	QMRF identifier (JRC Inventory): **Q13-410-0067**
	QMRF Title: **Toxtree QSAR 13: mutagenicity of alpha,beta unsaturated aliphatic aldehydes in Salmonella typhimurium TA100**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

Toxtree QSAR 13: mutagenicity of alpha,beta unsaturated aliphatic aldehydes in Salmonella typhimurium TA100

1.2.Other related models:

1.3.Software coding the model:

Toxtree

Standalone software application downloadable from the Joint Research Centre (JRC) website http://ecb.jrc.ec.europa.eu/qsar/qsar-tools/index.php?c=TOXTREE

2.General information

2.1.Date of QMRF:

June 2010

2.2.QMRF author(s) and contact details:

[1]Romualdo Benigni Experimental and Computational Carcinogenesis Unit, Environment and Health Department, Istituto Superiore di Sanita Viale Regina Elena 299, 00161 Rome, Italy rbenigni@iss.it

[2]Cecilia Bossa Experimental and Computational Carcinogenesis Unit, Environment and Health Department, Istituto Superiore di Sanita Viale Regina Elena 299, 00161 Rome, Italy cecilia.bossa@iss.it

2.3.Date of QMRF update(s):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

Romualdo Benigni rbenigni@iss.it

2.6.Date of model development and/or publication:

2003; external validation 2005

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

[1]Benigni R, Passerini L & Rodomonte A (2003), Structure-activity relationships for the mutagenicity and carcinogenicity of simple and --unsaturated aldehydes. Environmental and Molecular Mutagenesis 42, 136-143.

[2]ture-activity relationship models. Environmental and Molecular Mutagenesis 46, 268-280.

[3]Benigni R, Bossa C, Netzeva TI & Worth AP (2007). Collection and evaluation of (Q)SAR models for mutagenicity and carcinogenicity. JRC report EUR 22772 EN. Luxembourg, Office for the Official Publications of the European Communities. http://ecb.jrc.ec.europa.eu/documents/QSAR/EUR_22772_EN.pdf

2.8.Availability of information about the model:

The model is non-proprietary; the training and test sets are available

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

None to date.

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Salmonella typhimurium TA100 strain

3.2.Endpoint:

4.Human Health Effects 4.10.Mutagenicity

3.3.Comment on endpoint:

3.4.Endpoint units:

Nonmutagen / Mutagen

3.5.Dependent variable:

Nonmutagen / Mutagen

3.6.Experimental protocol:

Training set data retrieved from literature (different laboratories); mostly using pre-incubation at 37° without exogenous metabolic activation. Test set data from one laboratory using pre-incubation at 37° without exogenous metabolic activation.

3.7.Endpoint data quality and variability:

Literature sources were critically reviewed by the authors of the model. The yes/no call was established by the authors.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

QSAR

QSAR13

The algorithm was originally derived through Linear Discriminant Analysis. The present form is mathematically equivalent to the original one, and has been generated through Canonical Discriminant Analysis for the sake of simplicity and ease of implementation. W is a mutagenicity probability score. W(mean, Class) is the average value for each of the two classes in the training set. When calculated for an individual chemical, W indicates to which class the chemical should be assigned, based on the Threshold (given below) that separates the two classes.

w = 0.387 MR – 3.12 logP + 3.23 LUMO

w(mean,Class1) = 9.689

N1 = 3 (non-mutagens)

w(mean,Class2) = 6.373

N2 = 17 (mutagens)

Threshold = 8.032

4.3.Descriptors in the model:

[1]MR Molar Refractivity

[2]logP logarithm of the partition coefficient between octanol and water

[3]LUMO eV energy of the Lowest Unoccupied Molecular Orbital

4.4.Descriptor selection:

Descriptors initially screened: MR, logP, LUMO, energy of the Highest Occupied Molecular Orbital (HOMO), partial charge on the carbonilic carbon; partial charge on the ? carbon. The descriptors were selected by testing different combinations based on mechanistic hypotheses, together with the use of Stepwise Linear Discriminant Analysis.

4.5.Algorithm and descriptor generation:

4.6.Software name and version for descriptor generation:

Daylight

MR and logP calaculated by Daylight software freely available on-line in 2001 – 2006. Values transformed by multiplying X 10.

http://www.daylight.com/

MOPAC 6.0

Calculation of LUMO, HOMO and partial charges

Quantum Chemistry Program Exchange, Bloomington

http://openmopac.net/

4.7.Chemicals/Descriptors ratio:

6.67 = 20 chemicals / 3 descriptors

(6 descriptors originally screened)

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The applicability domain of the model is defined in terms of:

a) structures to which it applies;

b) range of values of the descriptors in the model.

In the comparison of the test to the training set, mathematical chemical structural similarity was also applied.

