Moroccan Journal of Chemistry
ISSN: 2351-812X
http://revues.imist.ma/?journal=morjchem&page=login
AdedirinOluwaseye & al. / Mor. J. Chem. 6 N°4 (2018) 781-794

QSAR modeling of some anticonvulsant molecules as γ-aminobutyrate-aminotransferase inhibitors

AdedirinOluwaseye[a*], AdamuUzairu[b], Gideon A. Shallangwa[b] and Stephen E. Abechi[b]

[a]Chemistry Advance Research Center, Sheda Science and Technology Complex, P.M.B. 186, Garki, Abuja, Nigeria
[b]Chemistry Department, Ahmadu Bello University,P.M.B. 1044, Zaria, Nigeria

** Corresponding author:*
adedirinoluwaseye@yahoo.com
Received 13 Feb 2018,
Revised 18 Nov 2018,
Accepted 28 Dec 2018

Abstract

Quantitative structure-activity relationship study was done on 62 compounds with anticonvulsant activity in maximal electroshock-induced seizures test. The molecular structure of the compounds was optimized with parametric semi-empirical PM3 method available in Spartan 14 software. Quantum mechanical descriptors were extracted from the property and output module of the software. Combination of activity based-clustering and genetic function algorithm chemo-metric techniques were used to map molecular descriptors to activity values. A well-validated and robust quantitative structure-activity model was obtained with R^2 = 0.947, Q^2= 0.924, F = 91.42 and $R^2_{pred(test)}$ = 0.881. The descriptor contained in the model suggested anincrease in the number of O and N atoms in the molecule augments the activity of the studied compounds. Also, theintroduction of electron donating substituents is beneficial to the activity of the studied compounds. Armed with these, information, new hypothetical1H-pyrazole-5-carboxylic acid derivatives were designed using template approach and screened *in silico*. Compounds with hypothetic anticonvulsant activity better than the template were docked with γ-aminobutyrate-aminotransferase and their binding affinity was found to be comparable and even superior to that of 4-aminohex-5-enoic acid (vigabatrin), a known inhibitor of γ-aminobutyrate-aminotransferase.

Keywords: *Γ-aminobutyrate-aminotransferase; Quantitative structure-activity relationship; Molecular Molecular docking; Genetic function algorithm*

1.Introduction

Γ-aminobutyric acid (GABA) is one of most widely distributed inhibitory neurotransmitter in the mammalian central nervous system and convulsion occurs when its concentration drops below acertain threshold level in thebrain. Raising it concentration level in the brain terminates convulsion. One of the enzymes that degrade its concentration in the brain is γ-aminobutyrate aminotransferase (GABA_AT). It does this by converting to succinic semialdehyde. This action had been implicated in convulsion associated with epilepsy[1]. Therefore, any compounds that can reduce or entirely terminate the action of GABA_AT are agold star. A number of molecules having activity in maximal electroshock seizure (MES) test had been reported to inhibit GABA_AT including 4-aminohex-5-enoic acid, 3,3-diphenylpyrrolidine-2,5-dionee.t.c.[2] Usage of these molecules and other antiepileptic drugs being market has failed to totally provide ananswer to the problem of this disorder. Because there are about 50 million clinical cases of the disorder worldwide and approximately 20% to 30% of the patients do not respond to any marketed anti-epileptic drugs. Therefore, the need for anew anticonvulsant chemotherapeutic agent is of great importance [3]. Generally, drug discovery process is very challenging, however, with the aid of computer-aided drug design methodologies like quantitative structure-activity relationships and molecular dockinghigh-quality leads which are more likely to succeed in clinical trials can be developed. The objective of this work was to use these two strategies to proposed series of 1H-pyrazole-5-carboxylic acid derivatives*in silico* with improved anticonvulsant activity in MES test in view which can later be screen*in vivo* and *in vitro* to check the validity of the proposal. QSAR is a technique that relates quantitative molecular descriptors to the biological activities of compounds. It offers an *in silico* tool to propose the activities of known and hypothetical chemical compounds[4].Molecular docking, on the other hand, is a technique used to exploring the binding mode of two interacting molecules revealing key elements and mechanism of their interaction[5]

2. MATERIAL AND METHODS

Dataset

The dataset for the study was made up of 62 derivatives of 1H-pyrazolo[3, 4]pyridine, 1H-pyrazole-5-carboxylic acid and hydrazine carboxamide reported literature [6]. to possess anticonvulsant activity in maximal electroshock seizure (MES) test. The anticonvulsant activity of these compounds reported in ED_{50} (mg kg^{-1}) was converted to ED_{50} (mol kg^{-1}) and later to Log $\left(\frac{1}{ED_{50}\,(\mathrm{mol\,Kg^{-1}})}\right)$ in other to reduce the skewness in the biological activity and move it to nearly normal distribution [7]. **Table** 1 presents the names and anticonvulsant activity values of the dataset compounds in logarithm unit (pED_{50}).

