



QSAR STUDIES ON DERIVATIVES OF QUINAZOLINE-4(3H)-ONES WITH ANTICONVULSANT ACTIVITIES

O. ADEDIRIN^{1&2}, A. UZAIRU², G.A. SHALLANGWA² and S.E. ABECHI²

¹Chemistry Advance Research Center, Sheda Science and Technology Complex, FCT, Nigeria

²Chemistry Department, Ahmadu Bello University, Zaria, Nigeria

Corresponding author: Chemistry Advance Research Center, Sheda Science and Technology Complex, FCT, Nigeria, Phone: +234800593145

E-mail address: adedirinoluwaseye@yahoo.com; senguade@gmail.com

ARTICLE INFO

Article history:

Received 2018-01-31

Accepted 2018-02-23

Available online 2018-06-30

Keywords

Epilepsy

Quantitative structure activity relationship

Kennard-Stone algorithm

Genetic function algorithm

Quinazolinone ring system

ABSTRACT

Quantitative structure-activity relationship study was done on some quinazoline-4(3H)-ones derivatives with anticonvulsant activity against maximal electroshock-induced seizure. The quinazoline derivatives used as dataset and their anticonvulsant activity value were obtained from the literature. The molecular structure of the dataset compounds was generated with Spartan 14 software. This was optimized with PM3 semi-empirical quantum mechanical method available in the software. Molecular descriptors were obtained from the optimized structures using the PaDEL-Descriptor software. Activity values of the compounds and molecular descriptors obtained from the optimized structure made up the database for the study. The database was divided into training and test sets with Kennard Stone algorithm. Genetic function algorithm was used to develop quantitative structure-activity relationship models. The best model obtained was stable, robust and had good statistical parameters including determination coefficient R^2 (0.899), adjusted determination coefficient R^2_{adj} (0.888), variance ratio F (82.03), leave one out cross-validated determination coefficient Q^2 (0.866) and predicted determination coefficient for the test set R^2_{pred} (0.7406). The model indicated that the anticonvulsant activity of the studied compounds was dependent on Broto-Moreau autocorrelation-lag2/weighted by Vander Waals volume (ATS2v), average coefficient sum of the last eigenvector from Barysz matrix/weighted by Vander Waals volume (VE2_DZv), largest absolute eigenvalue of Burden matrix-6/weighted by relative atomic mass (SpMax6_Bhm), average valence path of order 6 and radial distribution function at 4.5 interatomic distance weighted by first ionization potential (RDF45i).

1. INTRODUCTION

Epilepsy comprises a group of disorders characterized by the periodic and abnormal discharge of nervous tissue. It often manifests as violent involuntary muscle contractions and convulsions a phenomenon known as a seizure. Seizures can have many causes and constitute evidence of an underlying neurologic disorder. It is reported to affect about 50 million individuals worldwide of which about 10.5 million are children and almost 90% of these people are in the developing countries (THIRUMURUGAN *et al.*, 2006). Pharmacological management remains the primary method to treat epilepsy and many chemical agents known as antiepileptic drugs (AEDs) have been developed. However, despite the development these AEDs over 30% of people with epilepsy do not have seizure control and others do so only at the expense of significant dose-related toxicity and other adverse effects (BROWN AND HOLMES, 2001; KRAMER, 2001). Therefore, the search for new anticonvulsant drugs with fewer side effects has been continuous.

Quinazolinone ring system has been reported to have wide range of biological properties including antitumor, anti-HIV, selective estrogen beta modulator, anti-inflammatory, antibacterial, antidepressant and central nervous system depressant activities (BAVETSIAS *et al.*, 2007; AL-RASHOOD *et al.*, 2006; ALAGARSAMY *et al.*, 2007; GÜNGÖR *et al.*, 2006). This ring system is considered as an interesting moiety and many derivatives substituted by different heterocyclic moieties at 3rd position of their heterocyclic system has been reported to exhibit anticonvulsant property with reduced toxicity (EL-HAKIM *et al.*, 1994; EL-NASER-OSSMAN *et al.*, 1994).

The objective of the present study is to perform quantitative structure-activity relationship study (QSAR) on some newly synthesis quinazoline-4(3H)-ones derivatives reported in literature (GEORGEY *et al.*, 2008; MOHAMED, 2014) to possess anticonvulsant activity against maximal electroshock-induced seizure (MES). This was done to gain insight into factors that influence the activity of the studied compounds and provide a rationale for designing new compounds with improved activity value. QSAR analysis is an area of computational research concern with establishing a mathematical relationship between a quantitative measure of chemical structure (molecular descriptors) and a biological activity (ROY AND GHOSH, 2010). It has played an essential role in the development of compounds in medicinal chemistry, pharmaceutical. designs, drug metabolism and optimization of the molecular structure with defined purpose (SPECK-PLANCHE AND CORDEIRO, 2012).

