Computational Modeling of the Anticonvulsant Activity of 3-Aminopropane-1,2-Diol and 1-Aminoethane-1,2-Diol Derivatives

Adedirin Oluwaseye[1,2*], Adamu Uzairu[2], Gideon A. Shallangwa[2], Stephen E. Abechi[2]

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

[2]Chemistry Department, Ahmadu Bello University, Zaria, Nigeria

[*]**Corresponding author:** Adedirin Oluwaseye, Chemistry Advance Research Center, Sheda Science and Technology Complex, FCT, Nigeria. Tel: +234800593145; Email: adedirinoluwaseye@yahoo.com ; senguade@gmail.com

Abstract

Quantitative Structure-Activity Relationship (QSAR) modeling was conducted on some 3-aminopropane-1, 2-diol and 1-aminoethane-1, 2-diol derivatives with anticonvulsant activity against maximal electroshock induced seizure using Genetic Function Algorithm-Multiple Linear Regression (GFA-MLR) method. The data set (37 molecules), was divided into 26 training and 11 test subsets by Modified-K-mediods clustering method. The models built by the GFA-MLR method provided satisfactory statistical results with LOF (0.087 to 0.097), R^2 (0.963 to 0.980), Q^2 (0.948 to 0.971), F (139.3 to 258.3), R^2_{pred} (0.861 to 0.931) and MAE (95%) (0.059 to 0.066). Descriptors contained in these models suggested that increment in molecular mass and polarizability of dataset molecules was favorable for improving their anticonvulsant activity values. Intelligent consensus modeling applied to the models gave a representative model with improved MAE (95%) of 0.054. Applicability domain of the models was well defined and therefore, the models can be used to screen molecules for anticonvulsant activity.

Introduction

The continuous effort to investigate new molecules with anticonvulsant properties is important because about 30% of epileptic patients with convulsion as a major symptom do not respond to marketed Antiepileptic Drugs (AEDS) [1]. In addition, almost all marketed AEDs had attendant side effects [2]. Therefore, developing new anticonvulsant improved quality in term of potency and safety is a continuous task for medicinal scientist. Modern computational chemistry as an evolving discipline provided rational approaches to drug design. It accelerates and reduces the cost of drug discovery process via obliteration of classical trial and error approach [3]. An example of computational approach to drug discovery is Quantitative Structure-Activity Relationship (QSAR) modeling which aids in identifying structure feature of a molecule which correlates mathematically with the observed biological activity of the molecule [7-8] [4-6]. This approach is now widely used as an aid to or an outright substitute for experimental studies to predict the activity of the molecules from their structure [7]. It reduces the number of animals needed for experiment and reduces cost in term of funds and time [8].

3-aminopropane-1, 2-diol and 1-aminoethane-1, 2-diol derivatives were reported to have anticonvulsant activity in Maxima Electroshock Seizure (MES) test, which one of the golden test for preliminary screening of molecules for anticonvulsant activity. To the best of our knowledge there were no QSAR reports on this group of molecules using a combination of density function theory quantum mechanical method and chemo metric principles. The objective of this study is to explore the structural features that are responsible for observed anticonvulsant activity of these groups of molecules through QSAR methodology.
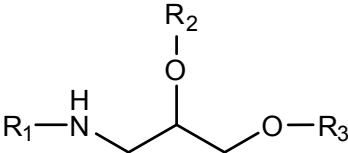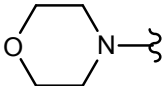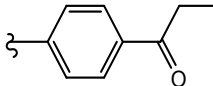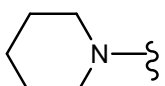
Materials and Method

Data Set
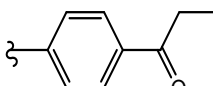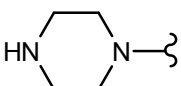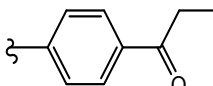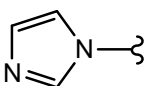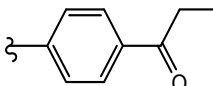
The data set were derivatives of 3-aminopropane-1, 2-diol and 1-aminoethane-1, 2-diol whose IUPAC name and anticonvulsant activity (against Maxima Electroshock (MES) induced seizure) values were obtained from literature [9]. The activity value reported as amount of molecule (mg kg^{-1}) that is effective to prevent convulsion in fifty percent of the tested animals (ED_{50}) was transformed to ED_{50} (mol kg^{-1}) and later to log ($1/ED_{50}$) to abate the deviation to the normal distribution of the data set activity values [10].

