	QMRF identifier (JRC Inventory): **Q13-410-0046**
	QMRF Title: **QSAR for Ames test of alpha, beta-unsaturated carbonyl compounds**
	Printing Date: **Dec 11, 2019**

1.QSAR identifier

1.1.QSAR identifier (title):

QSAR for Ames test of alpha, beta-unsaturated carbonyl compounds

1.2.Other related models:

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]Francisco Girón Rodríguez 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

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

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 A M, Girón-Rodríguez F & Cordeiro MNDS (2009). Qsar models to predict mutagenicity of acrylates, methacrylates and alpha, beta-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 = -0.499 nCb-+2.400 nRCHO+8.099 nArC=N + 5.533 nArNO2+4.671 nPO4+4.795 nCH2RX-3.830 nCXr +2.861 nCconjX-1.969

4.3.Descriptors in the model:

[1]nCb- number of substituted benzene C(sp2)

[2]nRCHO number of aldehydes (aliphatic)

[3]nArC=N number of imines (aromatic)

[4]nArNO2 number of nitro groups (aromatic)

[5]nPO4 number of phosphates/thiophosphates

[6]nCH2RX number of CH2RX

[7]nCXr number of X on ring C(sp3)

[8]nCconjX number of X on exo-conjugated C

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:

Descriptors were generated by the Dragon software and are based on the counting of 120 atom-cantered fragments, as defined by Ghose-Crippen

(Viswanadhan et al., 1989).

4.6.Software name and version for descriptor generation:

DRAGON

Calculation of several sets of molecular descriptors from molecular geometries (topological, geometrical, WHIM, 3D-MoRSE, molecular profiles, etc.)

Prof. R.Todeschini - distributed by Talete srl, via Pisani 13, 20124 Milano, Italy

http://www.disat.unimib.it/chm

4.7.Chemicals/Descriptors ratio:

176/8=22

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. (2009).

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:

StatSoft STATISTICA v 7.0

http://www.statsoft.com/

5.4.Limits of applicability:

Substances that had a higher leverage value than the threshold (h*=0.136) 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:

176 compounds: 83 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
?=0.493; p<10^{-5}; F=21.431(Fisher function); FIT(?)=0.714;
?=0.772 , D^2=4.073;
Sensitivity: 80.72%; Specificity: 91.39%; Accuracy: 86.36%; False positives=8.6%; False negatives=19.27%

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

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:

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. The training set contained 80% (176/220) of the original data whereas the test set the remaining 20%. The k-MCA analysis was separately made for each group: mutagenic and non-mutagenic. Selection of the training and test sets was then carried out by taking compounds belonging to each cluster, proportionally to the size of the cluster. The pool of descriptors was formed for the entire Dragon descriptors family. We also made an inspection of the standard deviation between and within clusters, the respective Fisher ratio and p level of significance (ought to be lower than 0.05) (McFarland and Gans, 1995, Johnson and Wichern, 1988). Table 1

Table 1. Standard deviation between and within clusters, degrees of freedom (df), Fisher ratio (F) and level of significance (p) of the variables in the k-means cluster analysis: see attached file

7.7.Predictivity - Statistics obtained by external validation:

Sensitivity: 85.71%; Specificity: 91.30%; Accuracy: 88.63%; False positives=8.69%; False negatives=14.28%

7.8.Predictivity - Assessment of the external validation set:

All compounds in the test set are within the limits of aplicability.

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:

Well-known structural alerts such as aromatic nitro groups (nArNO2) and aromatic amines, more specifically imines (nArC=N), aliphatic halogenated derivatives, especially the primary, known alkylating agents (nCH2RX), making such compounds mutagenic in the Ames test. Moreover, the halogen atom presence in the alpha or beta position of the double bound (nCconjX) increases the mutagenicity in the Ames test. Furthermore, as expected, aldehydes (nRCHO) are more mutagenic than other carbonyl groups, and the presence of this functional group increases the mutagenicity of the substance.

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.

9.3.Supporting information:

AMES Training_176.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-410-0046/attachment/A707
AMES Test_44.sdf	http://qsardb.jrc.ec.europa.eu/qmrf/protocol/Q13-410-0046/attachment/A708

Test set(s)

10.Summary (JRC QSAR Model Database)

Supporting information

10.1.QMRF number:

Q13-410-0046

10.2.Publication date:

2013-06-28

10.3.Keywords:

QSAR for Ames mutagenicity of alpha;beta-unsaturated carbonyl compounds;

10.4.Comments:

former Q14-26-8-158