Molecular structure optimization and descriptor calculation

Molecular structures of the dataset compounds were created with Spartan 14 [8] and their geometry was optimized with semi-empirical parametric PM3 quantum mechanical method Molecular properties for the compounds were obtained from the properties and output modules ofSpartan 14. These properties extracted included various electronic properties, atomic charges, thermodynamic properties and frontier orbital energy values. In addition, the following descriptors were derived from those extracted from Spartan 14:

a. Energy gap ( ) defined as the difference between ϵLUMO and ϵHOMO[9]

$$\Delta\in = \epsilon\mathrm{LUMO} - \epsilon\mathrm{H(} \qquad (1)$$

b. Ionization energy (I) defined as the difference in total energy between one electron deficient and form of a molecule[9]:

$$I \equiv \chi^{-} = E(N_0 - 1) - E \qquad (2)$$

c. Electron affinity (A) defined as difference in total energy between neutral and one electron rich form of a molecule[9]:

$$A \equiv \chi^{+} = E(N_o) - E(N_0 \quad (3)$$

d. Electronegativity index (χ) is the negative value of chemical potential (μ) which is the first partial derivative of the energy of a system with respect to the number of electron N in the system at constant external potential [9]

Table 1 Molecular formula of dataset compounds and their anticonvulsant activities

No	Molecular formula	-Log ED_{50}	Sr.No	Molecular formula	-Log ED_{50}
1	$C_{10}H_{13}ClN_2O$	3.447	32	$C_{12}H_{10}FN_5$	4.325
2	$C_9H_{10}N_2O_2$	3.242	33*	$C_{13}H_{12}FN_5$	4.747
3	$C_{11}H_{13}ClN_2O_2$	3.633	34	$C_{14}H_{14}FN_5$	4.433
4	$C_{11}H_{14}N_2O_2$	3.253	35	$C_{15}H_{14}FN_5$	4.653
5	$C_{15}H_{17}NO_3$	3.181	36	$C_{14}H_{14}FN_5$	4.613
6	$C_{11}H_{11}NO_3$	3.011	37	$C_{12}H_9F_2N_5$	4.849
7	$C_{11}H_{10}ClNO_3$	2.952	38*	$C_{13}H_{11}F_2N_5$	4.716
8	$C_{11}H_{12}ClNO_3$	2.976	39	$C_{13}H_{11}F_2N_5$	4.807
9	$C_{12}H_{14}ClNO_3$	3.468	40	$C_{14}H_{11}F_2N_3O_2$ (2,3-difloro)	4.675
10	$C_{12}H_{14}ClNO_2$	3.283	41	$C_{14}H_{11}F_2N_3O_2$ (2,6-difloro)	4.721
11	$C_{11}H_{11}Cl_2NO_2$	3.667	42	$C_{14}H_{11}F_2N_3O_2$ (3,4-difloro)	5.089
12*	$C_8H_{16}O_2$	3.396	43	$C_{15}H_{15}N_3O_2$ (o-tolyloxy)	4.678
13	$C_8H_{17}NO$	3.199	44	$C_{15}H_{15}N_3O_2$ (m-tolyloxy)	4.943
14	$C_{11}H_{15}NO$	3.440	45	$C_{15}H_{15}N_3O_2$ (p-tolyloxy)	4.895
15*	$C_{12}H_{17}NO$	3.327	46	$C_{16}H_{17}N_3O_2$	4.640
16	$C_{11}H_{11}NO_2$	3.228	47	$C_{17}H_{17}N_3O_2$	5.053
17*	$C_{10}H_{12}ClNO$	3.034	48*	$C_{18}H_{21}N_3O_2$ (4-butyl)	4.986
18*	$C_{12}H_{10}FN_5$	4.307	49	$C_{18}H_{21}N_3O_2$ (4-tert-butyl)	5.268
19	$C_{12}H_{14}N_6$	4.180	50	$C_{15}H_{15}N_3O_2$	4.442
20	$C_{11}H_{11}FN_6$	4.488	51	$C_{15}H_{14}FN_3O_2$	4.930
21	$C_{12}H_{13}FN_6$	4.114	52*	$C_{16}H_{16}FN_3O_2$	4.851
22	$C_{12}H_{12}F_2N_6$	4.444	53*	$C_{15}H_{14}ClN_3O_2$	5.018
23*	$C_{11}H_{10}F_2N_6$	4.577	54	$C_{16}H_{16}ClN_3O_2$	5.041
24	$C_{13}H_{12}FN_5$	4.331	55	$C_{15}H_{14}BrN_2O_2$	4.900
25	$C_{14}H_{14}FN_5$	4.354	56	$C_{14}H_{13}N_3OS$	4.802
26	$C_{13}H_{13}N_5$	4.203	57	$C_{14}H_{12}FN_3OS$	4.764
27	$C_{13}H_{12}FN_5$	4.507	58	$C_{15}H_{15}N_3OS$	4.293
28*	$C_{12}H_{11}N_5$	4.207	59	$C_{10}H_{12}BrN_3O$	4.335
29	$C_{12}H_{11}FN_6$	4.509	60*	$C_{16}H_{17}ClN_2O_3$	4.449
30	$C_{12}H_{11}N_5$	4.052	61	$C_{17}H_{19}ClN_2O_3$	4.627
31	$C_{13}H_{13}N_5$	4.379	62	$C_{17}H_{19}BrN_2O_3$	4.554

*Molecule whose number is marked with * represent member of test dataset*

$$\chi = -\mu = -\left(\frac{\partial E}{\partial N}\right)_{v(r)} = \frac{1}{2}(I + A) \cong \frac{1}{2}(\epsilon HOMO + \epsilon LU \quad (4)$$

e. Chemical hardness (η) is the second partial derivative of energy of a system with respect to the number of electron N in the system at constant external potential

$$\eta = \left(\frac{\partial^2 E}{\partial N^2}\right)_{v(r)} = -\left(\frac{\partial \chi}{\partial N}\right)_{v(r)} = \frac{1}{2}(I - A) \cong \frac{1}{2}(\epsilon HOMO + \epsilon LU \quad (5)$$

f. Global softness (S) is the reciprocal of chemical hardness[9]:

$$S \quad (6)$$

g. Electrophilicity index (Ω) is defined as one-half of the ratio of the square of chemical potential to the chemical hardness

$$\Omega = \frac{\mu^2}{2\eta} = \frac{\chi^2}{2\eta} = \frac{(}{8} \quad (7)$$

h. Polarity parameter defined as thedifference between maximum and minimum atomic charge in a molecule.

