JOTCS A
Journal of the Turkish Chemical Society
Section: A
Chemistry
TURKISH CHEMICAL SOCIETY

A novel QSAR model for designing, evaluating, and predicting the anti-MES activity of new 1H-pyrazole-5-carboxylic acid derivatives

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

[1]Chemistry Advance Research Center, Sheda Science and Technology Complex, FCT, Nigeria
[2]Chemistry Department, Ahmadu Bello University, Zaria, Nigeria

Abstract: A quantitative structure–activity relationship (QSAR) study was performed to develop a model that relates the structures of 62 compounds, which have activity against maximal electroshock-induced seizure (MES), with their anti-MES activity. Molecular structures of the compounds were geometrically optimized and energetically minimized using a combination of modified Merck force field (MMFF) molecular mechanics, Austin model 1 (AM1) semi-empirical quantum mechanical and density functional theory (DFT) quantum mechanical method using the Becke's three parameter exchange functional (B3) hybrid with Lee, Yang and Parr correlation functional (LYP) and basis set of the double zeta split valence plus polarization quality 6-31G** i.e. B3LYP/6-31G**. Theoretically derived descriptors were obtained from the optimized structures, a genetic function approximation (GFA) algorithm was also applied to select the optimal descriptors and multiple linear regression (MLR) was used to establish a relationship between the anti-MES activity of the compounds and the optimal molecular descriptors. A six-parametric equation containing dipole moment (μ), energy of the lowest unoccupied molecular orbital (ϵLUMO), polar surface area (PSA), accessible surface area derived from wave function (WAA), sum of the square root of square of the charge on all atom of the molecule (QA) and sum of the square root of square of the charge on all fluorine atoms in the molecule was obtained as the QSAR model in the present study with good statistical qualities (R^2=0.937, R^2_{adj}=0.928, F=104.11, R^2_{pred}=0.929 and Q^2=0.913). The QSAR model was used to study estimate the anti-MES activities of 1H-pyrazole-5-carboxylic acid derivatives not yet synthesized. 10 out of the 101 screened compounds had improved anti-MES activity when compared to the template (*i.e.* ethyl 4-(4-chlorophenyl)-3-morpholino-1H-pyrrole-2-carboxylate, which is compound number 61 in the dataset) used to design the 101 derivatives. These 10 compounds were docked with voltage-gated sodium channel (PDB code: 2KaV) and their binding affinity were comparable to that of phenytoin (a standard drug known to possess anti-MES activity).

Keywords: MES, QSAR, GFA, Molecular docking, Kennard-Stone algorithm.

Submitted: April 08, 2017**. Accepted:** July 20, 2017**.**

Cite this: Oluwaseye A, Uzairu A, Shallangwa G, Abechi S. A novel QSAR model for designing, evaluating, and predicting the anti-MES activity of new 1H-pyrazole-5-carboxylic acid derivatives. JOTCSA. 2017 Jul;4(3):739–74.

DOI: 10.18596/jotcsa.304584

Corresponding author. E-mail. adedirinoluwaseye@yahoo.com; senguade@gmail.com

INTRODUCTION

Epilepsy is not a specific disease but a syndrome characterized by excessive discharges of large number of neurons altering the normal electrochemical balance in the brain (1). Its episodes are often characterized by seizures which occur when a neuron or groups of neurons in the brain become hyper-excitable or irritable, due to a number of reasons, *e.g.* hypoxia, ischemia, hypoglycemia, or electrolyte abnormalities, causing these nerve cells to discharge action potentials irregularly without adequate suppression and attenuation. Action potential is the electro-physiologic voltage change manifested in the axon of neurons due to a transient variation in the sodium and potassium permeability of the axon. Depending on the location (focus) of the aberrant discharges in the brain, their resultant effect could manifest as a motor symptom (*e.g.* tonic-clonic contractions) or sensory manifestations (*e.g.* paresthesia and hallucinations). If these foci spread to various areas of the brain, it leads to chaotic, uninhibited discharge of electrical activity and the resultant motor and/or sensory activity manifested by patient is clinically described as a seizure (2). Control of seizure can be accomplished by suppressing the action potential via manipulation of sodium and potassium ion permeability, rendering the axon refractory to the action potential, or blocking transmission of impulses at the synapse by blocking the neurotransmitter from binding to its receptor site, or preventing its release and/or synthesis (2). Researchers over the years have concentrated their effort on developing therapies to control and prevent these seizures, mainly with the use of medications which had led to the development many antiepileptic drugs (AEDs) (3). However, with optimal usage of the available AEDs about 25% of patients continue to have epileptic episode and usually, AED treatment require several years and side effects may appear (4-5). Therefore, the need for the development of new, more effective and safer antiepileptic drugs cannot be overemphasis.

