QSAR AND MOLECULAR DOCKING BASED DESIGN OF SOME N-BENZYLACETAMIDE AS Γ-AMINOBUTYRATE-AMINOTRANSFERASE INHIBITORS

O. ADEDIRIN [1], A. UZAIRU[2], G.A. SHALLANGWA[2] and S.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: Chemistry Advance Research Center, Sheda Science and Technology Complex, FCT, Nigeria, Phone: +234800593145
E-mail address: adedirinoluwaseye@yahoo.com; senguade@gmail.com

ABSTRACT

Quantitative structure activity relationship study (QSARs) and molecular docking were used to design and virtually screen some new N-benzylacetamide derivatives for their ability to inhibit γ-amino butyrate-aminotransferase. Ninety compounds with anticonvulsant activity against maximal electroshock induced seizures were used for QSAR study. HF/DFT B3LYP/6-31G** quantum mechanical method was employed to optimize/minimize the molecular structure of these compounds. Genetic Function Algorithm (GFA) method was used to develop the QSAR models. Each model gave an octa-parametric equation with good statistical qualities (R^2 ranged from 0.823 to 0.893, Q^2 from 0.772 to 0.854, F from 36.53 to 37.10, R^2_{pred} (test) from 0.768 to 0.893). Information obtained from the parameter contained in the models suggested that increasing the molecular mass and linearity of molecule would lead to increase in anticonvulsant activity of studied compounds. These informed the design and virtual screening of 118 new N-benzylacetamide derivatives using 2-acetamido-N-benzyl-2-(5-methylfuran-2-yl)acetamides as the template. The designed molecules were docked with γ-amino butyrate-aminotransferase (GABA_AT; PDB: 1OHV) using Internal Coordinate Mechanics Program (ICM-pro 3.8-3). The binding affinity of the designed compounds with GABA_AT were better to that of 4-aminohex-5-enoic acid (vigabatrin); 3, 3-diphenylpyrrolidine-2, 5-dione (phenytoin) and comparable to that of 5H-dibenzo [b,f]azepine-5-carboxamide (carbamazepine), which are known inhibitors of GABA_AT. Therefore, the designed molecules have potential as inhibitors of GABA_AT and consequently as anticonvulsant agent.

1. INTRODUCTION

One of the major inhibitory amino acid neurotransmitter of the mammalian central nervous (CNS) system is γ-amino butyric acid (GABA). Reduction of its concentration in the brain has been implicated not only in symptoms associated with epilepsy but also several other neurodegenerative/psychiatric conditions like stroke, anxiety, schizophrenia and so on (WHITING, 2003). The enzyme γ-amino butyrate-aminotransferase (GABA_AT) catalyzes the degradation GABA to succinic semi aldehyde thereby reducing its level in the brain. When the level diminishes below certain threshold, convulsions result (PAOLA *et al.*, 2004). Direct administration of GABA peripherally to remedy this has been reported not feasible because of difficulty in crossing the blood-brain barrier (DAVID *et al.,* 2000; PHYLLIS, 2011). Therefore, inhibition of the activity of GABA_AT is a way to raise cerebral concentrations of GABA and this has become the target for many anticonvulsant drugs. Numerous anticonvulsant molecules (antiepileptic drugs, AEDs) that inhibit the action of GABA_AT have been developed over the years e.g. 4-aminohex-5-enoic acid (vigabatrin) (PAOLA *et al.*, 2004). However, with optimal usage of the available AEDs, epilepsy still threatens more than 50 million people worldwide (USMAN *et al.,* 2017). Roughly 20% to 30% of patients don't react to marketed AEDs and those that responded do so at the risk of other side effects like depression, agitation, tremor, double vision, poor vision and etc. (MATTSON, 1995). In this light, developing new molecule with enhanced antiepileptic activity and lessened side effect is a fundamental task for medicinal science.

Drug discovery and development is an arduous task, yet with the assistance of novel drug discovery methods like computer aided drug design (CADD), excellent leads which will probably prevail in clinical trials can be developed (IBEZIM *et al.*, 2009). The objective of the present study is to utilize quantitative structure activity relationship study (QSARs) strategy to design new 2-acetamido-N-benzyl-2-(5-methylfuran-2-yl)acetamides with enhanced anticonvulsant activity values against maximal electroshock induced seizure (MES) in view and other use molecular docking method to investigate the capacity of the designed compounds to interact with a known crystal structure of GABA_AT obtained from protein data bank (PDB: 1OHV).QSAR is a CADD technique that relates quantitative measure of chemical structure (i.e. molecular descriptors) of compounds to their activities employing regression or classification based approaches. It offers an in silico tool for the development of predictive models that can be used to propose the activities of known and hypothetic chemical entities (IBEZIM *et al.*, 2009; SING, 2013). Molecular docking is a CADD technique that explores the binding mode of two interacting molecules relying on their topographic features or energy consideration (LESK, 2002; PEDRO AND JOHN, 2010).

2. MATERIAL AND METHODS

2.1 Dataset

Ninety derivatives of N-benzylacetamide were used as dataset. Their structural formula and anticonvulsant activity values against maximal electroshock seizure test were taken from literatures (KING, 2010). The activity values were expressed as ED_{50} (mg/kg) (concentration of compound that is effective on 50% tested animals). ED_{50} (mg/kg) was converted to molar unit and subsequently to logarithm unit by taking the logarithm of the inverse of ED_{50} (mol/kg). The resulting scaled anticonvulsant activity values (pED_{50}) and the corresponding molecular structure in the dataset are presented in Table 1.

2.2 Dataset splitting

For modeling purpose, the complete dataset was divided into training sets and test sets using Kennard and Stone algorithm (KENNARD AND STONE, 1969) 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:

$$d_{ij} = \|\mathbf{X_i} - \mathbf{X_j}\| = \sqrt{\sum_{k=1}^{m}(x_{ik} - x_{jk})^2} \qquad (1)$$

In Equation 1, 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 largest value of d_{ij}. Compounds that exhibits 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.

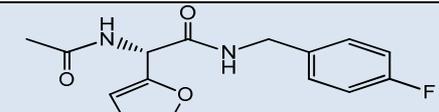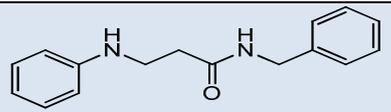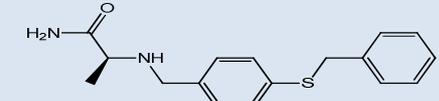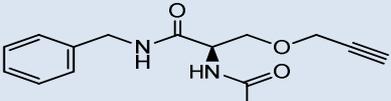## 2.3 Molecular descriptors calculation

Build module available in Spartan 14 (SHAO et al, 2006) was used to generate the 2D molecular structures of dataset compounds. These were subsequently converted to 3D by view module in the software. Conformation of structures were optimized using HF/DFT B3LYP/6-31G** quantum mechanical method to ensure a well-defined conformer relationship across the compounds. The energy minimized structures were ported to PaDEL-Descriptor (YAP, 2011) used to compute various 0D, 1D, 2D and 3D-classes of chemometric molecular descriptors.

2.4 Model development

Genetic function algorithm (GFA) module available in Material Studio 7.0 (ACCERLYS, 2007) was used for QSAR model development. GFA uses genetic algorithm to search over the entire data space for possible QSAR models and uses Friedman lack of fitness (LOF) function to estimate the fitness of each model. The activity values and descriptors of the training set only were used to generate the models. GFA module parameters were set as equation length (5 to 12), population (10000), maximum generation (500), number of top equation returned (5), mutation probability (0.1), and scaled LOF smoothness (0.5). The advantages of GFA over other techniques include: production of multiple models, automatic selection of the exact number of descriptors needed to build a full-size model, resistant to over-fitting, allowance of user control over the smoothness of fit and length of equation, allowance for the use of higher-order polynomials, and applicability when number of descriptors is more than the number of dataset (ROGERS AND HOPFINGER, 1994).

Table 1-Molecular structure and anticonvulsant activity values for dataset compounds

No.	Molecular structure	pED$_{50}$	No.	Molecular structure	pED$_{50}$
1[a]		5.009	14		4.682
2		5.003	15[a]		4.628
3		4.949	16		4.525
4		4.933	17		4.546
5		4.882	18		4.567
6		4.857	19		4.447
7[a]		4.720	20		4.479
8		4.592	21[a]		4.368
9		4.603	22		4.529
10		4.530	23		4.284
11		4.423	24		4.034
12[a]		4.275	25		4.228
13		4.428	26		4.268
27		4.383	42		4.218