2. MATERIAL AND METHODS

2.1 Database development

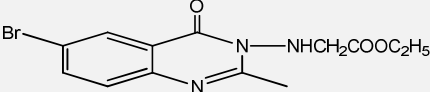
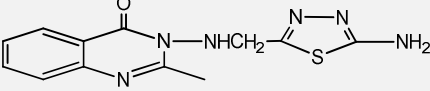
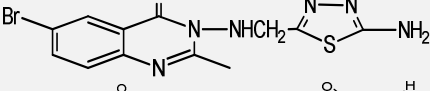
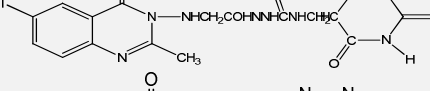
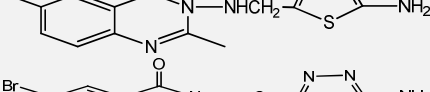
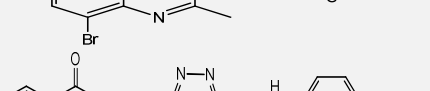
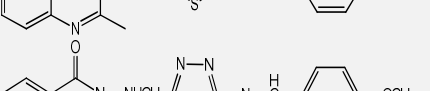
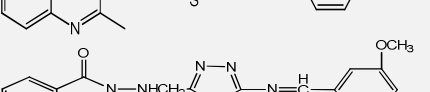
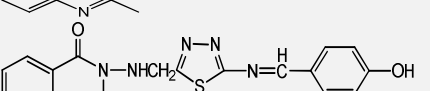
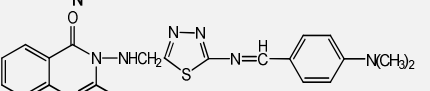
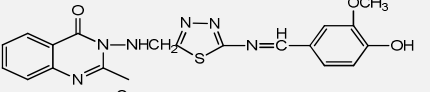
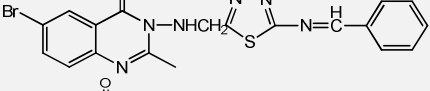
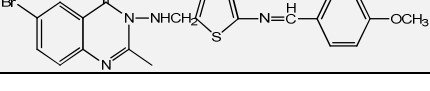

The dataset comprises derivatives of quinazoline-4(3H)-ones obtained from literature concerned only with the synthesis of the derivatives and their pharmacological test using similar assay (GEORGEY *et al.*, 2008; ARCHANA *et al.*, 2004; MOHAMED, 2014). The assay used was maximal electroshock seizure (MES) test on mice and compounds anticonvulsant activity values were reported as percent protection. This was recalculated to the logarithmic unit for comparability between

compounds and linear correlation of the activity value to free energy change (GRAMATICAL *et al.*, 2012; NOOLVI *et al.*, 2010). Logit transform was used to convert the activity value to logarithm unit:

$$BA = \text{Log} \left(\frac{MW}{d} \left[\frac{\% \text{ Protection}}{100 - \% \text{ Protection}} \right] \right) \quad (1)$$

In equation 1, MW is the molecular weight of the compound and d is the dose in mg/kg. Molecular structure and anticonvulsant activities in logarithm unit for the dataset compounds are presented in Table 1.

Table 1 Molecular structure of quinazoline-4(3H)-one derivative and anticonvulsant activities

S/N	Compound Structure/Name	BA
*1		6.687
2		6.807
3		7.018
4		7.281
5		6.964
6		7.172
*7		7.099
*8		7.499
9		7.308
10		7.293
11		7.321
12		7.517
13		7.358
14		7.577

Continued Table 1

S/N	Compound Structure/Name	BA
15		7.385
16		7.564
17		7.588
18		7.591
19		7.353
20		7.573
21		7.573
22		7.794
23		7.584
*24		7.821
25		7.483
*26		7.146
*27		7.873
*28		7.862
29		7.882
30		8.237
31		7.459
*32		7.352

Continued Table 1

S/N	Compound Structure/Name	BA
33		6.673
*34		7.424
*35		6.997
36		7.621
37		7.866
38		8.266
*39		8.262
*40		8.013
41		7.834
*42		7.306
43		7.860
44		7.172
45		7.222
46		7.622
*47		7.354
48		7.355
49		7.476

Continued Table 1

S/N	Compound Structure/Name	BA
50		7.497
*51		7.465
52		6.509
53		6.777
*54		6.983
55		7.611
56		7.672
57		7.364
58		7.469
59		7.624
*60		7.918
61		7.414
62		7.013
63		7.331
64		7.635
65		7.404

Continued Table 1

S/N	Compound Structure/Name	BA
66		7.659
67		7.171
68		7.647
69		7.844
*70		8.257

Data used as the test set are designated with asterisk *, BA is the anticonvulsant activities in logarithm unit

2.2 Molecular structure generation and descriptor calculation

With the specific end goal of developing QSAR model that can predict accurately the anticonvulsant activities of the data set molecules, the 3D structure of the dataset molecules was generated with Spartan 14 software (SHAO *et al.*, 2006) and their geometry optimized with the semi-empirical quantum mechanical method available in the software. Optimization was performed until the root mean square (RMS) gradient for each compound was smaller than 10^{-6} (au). This was done to give a better description of dataset compounds molecular orbital wavefunction and obtain reliable data on electronic properties of each compound (GAVERNET *et al.*, 2007; CHOUDHARY AND SHARMA, 2014). The optimized structures were ported to PaDEL-Descriptor software (YAP, 2011) to compute around 1875 different physicochemical, topological and structural molecular descriptors. The activity values and the descriptor constitute the database for the study.

2.3 Database division

The database comprises of activity value and descriptors for compounds arranged in a $n \times m$ matrix where n is the number of compounds and m the number of descriptors. It was divided into training and test sets with Kennard and Stone algorithm available in Datasetdivision 1.2 software (AMBURE *et al.*, 2015). This algorithm was reported to produce excellent dataset division results (ROY *et al.*, 2008; ROY, 2007). The algorithm proceeded by finding the Euclidean distances d_{ij} between the molecular descriptor vectors of each pair of compounds i, j :

$$d_{ij} = \|\mathbf{X}_i - \mathbf{X}_j\| = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad (2)$$

In equation 14, k is the number of descriptors, x_{ik} , x_{jk} were similar descriptor contained in the \mathbf{X}_i , \mathbf{X}_j descriptor vector respectively. Once the distances had been calculated, two compounds that were farthest apart in terms of the measured distance were selected i.e. the pair I, J with the largest value of d_{ij} . Compounds that exhibit the largest minimum distance with respect to the two previously selected compounds were selected and placed in the training set. These steps were repeated until

the desired number of compounds had been added to the training set and the remaining compounds were used as the test set (KENNARD AND STONE 1976).