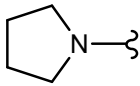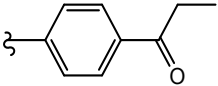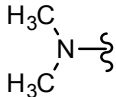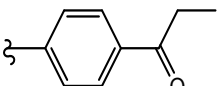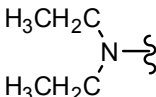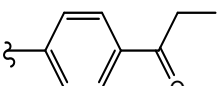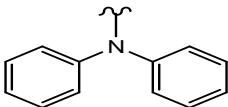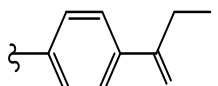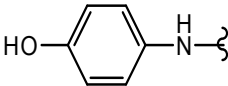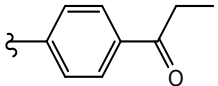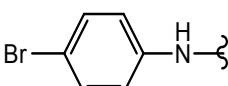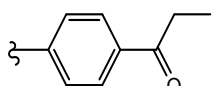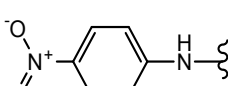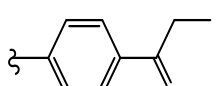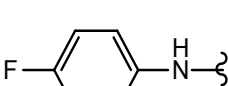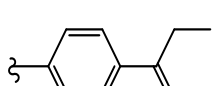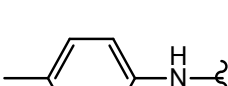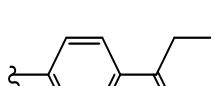### Molecular Structure Generation, Optimization and Descriptors Calculation
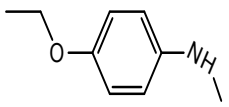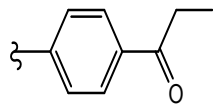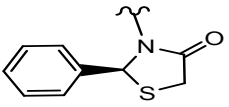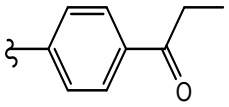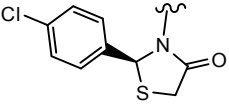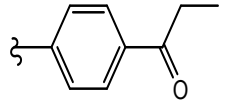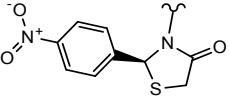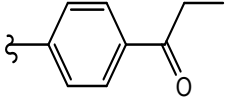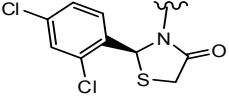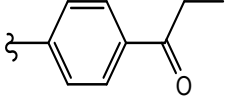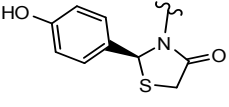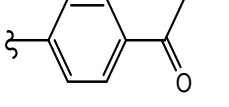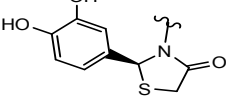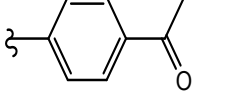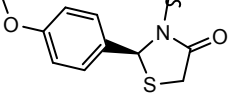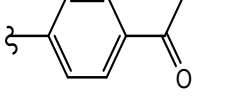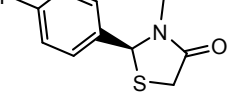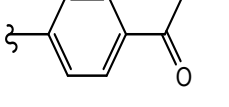
Spartan 14 [11] software was used to draw and optimize the equilibrium geometry of each molecule in the dataset. Density function theory B3LYP/6-31G** quantum mechanical method was employed for optimization calculation using Pulay DIIS algorithm and direct geometric minimization. This method gave the most stable molecule associated with absolute minima in the potential energy hypersurface which represents the most probable structure of the molecule [12]. DFT also gave reliable information on electronic properties of molecule [13].
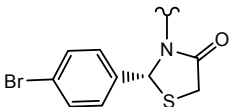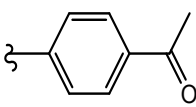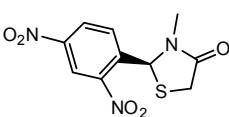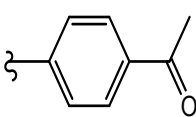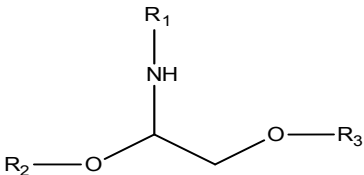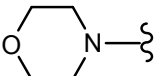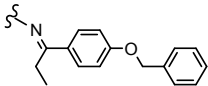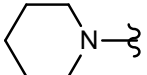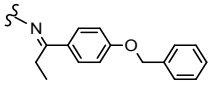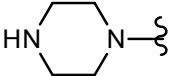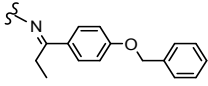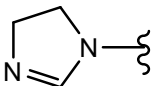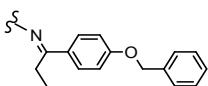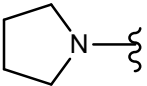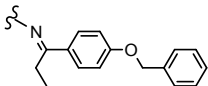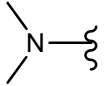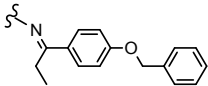
The optimized structures were ported to PaDEL-Descriptor software [14] to compute around 1875 different physicochemical, topological and structural molecular descriptors. Molecular structure and the corresponding anticonvulsant activity value of dataset molecules were presented in (Table 1). Datasets anticonvulsant activity values and calculated molecular descriptors arranged in a matrix constituted the database for the study.

ID	R1	R2	R3	pED_{50}	pED_{50}(pred.)	Residual
1		H		3.979	4.054	-0.075
2		H		3.486	3.755	-0.268
3		H		3.772	3.487	0.285
4		H		3.595	3.568	0.027

5*	(pyrrolidine-N)	H	(4-propanoylphenyl)	3.553	3.734	-0.181
6	H_3C–N, H_3C	H	(4-propanoylphenyl)	3.243	3.234	0.009
7*	H_3CH_2C–N, H_3CH_2C	H	(4-propanoylphenyl)	3.389	3.327	0.062
8*	(diphenylamino-N)	H	(4-propanoylphenyl)	3.966	3.872	0.094
9	HO–(phenyl)–N(H)	H	(4-propanoylphenyl)	3.639	3.704	-0.065
10	Br–(phenyl)–N(H)	H	(4-propanoylphenyl)	3.643	3.612	0.031
11	^-O–N^+(=O)–(phenyl)–N(H)	H	(4-propanoylphenyl)	3.231	3.233	-0.002
12*	F–(phenyl)–N(H)	H	(4-propanoylphenyl)	3.438	3.345	0.093
13	CH_3–(phenyl)–N(H)	H	(4-propanoylphenyl)	3.315	3.357	-0.042

14		H		3.874	3.825	0.049
15*			H	3.373	3.424	-0.052
16*			H	3.586	3.519	0.066
17			H	4.096	4.087	0.009
18*			H	3.532	3.553	-0.021
19			H	4.191	4.246	-0.054
20			H	3.409	3.495	-0.086
21*			H	3.504	3.494	0.010
22			H	3.458	3.413	0.046