#	Structure	Value	#	Structure	Value
28		4.360	43[a]		4.167
29		4.373	44[a]		4.234
30		4.333	45		4.173
31		4.093	46		4.171
32		4.174	47		4.077
33[a]		4.125	48[a]		4.032
34		4.144	49		3.923
35		4.073	50		3.774
36[a]		4.096	51		3.805
37		4.083	52		3.811
38		4.085	53[a]		4.092
39		4.225	54		4.107
40		4.113	55		4.158
41[a]		4.107	56		3.864
57		3.770	74		3.733
58		3.866	75		3.809
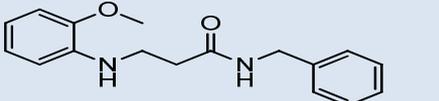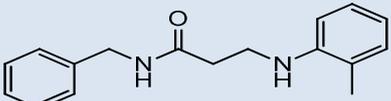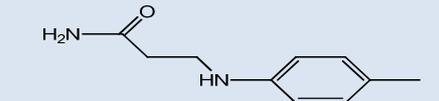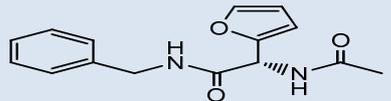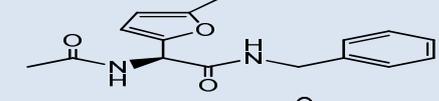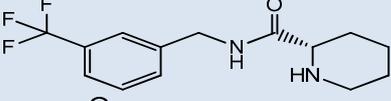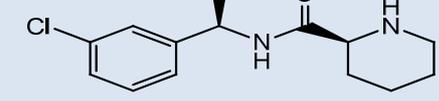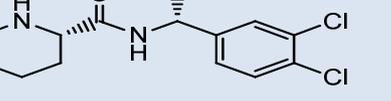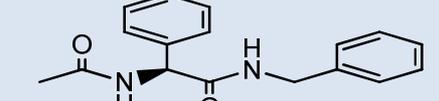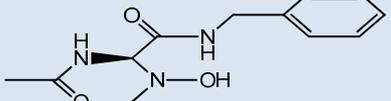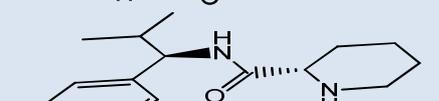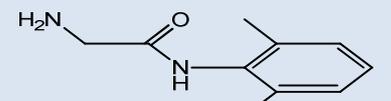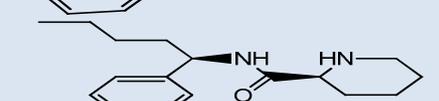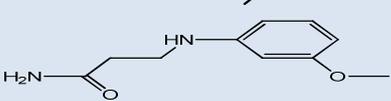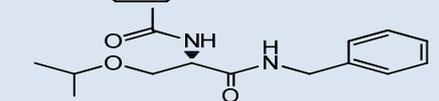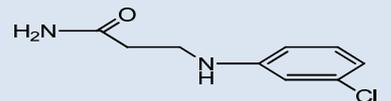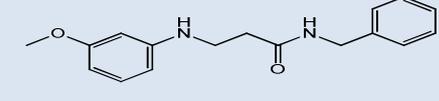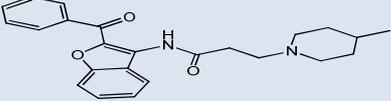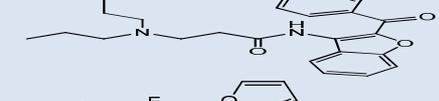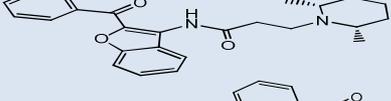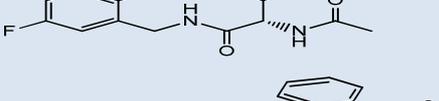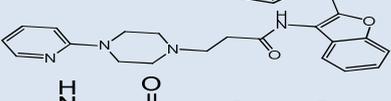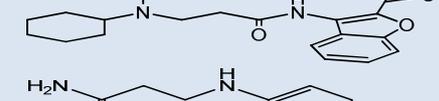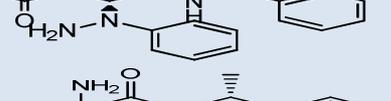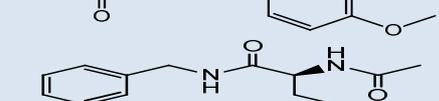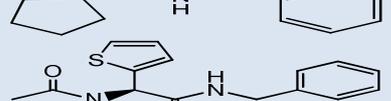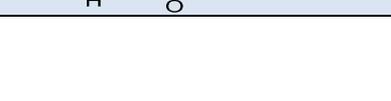
59		3.764	76[a]		3.741
60		4.013	77		3.821
61		3.861	78		3.694
62		3.834	79		3.898
63		3.672	80		3.728
64		3.770	81		3.593
65		3.839	82[a]		3.460
66		3.630	83		3.489
67		3.568	84		3.768
68		3.607	85[a]		3.472
69		3.545	86		2.933
70		3.690	87[a]		2.693
71		3.708	88		2.684
72		3.747	89[a]		2.610
73[a]		3.483	90		2.728

[a] compound in the test set

2.5 Model validation

The models produced were used to predict the anticonvulsant activity values for the test set compounds and the result obtained were used to check for the presence of systematic error in the models using the criteria of ROY et al. (2016). In the absence of systematic error, the models were validated with various internal and external validation parameters. Internal validation was done with the training set data only using the leave-one-out (LOO) cross-validation, y-randomization and other metrics obtained from multiple linear regressions (MLR) carried out on each models. In the leave-one-out cross-validation techniques, the training set is primarily modified by eliminating one compound from the data set and using the remaining data to construct a QSAR equation using the same descriptor combination contained in the model being validated. The new equation obtained was then used to predict the activity of the eliminated compound. The cycle was repeated until all the molecules of the training set had been eliminated once. The obtained LOO predicted activity was used to calculate various parameters including predicted error sum of

square (PRESS), standard deviation of error of prediction (SDEP) and LOO square correlation coefficient (Q^2) using the following equation

$$PRESS = \sum \left(Y_{obs(train)} - Y_{pred(train)} \right)^2 \qquad (2)$$

$$SDEP = \sqrt{\frac{PRESS}{n}} \qquad (3)$$

$$Q^2 = 1 - \frac{PRESS}{\sum \left(Y_{obs(train)} - \overline{Y}_{train} \right)^2} \qquad (4)$$

In equation 2 - 4, $Y_{obs(train)}$ is the observed activity values for the training set data, $Y_{pred(train)}$ and is the predicted activity values of the training set data based on the LOO technique, $\overline{Y}_{train}$ is the average of the observed activity value for the training set and n is the number of observation in the training set. The threshold value of Q^2 is 0.5.

In y-randomization techniques, the observed activity values were permuted several times keeping the descriptor matrix unchanged. For each permutation, a new model was developed at the same confidence level as the original model. Then square correlation coefficients for the randomized models R^2_r were estimated (TROPSHA et al., 2003). Deviation in the value of the average of square correlation coefficient of the randomized models $\overline{R}^2_r$ from the square correlation coefficient of non-randomized models R^2 was used to calculate the Y-randomization parameter $^cR^2_p$:

$$^cR^2_p = R \times \sqrt{R^2 - \overline{R}^2_r} \qquad (5)$$

Threshold value of $^cR^2_p$ is 0.5. The MLR metrics used included determination coefficient (R^2), adjusted determination coefficient (R^2_{adj}), standard error of estimation (SEE) and variance ratio (F). Only the test set data was used to validate the model externally using predicted square correlation coefficient R^2_{pred} and modified square correlation coefficient R^2_m (ROY et al., 2012). R^2_{pred} reflect the degree of correlation between the observed and predicted activity data for the test set and its defined:

$$R^2_{pred} = 1 - \frac{\sum \left(Y_{obs(test)} - Y_{pred(test)} \right)^2}{\sum \left(Y_{obs(test)} - \overline{Y}_{training} \right)^2} \qquad (6)$$

In equation 6, $Y_{obs(test)}$ and $Y_{pred(test)}$ are the observed and predicted activity data for the test set compounds, while $\overline{Y}_{training}$ indicates the mean observed activity of the training set. The threshold value for R^2_{pred} is 0.5. Modified square correlation coefficient (R^2_m) is defined:

$$R^2_m = r^2 \times \left(1 - \sqrt{(r^2 - r_0^2)} \right) \qquad (7)$$

In the equation 6, r^2 and r_0^2 are the square correlation coefficients of the plot of observed against predicted activity values for test set data with and without intercept respectively. The threshold value for R^2_m is 0.5. Other criteria used for external predictive capacity validation were (a) $Q^2 > 0.5$; (b) $R^2_{pred} > 0.6$; (c) $r^2 - r_0^2/r^2 < 0.1$ and $0.85 \le k \le 1.15$ or $r^2 - r'^2/r^2 < 0.1$ and $0.85 \le k' \le 1.15$ and (d) $|r_0^2 - r_0'^2| < 0.3$. In the criteria Q^2, R^2_{pred}, r^2 and r_0^2 were as discussed before. While, 'k' is the slope of the graph of observed versus predicted activity, r'^2_0 is the square correlation

coefficients between predicted versus observed activity for test set data without intercept and k′ is the slope (GOLBRAIKH AND TROPSHA, 2002). Furthermore, error-based criteria were used to check if the predictions of the models are good, bad or moderate. These criteria were based on the mean absolute error (MAE) defined as:

$$MAE = \frac{1}{n} \times \sum |Y_{obs} - Y_{pred}| \qquad (8)$$

In equation 8, Y_{obs} and Y_{pred} are experimental and predicted response values for the test set data only. The criteria stated that when the number of test set data point is at least 10 (for statistical reliability) and there is no systematic error in the model (for statistical applicability) then, (a) for good prediction, MAE $\le$ 0.1 × training set response range or MAE + (3 × σ) $\le$ 0.2 × training set response range, (b) for bad prediction MAE > 0.15 × training set response range or MAE + (3 × σ) > 0.25 × training set response range, and (c) any prediction which does not fall condition a and b may be considered as of moderate quality. In the criteria, σ denotes the standard deviation of the absolute error values for the test set data (ROY et al., 2016).

2.6 Models applicability domain (AD)

The AD of a QSAR model is the physical-chemical, structural or biological space, knowledge or information on which the training set of the model has been developed, and for which it is applicable to make predictions for new compounds (JAWORSKA et al., 2005). In the study, model AD was investigated with extent of extrapolation method which is based on the leverages value (h) and the standardized residual (SDR) for each molecule in the dataset, produced by the model. The plot of SDR versus h (Williams plot) gave simple pictorial representation of the AD. In the plot, the AD is established inside a squared area within ± 3 SDR and a threshold leverage (h*), which is generally fixed at 3 (k + 1)/n, where n is the number of training-set molecules and k is the number of descriptors in the model. Any prediction by the model for any compounds whose leverage value is higher than the threshold leverage is considered unreliable. On the other hand, when the leverage value of a compound is lower than the threshold value, the probability of accordance between predicted and observed values is as high as that for the training-set compounds.

S.7 In silico design of title compound

(S)-2-acetamido-N-benzyl-2-(5-methylfuran-2-yl) acetamide (Molecule 37) in the dataset (Table 1) was used as template to design about 119 hypothetic derivatives. The template was chosen because of it relatively high activity i.e. pED_{50} = 4.174. The design of the derivatives was guided by the information obtained from the descriptors contained in the models. Modification to the template was done by simple addition and removal of substituent. Molecular geometries of the designed molecules were optimized, descriptors obtained and their leverage values calculated as described for the training dataset. The threshold leverage (h*) of the models was used to screen designed molecules that were 'X-outlier' to the descriptor space of the models. Hypothetical anticonvulsant activity values of the designed molecules within the AD of the models were predicted

with the models and those found with improved activity value were selected for molecular docking study.

2.8 Molecular docking studies

Molsoft Internal Coordinate Mechanic Program (ICM-pro 3.8-3) was used for the docking study. High resolution 2.3 Å crystal structure of GABA_AT (PDB: 1OHV) was downloaded from protein data bank. It was loaded into the ICM-pro workspace where it was prepared prior to docking. PDB: 10HV existed as two asymmetric physiologically identical dimmers containing four chains of amino acids (PAOLA et al., 2004). A dimer (represented by chains C and D), water molecules, intrinsic ligands and heteroatom contained in the pdb files were removed. Then, the remaining dimer (chains A and B) were converted to ICM-object. During conversion, hydrogen and missing heavy atoms were added were automatically added to the protein structure. Also, atom type and partial charges were assigned and the orientations of His, Asn, Glu and Cys residue were optimized. The available pockets in the receptor was located, binding site and grid map were created around the best pocket identified by the software.

The design molecules optimized geometry were saved in a tabular form (chemical table) and used as the ligands molecules. The chemical table was imported into ICM-Pro and dock chemical table module was invoked. Docking thoroughness (10) and number of conformation returned (10) parameters were set. The ligands were flexible during the docking process and their binding pose and internal torsions were sampled by biased probability Monte Carlo (BPMC) global-energy optimization/minimization procedure. Energy calculations were based on the ECEPP/3 force field with a distance-dependent dielectric constant (MOLSOFT, 2012).