2.4 Variable selection/model development

The training set data only were used for selection of optimal descriptors. Prior to variable selection, descriptors of the training set data were scaled with equation 3 below. This gave the descriptors equal opportunity to possibly influence the models (TROPSHA *et al.*, 2003; TROPSHA 2010).

$$X^1 = \frac{(X_i - \bar{X})}{\sigma} \quad (3)$$

In equation 3, X^1 is the standardized descriptor values, X_i is the un-standardized descriptor values, \bar{X} and σ are the average and standard deviation value for each column of descriptors X respectively.

To decide the best descriptor variables combination that will produce the most statistically robust linear QSAR models, GFA strategy accessible in Materials Studio 7.0 software was utilized. GFA is a popular method that combines Holland's genetic algorithm to scan over the descriptors space for conceivable QSAR models and uses lack of fit (LOF) function from Friedman's multivariate adaptive regression splines algorithm to appraise the fitness of each model (ROGERS AND HOPFINGER, 1994). In the software, GFA analysis condition was set such that the equation length range (5 - 12), population (10000), maximum generation (500), number of top equation returned (3), mutation probability (0.1) and scaled LOF smoothness parameter (0.5). The following steps were performed: (1) initial population of equations were generated by random number of descriptors, (2) pairs from the population of equations were chosen at random, crossovers were performed and progeny equations were generated, (3) the fitness of each progeny equation was assessed by lack of fit (LOF) score that automatically penalizes models with too many features. The fitness function, i.e., lack-of-fit is calculated by

$$LOF = \frac{LSE}{(1 - \frac{c + dp}{M})^2} \quad (4)$$

In equation 4, c is the number of basis functions, d is the smoothing parameter, M is the number of samples in the training set, LSE is the least square error and p is the total number of features contained in all basis functions.

2.5 Model validation

The quality, reliability and predictive ability of the models reported by GFA module were evaluated with different internal and external validation techniques and parameters. On each combination of descriptors, multiple linear regressions (MLR) and correlation analyses were performed at 0.05 level of significance. The quality of each model was then examined by the following parameters:

Determination coefficient or square of the correlation coefficient (R^2): describes the fraction of the total variation attributed to the model. The closer the value of R^2 is to 1.0, the better the regression equation explains the Y variable (TROPSHA *et al.*, 2003).

$$R^2 = \frac{[\sum\{(Y_{obs} - \bar{Y}_{obs}) \times (Y_{pred} - \bar{Y}_{pred})\}]^2}{\sum(Y_{obs} - \bar{Y}_{obs})^2 \times \sum(Y_{pred} - \bar{Y}_{pred})^2} \quad (5)$$

In Equation 5, Y_{obs} and Y_{pred} correspond to the observed (i.e. experimental) and predicted response values respectively of the training compounds

Explained variance (R^2_{adj}): a modified form of determination coefficient which accounts for the effect of new explanatory variables in the model, by incorporating a degree of freedom to the model (TROPSHA *et al.*, 2003).

$$R^2_{adj} = \frac{(N-1) \times R^2 - p}{N-1-p} \quad (6)$$

In Equation 6, N is the number of the molecule in the data, R^2 is the determination coefficient, p is the number of descriptors in the model and $N-1-p$ is the degree of freedom.

Variance ratio (F): the ratio of regression mean square to the deviation mean square.

$$F = \frac{\frac{\sum(Y_{pred} - \bar{Y}_{obs})^2}{p}}{\frac{\sum(Y_{obs} - Y_{pred})^2}{N-p-1}} \quad (7)$$

In Equation 7, Y_{obs} and Y_{pred} represent experimental and predicted response values respectively, \bar{Y}_{obs} is the mean response value for the training set data, N is the number of the molecule in the data, p is the number of descriptors in the model and $N-1-p$ is the degree of freedom. The p -value for every regression coefficient should be significance at $p < 0.05$ and the overall significance of the regression equation F -value should be high (ROY *et al.*, 2016).

The standard error of estimate (SEE): standard deviation of the model defined. Low SEE is an indication of a good model (TROPSHA, 2010).

$$SEE = \sqrt{\frac{(Y_{obs} - Y_{pred})^2}{N-p-1}} \quad (8)$$

Variance inflation factor (VIF): the inverse of tolerance used to detect the presence of multicollinearity in a model.

$$\text{Tolerance} = 1 - R_j^2 \quad (9)$$

$$VIF = \frac{1}{1 - R_j^2} \quad (10)$$

In equation 8–10, R_j^2 is the coefficient of determination of a regression of descriptor j on all the other descriptors. If the tolerance value is less than a preset cut-off value (e.g., 0.1) or the VIF is higher than a cut-off value (e.g., 10), a multicollinearity problem exists in the descriptor set (SIVAKUMAR *et al.*, 2007).