23			H	3.971	3.929	0.042
24			H	3.967	3.961	0.006
25		H		3.485	3.527	-0.042
26		H		4.110	4.158	-0.048
27		H		3.934	3.876	0.058
28		H		3.875	3.894	-0.019
29		H		4.135	4.102	0.033
30*		H		3.473	3.559	-0.086
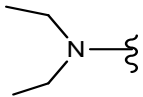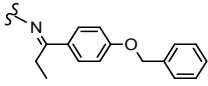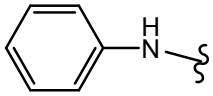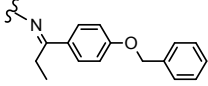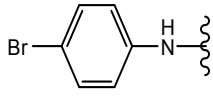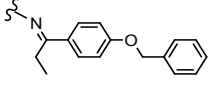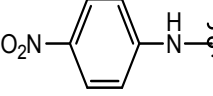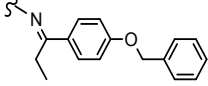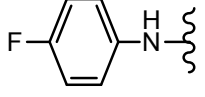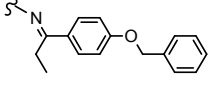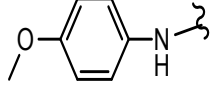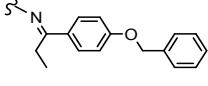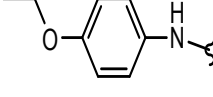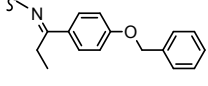
No.	Structure 1		Structure 2			
31		H		3.366	3.405	-0.040
32		H		3.469	3.436	0.032
33*		H		3.401	3.454	-0.053
34*		H		3.508	3.568	-0.061
35		H		4.122	4.069	0.052
36		H		3.546	3.506	0.040
37		H		3.967	3.944	0.023

Table 1: Molecular structure and anticonvulsant activities of dataset molecules.

Dataset Division

Modified K-Medoid clustering algorithm proposed by (Park & Jun, 2009) [15] available in Modified KMedoid version 1.2 was used to divide the database into training set for model development and test set for model validation. The algorithm proceeds via three main steps which are, selection of initial Mediod, update of selected mediods and assignment of object to mediods. In the first step, given n numbers of objects having p number of variables (descriptors) each, they were grouped into given k clusters, where k < n. Defining the variable of object i as X_{ia} (i = 1,. . .,n; a = 1,. . .,p). The Euclidean distance (d_{ij}) between two object i and object j was calculated:

$$d_{ij} = \sqrt{\sum_{a=1}^{p}(X_{ia} - X_{ja})^2} \quad \text{i = 1, ..., n and j = 1 ..... n} \quad (1)$$

Scaled Euclidean distance (v_j) for each object was calculated by dividing the distances by sum of the entire distances. The v_j for objects in each cluster k was arranged in ascending order and objects with smallest v_j values in each group are selected as the initial most middle objects in a cluster (mediods). The objects were re-shuffled to obtain initial cluster by assigning each object to the nearest medoid. The sum of distances from all objects to their medoids was calculated and kept for comparison.

To update the mediods, a new medoid of each cluster was found, which is the object minimizing the total distance to other objects in its cluster. Then the current medoid in each cluster is updated by replacing with the new medoid. Then (the third step), each object is assigned to the nearest medoid resulting to the formation of new k clusters. The sum of distance of objects to their mediods (total cost distance) was re-calculated. Now, if the total cost distance is equal to the previous one, the algorithm stops, otherwise, it goes back to the second step [15].

Transformation of Descriptor Values

Molecular properties are often measured in different unit and regression analysis frequently produces equation that favors property with higher magnitude of measurement [10]. To give all properties (descriptors) equal chance of appearing in the models produced in the study, descriptor values were transformed by normalization method using the equation below:

$$X^1 = \frac{(X - X_{min})}{(X_{max} - X_{min})} \quad (4)$$

Where X^1 is the normalized descriptor, X_i is the original descriptor value, X_{max} and X_{min} are the maximum and minimum descriptor value respectively in a descriptor column of the database [16].

Selection of Most Desirable Descriptor

Using the training set data only, combinations of descriptors that were optimally correlated with the anticonvulsant activity of the dataset molecules were selected using Genetic Function Algorithm (GFA) available in Materials Studio 7.0. GFA is a frequently used method that utilizes genetic algorithm to select combinations of descriptors that can be used to produce models and multivariate adaptive regression splines algorithm to evaluate the fitness of the models [17]. It has the advantage of producing of multiple models via repeated runs and automatically selects and determines the exact number of descriptors needed to build a full-size model.

Construction and Validation of QSAR Models

The combinations of training set descriptors reported by the GFA variable selection were collected in separate spreadsheets for both training and test sets. These spread sheets were imported into the MLRplusValidation1.3 [18] to calculate various internal and external validation parameters. Furthermore, the presence of multicolinearity problem among descriptor blend that made up a model was checked with Variance Inflation Factor (VIF) value for each descriptor i:

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

Where R_i^2 is the coefficient of determination of the regression of descriptor i on all the other descriptors. VIF value greater than 10 indicates high degree of correlation among descriptors (multicolinearity problem) [19]. Full explanations of the various validation parameters used were presented in (Table 2).