There are two main strategies often employed in the design and development of new AED including (a) the search of new compounds that cause a modification of a certain stage of the cellular mechanism of epilepsy (mechanism-based design) and (b) the structural modification of pre-existing compounds (structure-based design) (6). *In silico* studies had contributed its quota to the design and development of AEDs using these approaches (7, 8). *In silico* studies are alternatives to the real world of synthesis and screening of compounds in the laboratory involving virtual world of data analysis, hypothesis, and design that reside inside a computer. Its application ensures that the expensive commitment to actual synthesis and bioassay is made after exploring the initial concepts with computational models and screens (9). Precise classification of the AEDs according to their mechanisms of action is not presently possible because some of them do not act on a specific binding site, and most of them interact with more than one receptor (10, 11). However, some of the cellular mechanisms, which may occur during drug action in the epileptic patient, include, among others,

voltage-dependent blockade of Na^+ channels, modulation of γ-aminobutyric acid (GABA) synthesis, or degradation, inhibition of cellular GABA uptake, modulation of GABA (A) receptors, modulation of various excitatory amino acid receptors, and modulation of adenosine metabolism (12-14). Moreover, different experimental animal models have been used to study and evaluate the anticonvulsant activity of drug molecules including which maximal electroshock seizure (MES) test and the subcutaneous pentylenetetrazole (PTZ) test are the most popular. MES test electrically induced seizure in animals producing hind limb tonic extensor while drug molecules were used to abolish this effect and any drug that is effective against MES had been reported to act as inhibitors to neuronal voltage-gated sodium channel (VGSC). Also, MES test had been reported to represents a good grand mal seizure model and identified compounds that prevent seizure spread (14).

The main objective of this present work is to find rationality in the design of new 1H-pyrazole-5-carboxylic acid derivatives which are active against MES induced seizure using quantitative structure activity relationship study (QSAR) and molecular docking strategies. QSAR is a computational approach that relates quantitative measure of chemical structure of compounds with their activities employing series of computer-based processes in order to predict a relationship, model or equation that will help to propose the activity of known compounds with unknown activities or unknown compounds and their activities (15). This approach has been implicated in the development process of many anticonvulsant molecules (16-19). However, there is no report on the rational design of novel 1H-pyrazole-5-carboxylic acid derivatives using QSAR strategy. Molecular docking on the other hand is a technique that is used to explore the binding mode of two interacting molecules depending upon their topographic features or energy consideration, in order to fit them into conformation that lead to favorable interactions. Docking is often used in revealing key elements and mechanism of protein-ligand interaction and consequently in rational drug design as starting point for finding new lead compounds or drug candidates (20, 21). These strategies therefore permit the rational design of novel 1H-pyrazole-5-carboxylic acid derivatives, quantitative estimates of their potencies using the information on the molecular descriptors contributions to anti-MES activity and elucidate the interaction between the designed molecules and the anticonvulsant molecular target.