$$PP = Q_{max} - \quad (8)$$

i. The polarization of the molecule is defined as the ratio of the sum of the absolute value of the charges of all atoms in the molecule to the number N of atoms in the molecule[9].

$$PM = \sum_{A=1}^{N} (\quad (9)$$

j. Local dipole index (LDI) defined as the sum over all connected atoms in the molecule of the absolute value of the charge difference between bonded atom $|Q_A -$divided by the number of bond [9]

$$LDI = \sum_{AB} \left(\frac{|Q_A -}{N_i}\right. \quad (10)$$

k. Topological electronic index (T^E) defined as the sum of the absolute differences in electronic excess charges on all atomic pairs $(|q_i -$ in a given, molecule divided by the squares of the respective interatomic distances[9].

$$T^E = \sum_{ij, i=j} \frac{|}{} \quad (11)$$

l. Mean polarizability (α_m) and anisotropy of the polarizability (β^2).

$$= (\quad \alpha_{xx} + \alpha_{yy} +) \quad (12)$$

$$= \quad \frac{1}{2}\left[(\alpha_{xx} - \alpha_{yy})^2 + (\alpha_{yy} - \alpha_{zz})^2 + (\alpha_{zz} - \alpha_{} \right. \quad (13)$$

Dataset pretreatment

The compounds activity values and their corresponding descriptors were arranged in a matrix format. All descriptor columns containing a constant value or whose variance is less than 0.001 were discarded. Correlation analysis was performed on the dataset and one of any pair of the descriptor with a correlation coefficient greater than 0.9were discarded. The descriptor retained, in this case, had a higher correlation coefficient with the activity value compare to the discarded one. The pretreatment was done to reduce redundancy in the dataset and to aid in selecting optimal descriptors that explain the variation in activity value well.

Dataset division and descriptor transformation

The pretreated dataset matrix was imported into DatasetDIvision 1.2[10] and activity-based clustering (KS) available in the software was used to divide the data into a training set and test set. Only training set data was used in all process involved in the construction of QSAR model, the test set data is used to assess the predictive capability of the model. Therefore, training set descriptors were transformed with the auto-scaling equation:

$$X^{!} = \frac{X - \bar{X}}{\sigma} \quad (14)$$

In equation 15, $X^{'}$ is the auto-scaled descriptor, X is the value of each descriptor for a given molecule, $\bar{X}$is the average for each column of descriptors and σ is the standard deviation value for each column of descriptors. Auto-scaling is a combination of mean centering and variance scaling. It produces descriptor column with variance value of 1. This gives all the descriptor equal weight, importance and opportunity to influence the model being produced [11].

Selection of optimal descriptor

Training set data matrix was imported into Material Studio 8.0 study **Table** and genetic function algorithm (GFA) module available in the software was used to generate combinations of descriptors blend that better explain the variation in the anticonvulsant activity value of the studied compounds. GFA is a mixture of Holland's genetic algorithm and Friedman's multivariate adaptive regression splines (MARS) algorithm. It is a search method that finds exact or approximate solutions to any optimization and search problems[4]. It produces numbers of descriptor blends rather than a single equation as do most other statistical methods. It also gave the user control over the equation length. It uses fitness lack-of-fit (LOF) as it fitness function to forbid over-fitting and reduce redundancy in a model. LOF is calculated with the equation below and the smaller the LOF of a descriptor blend the better the quality of the equation produced from it:

$$LOF = \frac{LSE}{\left(1+\frac{C+d_p}{M}\right)^2} \quad (15)$$

In equation 2, c is the number of basic 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.

Multi-co-linearity analysis

The presence of high degree of correlation i.e.multi-co-linearity among the descriptors contained in the best descriptors blend reported by GFA was evaluated with variance inflation factor (VIF) value for each descriptor. The $VIF_i$for a given descriptor i in a model was calculated with the equation below:

$$VIF_i = \frac{1}{1-R_{ij}^2} \quad (16)$$

In equation 16, R^2_{ij} is the correlation coefficient of the multiple regression between the descriptor i and the remaining j descriptors in the model[12].

QSAR model and validation

The descriptors that constitute the best blend reported by the GFA were selected into a separatespreadsheet for both training and test sets. The activity value for their compound was placed as the last column in the matrix. The descriptors for the test set data were also auto-scaled using equation 14. Then, training and test set data matrices were imported into the MLRplusValidation1.3[10] software to calculate various internal and external validation parameters.

Models applicability domain

The extent of extrapolation approach based on compounds leverage (hi) values and standardized residual (SDR) produced by the model was used to define the applicability domain (AD) of the QSAR model. Compounds h_i are obtained as the diagonal element of hat matrix **H**:

$$\boldsymbol{H = X(X^T X)^{-1}.X^T} \quad (17)$$

In equation 17,**X** is the descriptor matrix and $\mathbf{X^T}$ is the transpose of X, and SDR was obtained as follows:

$$SDR = \frac{\hat{y} - y}{\sqrt{\frac{\sum_{i=1}^{n}(\hat{y} - y)^2}{n}}} \quad (18)$$

In equation 18, y is the observe activity value for either the training or the test set, $\hat{y}$ is the predicted activity value ofthe model and n is the number of molecules in the set considered. The model AD was defined by the boundary $0 < h_i < h^*$ and $-3 < SDR < 3$. Where h* is calledwarning leverage h* computed by:

$$h^* = \frac{3(k+1)}{n} \quad (19)$$

In equation 19, k is the number of descriptors in the model and n is the number of compounds that made up the training set. A quick visual assessment of the model AD is a plot of SDR versus h_i known as Williams plot was made.