2.6 Leave one out cross-validation test

To further investigate the internal robustness of the models, leave one out (LOO) cross-validation was done on the training set data. In perspective, various modified data sets were obtained from the original training set data by removing one molecule for each case. For every new data set generated, a QSAR model is developed using the same statistical method. This new model is applied to predict the activity value of the removed molecule. The obtained LOO predicted activity values are used to calculate the following parameters:

Predicted error sum of square (PRESS)

$$\text{PRESS} = \sum (Y_{\text{obs}} - Y_{\text{pred}})^2 \quad (11)$$

Standard deviation of error of prediction (SDEP)

$$\text{SDEP} \equiv \text{RMSEP} = \sqrt{\frac{\text{PRESS}}{n}} \quad (12)$$

Cross-validated squared correlation coefficient Q^2

$$Q^2 = 1 - \frac{\text{PRESS}}{\sum (Y_{\text{obs}} - \bar{Y}_{\text{obs}})^2} \quad (13)$$

In Equation 11 – 13, Y_{obs} is the experimental response values for the removed data, Y_{pred} is the predicted response value for the removed data using the new model obtained from the reduced training set and \bar{Y}_{obs} is the mean of the experimental response value for the training set data. A predictive model should have Q^2 value greater than 0.5 (TROPSHA *et al.*, 2003).

2.7 Y-randomization test

This strategy was used to check if the reported models are products of chance correlation. During this process, activity values of the training set data were randomly permuted without making any change to the descriptor matrix and this was followed by multiple linear regression analysis. Each randomization runs yield estimates of R^2 and Q^2 values, which were expected to be low than that of the un-randomized data (TROPSHA, 2010).

Randomization parameter designated ${}^cR_p^2$:

$${}^cR_p^2 = R^2 \times \left(1 - \sqrt{|R^2 - \bar{R}_r^2|} \right) \quad (14)$$

In Equation 14, R^2 is the square correlation coefficient for the regression analysis of non-randomized model and \bar{R}_r^2 is the average of the squared correlation coefficient for the regression analysis of all randomization runs. A model is said to be acceptable if ${}^cR_p^2$ is greater than 0.5.

2.8 Validation through the test set

The test set descriptors were scaled as described for the test set. The scaled values were consolidated into the proposed models and predicted activity for the test set data were calculated ($Y_{\text{pred}(\text{test})}$). Observed activity values ($Y_{\text{obs}(\text{test})}$) were regressed with predicted ($Y_{\text{pred}(\text{ext})}$) with different conditions. The resultant regression coefficients (r^2) and slopes (k) were used to judge the predictive ability of the models according to GOLBRAIKH and TROPSHA (2002) criteria which stated that a model as good predictive capacity if the following conditions are true:

$$R_{\text{pred}}^2 = 1 - \frac{\sum (Y_{\text{obs}(\text{test})} - Y_{\text{pred}(\text{test})})^2}{\sum (Y_{\text{obs}(\text{test})} - \bar{Y}_{\text{obs}(\text{train})})^2} > 0.6 \quad (15)$$

$$Q^2 > 0.5$$

$$r^2 - r_0^2/r^2 < 0.1 \text{ and } 0.85 \leq k \leq 1.15 \text{ or } r^2 - r'^2/r^2 < 0.1 \text{ and } 0.85 \leq k' \leq 1.15$$

$$|r^2_0 - r'^2_0| < 0.3$$

In equation 15, $\bar{Y}_{\text{obs}(\text{train})}$ is the mean response value for the training set data. In the criteria, r^2 and r_0^2 are regression coefficient of the plot of observed against predicted response for the test set data with and without intercept respectively, while k is the slope of this plot without intercept. r'^2 and r'^2_0 are regression coefficient of the plot of predicted against observed response for the test set data with and without intercept respectively and k' is the slope of this plot without intercept.

2.9 Applicability Domain (AD)

QSAR models must always be verified for their applicability with regard to the chemical domain of dataset compounds (NETZEVA *et al.*, 2005). In this study, AD of the model was defined by extrapolation leverage approach. Hat matrix was used to calculate the leverage value for each molecule in order to identify structural outlier in the dataset compounds. Identification of response outliers was done by calculating cross-validated standardized residuals for each compound and ± 3 was used as a threshold. The plot of the standardized residuals versus leverages (Williams plot) was used to show the AD of the model pictorially. The threshold value for the leverages h^* was set to $3(m + 1)/n$ where m is the number of descriptors in the model and n is the number of the training set compounds. Generally, any compound in the data set with leverage value h_i greater than the threshold h^* is outside the structural domain of the training set, therefore, it is an influential compound, predicted data for that compound is extrapolated by the model and could thus be less reliable. Furthermore, any compound in the data set with standardized residual greater ± 3 is outside the response domain of the training set, therefore, is an outlier to the model.

3. RESULT AND DISCUSSIONS**3.1 Dataset**

Kennard and Stone algorithm applied in the study divide the database into 52 training set and 18 test set compounds designated with an asterisk in Table 1. Table 2 presents the result of single column statistics performed on the activity values for both training and test set. It indicated that the maximum activity value in test sets was less than training set maximum value. Also, the minimum activity value in test sets was greater than the training set minimum. This indicated that the test set is interpolative of the training set i.e. obtained within the training set range. The similarity in mean and standard deviation of both sets showed that point distribution and spread in both sets were comparable.