MATERIAL AND METHODS

Data set

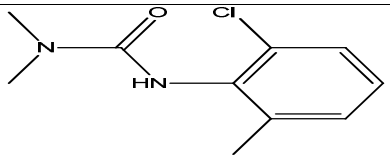
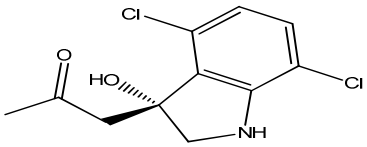
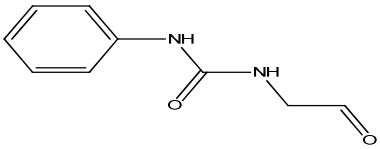
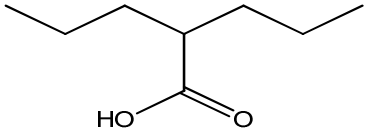
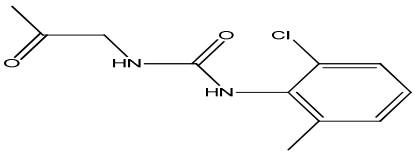
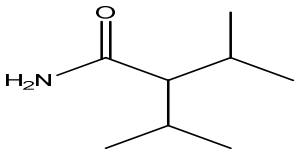
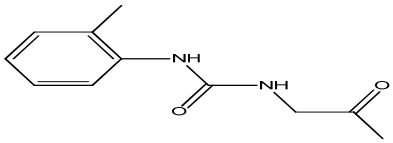
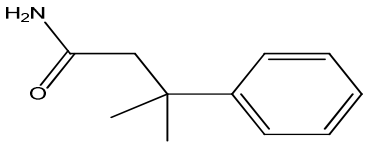
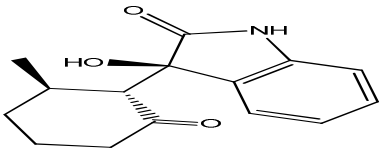
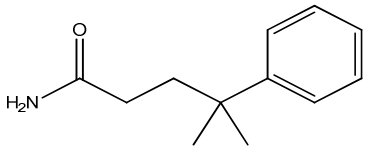
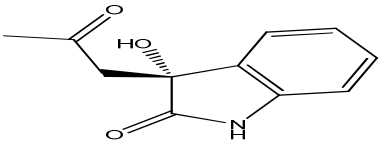
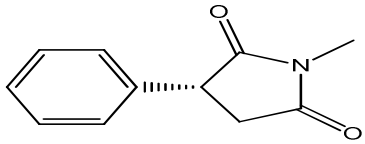
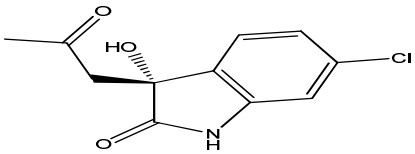
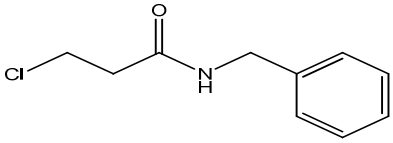
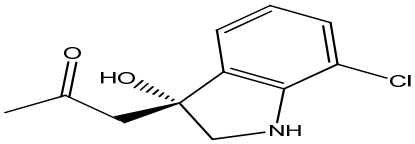
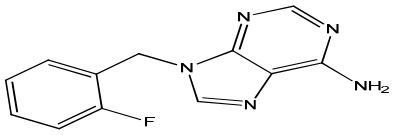
The data set was made up of derivatives of 1H-pyrazolo [3, 4-d]pyrimidine, 1H-pyrazole-5-carboxylic and hydrazine carboxamide obtained from literatures (22-34) with their anticonvulsant activity against MES-induced seizure expressed as ED_{50} (mg/kg) (a measure of the dose quantity that is effective in 50% of the tested animals). The reported anti-MES activities of the selected compounds were recalculated to molar unit for easy comparison between molecules and