In silico design and screening of hypothetical anticonvulsant compounds

Ethyl-4-(4-chlorophenyl)-3-morpholino-1H-pyrrole-2-carboxylate (compound 61 in the dataset) was used as template. It had relatively high anticonvulsant activity value (pED50 = 4.627).**Fig**. 1 depicts the2D structure of the template segmented into four: 1H-pyrrole, ethyl formate, morpholine, and chlorobenzene. Based on the information obtained from the QSAR model modifications were made around these fragments via insertion and deletion of asubstituent to obtain the new hypothetical molecules. Molecular geometries of the hypothetical molecules were optimized, descriptors obtained and their leverage values calculated as described for the training dataset. The hypothetical anticonvulsant activity of the designed hyopethical molecules was predicted with the QSAR model for those with leverage value less than the model warning leverage. Those with predicted activity higher than theactivity of the template were reported in the study and docked with γ-aminobutyrate-aminotransferase.

Molecular docking

Crystal structure of GABA_AT (PDB: 1OHV) was obtained from the Brookhaven ProteinDatabase (PDB http://www.rcsb.org/pdb) and imported into Internal Coordinate Mechanic Program (ICM-pro version 3.8-3) workspace where it was converted from PDB file to ICM object. The conversion process prepared the protein molecule by adding missing hydrogen; assigning atom type and partial charges; optimizing the orientation of His, Asn, Glu and Cys and removing water molecule. The optimized structures of the hypothetical compounds were arranged in a tabular form (Chemical **Table**) with Material Studio software and saved in SD format. The chemical **Table** was imported into the ICM workspace to be used as aligand for the docking. Dock chemical **Table** module available in the ICM software was used to dock the prepared protein to the ligands. The docking process was based on biased probability Monte Carlo (BPMC) global-energy-optimization/minimization procedure. And the binding affinity was estimated with ECEPP/3 force field [13].

Fig. 1 structural fragment of the chosen scaffold

3. RESULTS AND DISCUSSION

Dataset structure

Activity-based clustering used divided the 62 datasets into 50 training and 12 test set data. The test set molecules are marked with asterisks in **Table** 1. Descriptive statistics were performed on the activity values of the sets and the result obtained is reported in **Table** 2. It showed that there is asimilarity in their maximum value. The test set minimum was greater than the training set minimum. The training set range was greater than that of test set data. In addition,there is asimilarity in their mean, standard deviation, and sample variance with differences less than 0.5. These observations showed that the two sets had similar data point distribution. Dissimilarity measure was done on the two set data using normalized mean Euclidean distance to elucidate the dissimilarity in their descriptors space. **Fig**. 2 showed that the test set compounds were a representation of the entire dataset i.e. both sets had similar descriptors space. This indicated the activity based-clustering produced good dataset division.

Table 2-Training and test set data descriptive statistics

Mean	4.266	4.311
Standard Error	0.095	0.183
Median	4.444	4.379
Standard Deviation	0.664	0.635
Sample Variance	0.441	0.403
Range	2.317	2.155
Minimum	2.952	3.125
Maximum	5.268	5.280

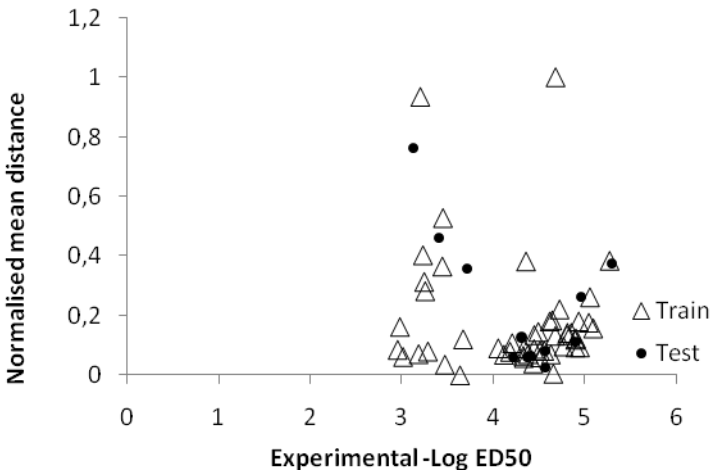

Figure 2-Diversity analysis of database compounds

QSAR model and quality

The model reported by the GFA used in the in the study is presented below

$$-\text{LogED}_{50} = 0.218(\pm0.215) - 0.079(\pm0.026)\mu + 0.429(\pm0.077)\varepsilon\text{LUMO} + 0.011(\pm0.004)\text{PSA} + 0.025(\pm0.001)\text{WAA} + 0.008(\pm0.007)\text{APA} - 0.187(\pm0.042)\text{QA} + 0.602(\pm0.350)\text{SQH} + 1.045(\pm0.196)\text{SSQF} \quad (20)$$

In equation 20, values in the parenthesis are the standard deviation. The equation contained m = 8 descriptors and was obtained from n = 52training compound, then, its Topliss ratio (n/m) is 6.5 which is greater than 5.This showed the model obeyed the QSAR semi-empirical rule of thumb[14]. The model was used to predict the activity values for training set compounds and the residual produced by the model was standardized using equation 18. The plot of its SDR against the predicted activity value (**Fig**. 3a) showed that the residuals wereevenly distributed around the line SDR equal zero, indicating the absence of systematic error in the model[4]. In addition, the plot of residuals against experimental value gave a determination coefficient far less than 0.5 (**Fig** 3b).Furthermore, a linear relation that existed between the models predicted activity and experimental activity (**Fig**.4), indicating concordance between the two variables and a good model.