3.2 Model

The best model produced by the GFA method used for model construction is presented below:

$$\begin{aligned} \text{BA} = & 7.441(\pm 0.017) - 0.478(\pm 0.050) \text{ATS2v} \\ & - 0.465(\pm 0.031) \text{VE2_DZv} \\ & + 0.324(\pm 0.028) \text{SpMax6_Bhm} \\ & - 0.237(\pm 0.020) \text{ASP-6} \\ & + 0.228(\pm 0.028) \text{RDF45i} \end{aligned} \quad (16)$$

In equation 16, values in the parenthesis are the standard deviation. The model was obtained from 52 training set compounds and contains 5 molecular descriptors. Therefore, it obeys the QSAR rule of thumbs which state that for the problem of chance correlation to be acceptably low, the ratio of

descriptors in the model to compounds used to build the model should not exceed 1:5 (TOPLISS AND CASTELLO, 1972). Also contained in Table 2 are the validation parameters for the model. As reported in the table, models determination coefficient R^2 , adjusted determination coefficient R^2_{adj} , leave one out cross-validated determination coefficient Q^2 and modified leave one out cross validation coefficient $R^2_{m(loo)}$ were greater than 0.5. This indicated that the model was not over parameterized and has good internal predictive ability. The value of R^2_{adj} (0.899) showed that model explained over 80% of the variation in training set data. The model variance ratio F-value (82.03) is far greater than critical F-value (2.39) for (5, 52) degree of freedom at $p < 0.05$. This suggested that variation in the activity explained by combinations of descriptors in the model is not attributed to chance. Furthermore, randomization parameter $^cR^2_p$ value is greater than 0.5, average randomization runs determination coefficient \bar{R}^2_{rand} and average randomization runs leave one out cross-validated determination coefficient \bar{Q}^2_{rand} are far less than 0.5. This further confirmed that the model is not a product of chance (ROY et al., 2012; ROY et al., 2015).

External validation result showed that models predicted square correlation coefficient R^2_{pred} , modified square correlation coefficient for test set $R^2_{m(test)}$ and overall modified square correlation coefficient $R^2_{m(overall)}$ were greater than 0.6, indicating the model has good external predictive ability. Also reported in the table, the model passed the GOBRAIKH and TROPSHA (2002) criteria for an externally predictive model. Furthermore, the model's descriptors variance inflation factor (VIF), t-statistics, p-values are presented in Table 3. The table showed that p-values for the descriptors were less than 0.05 and their t-statistics values were greater than 2 at 95% confidence level, indicating all descriptors are contributed significantly to the model. In addition, their VIF values were less than 10, meaning the model is void of multi-co-linearity problem and descriptors were reasonably orthogonal to each other (PABLO et al., 2012). Fitness plot for the model showing a linear relationship between predicted activity values by the model and observed experimental activity is presented in Fig 1.

Table 2-model's quality and validation statistics

	Single column statistics			
	Max	Min	Mean	St. dev.
Train set	8.266	6.508	7.449	0.344
Test set	8.263	6.687	7.518	0.455
	Validation parameters			
	Internal validation		External validation	
R^2	0.899	R^2_{pred}	0.706	
R^2_{adj}	0.888	r^2	0.713	
F	82.03	r^2_0	0.702	
Q^2	0.866	r'^2_0	0.671	
SDEP	0.125	$R^2_{m(test)}$	0.626	
SEE	0.115	$R^2_{m(overall)}$	0.764	
PRESS	0.809	$ r^2_0 - r'^2_0 $	0.031	
LOF	0.066	k	1.003	
$R^2_{m(loo)}$	0.845	$r^2 - r^2_0/r^2$	0.016	
\bar{R}^2_{rand}	0.113	k'	0.996	
\bar{Q}^2_{rand}	-0.165	$r^2 - r'^2/r^2$	0.059	
$^cR^2_p$	0.847	R	0.948	

Williams plot presented in Fig. 2 depicts the model's applicability domain. The plot showed that almost all dataset compounds had leverage values less than the warning/threshold leverage h^* (0.35) except for a compound in the training set.

This compound is an influential point but not an outlier with respect to the standardized residual (OPREA, 2005; TODESCHINI and CONSONNI, 2009). In general, the model presented in the study was robust and had good internal and external predictive ability and thus, can be utilized to predict the anticonvulsant activities of compounds in its applicability domain.

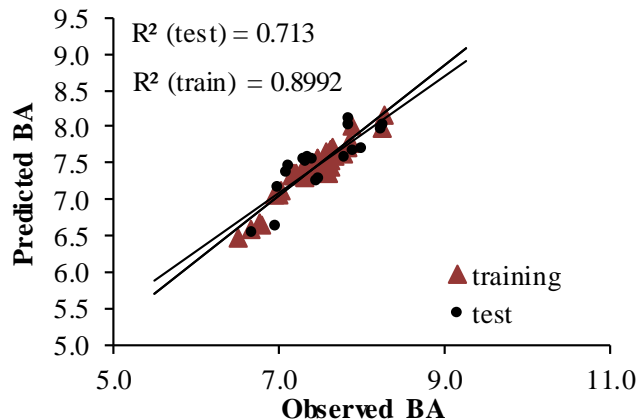


Figure 1 Fitness curve for the model

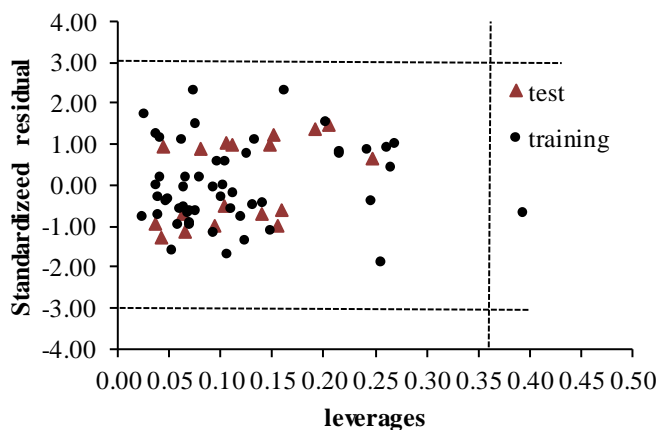


Figure 2 Williams plot for the model

3.3 Interpretation of descriptors

Mean effect was used to evaluate the relative importance and contribution of each descriptor to the model (POURBASHEER et al., 2009; RIAHI et al., 2009). It was calculated using the equation below (HABIBI AND DANADEH, 2009):

$$MF_j = \frac{\beta_j \sum_{i=1}^n d_{ij}}{\sum_j^m \beta_j \sum_i^n d_{ij}} \quad (17)$$