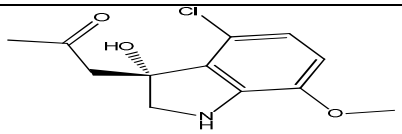
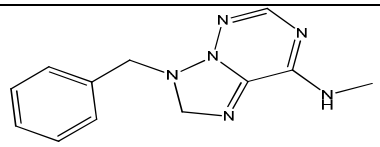
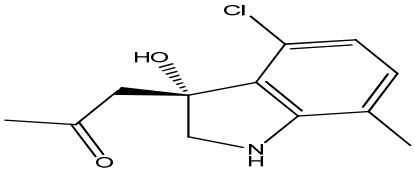
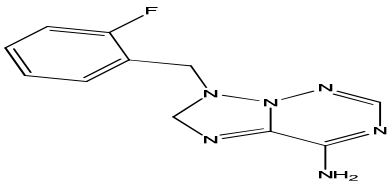
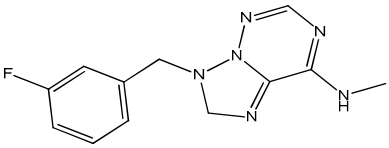
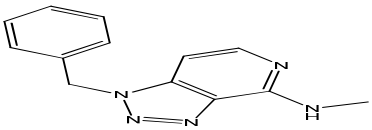
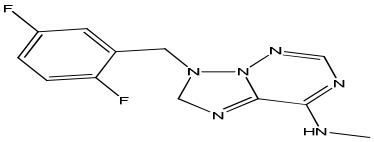
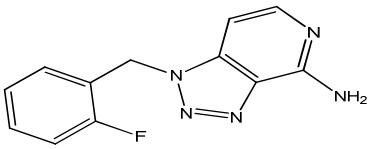
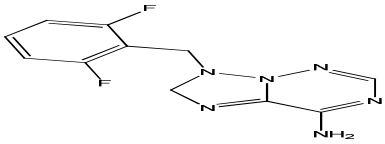
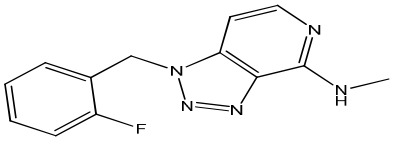
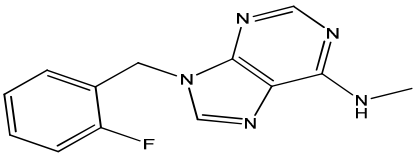
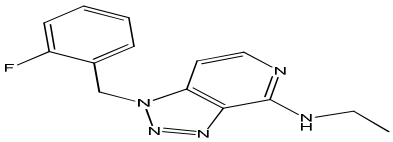
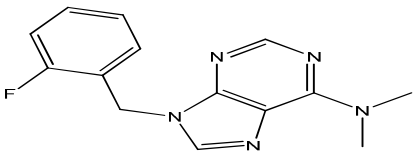
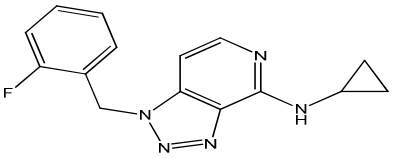
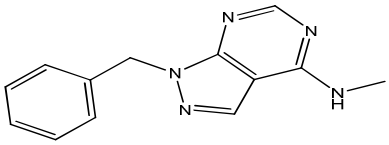
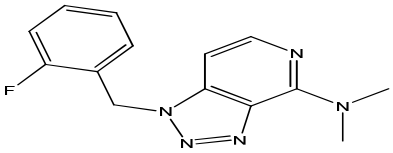
subsequently, they were converted to logarithmic unit (*i.e.* -log ED_{50} designated pED_{50}) to increase the linearity in the activity values as presented in Table 1 with their corresponding molecular structures.