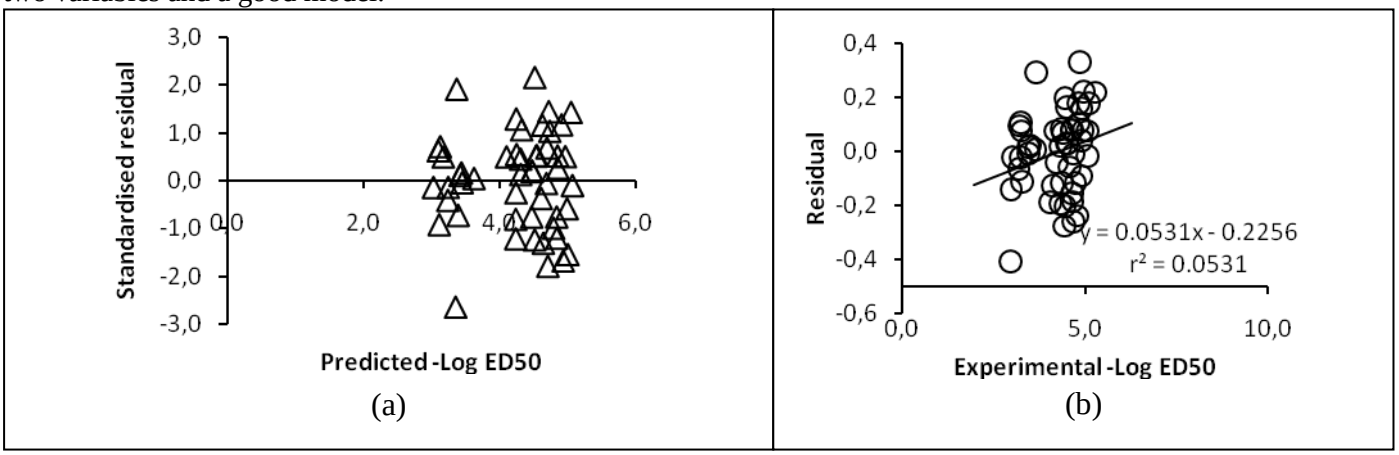

Fig.3 (a) plot of standardized residual against predicted anticonvulsant activity values in logarithm unit for the training set data, (b) plot of residual against experimental anticonvulsant activity values in logarithm unit for the training set data

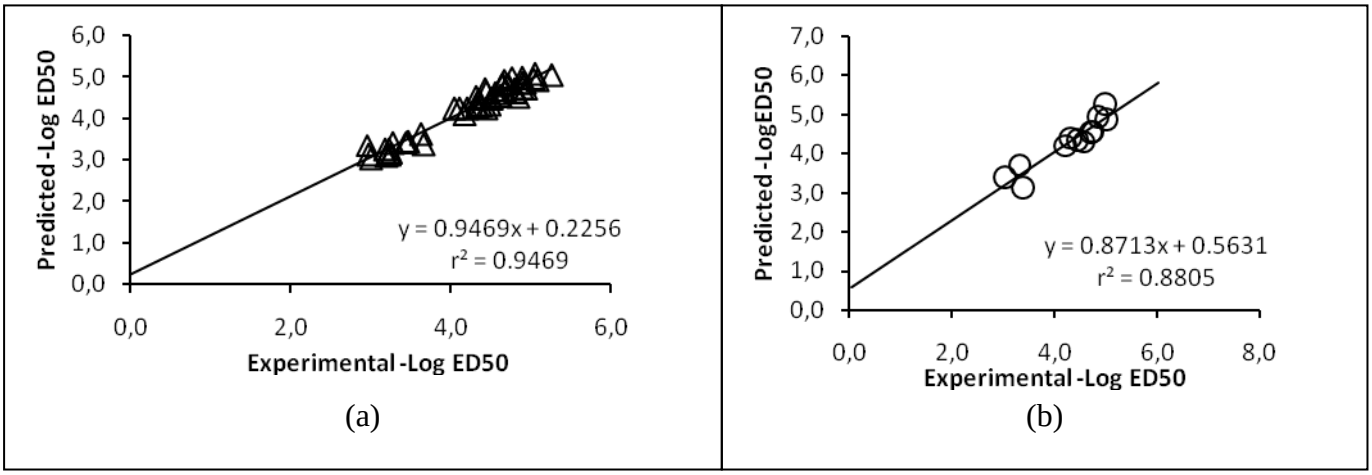

Fig. 4-plot of predicted versus experimental anticonvulsant activity values in logarithm unit for (a) training set and (b) test set data

Table 3 presents the results of correlation and multi-co-linearity analysisperformed on the training set data. It showed that the absolute value for highest correlation coefficient between any two descriptors in the model was 0.822 which is accep**Table** in terms of the data pretreatment condition used in the study. Also, the highest VIF value for descriptors in the mode was 5.149, indicating the model was accep**Table** and void of the multi-co-linearity problem [12]

Table 3- Descriptors correlation matrix and variance inflation factor

	μ	*εLUMO*	*PSA*	*WAA*	*APA*	*QA*	*SQH*	*SSQF*	*VIF*
μ	1								2.866
εLUMO	-0.303	1							1.638
PSA	0.641	-0.475	1						5.149
WAA	0.434	-0.415	0.461	1					2.760
APA	0.634	-0.535	0.822	0.313	1				4.608
QA	0.483	-0.228	0.509	0.682	0.329	1			3.645
SQH	0.558	-0.126	0.565	0.290	0.329	0.650	1		3.557
SSQF	-0.370	-0.111	-0.054	0.162	-0.047	0.035	-0.389	1	1.787