Symbol	Definition and allowed threshold	Ref.
Internal validation (validation with the training set data)		
LOF	Lack of fit, the smaller the value the better the model	Arthur 2016 [16]
R^2	Determination coefficient for training dataset, $R^2 > 0.5$ indicate goodness of fit	Tropsha, 2010 [10]
R^2_{adj}	Adjusted determination coefficient for training dataset. $R^2_{adj} > 0.5$ indicate good internal robustness	
F	Variance ratio	
Q^2_{LOO}	Square of correlation coefficient for leave one out cross-validation. $Q^2_{LOO} > 0.5$ indicate good internal robustness	
PRESS	Predicted error sum of square	
RMSEP	Root mean square error of prediction	
$^cR^2_p$	Y-randomization(scrambling) parameter, $^cR^2_p > 0.5$ indicate the model is not by chance correlation	Roy. 2007 [20]
External validation (validation with test set data) based on regression coefficient R		
$R^2_{(pred)}$	Predicted determination coefficient for the test set data, $R^2_{(pred)} > 0.6$ indicate good predictive ability	(Golbraikh and Tropsha, 2002) [21]
$\dfrac{r^2 - r_0^2}{r^2}$	r^2, r_0^2 are the square of correlation coefficient for the plot of predicted versus observed activity for test set with and without intercept respectively. If the value of the parameter is < 0.1 then, the model is predictive	
k	Slope for the plot of predicted versus observed activity for test set data. $0.85 \leq k \leq 1.15$ indicate model is predictive	
$\dfrac{r^2 - r_0'^2}{r^2}$	$r_0'^2$ is the square of correlation coefficient for the plot of observed versus predicted activity for test set data. If the value of the parameter is < 0.1 then, the model is predictive	
k'	Slope for the plot of observed versus predicted activity for test set data. $0.85 \leq k' \leq 1.15$ indicate model is predictive	
$\lvert r_0^2 - r_0'^2 \rvert$	r_0^2 and $r_0'^2$ are as defined above, $\lvert r_0^2 - r_0'^2 \rvert$ indicates the model is predictive	
External validation based on error measure		
AE	Average error for the test set data. $(AAE - \lvert AE \rvert) < (0.5 \times AAE)$ indicate presence of systematic error in the model	(Roy et al., 2016) [22]
AAE	Average absolute for the test set data,	
R^2_{res}	Square correlation coefficient for the plot of residual against measured activity values of the test set data, $R^2_{res} > 0.5$ indicate presence of systematic error in the model	
MAE	Mean absolute error for the test set data. (a) $MAE \leq 0.1 \times$ training set response range or $MAE + (3 \times \sigma) \leq 0.2 \times$ training set response range indicate good prediction. (b) $MAE > 0.15 \times$ training set response range or $MAE + (3 \times \sigma) > 0.25 \times$ training set response range indicate bad prediction. (c) Any prediction that does not fall into condition (a) and (b) may be considered as of moderate quality. Note, σ denotes the standard deviation of the absolute error values for the test set data.	

Table 2: QSAR model validation parameters.

Models Applicability Domain (AD)

The AD is the structural and chemical space of a QSAR model where it can make a reliable prediction [23]. Degree of extrapolation method was used to define AD in the study. It uses leverage (h_i) values for each compound obtained as the diagonal elements of a hat matrix and standardized residual (SDR) produced by the models. Williams plot (graph of SDR versus h_i) gives a quick pictorial representation of AD in this method. Hat matrix **H** was computed with the equation below:

$$H = X(X^TX)^{-1}X^T \qquad (17)$$

Where **X** represents the descriptors matrix and X^T is the transpose of the matrix. The diagonal elements of **H** are leverages for each compound. Leverage threshold (h*) for a model was computed with the equation below:

$$h^* = \frac{3(1+k)}{n} \qquad (18)$$

where n is the number of compounds in the training set only and k is the number of descriptors in the model. SDR was computed with the equation below:

$$SDR = \frac{Y_{pred} - Y_{obs}}{\sqrt{\frac{\sum_{j=1}^{n}(Y_{pred} - Y_{obs})^2}{n}}} \qquad (19)$$

Where n is the number of compounds in the training set. Y_{pred} and Y_{obs} are the predicted and experimentally observed activity values respectively. A compound with $h_i > h^*$ for a model is structurally dissimilar to other members of the model training set i.e. an influential data and prediction for such compound by the model are not reliable. A compound with SDR $> \pm 3$ is an outlier in the response space of the model [23].

Results and Discussion

Training Set and Test Set Data Structure

The clustering method used divivded the entire data into 26 training set (70% of the entire dataset) and 16 test set (30% of the entire dataset). The test compounds were marked with asterisk in Table 1. The plot of normalized mean distance against the observed anticonvulsant activity for both training and test set (Figure 1) showed that test set data was distributed within the descriptor space of the training set data. This showed that the data division method used performed well.
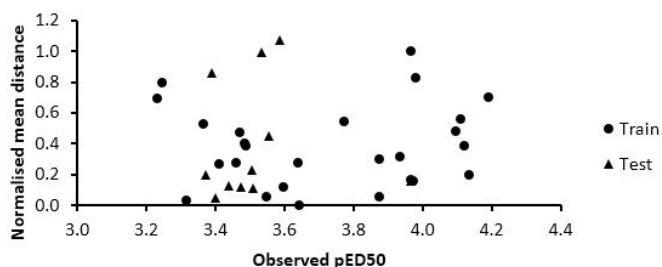
Figure 1: Diversity analysis.

QSAR Models and Validation Parameters

The top three models produced by the GFA-MLR method used in the study were presented in Equation 4 to 6. These QSAR models were obtained from 26 training set data and contained 4 descriptors each, meaning their Toplis ratio was 6.5. Hence, they do not violate the QSAR semi-empirical rule of thumb [24].

pED_{50} = 2.82476(+/-0.04039) + 0.34501(+/-0.04531) AATSC8m - 0.58205(+/-0.04129) MATS3m + 1.52741(+/-0.05204) Mp + 1.07807(+/-0.04242) SHCsats   (4)

pED_{50} = 2.74577(+/-0.05619) + 1.64954(+/-0.07662) AATS2p + 0.45055(+/-0.06123) AATSC8m - 0.77332(+/-0.0607) MATS3m + 1.13355(+/-0.05959) SHCsats   (5)

pED_{50} = 2.68245(+/-0.05494) + 0.65755(+/-0.05892) AATSC8m - 0.74266(+/-0.05605) MATS3m + 1.11772(+/-0.0558) SHCsats + 1.55598(+/-0.06817) TDB2p   (6)

Correlation analysis carried out on the descriptors contained in each model showed the highest absolute correlation coefficient between descriptors was 0.667 (Table 3). This indicated that descriptors contained in the models were relatively orthogonal to one another. Their VIF values (Table 3) further confirmed there was no multi-co-linearity problem in the models reported. Detailed quality and validation parameters values for the models were presented in (Table 4) These results showed that the models had good internal and external predictive ability and were void of systematic error.

Equation 4					
	AATSC8m	MATS3m	Mp	SHCsats	VIF
AATSC8m	1				1.077
MATS3m	0.194	1			1.252
Mp	0.001	0.362	1		2.240
SHCsats	0.156	-0.085	-0.677	1	2.011
Equation 5					
	AATS2p	AATSC8m	MATS3m	SHCsats	VIF
AATS2p	1				2.636
AATSC8m	-0.010	1			1.066
MATS3m	0.459	0.197	1		1.485
SHCsats	-0.677	0.135	-0.086	1	2.121
Equation 6					
	AATSC8m	MATS3m	SHCsats	TDB2p	VIF
AATSC8m	1				1.114
MATS3m	0.228	1			1.399
SHCsats	0.151	-0.026	1		2.060
TDB2p	-0.126	0.364	-0.677	1	2.442

Table 3: Models correlation matrix and variance inflation factor.