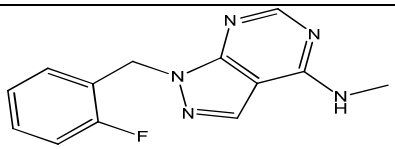
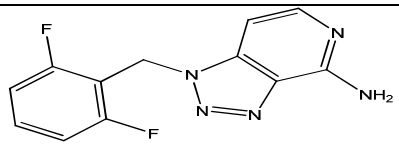
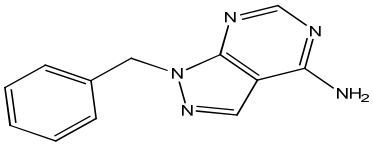
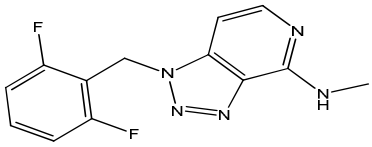
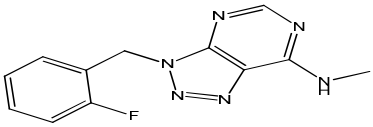
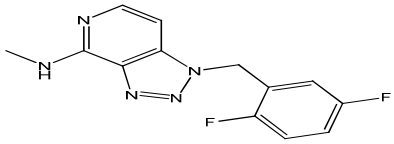
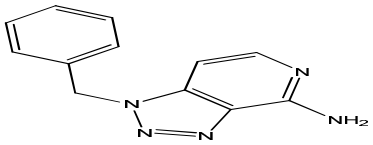
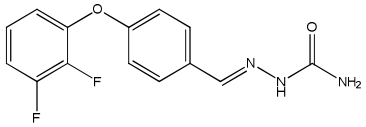
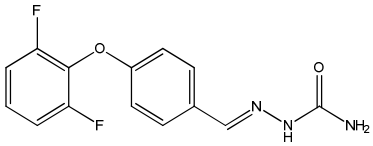
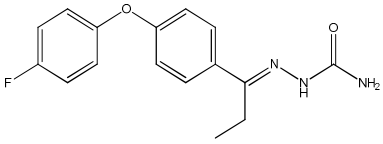
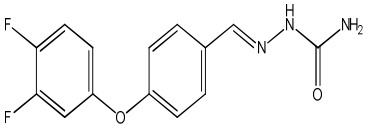
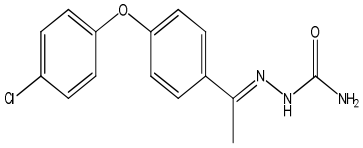
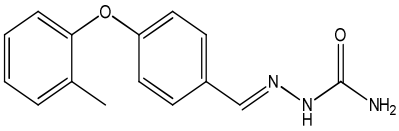
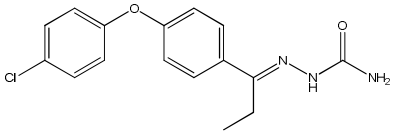
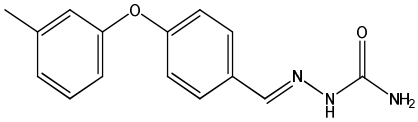
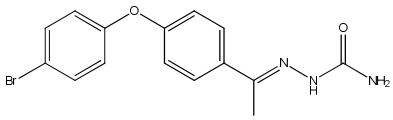
Calculation of the molecular descriptors

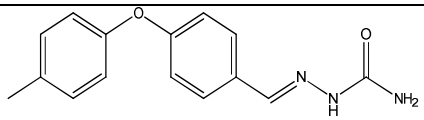
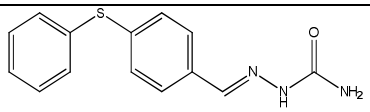
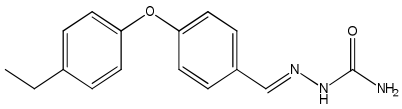
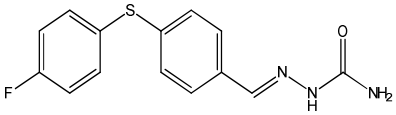
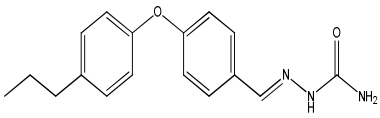
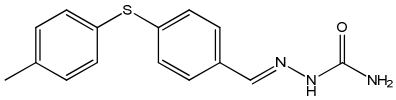
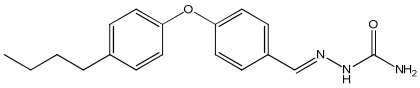
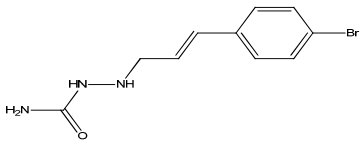
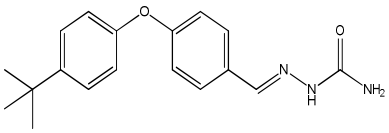
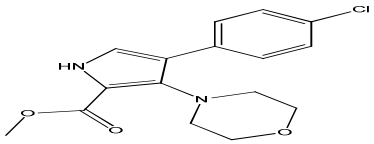
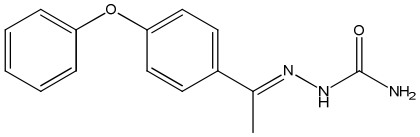
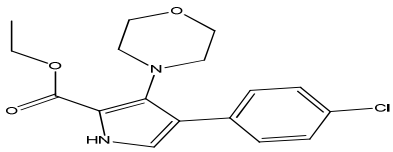
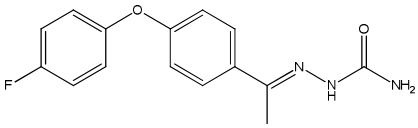
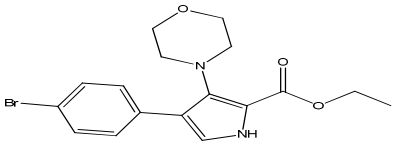
2D molecular structure of each molecule and subsequent conversion to 3D were drawn using sketch and view tools in Spartan 14 package (35). Using the same package, pre-optimization and energy minimization of the molecules were carried out with modified Merck force field (MMFF) molecular mechanics (MM) followed by Austin model 1 (AM1) semi-empirical quantum mechanical (QM) method until the root mean square (RMS) gradient value was smaller than 10^{-6} atomic unit (au). Thereafter, density function calculation were performed on the molecules using the Becke's three parameter exchange functional (B3) hybrid with Lee, Yang and Parr correlation functional (LYP) and basis set of the double zeta split valence plus polarization quality (6-31G**) was used *i.e.* B3LYP/6-31G** (35). These MM and QM calculations were done to obtain reliable energetic and accurate data on electronic properties of the molecules. In particular, DFT allows the identification of the most stable conformer of the molecules associated with the absolute minima in the potential energy hypersurface, which represents the most probable structures of the molecules when far enough from the receptor (3). Electronic, thermodynamic, and QSAR properties were extracted from the display properties module of Spartan 14 package. Also, Mulliken atomic charges for all the atoms were extracted from the display output module of the software. A total of 50 molecular descriptors were calculated for the data set as listed in Table 2.