Model validation parameters

Detailed of the validation parameter for the model computed by the MLRplusValidation1.3 is presented in **Table** 4. The result showed that R^2 and R^2_{adj}> 0.5 indicating the explained variance by the model is higher than the unexplained one andlinear relationship existed between the descriptors and the anticonvulsant activity of the studied compounds. The variance ratio $F_{(8,41)}$ reported for the model is greater the critical-F(2.17) for 8,41 degree of freedom at 95% level. Hence, the probability that the model is significant at 95% is high. Average R^2_r and average Q^2_r for the randomized model were less than 0.5, indicating the model was not a product of chance correlation[15]. Randomization parameters $^cR_p^2$> 0.5 confirmed this claim. Furthermore,determination coefficient R^2 reported for the training set andsquared correlation coefficient r^2 reported for test set showed the model hadgood predictive ability for both training and test set data. The result also indicated that the model passed all[16] criteria for a predictive model.

Model applicability domain

The extent warning leverage for the model h* was 0.54. Therefore, the AD of the model is defined by a square area bounded by 0< h < 0.54 and -3 < SDR < 3 as presented pictorially by the models Williams plot (**Fig**. 5). At a glance, the **Fig**ure showed that almost all the datasetcompounds were within the AD of the model except for compound 13 and 14 of the training set . Y is the observed activity value for training set, $\bar{Y}$, the average of the observed activity for training set $\hat{Y}$, Predicted activity for training set, $\hat{Y}_{loo}$ leave one out cross-validation predicted activity for training, Y_{ext}observed activity for thetest set,and$\hat{Y}_{ext}$ predicted activity for the test set

with leverage value greater than the warning leverage. Hence they are influentialcompounds[17]. The model had high-quality parameters and can be used to predict compounds that are alien to the model building process provided it is within the descriptors space of the model.

Table 4- Model Validation parameters and their threshold values

Parameter	Formula	Threshold	Model score	comment	Ref.
Internal validation					
R^2	$\frac{\left[\sum\left\{(Y-\overline{Y})\times(\hat{Y}-\overline{\hat{Y}})\right\}\right]^2}{\sum(Y-\hat{Y})^2\times\sum(\hat{Y}-\overline{\hat{Y}})^2}$	$R^2 > 0.6$	0.947	passed	[11]
R^2_{adj}	$\frac{(N-1)\times R^2-p}{N-1-p}$	$R^2_{adj} > 0.6$	0.936	passed	
Q^2	$1-\frac{\sum(Y-\hat{Y}_{loo})^2}{\sum(Y-\overline{Y})^2}$	$Q^2 > 0.6$	0.924	Passed	
$F_{(8,41)}$	$\frac{\sum(Y-\overline{Y})^2}{p}\Big/\frac{\sum(Y-\hat{Y})^2}{N-p-1}$	$F_{(8,63)} > 2.09$	91.42	Passed	
Random model					
$\overline{R}_r$	an average of the correlation coefficient for randomized data	$\overline{R} < 0.5$	0.392	Passed	[11]
$\overline{R}_r^2$	an average of determination coefficient for randomized data	$\overline{R}_r^2 < 0.5$	0.194	Passed	
$\overline{Q}_r^2$	an average of leave one out cross-validated determination coefficient for randomized data	$\overline{Q}_r^2 < 0.5$	-0.263	Passed	
${}^cR_p^2$	$R^2\times\left(1-\sqrt{\lvert R^2-\overline{R}_r^2\rvert}\right)$	${}^cR_p^2 > 0.6$	0.866	Passed	[15]
External validation					
R^2_{Pred}	$1-\frac{\sum(Y_{ext}-\hat{Y}_{ext})^2}{\sum(Y_{ext}-\overline{Y})^2}$	$R^2_{pred} > 0.6$	0.881	Passed	
r^2	Coefficient of determination for the plot of predicted versus observed for test set	$r^2 > 0.6$	0.881	Passed	[16]
r_0^2	r^2 at zero intercept		0.880	Passed	
$r_0'^2$	r^2 for the plot of observed versus predicted activity for thetest set at zero intercept		0.861	Passed	
$\lvert r_0^2-r_0'^2\rvert$		$\lvert r_0^2-r_0'^2\rvert < 0.3$	0.019	Passed	
k	Slope of the plot of predicted versus observed activity for test set at zero intercept	$0.85 < k < 1.15$	0.998	Passed	
$\frac{r^2-r_0^2}{r^2}$		$\frac{r^2-r_0^2}{r^2} < 0.1$	0.000	Passed	
k'	Slope of the plot of observed versus predicted activity at zero intercept	$0.85 < k' < 1.15$	0.999	Passed	
$\frac{r^2-r_0'^2}{r^2}$		$\frac{r^2-r_0'^2}{r^2} > 0.1$	0.022	Passed	