In equation 17, MF_j is the mean effect of a descriptor j , β_j is the coefficient of the descriptor J in the model, d_{ij} is the value of the descriptor in the data matrix, m is the number of descriptors that appear in the model and n is the number of molecules in the training set. Brief description descriptors contained in the model and their corresponding MF values were included in Table 3. The first descriptor in the model is ATS2v defined as Broto-Moreau autocorrelation-lag2/weighted by Vander Waals volume (TODESHINI AND COSONNI, 2009). It indicates that changing the van der Waals volume two vertices apart affect the biological activities of the molecules, thus changing the type of atoms or increase in the complexity of vertices that are two bonds apart will affect the anticonvulsant activity of the molecules. This descriptor had absolute MF (10.10) value - the highest among the descriptors. This implied

it had the highest contribution to the anticonvulsant activity of dataset molecule. However, its negative sign showed that increase in the magnitude of this descriptor increases the percent protection of studied compounds. Therefore, increase in

molecular linearity and taken into account the van der Waals volume of added substituent have great influence on the anticonvulsant activity of studied compounds.

Table 3- Model's descriptors definition, regression statistics and mean effect values

Symbol	Meaning	t-stat	p-value	VIF	MF
ATS2v	Broto-Moreau autocorrelation-lag2/weighted by Vander Waals volume	-9.470	2.2E-85	9.721	-10.10
VE2_DZv	Average coefficient sum of the last eigenvector from Barysz matrix/weighted by Vander Waals volume	-15.13	1.7E-19	3.091	9.254
SpMax6_Bhm	Largest absolute eigenvalue of Burden matrix-6/weighted by relative atomic mass	11.43	4.9E-15	3.039	5.467
ASP-6	Average valence path of order 6	-11.61	2.8E-15	1.649	-5.311
RDF45i	Radial distribution function at 4.5 interatomic distance weighted by first ionization potential	8.078	2.3E-10	3.553	1.685

The second descriptor is VE2_DZv defined as the average coefficient sum of the last eigenvector from Barysz matrix/weighted by Vander Waals volume (TODESHINI AND COSONNI, 2009). It is a spectral index obtained from the coefficient of the eigenvector associated with the largest negative eigenvalue of a given graph-theoretical matrix (Barysz) and the weighting scheme w is Vander Waals volume. It provides discrimination among graph vertices and lower values of this descriptor correspond to vertices of lower degree, farther from the center or from a vertex of higher degree. This implied it is an index of branching and its lower value corresponding to chain graph and a higher value corresponding to increased branching (TODESHINI AND COSONNI, 2009). The MF value of this descriptor was 9.25 (second among other descriptors). It contributed relatively high to the model and its positive sign implies a higher numerical value of this descriptor decreases the percent protection of the studied compounds. Therefore, increased branching in studied compounds may reduce their anticonvulsant activity values.

The third descriptor is SpMax6_Bhm defined as a largest absolute eigenvalue of Burden matrix-6/weighted by relative atomic mass (TODESHINI AND COSONNI, 2009). It's also a spectral index defined in terms of eigenvalues of a given graph-theoretical square matrix M (Burden matrix). Also, it is an index of molecular branching with smallest values corresponding to chain graphs and the highest to the most branched graphs (TODESHINI AND COSONNI, 2009). It had MF value of 3.039 (fourth position in relation to others) and its positive sign supported the second descriptor. Addition of highly branched substituent may increase the atomic mass of the studied compound and also increase the value of the descriptor. This effect also increases BA which means a decrease in percent protection of the molecule.

The fourth descriptor is ASP-6, defined as average valence path of order 6 (YAP, 2011; KIER AND HALL, 1976). It's a molecular connectivity descriptor obtained from hydrogen-suppressed valence-weighted graph defined for distinct edges and vertices equal to or greater than six in a sequence. It is also an index of branching, however, its value decreases with increased branching. That means opposite behavior compared to VE2_DZv and SpMax6_Bhm. This could inform its negative MF (-5.311) value and increase in the magnitude of the absolute value of this descriptor in the model increases the percent protection of studied molecule. It was

reported that presence of heteroatom like O and N increases the magnitude of the absolute value of this descriptor depending on the number of hydrogen atoms attached to them and their orbital type. Also, the presence of sulfur and halogen increases the magnitude of the absolute value of this descriptor (KIER AND HALL, 1976). Furthermore, the presence of multiple bonds also increases the magnitude of the absolute value of this descriptor.

The last descriptor is RDF45i defined as radial distribution function at 4.5 interatomic distance weighted by first ionization potential (YAP, 2011; TODESHINI AND COSONNI, 2009). It is a 3D descriptor which provides information about steric hindrance, or structure/activity properties, distribution of interatomic distances, bond distances, ring types, planar and non-planar systems, and atom types in a compound depending on the molecular property included in the function as the weighting scheme (TODESHINI AND COSONNI, 2009). RDF45i showed that a linear relationship exists between the anticonvulsant activity values of the studied compound and 3D distribution of the amount of energy required to remove the most loosely bounded electron, valence electron of isolated gaseous atoms in the molecule. It had positive MF (1.685) values (least in relation to others). Its positive MF implied its higher numerical value decreases percent protection of the studied compounds.

