	QMRF identifier (JRC Inventory): **Q13-410-0048**
	QMRF Title: **TOPS-MODE QSAR for Ames mutagenicity of alpha, beta-unsaturated carbonyl compounds**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

TOPS-MODE QSAR for Ames mutagenicity of alpha, beta-unsaturated carbonyl compounds

1.2.Other related models:

This model is related to "QSAR for Ames test of alpha, beta-unsaturated carbonyl compounds" QMRF (Q14-26-8-158).

1.3.Software coding the model:

2.General information

2.1.Date of QMRF:

December 2009

2.2.QMRF author(s) and contact details:

[1]Alfonso Pérez-Garrido Environmental Engineering and Toxicology Dpt., Catholic University of San Antonio, Guadalupe, Murcia,Spain aperez@pdi.ucam.edu

[2]Aliuska Morales Helguera Department of Chemistry, Faculty of Chemistry and Pharmacy, Central University of Las Villas, Santa Clara, Villa Clara, Cuba

[3]Gabriel Caravaca López Department of Food and Nutrition Technology, Catholic University of San Antonio, Guadalupe, Murcia, Spain

[4]M. Natália D. S. Cordeiro REQUIMTE, Chemistry Department, Faculty of Sciences, University of Porto, Porto, Portugal

[5]Amalio Garrido Escudero Environmental Engineering and Toxicology Dpt., Catholic University of San Antonio, Guadalupe, Murcia,Spain

2.3.Date of QMRF update(s):

2.4.QMRF update(s):

2.5.Model developer(s) and contact details:

Alfonso Pérez-Garrido Environmental Engineering and Toxicology Dpt., Catholic University of San Antonio, Guadalupe, Murcia,Spain aperez@pdi.ucam.edu

2.6.Date of model development and/or publication:

December 2009

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

Pérez-Garrido A, Helguera AM, Caravaca G, Cordeiro MNDS & Escudero AG (2010). A TOPological Substructural MOlecular DEsign approach for predicting mutagenesis end-points of alpha, beta-unsaturated carbonyl compounds. Toxicology 268, 64–77.-unsaturated carbonyl compounds. Dental material. Accepted manuscript.

2.8.Availability of information about the model:

Training and test sets are available. Algorithm available.

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

3.Defining the endpoint - OECD Principle 1

3.1.Species:

Salmonella typhimurium

3.2.Endpoint:

4.Human Health Effects 4.10.Mutagenicity

3.3.Comment on endpoint:

Salmonella typhimurium TA98, TA100, TA1535, TA1537, TA97 either with or without a metabolic activation mixture. In addition, strains TA102 and TA1538 have been applied in cases where the results of other strains were equivocal.

3.4.Endpoint units:

no units

3.5.Dependent variable:

Ames =1 positive results; Ames =-1 negative results.

3.6.Experimental protocol:

Salmonella typhimurium reversed mutation assay based on standard Ames test (Ames et al., 1975; Maron and Ames 1983; Mortelmans and Zeiger, 2000). The analysis has been restricted to the standard plate or preincubation tests of Salmonella typhimurium strains TA98, TA100, TA1535, TA1537, TA97 either with or without a metabolic activation mixture. In addition, strains TA102 and TA1538 have been applied in cases where the results of other strains were equivocal.

3.7.Endpoint data quality and variability:

The data set was extracted from Kazius et al. (2005). In the classification, a compound was categorized as a mutagen if at least one the Ames test result was positive while a compound was categorized as nonmutagen if exclusively negative Ames test results one or more were reported.

4.Defining the algorithm - OECD Principle 2

4.1.Type of model:

QSAR

4.2.Explicit algorithm:

QSAR derived by two-group Linear Discriminant Analysis

$AMES = 1.864 + 1.882?_1{}^{Dip2}? 3.757 \cdot 10^{?3}?_6{}^{Dip2}?1.734 \cdot 10^{?2}?_5{}^{Hyd} + 9.489?_1{}^{Gas} + 3.705 \cdot 10^{?2}?_5{}^{Ab}??2H?$

$7.534 \cdot 10^{?12}?_0?_{15}{}^{Pol} + 1.426 \cdot 10^{?2}?_1?_3{}^{Hyd}$

4.3.Descriptors in the model:

TOPS-MODE it is based on the calculation of the spectral moments of the so-called bond matrix (Estrada, 1996 and 1997).