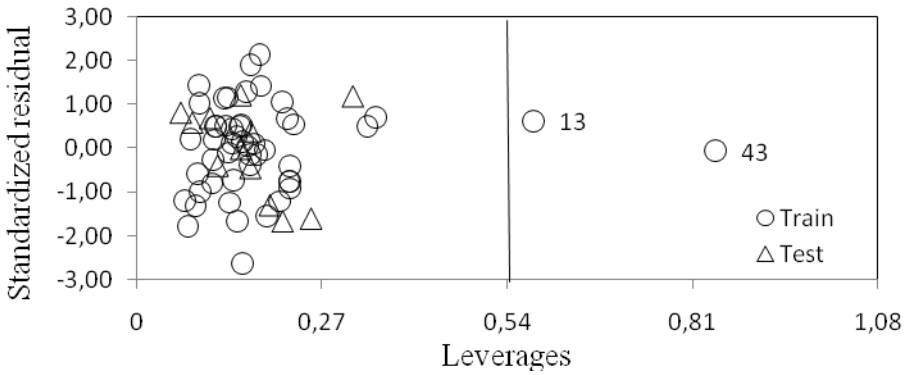

Fig. 5-William's plot for the model

Interpretation of descriptor and in silico design of molecules

A brief description of the descriptors contained in the model is given in **Table** 5. From the **Table** increase in the value of ϵLUMO, PSA, WAA, APA, SQH and SSQF increases the anticonvulsant activity of the dataset compounds and increase in the value of μ and QA reduces the anticonvulsant activity of the compounds. Addition of element with more electronegativity than C (e.g. F, Cl, O,and N) and increment in molecular size increases the values of ϵLUMO, PSA, WAA, APA, SQH and SSQF. Also, addition of electron donating substituent to the molecular system augments the value of these descriptors. With this information, modification was made around the pyrrole ring of the template (compounds 61). A number of compounds were designed and screened with the QSAR model and molecular docking with GABA_AT.

Table 5-A brief description of the descriptors contained in the model

Symbol	Definition	Coefficient
μ	the dipole moment	-0.079(±0.026)
εLUMO	the energy of the lowest unoccupied molecular orbital	0.429(±0.077)
PSA	polar surface area	0.011(±0.004)
WAA	molecule accessible area obtained from wave function calculation	0.025(±0.001)
APA	accessible polar area	0.008(±0.007)
QA	the sum of the absolute value of charges on all atoms in a molecule	- 0.187(±0.042)
SQH	the sum of charges on all hydrogen atoms in a molecule	0.602(±0.350)
SSQF	Square root of sum of square of charges on all fluorine atoms in a molecule	1.045(±0.196)

The newly designed hypothetical molecules with their hypothetical activity value are presented in **Table** 6. 21 hypothetical compounds were designed showing better activity value than the template. Substituting Cl atom of the chlorobenzene moiety of the template with more electronegative F gave compounds with improved activity value (compare molecules 1 and 2 to others). Increase in the molecular structure complexity by substituting the more linear carbonyl group of the ethyl formate moiety with less linear $-NOCH_3$, $-NNCH_3$, $-NN(CH_3)_2$ gave molecules with improved activity value (molecules 3 – 7). A similar observation was made when morpholine ring system of the template was substituted with a larger ring system (molecules 17 – 21). These observations demonstrated the applications of the QSAR model as a tool to optimize the activity of an existing molecule.Molecular docking analysis revealed that all designed molecule showed abetter binding affinity for GABA_AT compares to vigabatrin. However, the type and number of amino-acid each molecule binds to GABA_AT differs, implying that the mechanism of their

interaction with thetarget is different. In addition, correlation analysis between binding affinity and predicted activity of the designed molecules gave correlation coefficient of 0.062. This implied there is no correlation between the binding affinity of the designed molecule and their predicted hypothetical activity value.

Table 6.1-Docking score and predicted anticonvulsant activities of the designed molecules

No.	R1	R2	R3	R4	X	BA ($Kcalmol^{-1}$)	Amino acid	hi	Log (1/ED50)
1	H	$C_3H_5O_2$	C_3H_4NNH	C_6H_4Cl	CH	-21.91	HIE44, ILE72, PHE189,ARG192,GLU270, B:TRY348,PHE351	0.10	4.91
2	H	$C_3H_5O_2$	$C_6H_{12}N$	C_6H_4Cl	CH	-23.71	TRY69,LYS203, HIS206,ILE426, GLY438,ARG445,B:TRY348	0.45	4.97
3	H	C_3H_5OS	C_4H_8NO	C_6H_4F	CH	-25.90	HIE44,TRY69,LYS203,HIS206,GLU207,ASN423, B:TRY348	0.36	5.27
4	H	C_3H_5ONH	C_4H_8NO	C_6H_4F	CH	-18.12	HIE44,TRY69,LYS203,HIS206,GLU207,ASN423, B:TRY348	0.25	5.16
5	H	$C_3H_5ONOCH_3$	C_4H_8NO	C_6H_4F	CH	-20.78	HIE44,TRY69,HIS206,GLU207,ARG430,GLY438, TRY348	0.40	5.04
6	H	$C_3H_5ONNHCH_3$	C_4H_8NO	C_6H_4F	CH	-18.24	TRY69,LYS203, HIS206,GLU207, ARG430,GLY440,TRY348	0.36	5.02
7	H	$C_3H_5ONN(CH_3)_2$	C_4H_8NO	C_6H_4F	CH	-17.84	ILE77,LYS203,HIS206,GLU270,ASN423,GLY438, TRY348	0.35	5.36
8	H	$C_3H_5O_2$	C_4H_8NNH	C_6H_4F	CH	-21.41	TRY69,LYS203, HIS206,GLU270, ARG422,430, B:TRY348	0.34	5.36
9	H	$C_3H_5O_2$	$C_4H_8NNCH_3$	C_6H_4F	CH	-20.29	HIE44,TRY69,HIS206,GLU207,ASN423,GLY438, TRY348	0.41	5.37
10	H	$C_3H_5O_2$	C_4H_4NNH	C_6H_4F	CH	-21.55	TRY69,PHE189, HIS206,GLY438, B:TRY348,PHE351	0.45	5.23

Table 6.2 continued-Docking score and predicted anticonvulsant activities of the designed molecules