Table 1: Molecular structure of the data set and their anti-MES activity.

No.	Molecular structure	pED_{50}	No.	Molecular structure	pED_{50}
1[a]		3.447	11		3.667
2		3.242	12		3.396
3		3.633	13		3.199
4		3.253	14		3.440
5		3.181	15[a]		3.327
6		3.011	16		3.228
7		2.952	17		3.034
8		2.976	18		4.307

No.	Molecular structure	pED_{50}	No.	Molecular structure	pED_{50}
9		3.468	19		4.180
10		3.283	20		4.488
21		4.114	31		4.379
22		4.444	32		4.325
23		4.577	33		4.747
24		4.331	34		4.433
25		4.354	35		4.653
26		4.203	36[a]		4.613
27		4.507	37[a]		4.849

No.	Molecular structure	pED50	No.	Molecular structure	pED50
					
28[a]		4.207	38[a]		4.716
29		4.509	39		4.807
30		4.052	40		4.675
41		4.721	52[a]		4.851
42		5.089	53		5.018
43		4.678	54[a]		5.041
44		4.943	55[a]		4.900
45		4.895	56[a]		4.802

No.	Molecular structure	pED50	No.	Molecular structure	pED50
					
46		4.640	57		4.764
47[a]		5.053	58[a]		4.893
48		4.986	59		4.335
49		5.268	60[a]		4.449
50		4.442	61		4.627
51		4.930	62		4.554

[a] represent test set compound

Table 2 Density function theory calculated molecular descriptors

Symbol	Definition
Q^2	Sum of square of charges of all atom in the molecule
QA	Sum of absolute value of charges of all atom in the molecule
Q_{max}	Maximum atomic positive charge in the molecule
Q_{min}	Minimum atomic negative charge in the molecule
PP	Polarity parameter (difference between Q_{max} and Q_{min})