Parameters/Models	Eq. 4	Eq. 5	Eq. 6	Threshold value	Comment
Internal validation (validation with the training set data)					
LOF	0.087	0.092	0.097	Low value	
R^2	0.980	0.963	0.967	>0.6	
R^2_{adj}	0.976	0.957	0.961	>0.6	Robust models with good internal predictive ability (Tropsha, 2010) [10]
F	258.3	139.3	155.8		
Q^2_{LOO}	0.971	0.948	0.948	>0.5	
RMSEP	0.083	0.096	0.067		
PRESS	0.047	0.085	0.076		
$^cR^2_p$	0.909	0.886	0.891	>0.5	Model void of chance correlation (Roy, 2007) [20]
External validation (validation with test set data) based on regression coefficient R					
$R^2_{(pred)}$	0.897	0.861	0.931	>0.6	
r^2	0.736	0.739	0.837	>0.5	
r^2_0	0.723	0.617	0.809	>0.5	
r'^2_0	0.708	0.738	0.837	>0.5	
$\lvert r^2_0 - r'^2_0 \rvert$	0.015	0.121	0.027	<0.3	
$r^2 - r^2_0/r^2$	0.018	0.165	0.033	<0.1	Robust models with good external predictive ability
k	0.996	0.999	1.000	0.85<k<1.15	(Golbraikh & Tropsha, 2002) [21]
$r^2 - r'^2_0/r^2$	0.038	0.001	0.001	<0.1	
k'	1.0003	0.999	0.999	0.85<k<1.15	
External validation based on error measure					
R^2 (res. vs. obs.)	0.131	0.036	0.014	<0.5	Models was void of systematic error
nPE/nNE	0.833	1.200	1.750	<5	(Roy et al., 2016) [22]
MPE/MNE	0.995	0.987	0.692	<2	

MAE(95% data)	0.059	0.066	0.059		Models made good predictions
SD(95% data)	0.027	0.040	0.022		(Roy et al., 2016) [22]

Table 4: QSAR models validation scores.

Although the models reported were of good quality, the aim of all QSAR practitioners to improve the quality of prediction by reducing predicted residuals for test/query compounds to the barest minimum. To achieve this aim, intelligent consensus modeling [25] available in Intelligent Consensus Predictor version 1.1 software was applied on the models. Intelligent consensus modeling combined the proposed validated individual models (Equation 4 to Equation 6), and it carefully accounted for carefully accounting for the different assumptions characterizing each model. The optimized software setting for the study was without the entire additional criteria (i.e. Euclidean distance cutoff, applicability domain criteria and Dixon Q-test), a similar condition reported in literature [25].

The test-set validation parameters for individual models as well as consensus models obtained were reported in (Table 5). In the table, IM1, IM2 and IM3 represent the Eq. 4, Eq. 5 and Eq. 6 respectively. While, CM0 is the ordinary consensus model which uses simple average of prediction of individual model for all compounds in the test set; CM1 is the intelligent consensus model1 which uses the average of predictions from all qualified individual models; CM2 is the intelligent consensus model 2 which uses Weighted Average Predictions (WAPs) from all qualified individual models; and CM3 is the intelligent consensus model 3 which uses the best selection of predictions (compound-wise) from individual models [25].

Model	Q^2f_1	Q^2f_2	Q^2f_3	CCC	$\bar{R}^2_m$	ΔR^2_m	MAE	MAE (95%)	PRESS	PRESS (95%)	SDEP	SDEP (95%)
IM1	0.890	0.698	0.919	0.847	0.646	0.107	0.074	0.063	0.076	0.043	0.083	0.065
IM2	0.861	0.617	0.897	0.844	0.643	0.144	0.079	0.067	0.102	0.059	0.096	0.077
IM3	0.931	0.809	0.949	0.912	0.788	0.097	0.064	0.060	0.051	0.040	0.068	0.063
CM0	0.921	0.781	0.941	0.897	0.753	0.047	0.063	0.055	0.058	0.037	0.073	0.061
CM1	0.921	0.781	0.941	0.897	0.753	0.047	0.063	0.055	0.058	0.037	0.073	0.061
CM2	0.919	0.776	0.940	0.893	0.743	0.015	0.063	0.054	0.060	0.037	0.074	0.061
CM3	0.897	0.716	0.924	0.855	0.661	0.099	0.071	0.060	0.080	0.048	0.085	0.069

Table 5: Test set validation parameters for individual model and consensus model.

In the table, CM0 was ordinary consensus model in which simple average of prediction of individual model for test set compounds as used. CM1 was intelligent consensus 1in which the average of predictions from all qualified individual models for a given compounds were used. CM2 was intelligent consensus 2 in which uses weighted average predictions (WAPs) from all qualified individual models for a given test set compounds was used. Finally, CM3 was intelligent consensus 3 in which uses the best selection of predictions (compound-wise) from individual models was used [25].

Comparing the three individual models (IM1-IM3) with the three intelligent consensus models (CM1-CM3), it was obvious that the values of external validation parameters were better in almost all the cases for consensus models. The mean absolute error MAE (95%) metric for intelligent consensus models CM1 to CM3 were lower compare to that of individual models (Table 5). CM2 emerged as superior to all other models with MAE (95%) 0.054 (Table 5). CM2 was used to predict the activity of the entire data and the predicted activity values were reported in Table 1. The predicted test set activity values for the entire dataset by the individual models (IM1-IM3) and the intelligent consensus models (CM0-CM3) were presented in Table S2 of the Supplementary file.