4.4.Descriptor selection:

The replacement method (Duchowicz, 2006) was the algorithm employed for variable selection. This was used to select the variables (descriptors) with the highest influence on mutagenicity, but in contrast to regression analysis, which minimizes the standard deviation, we minimized the Wilk's Lambda.

4.5.Algorithm and descriptor generation:

The spectral moments of the edge adjacency matrix are defined as the traces. That is the sum of the main diagonal of the different powers of such matrix. Several bond weights such as standard bond distance (Std), standard bond dipole moments (Dip, Dip2), hydrophobicity (H), polar surface area (Pols), polarizability (Pol), molar refractivity (Mol), van der Waals radii (vdW), and Gasteiger–Marsilli charges (Gas) were used for computing the spectral moments of the bond matrix.

4.6.Software name and version for descriptor generation:

Modeslab

http://www.modeslab.com/

4.7.Chemicals/Descriptors ratio:

175 chemicals / 7 descriptors =25. The pool of original descriptors was 676.

5.Defining the applicability domain - OECD Principle 3

5.1.Description of the applicability domain of the model:

The applicability domain was assessed by using the Williams plot, i.e. the plot of standardized residuals versus leverage values (h). The applicability domain is established inside a squared area within ? x standard deviations and a leverage threshold h*=0.136 (h* is generally fixed at 3p/n, where n is the number of training compounds and p the number of model parameters, whereas x = 3). See Pérez-Garrido et al. (2009b).

5.2.Method used to assess the applicability domain:

Method based on leverage values (Gramatica, 2007)

5.3.Software name and version for applicability domain assessment:

STATISTICA v 7.0

StatSoft

http://www.statsoft.com/

5.4.Limits of applicability:

Substances that had a higher leverage value than the threshold (h*=0.12) are outside of the applicability domain.

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: No

INChI: No

MOL file: No

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:

175 compounds: 82 positives; 93 negatives

6.6.Pre-processing of data before modelling:

6.7.Statistics for goodness-of-fit:

The goodness-of-fit was evaluated by checking:
accuracy: the percentage of all chemicals correctly identified by the model;
sensitivity: the percentage of mutagenic (positive) chemicals correctly identified (calculated out of the total number of positives);
specificity: the percentage of non-mutagenic (negative) chemicals correctly identified (calculated out of the total number of negatives)
squared Mahalanobis Distances (D^2); Wilk's lambda (?),
Fisher function, FIT(?) and Kappa (?)

The parameter FIT(?) is similar to Kubinyi function in regression analysis, defined by: FIT(?)=(1-?)(n-k-1)/(n+k^2)?. where n is the number of compounds in the training set, k is the number of variables in the equation that describe the model, and ? is the Wilk´s Lambda. The FIT(?) criterion has a low sensitivity toward changes in k values, as long as they are small numbers, and a substantially increasing sensitivity for large k values.

The ? index (Cohen, 1960) excludes matching due solely to chance. However, a commonly cited scale is represented in by Landis and Koch (1977):

?<0 Less than chance agreement
? between 0.01 and 0.20 Slight agreement? between 0.21 and 0.40 Fair agreement? between 0.41 and 0.60 Moderate agreement
? between 0.61 and 0.80 Substancial agreement? between 0.81 and 0.99 Almost perfect agreement
?=0.0.451; $p<10^{-5}$; F=28.958(Fisher function);
FIT(?)=0.869; ?=0.707 , D^2=4.818;
Sensitivity: 86.59%; Specificity: 91.40%; Accuracy: 89.14%; False positives= 13.41%; False negatives= 8.60%

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

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

The leave-group-out (LGO) procedure was applied, leaving out 20% of the training set by random extraction and then recalculating the model and the statistics with the remaining chemicals. This LGO procedure was repeated 300 times. The mean values of the accuracy, sensitivity, and specificity for both training and test sets, as well as the mean values of Wilk's ? ($?_{Cross}$) and squared Mahalanobis distances (D^2_{Cross}) , are reported. $?_{Cross}$=0.448; D^2_{Cross}= 4.913;

Sensitivity Training: 86.34%; Specificity Training: 90.95%; Accuracy Training: 88.79%; Sensitivity Test: 86.13%; Specificity Test: 88.46%; Accuracy Test: 87.36%

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: No

INChI: No

MOL file: No

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:

44 compounds: 21 positives, 23 negatives

7.6.Experimental design of test set:

k-Means Cluster Analysis (k-MCA) was used to extract the test set. This partition was the same used in QMRF (Q14-26-8-158) and it explained in Pérez-Garrido et al. (2009b)

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity: 81.00%; Specificity: 78.26%; Accuracy: 79.54%; False positives= 14%; False negatives= 13% with all compounds. Sensitivity: 85.00%; Specificity: 85.71%; Accuracy: 85.37%; False positives=15%; False negatives=14.39% without considering the compounds that are outside the domain of applicability.