Symbol	Definition
T^E	Topological electronic index is the sum of the absolute difference between charges on all atomic pairs in a given molecule divided by the square of their relative distances)
QC	Sum of square of charges on all C atoms in the molecule
QH	Sum of square of charges on all H atoms in the molecule
QCl	Sum of square of charges on all Cl atoms in the molecule
QF	Sum of square of charges on all F atoms in the molecule
QBr	Sum of square of charges on all Br atoms in the molecule
QS	Sum of square of charges on all S atoms in the molecule
QN	Sum of square of charges on all N atoms in the molecule
QO	Sum of square of charges on all O atoms in the molecule
α	Polarizability
μ	Molecular dipole moment
V_{min}	Minimum molecular electrostatic potential
V_{max}	Maximum molecular electrostatic potential
I_{min}	Minimum value of local ionization potential mapped onto an electron density surface
PAA(75)	Polar accessible area corresponding to absolute value of electrostatic potential greater than 75 $KJmol^{-1}$
PAA(100)	Polar accessible area corresponding to absolute value of electrostatic potential greater than 100 $KJmol^{-1}$
PAA(125)	Polar accessible area corresponding to absolute value of electrostatic potential greater than 125 $KJmol^{-1}$
ϵHOMO	Energy of the highest occupied molecular orbital
ϵLUMO	Energy of the lowest unoccupied molecular orbital
Δϵ	Energy gap (difference between ϵHOMO and ϵLUMO)
HOMO(-1)	Energy of the energy level before the highest occupied molecular orbital
LUMO(+1)	Energy of the next energy level to the lowest unoccupied molecular orbital
α/Δϵ	Energy-weighted polarizability term
E_{tot}	Total energy of the molecule
IE	Ionization energy
EA	Electron affinity
η	Hardness (difference between IE and EA)
S	Softness (i.e. inverse of hardness)
χ	Chemical electro-negativity (average of the sum of IE and EA)
ω	Electrophilicity index ($\chi^2/2\ \eta$)
S^0	Standard entropy of the molecular system
H^0	Standard enthalpy of the molecular system
G^0	Standard Gibbs free energy of the molecular system
S^0/N	Standard entropy divided by the number of atoms in a molecular system
C_v	Heat capacity of the molecular system at constant volume
ZPE	Zero point energy of the molecular system

Symbol	Definition
Area	Molecular surface area
Volume	Molecular volume
PSA	Polar surface area
Ovality	Ovality
WAA	Accessible molecular surface area obtained from wave function
Log P	Lipophilicity parameter
HBA	Number of hydrogen bond acceptor
HBD	Number of hydrogen bond donor
W	Molecular weight

Normalization of descriptors value

Arranged in an n×m matrix, where n represent the number of molecules and m the number of descriptors in the data, the descriptors' value were pretreated and normalized using Equation 1 in order to give each variable the same opportunity at the onset to influence the model (36).

$$X^{1} = \frac{(X_i - X_{min})}{(X_{max} - X_{min})} \quad \text{(Eq. 1)}$$

Where X_i is the first of each descriptor for a given molecule, X_{max} and X_{min} are the maximum and minimum value for each column of descriptors X.

Splitting into training and test set

A very important step in QSAR analysis is the division of data into training set used to build the model and test set used to test evaluate the predictive ability of the build model. In the present study Nano BRIDGES software (37) was used to divide the data employing Kennard and Stone's algorithm (38). This algorithm has been applied with great success in many recent QSAR studies and has been highlighted as one of the best ways to build training and test sets (39-42). In this algorithm, two compounds with the largest Euclidean distance apart were initially selected for the training set. The remaining compounds for the training set were selected by maximizing the minimum distance between these two compounds and the rest of the compounds in the dataset. This process continues until the desired number of compounds needed for the training set have been selected then, the remaining compounds in the dataset would be used as the test set (38).

Selection of optimal descriptor

Selection of the combination of descriptors having good correlation with better explains the variability in anti-MES activity of studied compounds is the next crucial step towards building a predictive QSAR model. In the present study, this was done with Material studio 7.0 software using genetic function approximation (GFA) method. The process begins by arranging the training set data in an nm matrix format in the study table; where n represent the number of molecules and m

the number of descriptors in the data, the first column being the activity values and subsequent columns are the descriptors. Thereafter, the analysis condition was set as follows: The equation length range from 5 to 12 terms, population equal 10000, and maximum generation equal 500, number of top equation returned equal 5, mutation probability equal 0.1 and scaled LOF smoothness parameter equal 0.5. On completion of the heuristic search, five different combinations of descriptors were reported which were used for subsequent analysis.