Linear relationship existed between the experimental and predicted activity values by the CM2 (Figure 2) and there was even of its predicted activity residuals around the line standardized residual equal zero (Figure 3). These observations indicated that the model had good internal and external predictive capability and also void of systematic error. Therefore, it can be used to make prediction for known molecule without activity, provided the molecule is in the applicability domain (AD) of the developed models.
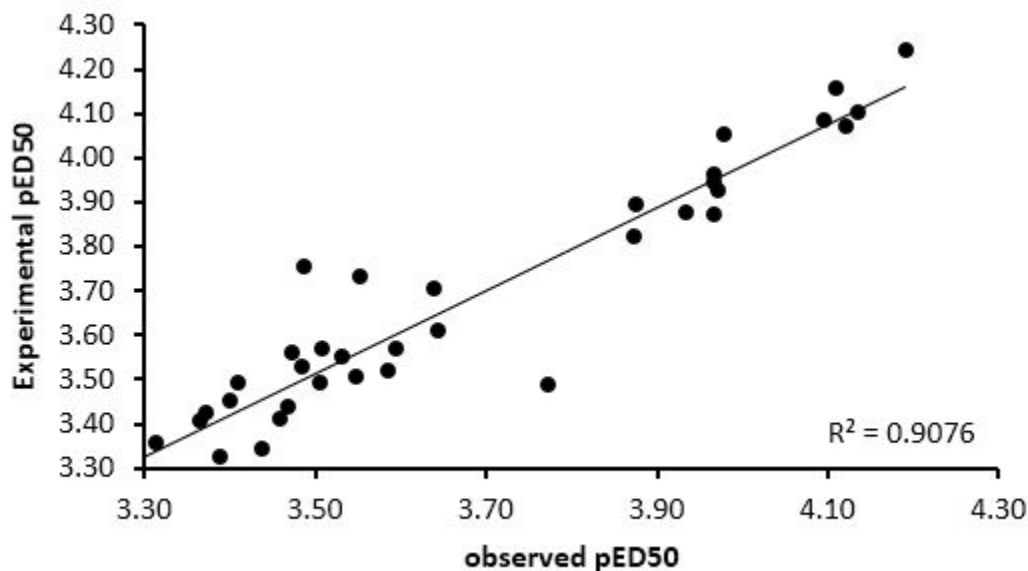
Figure 2: Models predicted versus experimental activity values for the data set molecules.
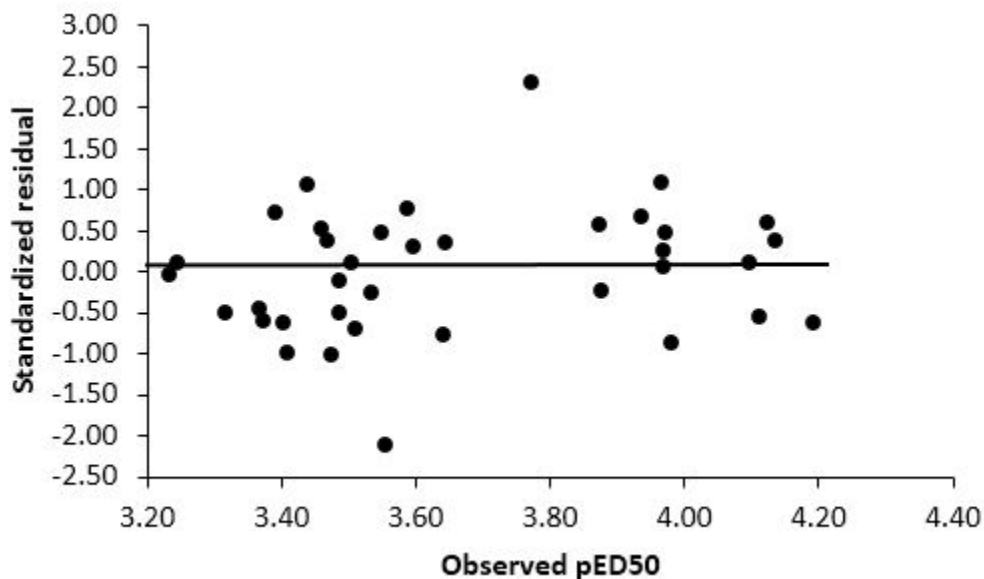
Figure 3: Models standardized residual against experimental anticonvulsant activity values.