3. RESULTS AND DISCUSSION

3.1 QSAR results

A total number of 1845 descriptors were calculated for the data set. The dataset compounds were divided into 72 training and 18 tests set. Only the training set was used to construct the QSAR models utilizing GFA. The test set was used to validate the constructed models externally. The models obtained and their validation parameters are reported in Table 2. Each model contained 8 descriptors and since they were obtained from 70 training set compounds, they do not violate the QSAR rule of thumb (TOPLIS AND COSTELLO, 1972). The rule recommended that the ratio of descriptors to compounds used for building a model should not exceed 1:5; otherwise the risk of chance correlation is high. The reported models are void of systematic error because, their AAE - |AE| are greater than $0.5 \times AAE$ and the square correlation coefficient of the plot of residual against observed activity for both training set $R^2_{(res-train)}$ (0.109 – 0.183) and test set $R^2_{(res-train)}$ (0.356 – 0.437) for them are less than 0.5 (ROY et al., 2016). The values of the R^2_{train} (0.823 - 893) for the models suggested they could explain up to 83.20 % of variances in the observed activities. The variance ratio F (36.54 – 65.37) for the models were greater than critical values of $F_{8, 72}$(2.07) for the 0.05 significance level. This indicated that the models remained significant at 95% level. R^2_{train}, R^2_{adj}, Q^2_{LOO} and $^cR^2_p$ for the models were greater than their threshold value 0.5 and the test set R^2_{pred} (0.830 – 0.893) and R^2_m (0.693 – 0.799) are also greater than threshold 0.5. Therefore, reported models have good internal and external predictive ability; they are robust because they have about 80% explained variance; and are not product of chance correlation (TROPSHA, 2003; ROY et al., 2012). Furthermore, the reported model passed the criteria for predictive model (GOLBRAIKH-TROPSHA, 2002) and criteria for good prediction (ROY et al., 2016).

Principal component regression (PCR) was conducted with the 16 descriptors shared by the GFA model. The result is included in Table 2 which identified eight components to be sufficient in explaining the variances in the data set. The relationship between the residual produced by the models and the observed activity is presented in Figure 1 and linear relationship was observed in the plot of observed versus predicted activity values by the models as presented in .
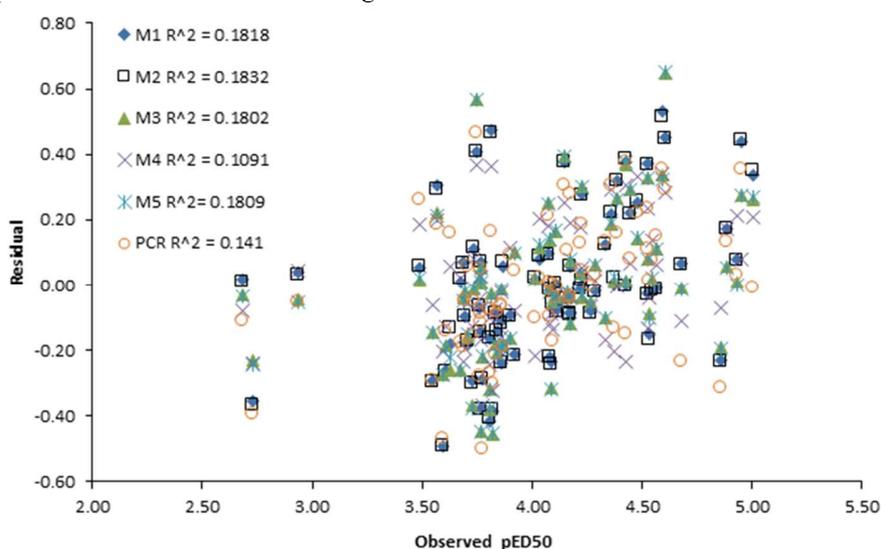
Figure 1-Plots of residual against the observed pED50 values for training set compounds (M1-M5 represent model 1-5 and PCR represent PCR model)

Figure 2. These are indication of the goodness of fit by the models. The observed and predict anticonvulsant activity values for the entire data set are presented in Table 3 and close agreement was observed between them. The statistical parameters for the reported models (Table 2) suggested absent of significant difference between the models. In other to investigate this claim, repeated measure ANOVA was performed on the result presented in Table 3. The result obtained indicated that Mauchly's test of sphericity F-value value under significant was less than 0.05 and Greenhouse-Geisser F-value under significant was greater than 0.05 (table not included). The result implied there is no statistical significant difference between predicted values by the models and observed values therefore; application of the models for consensus prediction will give robust prediction
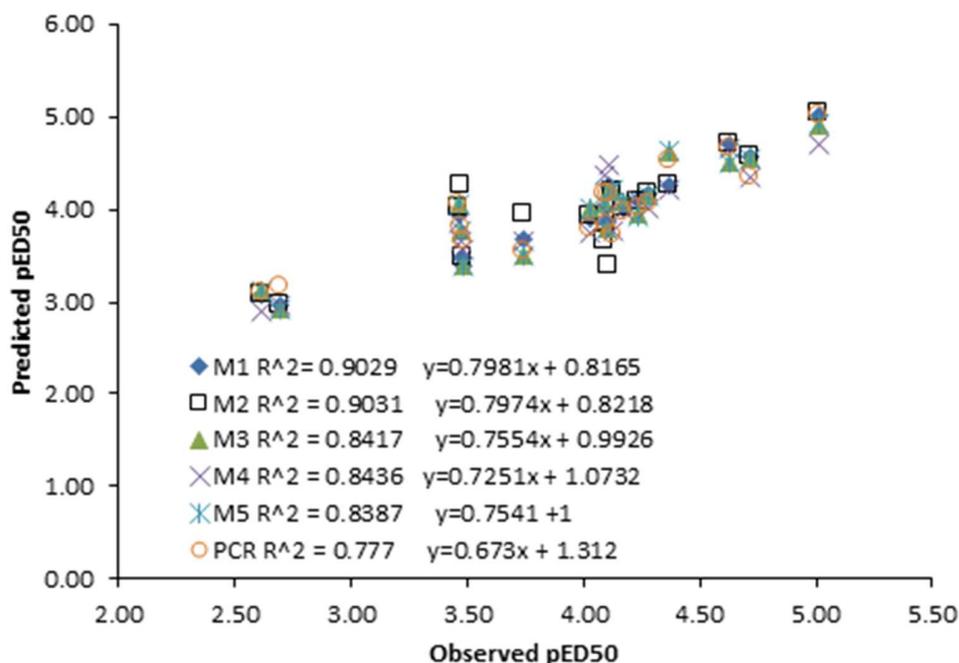
Figure 2-Plots of predicted against the observed pED50 values for test-set compounds (M1-M5 represent model 1-5 and PCR represent PCR model.

Table 2- GFA and PCR obtained QSAR models and their validation parameters

No	Models	Validation parameters
1	$pED_{50} = 4.057(\pm0.024) + 0.391(\pm0.090)AATSC7m - 0.314(\pm0.090)MATS7m + 0.425(\pm0.042)MATS5s + 0.133(\pm0.029)SC\text{-}5 - 0.403(\pm0.038)SdO + 0.249(\pm0.035)IC3 + 0.061(\pm0.031)E1p - 0.172(\pm0.028)\ E3s$	Internal $n = 70$, $R^2_{train} = 0.823$, $R^2_{adj} = 0.801$, $SEE = 0.207$, $F = 36.72$, $Q^2_{LOO} = 0.751$, $PRESS = 3.791$, $SDEP = 0.229$, $^cR^2_p = 0.771$, $R^2_{(res\text{-}train)} = 0.181$ External $R^2_{pred} = 0.893$, $R^2_m = 0.693$, $|r^2_0\text{-}r'^2_0| = 0.046$, $k = 0.998$, $r^2_0\text{-}\ r^2_0/r^2 = 0.014$, $k' = 0.999$, $r^2\text{-}r'^2/r^2 = 0.065$ $AAE\text{-}|AE| = 0.138$, $0.5 \times AAE = 0.069$, $R^2_{(res\text{-}test)} = 0.373$, $MAE = 0.155$, $0.1 \times Range(train) = 0.231$, $0.15 \times Range(train) = 0.347$, $0.2 \times Range(train) = 0.463$, $0.25 \times Range(train) = 0.579$, $MAE + (3 \times \sigma) = 0.466$
2	$pED_{50} = 4.057(\pm0.024) + 0.394(\pm0.091)AATSC7m - 0.323(\pm0.091)MATS7m + 0.424(\pm0.042)MATS5s + 0.133(\pm0.029)SC\text{-}5 - 0.402(\pm0.039)SdO + 0.247(\pm0.036)IC3 - 0.171(\pm0.029)E3e + 0.065(\pm0.031)E1p$	Internal $n = 70$, $R^2_{train} = 0.823$, $R^2_{adj} = 0.800$, $SEE = 0.207$, $F = 36.54$, $Q^2_{LOO} = 0.751$, $PRESS = 3.793$, $SDEP = 0.230$, $^cR^2_p = 0.781$, $R^2_{(res\text{-}train)} = 0.183$ External $R^2_{pred} = 0.892$, $R^2_m = 0.799$, $|r^2_0\text{-}r'^2_0| = 0.046$, $k = 0.998$, $r^2_0\text{-}\ r^2_0/r^2 = 0.015$, $k' = 0.999$, $r^2\text{-}r'^2/r^2 = 0.066$ $AAE\text{-}|AE| = 0.137$, $0.5 \times AAE = 0.078$, $R^2_{(res\text{-}test)} = 0.375$, $MAE = 0.156$, $0.1 \times Range(train) = 0.232$, $0.15 \times Range(train) = 0.348$, $0.2 \times Range(train) = 0.464$, $0.25 \times Range(train) = 0.579$, $MAE + (3 \times \sigma) = 0.466$