4. CONCLUSION

In the present study, QSAR analysis of some quinazoline derivatives with anticonvulsant activity against MES induced seizure was carried out and ATS2v, VE2_DZv, SpMax6_Bhm, ASP-6 and RDF45i molecular descriptors were identified to influence the activity of the studied compounds. The reliability and applicability of the model were defined and found to be good with R^2 (0.899), Q^2 (0.866), F (82.03) and R^2_{pred} (0.706). Attempt to interpret descriptors contained in the model showed that anticonvulsant activity of the studied compounds was influenced by the degree of branching and the electronic nature of the atoms in the molecules. The study provides insight into factors that influenced the activity of study compounds. It also provides tools that can serve as a knowledge generator applicable for *In silico* designed for new molecules within AD of the model that in principle would have improved anticonvulsant activities. Thus, serving as a guide before the more expensive *in vivo* and *in vitro* experiment is embarked upon.

REFERENCES

- ALAGARSAMY, V.; DHANABAL, K.; PARTHIBAN, P.; ANJANA, G.; DEEPA, G. MURUGESAN, B.; RAJKUMAR, S.; BEEVI, A. J. Synthesis and Pharmacological Investigation of Novel 3-(3-Methylphenyl)-2- substituted amino-3H-quinazolin-4-ones as an analgesic and anti-inflammatory agents. **Journal of Pharmacy and Pharmacology**, v. 59, p. 669-677, 2007.
- AL-RASHOOD, S. T.; ABOLDAHAB, I. A.; NAGI, M. N.; ABOUZEID, L. A.; ABDEL-AZIZ, A. A. M.; ABDELHAMID, S. G.; YOUSSEF, K. M.; AL-OBAID, A. M.; EL-SUBBAGH, H. I. Synthesis, dihydrofolate reductase inhibition, antitumor testing, and molecular Modeling Study of Some New 4(3H)-quinazolinone Analog, **Bioorganic and Medicinal Chemistry**, v. 14, p. 8608-8621, 2008.
- AMBURE, P.; AHER, R. B.; GAJEWICZ, A.; PYZYT, W. R. K. NanoBRIDGES software: open access tools to perform QSAR and nano-QSAR modeling. **Chemometry and Intelligent Laboratory System**, v147, p. 1-13, 2015.
- ARCHANA, S. V. K.; KUMAR, A. Synthesis of newer thiadiazolyl and thiazolidinonyl quinazolin-4(3H)-ones as potential anticonvulsant agents. **European Journal of Medicinal Chemistry**, v. 37, p. 873- 882, 2002.
- ARCHANA, S. V. K.; KUMAR, A. Synthesis of some newer derivatives of substituted quinazolinonyl-2-oxo/thiobarbituric acid as potent anticonvulsant agents. **Bioorganic and Medicinal Chemistry**, v. 12, p. 1257-1264, 2004.
- BAVETSIAS, V.; HENDERSON, E. A.; MCDONALD, E. Cyclopenta[g]quinazolinone-Based Inhibitors of Thymidylate Synthase Targeting α -Folate Receptor Overexpressing Tumours: Synthetic Approaches to 4-[N[(6RS)-2-Hydroxymethyl-4-oxo-3,4,7,8-tetrahydro-6H-cyclopenta[g]quinazolin-6-yl]-N-(prop-2-ynyl)amino}benzoic acid. **Tetrahedron**, v.63, p.1537-1543, 2007.
- BROWN, T. R.; HOLMES, G. L. N. Epilepsy, **English Journal of Medicine**, v. 344, p. 1145-1151, 2001.
- CHOUDHARY, M.; SHARMA, B. K. QSAR rationales for the 5-HT6 antagonistic activity of Epiminocyclohepta [b] indoles. **Journal of Pharmaceutical Chemistry**, v. 6, p.321-330, 2014.
- EL-HAKIM, A. E.; ABDEL-HAMIDE, S. G.; AL-AZHAR EL-HELBY, A. A. **Journal of Pharmaceutical Science**, v.14, p.156-163, 1994.
- EL-NASER, O.A.R.; EL-SAYED, B.S. Synthesis and anticonvulsant activity of some new 3-(p-Sulfamoylphenyl)-4(3H)-Quinazolinones. **Arzneimittelforsch**, v. 44, p. 915- 919, 1994.
- GAVERNET, L.; CABRERA, M. J. D.; BRUNO-BLANCH, L. E.; ESTIÚ, G. L. 3D-QSAR design of novel antiepileptic sulfamides. **Bioorganic and Medicinal Chemistry**, v. 15,n.3, p. 1556-1567, 2007.
- GEORGEY, H.; ABDEL-GAWAD, N.; ABBAS, S. Synthesis and anticonvulsant activity of some quinazolin-4-(3H)-one derivatives. **Molecules**, v. 13,p. 2557-2569, 2008.
- GRAMATICA, P.; CASSANI, S.; ROY, P.P.; KOVARICH, S.; YAP, C.W.; PAPA, E. QSAR Modeling is not “push a button and find a correlation”: A case study of toxicity of (Benzo)-triazoles on algae, **Molecular Informatics**, v.31, p. 817-835, 2012.
- GÜNGÖR, T.; CHEN, Y.; GOLLA, R.; MA, Z.; CORTE, J.R.; NORTHROP, J.P.; BIN, B.; DICKSON, J.K.; STOUCHE, T.; ZHOU, R.; JOHNSON, S.E.; SEETHALA, R.; FEYEN, J.H.M. Synthesis and Characterization of 3-Arylquinazolinone and 3-Arylquinazolinethione Derivatives as selective estrogen receptor beta modulators. **Journal of Medicinal Chemistry**,v. 49, p. 2440-2455, 2006.
- HABIBI, Y. A.; DANANDEH, J. M. Application of a genetic algorithm and an artificial neural network for global prediction of the toxicity of phenols to *Tetrahymena pyriformis*. **Monatsh Chemistry**, v. 140, p. 1279-1288, 2009.
- KEIR, L. B.; HALL, L. H. Molecular connectivity in chemistry and drug research. Academic press 111fifth Avenue, **New York**,v.14, p.47-89, 1976.
- KENNARD, R. W.; STONE, L. A. Computer aided design of experiments. **Technometrics**, v.11, p. 137-148, 1969.
- KRAMER, G. Epilepsy in the elderly: some clinical and pharmacotherapeutic aspects, **Epilepsia**, v. 42, p. 55-59, 2001.
- MOHAMED F. Z. New fluorinated quinazolinone derivatives as anticonvulsant agents. **Journal of Taibah University Medical Sciences**, v.9, n.2, p.104-109, 2014.
- NOOLVI, M. N.; PATEL, H. M.; BHARDWAJ, V. 2D qsar studies on a series of 4-anilino quinazoline derivatives as tyrosine kinase (egfr) inhibitor: an approach to design anticancer agents, **Digest Journal of Nonmaterial and Biostructures**, v. 5, n.2, p. 387 - 401, 2010.
- OPREA, T. I. (2005) Chemoinformatics in drug discovery, WILEY-VCH Verlag GmbH and Co, **Weinheim**, v. 23, p. 437-453, 2005.
- PABLO, R. D.; IBEZIM, E.; ERLINDA, V. O.; EDUARDO, A. C. QSAR on aryl-piperazine derivatives with activity on malaria, **Chemometrics and Intelligent Laboratory Systems**, v. 110, p. 81-88, 2012.
- POURBASHEER, E.; RIAHI, S.; GANJALI, M. R.; NOROUZI, P. Application of genetic algorithm-support vector machine (GA-SVM) for prediction of BK-channels activity. **European Journal of Medicinal Chemistry**, v. 44, p. 5023-5028, 2009.
- RIAHI, S.; POURBASHEER, E.; GANJALI, M. R.; NOROUZI, P. Investigation of different linear and non-linear chemometric methods for modeling of retention index of essential oil components: concerns to support vector machine. **Journal of Hazardous Material**, v. 166, p.853-859, 2009.
- ROGERS D.; HOPFINGER, A. J. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships, **Journal of Chemical Information and Computational Science**, v.34,p. 854-866, 1994.
- ROY, K. On some aspects of validation of predictive QSAR models. **Expert Opinion in Drug Discovery**, v.2, p.1567-1577, 2007.
- ROY, K.; KAR, S.; DAS, R. N. A primer on QSAR/QSPR modeling fundamental concepts, **Springer**, p. 48-58, 2015.
- ROY, K.; MITRA, I.; KAR, S.; OJHA, P. K.; DAS, R. N.; KABIR. H. Comparative studies on some metrics for external validation of QSPR models, **Journal of**