Models Applicability Domain

The William plots for the models (Figures 4-6) showed that all dataset molecules had leverage value less than less than the models threshold leverage (h* = 0.57) and their standardized residual (SDR) were less than ±2.5. This indicated that all molecules were within the applicability domain of the models defined by the square area 0 < hi < h* and -2.5 < SDR < 2.5. Hence, the models reported were able to predict the activity values for all dataset molecules with high level of reliability. In summary, the models had high-quality parameters and great predictive power for molecules within their AD.
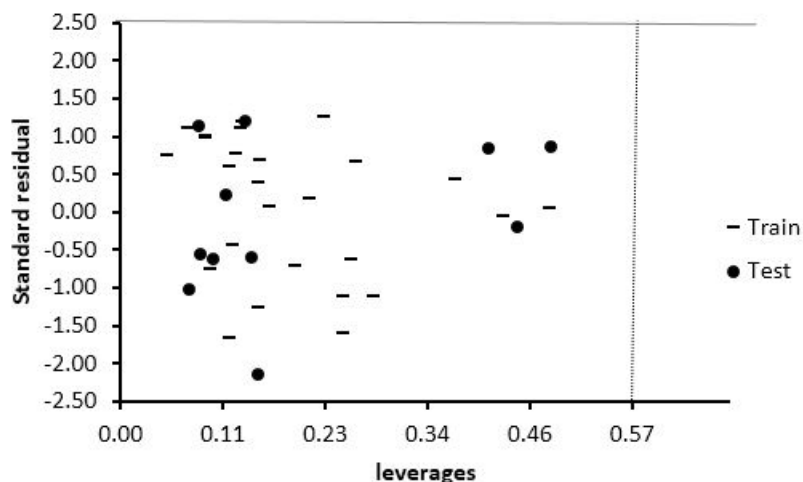
Figure 4: Williams plot defining the applicability domain of QSAR model represented by Equation 4.
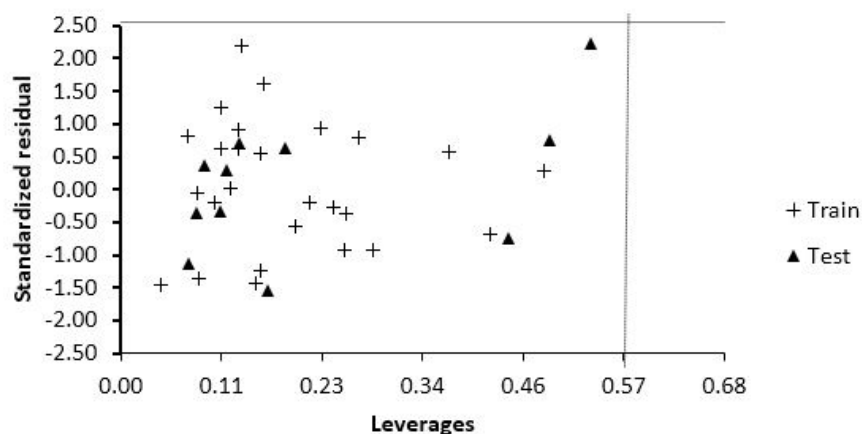
Figure 5: Williams plot defining the applicability domain of QSAR model represented by Equation 5.
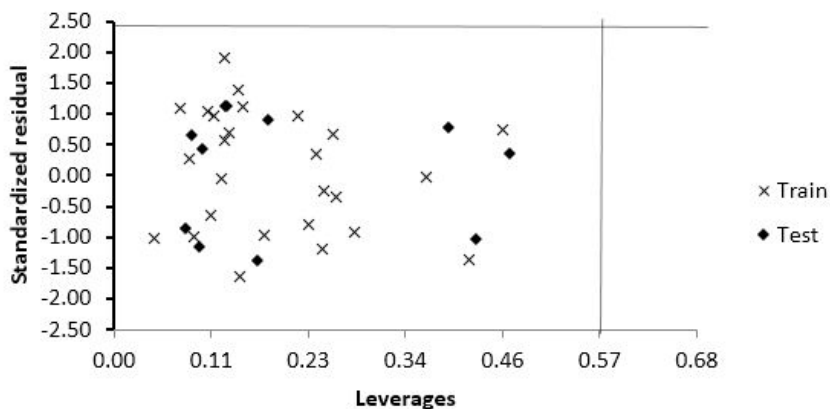
Figure 6: Williams plot defining the applicability domain of QSAR model represented by Equation 6.

Descriptors Interpretation

A QSAR model can be used as knowledge generator to improve the biological activity under consideration for any molecule. Interpretation of the model descriptors usually played a major role in this endeavor. Therefore, attempt was made in the study to a brief interpretation for descriptors contained in the reported QSAR models. Table 6 contained definition of descriptors shared by reported models; their average regression coefficient and incidence.

| No.	Descriptors	Physical meaning	ARC(I)
1	AATSC8m	Average/centered autocorrelation of topological structure -lag8/weighted mass	0.484(3)
2	AATS2p	Average autocorrelation of topological structure -lag2/weighted by polarizability	1.649(1)
3	MATS3m	Moran autocorrelation – lag 3/weighted by relative atomic mass	-0.699(3)
4	Mp	Mean atomic polarizability (scaled on Carbon atom)	1.527(1)
5	SHCsats	Sum of atom-type H E-State: H on C sp3 bonded to saturated C	1.109(3)
6	TDB2p	Topological distance based autocorrelation - lag 9 / weighted by polarizability	1.555(1)
Note: ARC (I) is average regression coefficient (incidence).			

Table 6: Molecular descriptors, their regression coefficient and incidence.

AATSC8m, AATS2p and MATS3m were 2D spatial-dependent autocorrelation descriptors calculated on a molecular graph with the use of Broto-Moreau coefficient (in the case AATSC8m and AATS2p) and Moran coefficient (in the case of MATS3m) [26]. AATSC8m measures the strength of the connection between relative atomic masses of two atoms in a molecular space separated by eight bonds (lag 8). It had positive average regression coefficient and appeared in the three models (Table 6). AATS2p measures the strength of the connection between polarizability of two atoms in a molecular space separated by two bonds (lag 2). Also had positive average regression coefficient and with one incidence in the entire models (Table 6). While MATS3m measures the strength of the connection between relative atomic masses of two atoms in a molecular space separated by three bonds (lag 3), it was negatively correlated with the anticonvulsant activity of the studied dataset (Table 6). It also appeared in the three models. Therefore, increment in values of AATSC8m and AATS2p augments the anticonvulsant activity value of dataset molecules, while, that of MATS3m diminishes the activity.

Mp was a 2D constitutional descriptor defined as mean atomic polarizability scaled on Carbon atom [14]. It was positively correlated to the anticonvulsant activity of the studied dataset and occurred in one of the model (Table 6). SHCsats is a 2D electrotopological-state index of an atom which unifies in a single index both electronic and topological description of a molecule

[27]. It is defined as Sum of atom-type H on C sp3 bonded to another saturated C. It had positive regression coefficient and incidence of three (Table 6). TDB2p is 3D topological distance based autocorrelation - lag 2 / molecular polarizability. It is a member of the 3D autocorrelation descriptors [28] which uses both Euclidean (geometric) and topological distances to encode information about molecular structure. TDB is an index of shape and branching of molecules [26]. It occurs in one of the model reported and it's positively correlated to the anticonvulsant activity of dataset molecules.

In summary, the descriptors contained in the reported models suggested that increment in the molecular mass and polarizability will improve the anticonvulsant activity of the dataset molecules. This can be achieved via chain elongation to increase the value of SHCsats, AATSC8m and addition of electronegative elements which will be favorable to the values of AATS2p, Mp and TDB2p.

Conclusion

Anticonvulsant activity of some 3-aminopropane-1,2-diol and 1-aminoethane-1,2-diol derivatives were successfully model via QSAR strategy. The QSAR models obtained had good statistical quality: LOF (0.087 to 0.097), R^2 (0.963 to 0.980), Q^2 (0.948 to 0.971), F (139.3 to 258.3), R^2_{pred} (0.861 to 0.931) and mean absolute error after removal of 5% data i.e. MAE (95%) (0.059 to 0.066). Intelligent consensus 2 (CM2) with MAE (95%)

of 0.054 was the golden model for making prediction in the study. The result in the study showed that AATSC8m, AATS2p, MATS3m, Mp, SHCsats and TDB2p descriptors had influence on the anticonvulsant activity values of dataset molecules. Therefore, increase in molecular mass and polarizability of dataset molecules is favorable for improving their anticonvulsant activity values. The models reported were robust and with good predictive ability. Their applicability domains were well defined and they can have used to virtually design and screen molecules for anticonvulsant activity.