3	pED_{50} = 4.057($\pm$0.024) + 0.213($\pm$0.040)AATSC7e -0.151($\pm$0.0409)AATSC7s +0.433($\pm$0.042)MATS5s - 0.328($\pm$0.042)SdO -0.166($\pm$0.035) maxHCsatu +0.276($\pm$0.035)IC3 +0.068($\pm$0.031) E1p -0.147($\pm$0.029) E3s	<u>Internal</u> n = 70, R^2_{train} = 0.825, R^2_{adj} = 0.803, SEE= 0.205, F = 37.10, Q^2_{LOO} = 0.773, PRESS = 3.463, SDEP = 0.219, $^cR^2_p$ = 0.769, $R^2_{(res-train)}$ = 0.180 <u>External</u> R^2_{pred}= 0.835, R^2_m = 0.751, $	r^2_0-r^{'2}_0	$ = 0.081, k = 0.996, r^2_0- r^2_0/r^2 = 0.012, k′ = 1.000, r^2-r'^2/r^2 =0.107 AAE-	AE	0.177, 0.5 × AAE = 0.089, $R^2_{(res-test)}$ = 0.358, MAE = 0.201, 0.1×Range(train) = 0.232, 0.15×Range(train) = 0.348, 0.2×Range(train) = 0.464, 0.25×Range(train) = 0.579, MAE + (3 × σ) = 0.553
4	pED50 = 4.057($\pm$0.019) + 0.233($\pm$0.033)MATS7e + 0.472($\pm$0.035)MATS5s - 0.122($\pm$0.035)MATS7s -0.357($\pm$0.037)SdO + 0.211($\pm$0.029)IC3 + 0.341($\pm$0.037)RDF25m - 0.279($\pm$0.037)RDF70m -0.172($\pm$0.024)E3s	<u>Internal</u> n = 70, R2train = 0.893, R2adj = 0.879, SEE= 0.161, F = 65.37, Q2LOO = 0.854, PRESS = 2.213, SDEP = 0.175, cR2p = 0.854, R2(res-train) = 0.109 <u>External</u> R2pred= 0.830, R2m = 0.738,	r20-r'20	= 0.099, k = 1.006, r20- r20/r2 = 0.019, k′ = 0.989, r2-r'2/r2 =0.136 AAE-	AE	0.215, 0.5 × AAE = 0.107, R2(res-test) = 0.437, MAE = 0.231, 0.1×Range(train) = 0.232, 0.15×Range(train) = 0.348, 0.2×Range(train) = 0.464, 0.25×Range(train) = 0.580, MAE + (3 × σ) = 0.544
5	pED50 = 4.056($\pm$0.024) +0.215($\pm$0.039) AATSC7e - 0.152($\pm$0.041)AATSC7s + 0.432($\pm$0.042)MATS5s - 0.327 ($\pm$0.042)SdO - 0.167($\pm$0.035)maxHCsatu +0.274($\pm$0.035)IC3 - 0.147($\pm$0.029)E3e +0.071($\pm$0.031)E1p	<u>Internal</u> n = 70, R2train = 0.824, R2adj = 0.802, SEE= 0.206, F = 36.964, Q2LOO = 0.772, PRESS = 3.468, SDEP = 0.219, cR2p = 0.765, R2(res-train) = 0.181 <u>External</u> R2pred= 0.832, R2m = 0.757,	r20-r'20	= 0.082, k = 0.995, r20- r20/r2 = 0.011, k′ = 1.000, r2-r'2/r2 =0.109 AAE-	AE	0.177, 0.5 × AAE = 0.089, R2(res-test) = 0.356, MAE = 0.203, 0.1×Range(train) = 0.232, 0.15×Range(train) = 0.348, 0.2×Range(train) = 0.464, 0.25×Range(train) = 0.579, MAE + (3 × σ) = 0.466
		<center>PCR MODEL</center>
	pED50 = 4.056($\pm$0.021) + 0.056($\pm$0.009)PC1 + 0.121($\pm$0.014)PC2 + 0.194($\pm$0.026)PC3 + 0.280($\pm$0.03573)PC4 -0.368($\pm$0.036)PC5 + 0.080($\pm$0.038) PC6 -0.333($\pm$0.046)PC7 -0.409($\pm$0.067)PC8	<u>Internal</u> n = 70, R2train = 0.859, R2adj = 0.841, SEE= 0.184, F = 41.96, Q2LOO = 0.809, PRESS = 2.916, SDEP = 0.201, cR2p = 0.799, R2(res-train) = 0.141 <u>External</u> R2pred= 0.768, R2m = 0.683,	r20-r'20	= 0.169, k = 0.998, r20- r20/r2 = 0.019, k′ = 0.997, r2-r'2/r2 =0.237 AAE-	AE	0.234, 0.5 × AAE = 0.126, R2(res-test) = 0.453, MAE = 0.251, 0.1×Range(train) = 0.232, 0.15×Range(train) = 0.348, 0.2×Range(train) = 0.464, 0.25×Range(train) = 0.579, MAE + (3 × σ) = 0.668

Table 3- Observed and modeled anticonvulsant activity of studied compounds.

Cpd. No.			pED$_{50}$ Models				Cpd. No.			pED$_{50}$ Models			
	Obs.	1	2	3	4	5	PCR	Obs.	1	2	3	4	5	PCR
1[a]	5.009	5.025	5.028	4.915	4.705	4.919	5.020	46	4.077	4.299	4.297	3.940	4.100	3.938	4.072
2	5.003	4.667	4.651	4.740	4.794	4.731	5.012	47[a]	4.032	3.912	3.933	4.002	3.748	4.022	3.785
3	4.949	4.509	4.506	4.676	4.870	4.673	4.661	48	3.923	4.137	4.142	3.820	4.003	3.824	3.889
4	4.933	4.851	4.856	4.924	4.723	4.932	4.908	49	3.774	4.063	4.060	3.995	4.142	3.993	4.231
5	4.882	4.707	4.711	4.822	4.711	4.828	4.774	50	3.805	4.221	4.215	4.187	4.050	4.180	4.055
6	4.857	5.089	5.092	5.046	4.928	5.052	5.111	51	3.811	3.335	3.344	3.837	3.449	3.848	3.667
7[a]	4.722	4.573	4.575	4.552	4.354	4.554	4.355	52[a]	4.092	4.020	4.019	4.056	4.365	4.055	4.165
8	4.592	4.060 [c]	4.077 [c]	4.256 [c]	4.251 [c]	4.268 [c]	4.388	53	4.107	4.102	4.101	3.942	4.189	3.940	4.140

9	4.603	4.151	4.154	3.954	4.318	3.953 b	4.346	54	4.158	4.203	4.203	4.181	4.195	4.181	4.064
10	4.682	4.616	4.619	4.692	4.794	4.697	4.894	55 a	4.107	4.264	4.262	4.198	4.480	4.196	4.178
11 a	4.628	4.708	4.710	4.504	4.659	4.504	4.649	56	3.77	3.911	3.918	3.715	3.847	3.717	3.866
12	4.525	4.152	4.156	4.195	4.290	4.197	4.310	57	3.866	3.807	3.795	4.051	4.079	4.043	4.038
13	4.546	4.567	4.566	4.526	4.410	4.524	4.477	58	3.764	3.834	3.830	3.743	3.809	3.738	4.016
14	4.567	4.577	4.579	4.456	4.507	4.457	4.429	59	4.013	3.994	3.991	3.996	4.229	3.994	4.107
15	4.447	4.224	4.229	4.152	4.381	4.154	4.375	60	3.861	3.972	3.983	4.048	3.885	4.058	3.916
16	4.479	4.229	4.223	4.338	4.146	4.333	4.279	61	3.834	3.919	3.922	4.014	3.911	4.020	4.006
17 a	4.368	4.262	4.265	4.624	4.216	4.627	4.532	62	3.672	3.661 c	3.654 c	3.935	3.650	3.935	3.833
18	4.529	4.558	4.560	4.450	4.663	4.451	4.433	63	3.77	3.699	3.698	3.764	3.743	3.763	3.723
19	4.530	4.682	4.699	4.618	4.571	4.632	4.518	64	3.864	4.105	4.105	3.878	4.024	3.878	3.924
20	4.423	4.424	4.426	4.412	4.657	4.414	4.560	65	3.733	3.624	3.619	3.745	3.655	3.754	3.777
21 a	4.275	4.165	4.159	4.142	4.016	4.137	4.055	66	3.809	3.975	3.972	4.127	3.911	4.127	3.878
22	4.428	4.053	4.045	4.059	4.121	4.051	4.081	67 a	3.741	3.669	3.665	3.493	3.652	3.491	3.541
23	4.383	4.063	4.065	4.118	4.386	4.121	4.242	68	3.821	4.206	4.202	4.277	4.146	4.274	4.083
24	4.360	4.141	4.136	4.175	4.068	4.171	4.072	69	3.694	3.792	3.792	3.714	3.887	3.712	3.738
25	4.373	4.371	4.349	4.364	4.579	4.346	4.494	70	3.898	3.993	3.995	4.062	3.783	4.063	3.821
26	4.333	4.209	4.209	4.433	4.503	4.435	4.222	71	3.728	4.019	4.029	4.098	3.859	4.106	3.877
27	4.093	4.162	4.152	4.145	4.230	4.134	4.258	72	3.593	4.083	4.084	3.869	3.795	3.866	4.007
28	4.284	4.302	4.305	4.223	4.271	4.224	4.255	73	3.839	3.982	3.981	4.047	3.940	4.047	3.916
29	4.034	3.952	3.946	3.917	3.830	3.911	4.011	74	3.63	3.812	3.763	3.891	3.571	3.850	3.508
30	4.228	4.227	4.216	4.268	4.187	4.259	4.185	75	3.568	3.262	3.275	3.346	3.369	3.355	3.397
31	4.268	4.349	4.355	4.321	4.336	4.325	4.272	76	3.607	3.867	3.871	3.793	3.736	3.796	3.743
32	4.218	4.230	4.231	4.183	4.042	4.183	4.098	77	3.545	3.838	3.843	3.692	3.607	3.693	3.800
33 a	4.167	4.014	4.015	4.091	3.992	4.091	3.976	78	3.69	3.619	3.625	3.731	3.923	3.740	3.832
34 a	4.234	4.102	4.089	3.942	4.111	3.929	4.014	79	3.708	3.891	3.882	3.867	3.767	3.860	3.660
35	4.173	4.117	4.114	4.096	3.982	4.093	3.928	80	3.747	3.341	3.339	3.182	3.381	3.179	3.367
36	4.171	4.259	4.259	4.289	4.204	4.290	4.211	81 a	3.483	3.386	3.385	3.384	3.561	3.383	3.646
37	4.174	4.260	4.266	4.210	4.216	4.215	4.208	82 a	3.460	3.952	3.948	4.058	3.868	4.055	4.065
38 a	4.125	4.175	4.196	4.190	3.778	4.206	3.729	83	3.489	3.435	3.429	3.476	3.304	3.468	3.303

39	4.144	3.768	3.767	3.748	3.892	3.748	3.879	84	3.768	4.144	4.148	4.214	3.935	4.218	3.843
40	4.073	3.977	3.981	3.821	3.893	3.823	3.877	85 [a]	3.472	3.469	3.479	3.762	3.668	3.776	3.82
41 [a]	4.096	3.863	3.867	3.795	3.976	3.797	3.829	86	2.933	2.892	2.901	2.976	2.887	2.985	2.978
42	4.083	4.100	4.098	4.095	4.201	4.092	4.175	87 [a]	2.693	2.968	2.974	2.912	2.975	2.918	3.163
43	4.085	4.320	4.330	4.399	4.174	4.405	4.087	88	2.684	2.668	2.674	2.714	2.762	2.719	2.772
44	4.225	3.948	3.951	3.922	4.170	3.925	4.060	89 [a]	2.610	3.079	3.082	3.112	2.905	3.114	3.102
45	4.113	4.187	4.198	4.154	4.142	4.165	4.117	90	2.728	3.083	3.094	2.958	2.967	2.970	3.017

[a] compound in the test set; [b] response outlier in model 5; [c] compound with leverages higher than the threshold leverage h* = 0.375

3.2. Applicability domain

On analyzing the applicability domain (AD) with the Williams plots (Figure 3) for the models based on the entire dataset, compound 9 was identified as y-outlier for model 5. The limit of normal values for the response variable was set as ±3σ (standard deviation) units. Compounds 8 was observed to have leverages (h) values greater than the threshold leverage (h* = 0.375) for all the models and compound 62 was found to have leverage values greater than threshold leverage for model 1 and 2.

