Descriptors ratio:

2643 descriptors / 248 chemicals

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

Fragment covered by the training set or not; whether in range (univariate analysis); Optimal Prediction Space (Multivariate analysis).

5.2.Method used to assess the applicability domain:

TOPKAT's Optimal Prediction Space (OPS)

5.3.Software name and version for applicability domain assessment:

DS TOPKAT v2.1

TOPKAT in Discovery Studio 2.1

Accelrys Software Inc., 10188 Telesis Court, Suite 100, San Diego, CA 92121

http://accelrys.com/products/discovery-studio/toxicology/

5.4.Limits of applicability:

All descriptors have to be covered by the training set.

6.Internal validation - OECD Principle 4

6.1.Availability of the training set:

Yes

6.2.Available information for the training set:

CAS RN: Yes

Chemical Name: Yes

Smiles: Yes

Formula: Yes

INChI: Yes

MOL file: Yes

6.3.Data for each descriptor variable for the training set:

No

6.4.Data for the dependent variable for the training set:

All

6.5.Other information about the training set:

The training set is stored with the model and can be retrieved through similarity search.

237 data points: 80 positive values; 157 negative values

6.6.Pre-processing of data before modelling:

no

6.7.Statistics for goodness-of-fit:

True Positive: 67

False Negative: 13

False Positive: 6

True Negative: 151

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

Specificity = 0.87

Sensitivity = 0.88

ROC score = 0.89

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:

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:

8.2.A priori or a posteriori mechanistic interpretation:

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

A QSAR-based system generates and validates accurate, rapid assessments of chemical toxicity solely from a chemical's molecular structure. Unique among SAR-based technologies, DS TOPKAT uses robust, cross-validated models based on experimental data of highly consistent protocol. The models are subjected to extensive diagnostics for accuracy and validity. DS TOPKAT uses patented Optimum Prediction Space (OPS) technology to assure that the compounds under investigation are well represented in the models. Included within DS TOPKAT are tools that allow you to easily build molecules or queries. DS TOPKAT can be used for tests including physical/chemical, environmental fate, ecotoxicity, toxicity, mutagenicity, and subchronic reproductive/developmental. DS TOPKAT is fast, cost-effective, and proven.

9.2.Bibliography:

[1]Hall LH, Mohney B & Kier LB (1991). The Electrotopological State: Structure Information at the Atomic Level for Molecular Graphs. Journal of Chemical Information and Computer Science 31, 76-82.

[2]Kier LB (1986). Shape Indices of Orders One and Three from Molecular Graphs. Quantitative Structure-Activity Relationships 5, 1-7.

[3]Gombar VK & Jain DVS (1987). Quantification of Molecular Shape and Its Correlation with Physiochemical Properties. Indian Journal of Chemistry 26A, 554-555.

9.3.Supporting information:

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

10.Summary (JRC QSAR Model Database)

10.1.QMRF number:

Q13-412-0039

10.2.Publication date:

2013-06-27

10.3.Keywords:

Accelrys;TOPKAT;rodent;female mouse;carcinogenicity;

10.4.Comments:

former Q11-25-20-154