References

1. Abdulfatai U, Uzairu A, Uba S (2017) Quantitative structure-activity relationship and molecular docking studies of a series of quinazolinon-yl analogues as inhibitors of gamma amino butyric acid aminotransferase. J Adv Res 8: 33-43.

2. Obniska J, Rzepka S, Kamiński K (2012) Synthesis and anticonvulsant activity of new N-Mannich bases derived from 3-(2-fluorophenyl)-and 3-(2-bromophenyl)-pyrrolidine-2, 5-diones. Part II. Bioorg Med Chem 20: 4872-4880.

3. Pourbasheer E, Aalizadeh R, Ganjali MR, Norouzi P (2014) QSAR study of IKKβ inhibitors by the genetic algorithm: multiple linear regressions. Medicinal Chemistry Research 23: 57-66.

4. Pourbasheer E, Aalizadeh R, Ganjali MR, Norouzi P, Shadmanesh J, et al. (2014) QSAR study of Nav1. 7 antagonists by multiple linear regression method based on genetic algorithm (GA-MLR). Medicinal Chemistry Research 23: 2264-2276.

5. Pourbasheer E, Aalizadeh R, Ganjali MR, Norouzi P, Banaei A (2014) QSAR study of mGlu5 inhibitors by genetic algorithm-multiple linear regressions. Medicinal Chemistry Research 23: 3082-3091.

6. Pourbasheer E, Beheshti A, Khajehsharifi H, Ganjali MR, Norouzi P (2013) QSAR study on hERG inhibitory effect of kappa opioid receptor antagonists by linear and non-linear methods. Medicinal Chemistry Research 22: 4047-4058.

7. Yousefinejad S, Hemmateenejad B (2015) Chemometrics tools in QSAR/QSPR studies: A historical perspective. Chemometrics and Intelligent Laboratory Systems 149: 177-204.

8. Ibezim E, Duchowicz P, Ibezim N, Mullen L, Onyishi I, et al. (2009) Computer-aided linear modeling employing QSAR for drug discovery. Scientific Research and Essays 4: 1559-1564.

9. Fegade JD, Chaudhari RY, Patil VR (2011) QSAR analysis of some aryloxypropanolamine analogues as anticonvulsants. Der Pharma Chemica 3: 96-109.

10. Tropsha A (2010) Best practices for QSAR model development, validation, and exploitation. Molecular Informatics 29: 476-488.

11. Shao Y, Molnar LF, Jung Y, Kussmann J, Ochsenfeld C, et al. (2006) Advances in methods and algorithms in a modern quantum chemistry program package. Physical Chemistry Chemical Physics 8: 3172-3191.

12. Siddiqui SA, Rasheed T, Bouarissa N (2013) Investigation of superhalogen behaviour of RuFn (n= 1-7) clusters: density functional theory (DFT) study. Bulletin of Materials Science 36: 743-749.

13. Choudhary M, Sharma BK (2014) QSAR rationales for the 5-HT6 antagonistic activity of Epiminocyclohepta [b] indoles. Der Pharma Chem 6: 321-330

14. Yap CW (2011) PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. J Comput Chem 32: 1466-1474.

15. Park HS, Jun CH (2009) A simple and fast algorithm for K-medoids clustering. Expert Systems with Applications 36: 3336-3341.

16. Arthur DE, Uzairu A, Mamza P, Abechi SE, Shallangwa G (2016) In-silco study on the toxicity of anti-cancer compounds tested against MOLT-4 and p388 cell lines using GA-MLR technique. Beni-Suef University Journal of Basic and Applied Sciences 5: 320-333.

17. Rogers D, Hopfinger AJ (1994) Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. Journal of Chemical Information and Computer Sciences 34: 854-866.

18. Ambure P, Aher RB, Gajewicz A, Puzyn T, Roy K (2015) "NanoBRIDGES" software: Open access tools to perform QSAR and nano-QSAR modeling. Chemometrics and Intelligent Laboratory Systems 147: 1-13.

19. Beheshti A, Pourbasheer E, Nekoei M, Vahdani S (2016) QSAR modeling of antimalarial activity of urea derivatives using genetic algorithm-multiple linear regressions. Journal of Saudi Chemical Society 20: 282-290.

20. Roy K (2007) On some aspects of validation of predictive QSAR models. Expert Opinion in drug discovery 2: 1567-1577.

21. Golbraikh A, Tropsha A (2002) Beware of q 2! J Mol Graph Model 20: 269-276.

22. Roy K, Das RN, Ambure P, Aher RB (2016) Be aware of error measures. Further studies on validation of predictive QSAR models. Chemometrics and Intelligent Laboratory Systems 152: 18-33.

23. Netzeva TI, Worth AP, Aldenberg T, Benigni R, Cronin MT, et al. (2005) Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. Altern Lab Anim 33: 155-173.

24. Damme SV, Bultinck P (2007) A new computer program for QSAR-analysis: ARTE□QSAR. J Comput Chem 28: 1924-1928.

25. Roy K, Ambure P, Kar S, Ojha PK (2018) Is it possible to improve the quality of predictions from an "intelligent" use of multiple QSAR/QSPR/QSTR models? Journal of Chemometrics 32: e2992.

26. Todeschini R, Consonni V (2009) Molecular descriptors for chemoinformatics, 41 John Wiley & Sons.

27. Kier LB, Hall LH (1976) Molecular connectivity VII: specific treatment of heteroatoms. J Pharm Sci 65: 1806-1809.

28. Klein CT, Kaiser D, Ecker G (2004) Topological distance based 3D descriptors for use in QSAR and diversity analysis. J Chem Inf Comput Sci 44: 200-209.