In summary, the produced models matched high quality parameters for both the training and test set data. They had good fitting power and capability for assessing external data. Furthermore, almost 99% of the studied compounds were within the applicability domain of the proposed model meaning they were evaluated correctly except for compounds 8, 9 and 62 that showed inconsistency.
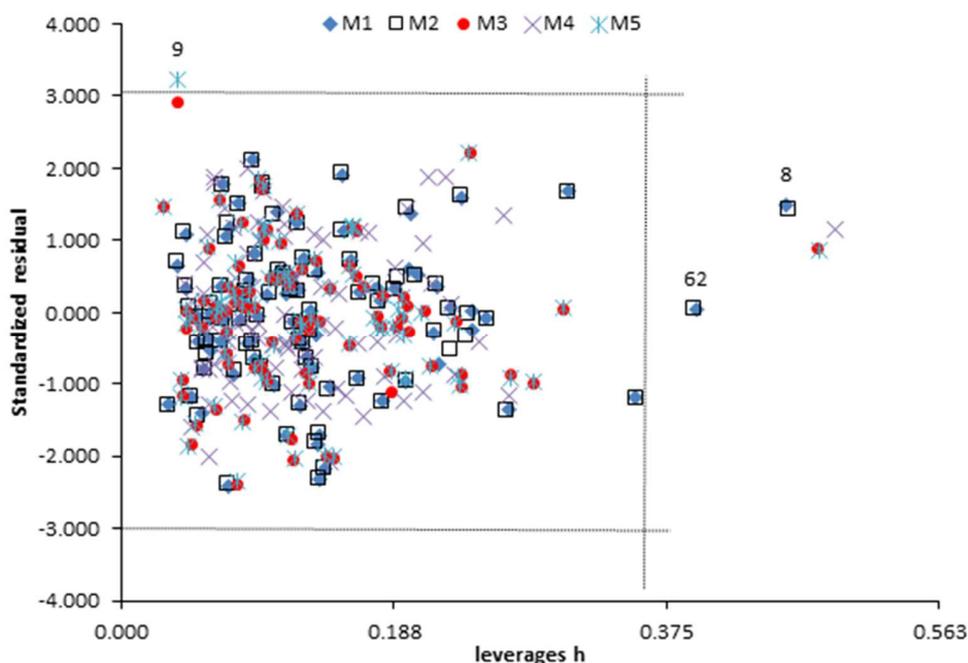
Figure 3-Williams plots for the models. In the charts, horizontal lines refer to the residual limit (± 3 standard deviation) and vertical line refers threshold leverage h* = 0.375(M1-M5 represent model 1-5)

3.3. Descriptor interpretation

Table 4 presents a brief definition, average regression coefficient and incidence of the 16 descriptors shared by the reported model in the study. AATSC7m, AATSC7e and AATSC7s are Moreau-Broto average/centered 2D autocorrelation descriptors. They are based on spatial-dependent autocorrelation function which measures the strength of the relationship between observations (atomic or molecular properties) and space separating them (lag) (TODESHINI AND COSONNI, 2009).These descriptors are obtained by taking the molecule atoms as the set of discrete points in space and an atomic property as the function evaluated at those points. When these descriptors are calculated on molecular graph, the lag coincides with the topological distance between any pair of vertices. AATSC7m, AATSC7e and AATSC7s were defined on the molecular graphs

using atomic masses (m), Sanderson electronegativities (e) and inductive effect respectively of pairs of atoms 7 bond apart as the weighting scheme. These observations suggested that atomic masses and electronic distribution of atoms that made up the molecule had significant effect on the anticonvulsant activity of the dataset. In addition, the signs of the regression coefficients for each descriptor indicated the direction of influence of descriptors in the models such that, positive regression coefficient associated to a descriptor will augment the activity profile of a compound while the negative coefficient will diminish the activity of a compound. Therefore, increase in the values of AATSC7m and AATSC7e augments the anticonvulsant activities of the compounds, while, that of AATSC7s diminishes it (Table 4). Therefore, chain elongation and introduction of more electronegative atom into the molecular system will be favorably disposed to increase in anticonvulsant activity.

MATS7m, MATS7s, MATS7e and MATS75s are also spatial-dependent 2D autocorrelation descriptors with the incorporation Moran coefficient (index) (TODESHINI AND COSONNI, 2009) in the measurement of the strength of the relationship between observations and space separating them. These Moran autocorrelation descriptors contained in the model reported in this study were defined on the molecular graphs using atomic masses (m), Sanderson electronegativity (e) and inductive effect of pairs of atoms 7 and 5 bond apart as the weighting scheme. These observations supported the claim that atomic masses and electronic distribution had significant effect on the anticonvulsant activity of the molecules. MATS7m and MAT7s were negatively correlated to the anticonvulsant activity values (Table 4).

SdO and maxHCsatu are 2D-electrotopological state (E-state indices) atom type descriptor. In general E-state indices

encodes the intrinsic electronic state of each atom as perturbed by the electronic influenced of all other atoms in the molecule within the context of the topological character of the molecule (HALL AND KIER, 1995). SdO is usually calculated for compounds containing a carbonyl group, i.e. ketones, carboxylic acids, esters, amides and urea, nitro and nitroso compounds, sulfones, and sulfoxides (HUUSKONEN *et al.,* 2000).

It is negatively correlated to the anticonvulsant activity of the dataset (Table 2). maxHCsatu favor the addition of –CH3 to unsaturated C atom e.g. in benzene ring. However, it's negatively correlated to the activity of the dataset molecules. SC-5 is a simple connectivity chi index which is a weighted count of sub-graph whose number of edges (bonds) is five (KIER AND HALL, 1976). Increase in molecule linearity increases it value and increasing the value of this descriptor augments the activity of the data set molecules (Table 2).

RDF25m and RDF70m are 3D radial distribution function at 2.5 and 7.0 inter-atomic distance weighted by atomic masses (TODESHINI AND COSONNI, 2009).

The presence of this descriptor in the model suggested the occurrence of a linear relationship between anticonvulsant activity and the 3D molecular distribution of atomic masses in the molecules calculated at radius of 2.0 Å and 7.0 Å from the geometrical centers of each molecule. It was observed that RDF25m is positively correlated to activity (Table 3) and reverse is the case for RDF70m. IC3 is a topological information index of a graph based on neighbor degrees and edge multiplicity. It is a measure of graph complexity (KIER AND HALL, 1976). It is positively correlated with anticonvulsant activities of dataset molecules.

Table 4- Models descriptors along with their physical meaning, average regression coefficient and incident

No	Descriptor	Descriptor class	Physical meaning	Av. reg. coeff. (Incidence)
1	AATSC7m	2D-autocorrelation	Average centered Broto-Moreau autocorrelation - lag 7 / weighted by mass	0.393(2)
2	AATSC7e	2D-autocorrelation	Average centered Broto-Moreau autocorrelation - lag 7 / weighted by Sanderson electronegativities	0.214(2)
3	AATSC7s	2D-autocorrelation	Average centered Broto-Moreau autocorrelation - lag 7 / weighted by I-state	-0.152(2)
4	MATS7m	2D-autocorrelation	Moran autocorrelation - lag 7 / weighted by mass	-0.323(2)
5	MATS5s	2D-autocorrelation	Moran autocorrelation - lag 5 / weighted by I-state	0.437(5)
6	MATS7e	2D-autocorrelation	Moran autocorrelation - lag 7 / weighted by Sanderson electronegativities	0.233(1)
7	MATS7s	2D-autocorrelation	Moran autocorrelation - lag 7 / weighted by I-state	-0.122(1)
8	SdO	2D- Electrotopological State Atom Type Descriptor	Sum of atom-type E-State: =O	-0.354(5)
9	SC-5	2D-Chi cluster descriptor	Simple cluster of order 5	0.133(2)
10	maxHCsatu	2D- Electrotopological State Atom Type Descriptor	Maximum atom-type H E-State: H on C sp3 bonded to unsaturated C	-0.167(2)
11	RDF25m	3D- RDF Descriptor	Radial distribution function - 025 / weighted by relative mass	0.341(1)
12	RDF70m	3D- RDF Descriptor	Radial distribution function - 070 / weighted by relative mass	-0.279(1)
13	IC3	Information Content Descriptor	Information content index (neighborhood symmetry of 3-order)	0.251(5)
14	E3e	3D-PaDEL WHIM Descriptor	3rd component accessibility directional WHIM index / weighted by relative Sanderson electro-negativities	-0.159(2)
15	E1p	3D-PaDEL WHIM Descriptor	1st component accessibility directional WHIM index / weighted by relative polarizabilities	0.066(4)
16	E3s	3D-PaDEL WHIM Descriptor	3rd component accessibility directional WHIM index / weighted by relative I-state	-0.163(3)

Table 5-Designed molecule and their leverage values and predicted anticonvulsant activity (pED$_{50}$)