No	R_1	R_2	R_3	R_4	X	BA (Kcalmol-1)	Amino acid	h_{ii}	Log (1/ED50)
11	H	$C_3H_5O_2$	C_4H_4NO	C_6H_4F	CH	-22.61	HIE44,TRY69,ARG192,LYS203, HIS206,GLU270, B:TRY348	0.34	4.90
12	H	$C_3H_5O_2$	C_4H_4NS	C_6H_4F	CH	-22.51	HIE44,ARG192, GLU270,ILE426, GLY438,440,B:ILE105	0.39	4.99
13	H	$C_3H_5O_2$	C_3H_4NNH	C_6H_4F	CH	-22.55	HIE44,ILE72,PHE189,ARG192,HIS206,GLU270,GLY438	0.51	5.13
14	H	$C_3H_5O_2$	C_2H_4NO	C_6H_4F	CH	-18.17	HIE44,TRY69,PHE189,HIS206,GLY438, B:TRY348, PHE351	0.43	5.12
15	H	$C_3H_5O_2$	$C_2H_5N_2$	C_6H_4F	CH	-16.04	TRY69,PHE189, ARG192,HIS206, GLY438, B:TRY348	0.04	5.21
16	H	$C_3H_5O_2$	$C_6H_{12}N$	C_6H_4F	CH	-23.16	TRY69,ILE72,GLY438,ASN423,ILE426, B:TRY348	0.06	4.90
17	H	$C_3H_5O_2$	$C_5H_{10}N$-20-O	C_6H_4F	CH	-15.41	HIE44,TRY69,ILE72, ARG192,LYS203, GLU207,GLY440,B:ILE105	0.04	4.89
18	H	$C_3H_5O_2$	$C_5H_{10}N$-18-O	C_6H_4F	CH	-21.14	TRY69,ILE72,ARG422,ASN423, GLY438 B:TRY348	0.12	4.91
19	H	$C_3H_5O_2$	$C_5H_{10}N$-19-O	C_6H_4F	CH	-13.84	TRY69,ILE72,LYS203,HIS206,GLU270,GLY438 B:TRY348,PHE351	0.04	4.92
29	H	$C_3H_5O_2$	$C_5H_{10}N$-18-S	C_6H_4F	CH	-19.45	TRY69,ILE72,ARG422,ASN423, GLY438 B:TRY348	0.10	5.16
21	H	$C_3H_5O_2$	$C_5H_{10}N$-19-NH	C_6H_4F	CH	-15.99	HIE44,TRY69,HIS206,ASN423,GLY438, B:TYR348	0.04	5.00
121	4-aminohex-5-enoic acid					-10.97	TYR69, HIS206, GLU270, B:TYR348		

4.CONCLUSION

A well statistically validated and robust QSAR model was constructed for compounds withanticonvulsant activity in maximal electroshock seizure test using quantum molecular descriptors obtained from PM3 semi-empirical quantum calculation. The study demonstrated the application of QSAR model as a knowledge generator by designing and screening, *in silico*, 21 hypothetical molecules with improved predicted hypothetical anticonvulsant activity values. Furthermore, docking of the hypothetical molecules with GABA_AT revealed they possess better hypothetical binding affinity for the target compares to vigabatrin which a real and known inhibitor of the target. The results of the studyis useful to screen hypothetical molecules before committing to actual synthesis and *in vivo* experiments to assess the capability of the hypothetical molecules or compounds to actually show real activity in MES test and inhibit GABA_AT.

References

[1] S. Bansal, B. Sinha, R. Khosa, Med. Chem. Res. 21 (2012) 3063–3072.

[2] B. LIPPERT, B.W. METCALF, M.J. JUNG, P. CASARA, FEBS J. 74 (1977) 441–445.

[3] H.S. White, Epilepsia 40 (1999).

[4] D.E. Arthur, A. Uzairu, P. Mamza, S.E. Abechi, G. Shallangwa, Beni-Suef Univ. J. Basic Appl. Sci. 5 (2016) 320–333.

[5] U. Abdulfatai, A. Uzairu, S. Uba, J. Adv. Res. 8 (2017) 33–43.

[6] E. Estrada, A. Peña, Bioorg. Med. Chem. 8 (2000) 2755–2770.

[7] M. Choudhary, B.K. Sharma, Pharma Chem. 6 (2014) 321–330.

[8] Y. Shao, L.F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S.T. Brown, A.T. Gilbert, L.V. Slipchenko, S.V. Levchenko, D.P. O'Neill, others, Phys. Chem. Chem. Phys. 8 (2006) 3172–3191.

[9] M. Karelson, V.S. Lobanov, A.R. Katritzky, Chem. Rev. 96 (1996) 1027–1044.

[10] P. Ambure, R.B. Aher, A. Gajewicz, T. Puzyn, K. Roy, Chemom. Intell. Lab. Syst. 147 (2015) 1–13.

[11] A. Tropsha, Mol. Inform. 29 (2010) 476–488.

[12] A. Beheshti, E. Pourbasheer, M. Nekoei, S. Vahdani, J. Saudi Chem. Soc. 20 (2016) 282–290.

[13] Y.A. Arnautova, R.A. Abagyan, M. Totrov, Proteins Struct. Funct.Bioinforma. 79 (2011) 477–498.

[14] S.V. Damme, P. Bultinck, J. Comput. Chem. 28 (2007) 1924–1928.

[15] K. Roy, Expert Opin. Drug Discov. 2 (2007) 1567–1577.

[16] A. Golbraikh, A. Tropsha, J. Mol. Graph. Model. 20 (2002) 269–276.

[17] T.I. Netzeva, A.P. Worth, T. Aldenberg, R. Benigni, M.T. Cronin, P. Gramatica, J.S. Jaworska, S. Kahn, G. Klopman, C.A. Marchant, others, ATLA 33 (2005) 155–173.