- Chemical Informatics and Modeling**, v. 52, p. 396–408, 2012.
- ROY, K.; DAS, R. N.; AMBURE, P.; AHER, R. B. Be aware of error measures. Further studies on validation of predictive QSAR models. **Chemometrics and Intelligent Laboratory Systems**, v.152, p.18–33, 2016.
- ROY, K.; GHOSH, G. Exploring QSARs with extended topochemical atom (ETA) indices for modeling chemical and drug toxicity, **Current Pharmaceutical Discovery**, v.16, p.2625-2639, 2010.
- ROY, P. P.; LEONARD, J. T.; ROY, K. Exploring the impact of training sets for the development of predictive QSAR models. **Chemometrics and Intelligent Laboratory Systems**, v.90, p.31-42, 2008.
- SHAO, Y.; MOLNAR, L. F.; JUNG, Y.; KUSSMANN, J.; OCHSENFELD, C.; BROWN, S.T. Advances in methods algorithms in modern quantum chemistry program package. **Physics Chemistry Chemistry Physics**, v. 8, p. 3172, 2006.
- SIVAKUMAR, P. M.; BABU, S. K. G.; MUKESH, D. QSAR studies on chalcones and flavonoids as anti-tuberculosis agents using genetic function approximation (GFA) method. **Bulletin of Pharmaceutical Chemistry (Tokyo)**, v.55, p 44–49, 2007.
- SPECK-PLANCHE, A.; CORDEIRO, M. N. Computer-aided drug design methodologies toward the design of anti-hepatitis C agents. **Current Topic on Medicinal Chemistry**, v. 12, p. 802-813, 2012.
- THIRUMURUGAN, R.; SRIRAM, D.; STABLES, J. 2,4-dimethoxyphenylsemicarbazones with anticonvulsant activity against three animal models of seizures: **Bioorganic and Medicinal Chemistry Journal**, v. 14, p. 3106– 3112, 2006.
- TODESCHINI, R.; CONSONNI, V. Molecular descriptors for chemo informatics, 2nd Edition, WILEY-VCH Verlag GmbH and Co, **Weinheim**, v.1 and 2, p. 27-37, 550-726, 2009.
- TOPLISS, J.G.; COSTELLO, R.J. Chance correlations in structure-activity studies using multiple regression analysis. **Journal of Medicinal Chemistry**, v.15, p. 1066–1068, 1972.
- TROPSHA A. Best practices for QSAR model development, validation, and exploitation. **Molecular Informatics**, v. 29, p. 476–488, 2010.
- TROPSHA, A.; GRAMATICA, P.; GOMBAR V.K.. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. **Molecular Informatics**, v. 22, p. 69–7, 2003.
- YAP, C.W. PaDEL-Descriptor. open source to calculate molecular descriptors and fingerprints. **Journal of Computational Chemistry**, v 32, n.7, p. 1466-1474, 2011.