No.	X1	X2	X3	X4		
1	CH$_3$	H	O	NH$_2$	0.123	4.453
2	CH$_3$	H	O	NHCH$_3$	0.148	4.724
3	CH$_3$	H	O	N(CH$_3$)$_2$	0.194	4.751
4	CH$_3$	H	O	NH-O-H	0.121	4.652
5	CH$_3$	H	O	NH-O-CH$_2$CH$_3$	0.17	4.702
6	CH$_3$	H	O	NHCOCH$_3$	0.294	3.625[b]
7	CH$_3$	H	O	NHCSCH$_3$	0.153	4.659
8	CH$_3$	H	O	NHCOCH$_2$CH$_3$	0.357	3.843[b]
9	CH$_3$	H	O	NHCSCH$_2$CH$_3$	0.156	4.778
10	CH$_3$	H	O	NHNH$_2$	0.129	4.573
11	CH$_3$	H	O	NHNHCH$_3$	0.161	4.661
12	CH$_3$	H	O	NHN(CH$_3$)$_2$	0.126	4.537
13	CH$_3$	H	O	Cl	0.188	4.839
14	CH$_3$	H	O	Br	0.328	4.803
15	CH$_3$	H	O	F	0.232	3.787[b]
16	CH$_3$	H	O	I	1.777[a]	5.493
17	CH$_3$	H	O	OH	0.13	4.339
18	CH$_3$	H	O	COOH	0.274	3.780[b]
19	CH$_3$	H	O	CH$_3$	0.108	4.633
20	CH$_3$	N(CH$_3$)$_2$	O	H	0.087	4.261
21	CH$_3$	CH2CH3	O	H	0.063	4.329
22	CH$_3$	NH2	O	H	0.059	4.294
23	CH$_3$	NHCH3	O	H	0.077	4.325
24	CH$_3$	NHOCH3	O	H	0.128	4.319
25	CH$_3$	NHOCH2CH3	O	H	0.198	4.321
26	CH$_3$	NHCOCH$_3$	O	H	0.224	3.293[b]
27	CH$_3$	NHCOCH$_2$CH$_3$	O	H	0.246	3.441[b]
28	CH$_3$	NHCSCH$_2$CH$_3$	O	H	0.1	4.42
29	CH$_3$	NHCSCH$_3$	O	H	0.086	4.437
30	CH$_3$	NHNH2	O	H	0.084	4.183
31	CH$_3$	NHNHCH3	O	H	0.111	4.241
32	CH$_3$	NHN(CH$_3$)$_2$	O	H	0.073	4.212
33	CH$_3$	Cl	O	H	0.152	4.591
34	CH$_3$	F	O	H	0.132	3.852[b]
35	CH$_3$	Br	O	H	0.219	4.24
36	CH$_3$	I	O	H	0.684[a]	4.727
37	CH$_3$	OH	O	H	0.078	4.054
38	CH$_3$	COOH	O	H	0.426[a]	3.214[b]
39	CH$_3$	H	S	H	0.047	4.205
40	CH$_3$	H	S	NH$_2$	0.089	4.437
41	CH$_3$	H	S	NHCH$_3$	0.125	4.673
42	CH$_3$	H	S	N(CH$_3$)$_2$	0.165	4.736
43	CH$_3$	H	S	NHOCH$_3$	0.097	4.591
44	CH$_3$	H	S	NHOCH$_2$CH$_3$	0.13	4.668
45	CH$_3$	H	S	NHSCH$_2$CH$_3$	0.139	4.688
46	CH$_3$	H	S	NHSCH$_3$	0.113	4.574
47	CH$_3$	H	S	NHCOCH$_2$CH$_3$	0.364	3.933[b]
48	CH$_3$	H	S	NHCSCH$_3$	0.109	4.585
49	CH$_3$	H	S	NHCSCH$_2$CH$_3$	0.14	4.692
50	CH$_3$	H	S	NHNH$_2$	0.109	4.65
51	CH$_3$	H	S	NHNHCH$_3$	0.089	4.354

52	CH$_3$	H	S	NHN(CH$_3$)2	0.098	4.527
53	CH$_3$	H	S	Cl	0.126	4.669
54	CH$_3$	H	S	F	0.161	3.994[b]
55	CH$_3$	H	S	Br	0.491[a]	4.936
56	CH$_3$	H	S	I	1.531[a]	5.195
57	CH$_3$	H	S	OH	0.128	4.061b
58	CH$_3$	H	S	COOH	0.423[a]	3.498[b]
59	CH$_3$	CH3	S	H	0.059	4.039[b]
60	CH$_3$	NH$_2$	S	H	0.067	4.005[b]
61	CH$_3$	NHCH$_3$	S	H	0.072	4.111[b]
62	CH$_3$	N(CH$_3$)$_2$	S	H	0.064	4.027[b]
63	CH$_3$	NHOCH3	S	H	0.121	4.161[b]
64	CH$_3$	NHCOCH3	S	H	0.25	3.355[b]
65	CH$_3$	NHCSCH$_3$	S	H	0.091	4.165[b]
66	CH$_3$	NHNH$_2$	S	H	0.106	4.125[b]
67	CH$_3$	NHN(CH$_3$)$_2$	S	H	0.094	4.108[b]
68	NHNH$_2$	H	O	H	0.119	4.354
69	NHNHCH$_3$	H	O	H	0.07	4.188
70	NHN(CH$_3$)$_2$	H	O	H	0.101	3.674[b]
71	OCH$_3$	H	O	H	0.197	4.455
72	NHOCH$_3$	H	O	H	0.108	3.791[b]
73	NHOH	H	O	H	0.114	4.141[b]
74	OH	H	O	CH$_3$	0.143	4.894

75	H	O	N	C	0.108	4.571
76	H	O	C	S	0.078	4.438
77	H	O	C	O	0.122	4.531
78	H	S	C	S	0.132	4.145b
79	NH$_2$	O	C	O	0.119	4.385
80	NHCH$_3$	O	C	O	0.097	4.536
81	NH$_2$	S	C	S	0.145	3.928[b]
82	NHCH$_3$	S	C	S	0.137	4.021[b]
83	N(CH$_3$)$_2$	O	C	O	0.209	4.421
84	N(CH$_3$)$_2$	S	C	S	0.067	4.032[b]
85	NHNH$_2$	O	C	O	0.144	4.321
86	NHNH$_2$	S	C	S	0.193	3.945[b]
87	NHNHCH$_3$	O	C	O	0.101	4.534
88	NHN(CH$_3$)$_2$	O	C	O	0.09	4.432
89	NHNHCH$_3$	S	C	S	0.087	4.343
90	NHN(CH$_3$)$_2$	S	C	S	0.085	4.254
91	OH	O	C	O	0.078	4.314
92	COOH	O	C	O	0.495	3.472[b]

93	NHNH$_2$	H	H	0.125	3.776[b]
94	CH$_3$	H	NH$_2$	0.156	4.096[b]
95	CH$_3$	H	CH$_3$	0.063	4.262
96	CH$_3$	H	OH	0.128	3.792[b]
97	CH$_3$	H	COOH	0.267	3.338[b]
98	CH$_3$	H	N(CH$_3$)$_2$	0.122	4.401
99	CH$_3$	H	NHOH	0.065	4.100[b]

100	CH3	H	CHCH2CH2	0.104	4.409
101	CH3	OH	CH3	0.21	4.016[b]
102	OH	H	CH3	0.09	4.406

103	H	0.059	3.862[b]
104	NH2	0.091	3.937[b]
105	NHCH3	0.094	4.145[b]
106	N(CH3)2	0.074	4.031[b]
107	NHOCH3	0.093	3.983[b]
108	NHCOCH3	0.358	3.101[b]
109	NHCOCH2CH3	0.374	3.160[b]
110	NHCSCH3	0.095	4.074[b]
111	NHCSCH2CH3	0.083	4.272
112	NHNH2	0.123	3.987[b]
113	NHNHCH3	0.115	4.080[b]
114	NHN(CH3)2	0.086	3.997[b]
115	Cl	0.066	4.150[b]
116	F	0.154	3.609[b]
117	OH	0.132	3.748[b]
118	COOH	0.348	3.118[b]

[a] leverage value greater than threshold; [b] activity values less than that of template (4.174)

E1p, E3e and E3s are 3D directional WHIM (Weighted Holistic Invariant Molecular) descriptors. These are indices developed to describe molecular structure in terms of size, shape, atom distribution and symmetry with respect to some invariant reference frame. They are univariate statistical indices calculated from the score of individual principal component (1, 2, 3) obtained by performing a principal component analysis (PCA) ont a centered molecular coordinate by different weighting scheme (TODESHINI AND COSSONI, 2009).

E1p was obtained from the first component and it is weighted by relative polarizabilities of atom that make up the molecule. E3e was obtained from the third component. It is weighted by Sanderson electronegativity values for all atoms that made up a molecule. While E3s was also obtained from the third component, however, weighted by relative I-state (inductive or I-effect). This suggested that the electronic distribution of atoms that made up the molecule had significant effect on the anticonvulsant activity of the molecules.

The two WHIM descriptors are obtained from the third component were negatively correlated to the anticonvulsant activity, while E1p was positively correlated. Generally, for directional WHIM descriptors, increase in chain length of an organic molecule increases both first and third component of directional whip descriptors. Cyclisation of the chain reduces both. Branching reduces first but increases the third a little. Increase in molecular mass increase both. Aromaticiation reduces it both drastically. Addition of halogen increases first but reduces the third component. Adding alkyl group to benzene increase both a little. And increase in unsaturation reduces the first such single > double > triple (TODESHINI AND COSSONI, 2009).

3.4 Designed molecule

Modification were made around the template (2-acetamido-N-benzyl-2-(5-methylfuran-2-yl)acetamide) structure to obtain about 118 new molecules which are presented in Table 5. Firstly, efforts were made to increase the linearity of the molecule by the addition of chain substituent at position 1, 9 and 17. However, the substituents added were not pure alkyl group because, addition of alkyl to benzene ring was reduces the value of first component directional WHIM descriptors and favors maxHCsatu (TODESHINI AND COSSONNI, 2009).

Additional ring system was avoided as much as possible because of their relationship with the directional WHIM descriptors and other descriptors that favors increase linearity (SC-5, IC3, AATSC and MATS). Increment in degree of unsaturation of the molecular system and number of carbonyl group was avoided in order to negate its incremental effect on SdO descriptor. Furthermore, additions of O, S, N, and halogen favors increase in molecular mass, linearity and polarizability the data set molecules. O, S, N addition also allows some degree compromise on relative inductive effect (I-state). Generally, relative inductive effects have been experimentally measured with reference to hydrogen, in decreasing order of -I effect or increasing order of +I effect, as follows: $-NH_3^+ > -NO_2 > -SO_2R > -CN > -SO_3H > -CHO > -CO > -COOH > -COCl > -CONH_2 > -F > -Cl > -Br > -I > -OR > -OH > -NH_2 > -C_6H_5 > -CH=CH_2 > -H$ (STOCK, 1972).

The average predicted activity values by the models and their leverages value are included in Table 5. It was observed that only 6 out of 118 designed molecules had leverage value greater than $h^* = 0.375$ and they were designated with superscript 'a'. Therefore, prediction made for these molecules by the models were considered unreliable [ROY et al, 2015; ERIKSON *et al.*, 2003; DEARDEN et al., 2007). About 50 of the designed

compounds had predicted activity value less than that of the template compounds (pED$_{50}$ of 4.174).

The predicted activities of these groups of molecules are designated with superscript 'b'. Substitution of the furan ring of the template molecule with piperidine or pyrole ring (molecules 93-118) does not lead to compounds with improved activity. This might be attributed to reduction in linearity of the molecules as pointed out by the models.

Addition of carbonyl compounds to the molecular system of the template, gave molecules with reduced activity value (Molecules 6, 8, 18, 26-27, 47, 58, 64, 92, 97, 108-109,118). Increment in the value of descriptor SdO and conjugation may contribute to this observation as pointed out by the models.

3.5 Docking simulation

X-Ray crystal structure of chains A and B of PDB: 1OHV is composed of 922 amino acids and 14608 atoms. It contained 15 pockets and binding site was defined around the pockets ranked 1, which is represented by blue mesh in Figure 4. The pocket had amino acids Ile72, Phe189, Arg192and Lys329 that were implicated in the inhibitory action of vigabatrin and acetate ion on GABA_AT (STORICI *et al.*, 2004). Also, its Merck's Drug-Like Density (DLID) score was greater than 0.5. This suggested the pocket could be a good site of action for any drug. Grid map size of 0.5 and dimensions 25.02Å X 24.82Å X 34.25Å were used.