5.2.Method used to assess the applicability domain:

Structures were checked by human experts. Ranges of descriptors values were calculated. Mathematical chemical similarity between test and training set was assessed as follows. The training and test sets were combined, and the overall Tanimoto similarity matrix was calculated with the Leadscope software. A Euclidian distance matrix was calculated from the similarity matrix and then subjected to Principal Component (PC) Analysis. Finally, the ranges of PC scores for the training and test sets were compared

5.3.Software name and version for applicability domain assessment:

Leadscope v. 2.4

Used for the calculation of the Tanimoto chemical similarity index.

http://www.leadscope.com/

5.4.Limits of applicability:

The models applies to ?,?-unsaturated aldehydes with both aliphatic and aromatic substituents. Chemicals with different functional groups, such

as carboxylates, should be excluded, as well as aldehydes with the unsaturated bond within a ring structure, since they may have a different type of reactivity.

Ranges of descriptors for the training set:

MR (10xDaylight): between 16.55 and 51.71
logP: between -0.01 and 2.95
LUMO: between -1.159 and -0.101

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: No
MOL file: Yes

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

All

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

All

6.5.Other information about the training set:

20 data points: 17 positive values; 3 negative values

6.6.Pre-processing of data before modelling:

6.7.Statistics for goodness-of-fit:

Squared Canonical Correlation = 0.61; 100% correct re-classification.

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

LOO 85% correct.

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

Since the training set includes only 3 negatives, no meaningful LMO cross-validation is possible.

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: Yes
Chemical Name: Yes
Smiles: Yes
Formula: Yes

INChI: No

MOL file: Yes

7.3.Data for each descriptor variable for the external validation set:

All

7.4.Data for the dependent variable for the external validation set:

All

7.5.Other information about the external validation set:

External validation set with 5 compounds appended.

7.6.Experimental design of test set:

Whereas the training set chemicals were retrieved from literature, the test set was generated by prospective experimental testing after modelling.

The applicability domain was taken into account by considering the factors of chemical structure, range of descriptors values, chemical similarity.

Chemicals with different functional groups, such as carboxylates, were excluded, as well as aldehydes with the unsaturated bond within a ring structure, since they may have a different type of reactivity.

Ranges of descriptors for the test set:

MR (10xDaylight): between 25.80 and 69.10

logP: between 0.59 and 3.49

LUMO: between -0.827 and 0.074

7.7.Predictivity - Statistics obtained by external validation:

Correct prediction 100%.

7.8.Predictivity - Assessment of the external validation set:

The test set is representative of the applicability domain; no further chemicals to test experimentally were available from commercial catalogues.

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:

High electrophilicity, low steric hindrance and high lipophilicity favor the activity. This result is concordant with the scientific evidence on mechanism of action of these chemicals, which react as direct electrophiles to form adducts with DNA (and proteins).

8.2.A priori or a posteriori mechanistic interpretation:

The descriptors to be screened were selected based on the knowledge on the mechanisms of action. The descriptors actually accepted in the model confirm the a priori hypotheses.

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

9.2.Bibliography:

[1]Benigni R, Passerini L & Rodomonte A (2003). Structure-activity relationships for the mutagenicity and carcinogenicity of simple and --unsaturated aldehydes. Environmental and Molecular Mutagenesis 42, 136-143.

[2]Benigni R, Conti L, Crebelli R, Rodomonte A & Vari MR (2005). Simple and alpha,beta unsaturated aldehydes: correct prediction of genotoxic activity through structure-activity relationship models. Environmental and Molecular Mutagenesis 46, 268-280.

[3]Benigni R, Bossa C, Netzeva TI & Worth AP (2007). Collection and evaluation of (Q)SAR models for mutagenicity and carcinogenicity. JRC report EUR 22772 EN. Luxembourg, Office for the Official Publications of the European Communities.
http://ecb.jrc.ec.europa.eu/documents/QSAR/EUR_23241_EN.pdf

[4]Woo YT, Lai DY, McLain JL, Manibusan M & Dellarco V (2002). Use of mechanism-based structure-activity relationships analysis in carcinogenic potential ranking for drinking water disinfection by-products. Environmental Health Perspectives 110, 75-87.

[5]Feron VJ, Til HP, de Vrijer F, Woutersen RA, Cassee FR & van Bladeren PJ (1991). Aldehydes: occurrence, carcinogenic potential, mechanism of action and risk assessment. Mutation Research 259, 363-385.

[6]Eder E & Deininger C (2001). Mutagenicity of 2-Alkylpropenals in Salmonella typhimurium strain TA 100: structural influences. Environmental and Molecular Mutagenesis 37, 324-328.

[7]Eder E & Deininger C (2001). Mutagenicity of 2-Alkylpropenals in Salmonella typhimurium strain TA 100: structural influences. Environmental and Molecular Mutagenesis 37, 324-328.

[8]Yang IY, Johnson F, Grollman AP & Moriya M (2002). Genotoxic mechanism for the major acrolein-derived deoxyguanosine adduct in human cells. Chemical Research in Toxicology 15, 160-164.

9.3.Supporting information:

Toxtree QSAR13 training_20.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-410-0067/attachment/A766
Toxtree QSAR13 test_5.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-410-0067/attachment/A767

Test set(s)

10.Summary (JRC QSAR Model Database)

Supporting information

10.1.QMRF number:

Q13-410-0067

10.2.Publication date:

2013-07-02

10.3.Keywords:

Toxtree;mutagenicity;Salmonella typhimurium;TA100;? ?-unsaturated aliphatic aldehyde;

10.4.Comments:

former Q19-35-35-292