7.8.Predictivity - Assessment of the external validation set:

Three compounds of the test set are outside of the aplicability domain: 96910-71-3, 23255-69-8 and 514-78-3 (6.81% of the test 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:

We note that the predominant mechanism is Michael type addition. Substituents in the alpha or beta-carbon atoms have a strong influence in mutagenicity just as for Michael acceptors.

8.2.A priori or a posteriori mechanistic interpretation:

A posteriori interpretation based on variables of the equation.

8.3.Other information about the mechanistic interpretation:

9.Miscellaneous information

9.1.Comments:

9.2.Bibliography:

[1]Duchowicz PR, Castro EA, Fernndez FM (2006). Alternative algorithm for the search of an optimal set of descriptors in qsar-qspr studies. MATCH Communications in Mathematical and in Computer Chemistry 55, 179–192.

[2]Gramatica P (2007). Principles of QSAR models validation: internal and external. QSAR & Combinatorial Science 26, 694-701.

[3]Cohen J (1960). A coefficient of agreement for nominal scales. Journal of Educational and Psychological Measurement 20, 37–46.

[4]Landis JR & Koch GG (1977). The measurement of observer agreement for categorical data. Biometrics 33, 159–174.

[5]McFarland JW & Gans DJ (1995). Chemometric methods in molecular design. pp. 295-307. VCH, Weinheim.

[6]Johnson RA & Wichern DW (1988). Applied MultiVariate Statistical Analysis. Prentice-Hall, New York.

[7]Ames BN, McCann H & Yamasaki E (1975). Methods for detecting carcinogens and mutagens with the Salmonella/mammalianmicrosome mutagenicity test. Mutation Research 31, 347-364.

[8]Maron DM & Ames B (1983). Revised methods for the Salmonella mutagenicity test. Mutation Research 113, 173-215.

[9]Mortelmans K & Zeiger E (2000). The Ames Salmonella/microsome mutagenicity assay. Mutatation Research 455, 29–60.

[10]Kazius J, McGuire R & Bursi R (2005). Derivation and validation of toxicophores for mutagenicity prediction. Journal of Medicinal Chemistry 48, 312–320.

[11]Estrada E (1996). Spectral Moments of the Edge Adjacency Matrix in Molecular Graphs. 1. Definition and Applications to the Prediction of Physical Properties of Alkanes. Journal of Chemical Information and Computer Sciences 36, 844–849.

[12]Estrada E (1997). Spectral Moments of the Edge-Adjacency Matrix of Molecular Graphs. 2. Molecules Containing Heteroatoms and QSAR Applications. Journal of Chemical Information and Computer Sciences 37, 320–328.

[13]Pérez-Garrido A, Helguera A M, Caravaca G, Cordeiro MNDS & Escudero A G (2009a). A TOPological Substructural MOlecular DEsign approach for predicting mutagenesis end-points of alpha, beta-unsaturated carbonyl compounds. Toxicology. Accepted manuscript.

[14]Pérez-Garrido A, Helguera A M, Girón-Rodríguez F & Cordeiro MNDS (2009b). Qsar models to predict mutagenicity of acrylates, methacrylates and alpha, beta-unsaturated carbonyl compounds. Dental material. Accepted manuscript.

9.3.Supporting information:

TOPS Mode Ames_training_176.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-410-0048/attachment/A713
TOPS Mode Ames_test_44.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-410-0048/attachment/A714

Test set(s)

10.Summary (JRC QSAR Model Database)

Supporting information

10.1.QMRF number:

Q13-410-0048

10.2.Publication date:

2013-06-28

10.3.Keywords:

Ames mutagenicity;TOPS-MODE;alpha;beta-unsaturated carbonyl compound;

10.4.Comments:

former Q14-37-8-303