44 out of the 118 designed molecules showed better predicted anticonvulsant activity compare to the template, they were docked with 1OHV receptor. The best result for each molecule after 10 conformations run is presented in Table 6. The molecules were ranked according to their predicted affinity for the target considered.

Hydrogen bond energy, hydrophobic interaction, electrostatic interaction, and the various amino acid involved in these interactions are also included in the table. The designed molecule had negative binding score. This indicated good binding affinity with the receptor. Their binding scores were better than that of the template, 4-aminohex-5-enoic acid (vigabatrin), 3, 3 diphenylpyrrolidine-2, 5-dione (phenytoin) and comparable to that of 5H-dibenzo [b, f] azepine-5-carboxamide (cabamezapine). Vigabatrin, phenytoin and carbamezapine are known inhibitors of GABA_AT. The designed molecules therefore have potential as inhibitor of GABA_AT and invariably as antiepileptic chemical agent.

Explaining the interaction between the ligands and the receptor molecules are a complex endeavor because various interactions are involved including: hydrophobic interaction, hydrogen bonding and electrostatic interaction. The correlation between this interactions and dock score (binding affinity) gave hydrogen bonding correlation value of 0.703, hydrophobic interaction correlation value of 0.248 and electrostatic interaction correlation value of -0.301. This was in agreement with the report that the specificity of the binding between ligand and receptor molecule is controlled by hydrogen bonding (YEH *et al.,* 2002).
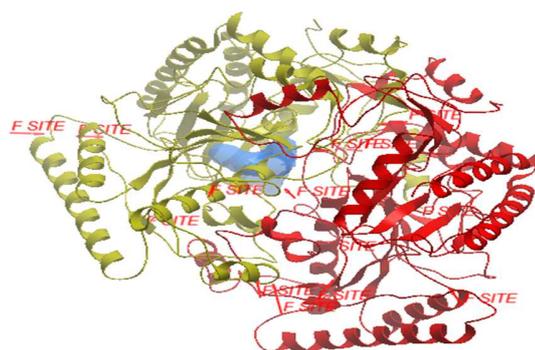

Figure 4-3D structure of chain A (red) and B (magenta) of PDB: 1OHV receptor, pocket ranked 1 (blue mesh) and the F-sites (red) are the remaining pocket positions.

Figure 5 presented the interaction between designed molecule 30 ((S)-2-acetamido-N-benzyl-2-(5-hydrazinylfuran-2-yl)acetamide) and the receptor pocket. From the diagram, it was observed that the molecule interacted with the receptor chain A: Tyr69, Arg192, Glu270, Gly438 via conventional hydrogen bond and Lys203 and Gly438 via carbon hydrogen bond. It also reacted with chain B Try348 via pi-donor hydrogen bond and hydrophobic pi-pi T-shape interaction. The number and types of interactions in this molecule were similar to, but more than that of 4-aminohex-5-enoic acid (i.e. vigabatrin a known GABA_AT inhibitors) (Figure 6).
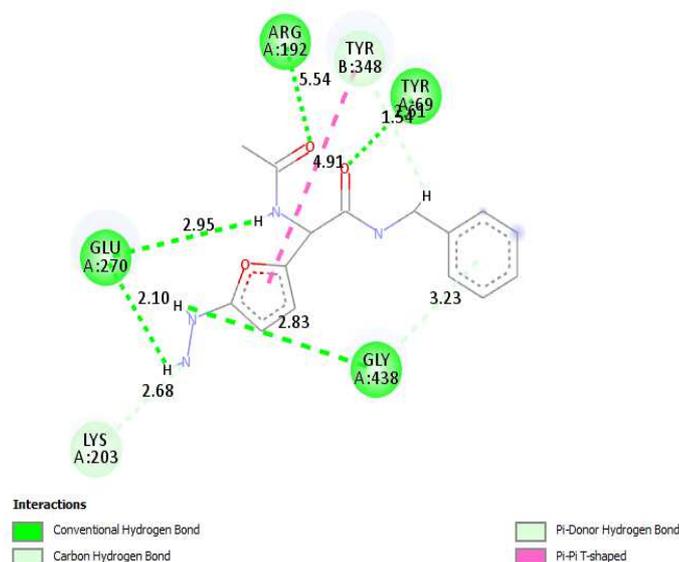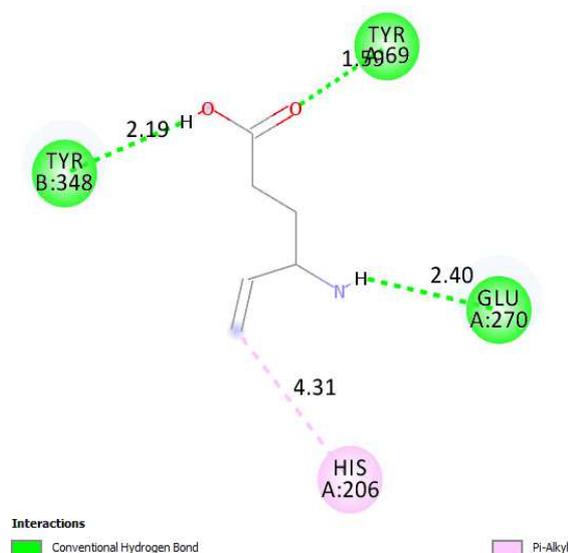

Figure 5-Intermolecular interaction between 2-acetamido-N-benzyl-2-(5-hydrazinylfuran-2-yl)acetamide (molecule No 30) and the pocket amino acid residue of GABA_AT

This could inform it's observed higher binding affinity compared to vigabatrin. In summary, the observed high binding affinity for the designed molecules when compare with vigabatrin, phenytoin and carbamazepine could be attributed to the number and types of interaction between them and the receptor. The docking result suggested that the designed molecules might be potential anticonvulsant molecules for the treatment of epilepsy episodes.

Figure 6-Intermolecular interaction between 4-aminohex-5-enoic acid (vigabatrin) and the pocket amino acid residue of GABA_AT

Table 6-Molecular docking result for the designed molecules showing various interactions with GABA_AT

Rank	No.	Score	H_b	H_{ph}	E_{in}	H_b	H_{ph}	E_{in}
			Energy(kcal/mol)			Intermolecular interaction and the amino acid involved		
1	30	-34.3	-9.91	-5.20	4.80	Tyr69, Arg192, Lys203, Glu270, Tyr348, Gly438	Tyr348	
2	31	-31.59	-8.19	-6.20	4.96	Tyr69, Arg192, Glu270, Tyr348, Gly438, Gly440	Tyr348	Glu270
3	7	-29.07	-5.83	-5.89	6.50	Tyr69,Arg192,His206,Glu270,Tyr348, Asn423	Tyr348	Arg192
4	9	-28.93	-5.73	-5.92	19.6	Tyr69,Arg192,His206,Glu270,Tyr348, Arg430,Asn423	Tyr348	Arg192
5	14	-28.9	-5.45	-6.20	5.40	Tyr69,Arg192,His206,Glu270,Tyr348, Gly438	Tyr348	Arg192
6	10	-28.75	-8.17	-5.22	5.96	Tyr69,Arg192,His206,Glu270,Tyr348, Gly440	Tyr348	Arg192
7	24	-28.51	-5.53	-5.93	6.54	Tyr69,Arg192,lys203,Glu270,Tyr348, Gly438	Tyr348,His206,	Arg192
8	13	-27.96	-5.41	-6.11	5.12	Tyr69,Arg192,Glu270,His206,Tyr348	Tyr348	Arg192
9	45	-27.79	-5.77	-6.27	7.75	Tyr69,Arg192,Glu270,Tyr348, Asn423	Tyr348, Ile426	His206, Arg192
10	5	-27.44	-5.91	-6.02	6.99	Tyr69,Arg192,His206,Glu270,Tyr348, Asn423	Tyr348, Ile426, Gly438	Arg192
11	33	-27.42	-5.33	-6.15	4.56	Tyr69,Arg192,Glu270,Tyr348, Gly438	Tyr348	
12	32	-27.37	-5.61	-6.86	6.86	Tyr69,Arg192,Glu270,Tyr348, Gly438	Tyr348, His206	Arg192
13	22	-27.28	-7.34	-5.29	5.13	Tyr69,Arg192,Glu270,Tyr348	Tyr348	Arg192
14	53	-27.21	-5.21	-6.44	5.18	Tyr69,Arg192,Glu270Tyr348	Tyr348	Arg192, His206
15	21	-27.11	-5.63	-6.24	5.42	Tyr69,Arg192,Glu270,Tyr348	Tyr348	
16	23	-26.97	-6.57	-6.00	4.80	Tyr69,Arg192,Glu270,Tyr348, Gly438, Gly440		Glu270
17	35	-26.79	-5.40	-6.18	4.53	Tyr69,Arg192,Glu270,Tyr348, Gly438	Tyr348	Glu270
18	2	-26.45	-5.94	-6.10	5.40	Tyr69,Arg192,Glu270,His206,Asn423	Tyr348	Arg192

19	12	-26.39	-5.93	-6.30	5.54	Tyr69,Arg192,Glu270,Tyr348,Asn423	Tyr348	His206, Arg192
20	50	-26.35	-7.61	-5.47	6.00	Tyr69,Arg192,Glu270,Gly449	Tyr388	His206, Arg192
21	1	-26.02	-5.93	-5.50	5.12	Tyr69,Arg192,Glu270,His206,Asn423	Tyr348	Arg192
22	4	-25.99	-5.84	-5.79	6.28	Tyr69,Arg192,Glu270,His206,Asn423	Tyr348, Gly438	Arg192
23	29	-25.91	-5.56	-5.60	14.2	Tyr69,Arg192,Glu270,Tyr348, Gly438	His206,Tyr348	Arg192
24	11	-25.89	-5.93	-5.926	5.81	Tyr69,Arg192,His206,Glu270,Tyr348, Asn423	Tyr348	Arg192
25	46	-25.64	-5.83	-6.448	6.02	Tyr69,Arg192,Glu270,Tyr348	Tyr348	
26	48	-25.58	-5.38	-6.12	5.03	Tyr69,Arg192,Glu270,Tyr348,Arg445 ,Arg422	Tyr348	
27	41	-25.5	-7.74	-6.13	7.92	Tyr69,Arg192,Glu270,Tyr348,Leu436	Tyr69,Tyr348	His206
28	49	-25.38	-5.77	-6.277	14.87	Tyr69,Arg192,Glu270,Tyr348,Arg422 ,Asn423	Tyr348	
29	40	-25.26	-5.68	-5.799	5.21	Tyr69,Arg192,Glu270,Tyr348,Asn423	Arg192,Tyr348	His206
30	20	-24.54	-5.08	-6.448	6.10	Tyr69,Arg192,Glu270,Tyr348,Gly438 , Gly440	Hie44	Glu270, Arg445
31	17	-24.42	-5.80	-5.499	5.06	Tyr69,Arg192,His206,Glu270,Tyr348, Asn423	Tyr348	Arg192
32	44	-24.02	-5.73	-6.599	5.63	Tyr69,Arg192,Glu270,Tyr348,Gly438 , Arg445	Tyr348,Gly438	
33	3	-24.01	-5.28	-6.474	5.73	Tyr69,Arg192,Glu270,His206,Asn423	Tyr348	Arg192
34	52	-23.81	-5.64	-6.763	5.04	Try69,Arg192,Glu270,Tyr348	Tyr348	
35	43	-23.45	-5.47	-6.17	7.20	Tyr69,Arg192,Glu270,Tyr348, Glu419,Arg422,Arg445	Tyr348	
36	51	-22.08	-4.10	-6.30	6.04	Tyr69,Arg192,Glu270,Tyr348, Asn423	Arg192,His206, Tyr348	
37	25	-22.06	-5.44	-6.17	8.98	Tyr69,Arg192,Glu270,Tyr348,Cys439	Hie44,Tyr348	Glu270
38	74	-21.73	-5.29	-5.15	4.83	Tyr69,Arg192,Glu270,Tyr348	Tyr348	Arg69
39	42	-21.48	-3.41	-6.63	6.48	Tyr69,Arg192,Glu270,Tyr348,Asn423	Tyr348	His206
40	39	-20.13	-3.40	-6.10	5.64	Tyr69,Arg192,Glu270,Tyr348	Tyr348	His206
41	28	-18.16	-3.36	-5.35	11.58	Glu270,Arg422,Asn423,Gly438	Phe351,Tyr348,Ile426	Arg430,Arg445
42	68	-17.05	-5.28	-4.73	5.78	Arg192,His206,Glu270	Ile72,Phe189,	Lys203,Glu270
43	71	-17.15	-3.24	-5.83	5.07	Tyr69,Arg192,Glu270,His206,Tyr348	Phe189,Tyr348	Arg192
Template		-26.08	-5.40	-6.20	4.75	Tyr69,Arg192,Glu270,Tyr348,Gly438	Try348	
Carbamazepine		-24.76	-1.61	-4.38	3.02	His206	Tyr348,His206	Arg422
Phenytoin		-16.69	-2.99	-4.17	3.91	Tyr69,Gly438	Tyr348, Gly438	Arg422,Arg445
Vigabatrin		-10.97	-3.99	-2.57		Tyr69, TYR 348, Glu270	His206	

H_b is hydrogen bond, H_{ph} is hydrophobic interaction and E_{in} is electrostatic interaction

.

4. CONCLUSION

Thoroughly validated QSAR models were developed for the anticonvulsant activity of some N-benzyl actetamide derivatives. The QSAR models obtained obeyed the "rule of thumb" and there is no redundancy among the chemometric molecular descriptors included in the model. The models were predictive (R^2 ranged from 0.823 to 0.893, Q^2 from 0.772 to 0.854, F from 36.53 to 37.10, $R^2_{pred(test)}$ from 0.768 to 0.893) and statistically significant at 95% confidence level. Models proposed that increase in molecular mass, molecular linearity and polarizability of the molecules that is within the AD of the models would increase the activity of the molecules. 2-acetamido-N-benzyl-2-(5-methylfuran-2-yl)acetamides, member of the training set within the AD of the proposed models was therefore chosen as scaffold to designed about 118 hypothetical molecules. Molecular descriptors of the designed molecules were used to extrapolate their inclusion in the AD of the proposed models. The models were

employed to predict hypothetical anticonvulsant activities for the designed molecules. Designed molecules found within the AD of the models with hypothetical anticonvulsant activity value greater than the observed activity value for the scaffold were then docked to GABA_AT (PDB: 1OHV). The binding affinities of these molecules for GABA_AT were better than that of scaffold, vigabatrin, and phenytoin and are comparable to that of carbamazepine. Since vigabatrin, phenytoin and carbamazepine are known GABA_AT inhibitors, the design molecules have potential as GABA_AT inhibitors. Therefore, the results of the study may be useful in future in vivo experiments for the ability of the designed compound to inhibit GABA_AT.

R E F E R E N C E S

ACCERLYS SOFTWARE Inc. Material Studio 7.0 modeling and simulation software. 2007 available from: http//:accerlys.com/product/.

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.

DAVID, T.T.; PETER, M.J.M.; SIMON, M.C.; NICHOLAS, L.S.Gamma-aminobutyric acid (GABA) transport across human intestinal epithelial (Caco-2) cell monolayers. Journal of Pharmacolology, v.129, n.3, p.457-464, 2000.

DEARDEN, J.C.; CRONIN, M.T.D.; KAISER, K.L.E. How not to develop a quantitative structure–activity or structure–property relationship. SAR and QSAR in Environmental Research, v.20,p.3-4, 2009,

ERIKSSON, L.; JAWORSKA, J.; WORTH, A.P.; CRONIN, M.T.; MCDOWELL, R.M. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environment and Health Perspective, v.111, p.1361-1375, 2003.

GOLBRAIKH, A.; TROPSHA, A. Beware of q2! Journal of Molecular Graphic and Modelling, v.20, p.269 – 276, 2002.

HALL, L.H.; KIER, L.B. Electrotopological state indices for atom types: A novel combination of electronic, topological, and valence state information. Journal of Chemical Informatics and Computational Science, v. 35, p. 1039 – 1045, 1995.

HUUSKONEN, J.J.; LIVINGSTONE, D.J.; TETKO, I.V. Neural network modeling for estimation of partition coefficient based on atom-type electrotopological state indices. Journal of Chemical Informatics and Computational Science, v.40, p. 947-955, 2000.

IBEZIM, E.C.; DUCHOWICZ, P.R.; IBEZIM, N.E.; MULLEN, L.M.A.; ONYISHI, I.V.; BROWN, S.A.; CASTRO, E.A. Computer aided linear modeling employing QSAR for drug discovery. Scientific Research and Essay, v.4, n.13, p.1559-1564, 2009.

JAWORSKA, J.; NIKOLOVA-JELIAZKOVA, N.; ALDENBERG, T. QSAR applicability domain estimation by projection of the training set descriptor space: a review. Alternative to Laboratory Animals, v. 33, p.445, 2005.

KENNARD, R.W.; STONE, L.A. Computer aided design of experiments. Technometrics, v.11, p. 137-148, 1969.

KIER, L.B.; HALL, L.H. Molecular connectivity in chemistry and drug research. New York: Academic Press. 1976, ch.4, p. 46.

KING, A.M. Synthesis and pharmacological evaluation of primary amino acid derivatives (PAADs): novel neurological agents for the treatment of epilepsy and neuropathic pain. Thesis submitted to Division of Medicinal Chemistry and Natural Products University of North Carolina at Chapel Hill, 2010. Retrieved 5th May 2016 from https://cdr.lib.unc.edu/.

LESK, A.M. Introduction to bioinformatics. Oxford University Press; 2002, ch.3. p.55

MATTSON, R.H. Efficacy and adverse effects of established and new antiepileptic drugs. Epilepsia, v.36, p.13-26, 1995.

MOLSOFT INTERNAL COORDINATE MECHANICS PROGRAM (ICM-pro 3.8.3) for bioinformatics. 2012, available from www.molsoft.com.

MOREAU, G.; TURPIN, C. Use of similarity analysis to reduce large molecular libraries to smaller sets of representative molecules. Analusis, v.24, p. m17-m21, 1996.

PAOLA, S; DE BIASE,, D.; FRANCESCO, B.; STEFANO, B.; ANDREA, M.; CAROLINE, P.; RICHARD, B.S; TILMAN, S. Structures of γ-aminobutyric acid (GABA) aminotransferase, a pyridoxal 5′-phosphate, and [2Fe-2S] Cluster-containing Enzyme, complexed with γ-ethynyl-GABA and with the antiepilepsy drug vigabatrin. Journal of Biological Chemistry, v.279, p.363-373, 2004.

PEDRO, J.B.; JOHN, B.O.M. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. Bioinformatics, v.26, n.9. p.1169-1175. 2010.

PHYLLIS, J.B. A Biochemist's experience with GABA. Journal of Orthomolecular Medicine, v. 26. p.1, 2011.

ROGERS, D. HOPFINGER, A.J. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. Journal of Chemical Informatics 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.; 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.; KAR, S.; AMBURE, P. On a simple approach for determining applicability domain of QSAR models. Chemometrics and Intelligent Laboratory Systems, v. 145, p. 22-29, 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 Chemical Information and Modeling, v. 52, p. 396-408, 2012.

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 and algorithms in modern quantum chemistry program package.

Physics Chemistry Chemistry Physics, v. 8, p. 3172, 2006.

SINGH P. Quantitative structure-activity relationship study of substituted-[1, 2, 4] oxadiazoles as S1P1 agonists. Journal Current Chemical and Pharmaceutical Science, v. 3.n.1, p.64-67, 2013.

STOCK, L.M. The origin of the inductive effect. Journal of Chemistry Education, v.46, n. 6, p.400, 1972.

STORICI, P.; QIU, J.; SCHIRMER, T.; SILVERMAN, R.B. Mechanistic crystallography. Mechanism of inactivation of gamma-aminobutyric acid aminotransferase by (1R,3S,4S)-3-amino-4-fluorocyclopentane-1-carboxylic acid as elucidated by crystallography. Biochemistry, v. 43, n.44, p. 14057-14063, 2004.

TODESCHINI, R.; CONSONNI, V. Molecular Descriptors for Chemoinformatics, Vol I and II WILEY-VCH Verlag GmbH and Co, Weinheim. 2009: 6-745.

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.; GRAMATICA, P.; GOMBAR, V.K. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR and Combinatorial Science, v. 22, p. 69-77, 2003.

USMAN, A.; ADAMU, U.; SANI, U. Quantitative structure-activity relationship and molecular docking studies of a series of quinazolinonyl analogues as inhibitors of gamma amino butyric acid aminotransferase. Journal of Advance Research, v.8, p. 33-34, 2017.

WHITING, P.J. GABA-A receptor subtypes in the brain: a paradigm for CNS drug discovery. Reviews in Drug Discovery Today, v.8, n. 10, p. 445-449, 2003.

YAP C.W. PaDEL-Descriptor: open source software to calculate molecular descriptors and fingerprints. Journal of Computational Chemistry, v.32, n.7, p.1466-1474, 2011.

YEH, S.; TSAI, M.Y.; XU, Q.; MU, X.M.; LARDY, H.; HUANG, K.E. Generation and characterization of androgen receptor knockout (ARKO) Mice: an in vivo model for the study of androgen Functions in Selective Tissues. Proceedings of the National Academy of Science of the United State, v.99. n.21, p. 13498 – 13503, 2002.