GFA is a combination of Holland's genetic algorithm and Friedman's multivariate adaptive regression splines algorithm (43). It uses a genetic algorithm to perform a search over the entire descriptor space for possible combination of descriptor that will produced a good QSAR model and uses certain fitness function score obtained via multivariate adaptive regression splines algorithm to estimate the fitness of each model. This method has the following advantage: (a) generation of multiple combination of descriptor that can be utilized to it builds multiple models rather than a single model, (b) automatically selects and determines the exact number of descriptors needed to build a full-size model, (c) it incorporates the lack of fit (LOF) error measure to resists over-fitting, (d) allows user control over the smoothness of fit and length of equation, (e) it can be use on a very large pool of descriptors, (f) allow the use of different function including splines, gaussians, or higher-order polynomials to construct model.

Mapping to descriptor with activity

Relating the molecular descriptor to the activities value leads to the construction of QSAR models. The process started by performing correlation analyses on each of the five groups of descriptor combination returned by the GFA analysis and utilizing their corresponding correlation matrix to evaluate the variance inflation factors (VIF) value for each descriptor in a group. VIF are the diagonal elements of the inverse of correlation matrix (44) and they are used to reveal the extent of multi co-linearity between descriptors. Any group with any descriptor having VIF value greater 10 was discarded and the remaining groups were used to build QSAR equations (models) using multiple linear regression (MLR) method.

Model quality and validation

The quality of the QSAR models produced by the GFA-MLR method was judged with the statistical metrics produced by the method: determination coefficient (R^2), adjusted determination coefficient (R^2_{adj}), standard error of estimation (SEE), variance ratio (F) and t-statistics or p-value for each descriptor. Determination coefficient R^2 is defined by Equation 2:

$$R^2 = 1 - \frac{\sum(Y_{obs} - Y_{calc})^2}{\sum(Y_{obs} - \bar{Y}_{obs})^2} \quad \text{(Eq. 2)}$$

where Y_{obs}, Y_{calc} and are the observed, calculated and average of the observed anticonvulsant activity respectively (*i.e.* dependent variable). R^2 is a measure of the explanatory power of the model used to explain the variation in the activity value of molecules used in building the model. An ideal model has an R^2 value of 1 and as the value deviate from 1, the fitting quality of the model deteriorates. Adjusted determination coefficient R^2_{adj} is a modified form of determination coefficient which accounts for the effect of new explanatory variables in the model, since it incorporates degree of freedom of the model (45). To reflect the explained variance in a better way R^2_{adj} is the candidate of choice because the inclusion of some other independent variables (either relevant or irrelevant) in multiple regression models mostly generating a non-decreasing R^2 value. R^2_{adj} is defined by Equation 3:

$$R^2_{adj} = \frac{(N-1)\times R^2 - p}{N-1-p} \quad (3)$$

where N is the number of data point (number of molecules in the data), R^2 is the determination coefficient and p is the number of explanatory variable (descriptors) in the model. N-1-p is the degree of freedom. Variance ratio F is the ratio of regression mean square to the deviation mean square. It is used to judge the overall significance of the regression coefficients and its value should be significance at $p < 0.05$ for all the regression coefficients (*i.e.* the p-value for all the descriptor in a model should be less than 0.05 or all t-statistics should be greater than 2). For overall significance of the regression coefficients, the F value should be high. F is defined by Equation 4:

$$F = \frac{\frac{\sum(Y_{calc}-\bar{Y}_{obs})^2}{p}}{\frac{\sum(Y_{obs}-Y_{calc})^2}{N-p-1}} \quad \text{(Eq. 4)}$$

Standard error of estimation (SEE): is equivalent to the models standard deviation, it's a measure of model quality and a model is said to be a better model if it has low SEE value. SEE is defined by equation 5:

$$SEE = \sqrt{\frac{(Y_{obs}-Y_{calc})^2}{N-p-1}} \quad (5)$$

Model validation

To further evaluated the robustness, accuracy and reliability of the models constructed, various validation techniques including the leave-one-out (LOO) cross-validation procedure, validation through an external test set, Y-randomization, Golbraikh and Tropsha criteria for a predictive model (46) and modified square correlation coefficient (R^2_m) introduced by Roy *et al* (47) were used. In the leave-one-out cross-validation techniques, the training set is primarily modified by eliminating one compound from the set. QSAR model is then rebuilt based on the remaining molecules of the training set using the descriptor combination originally selected, and the activity of the deleted compound is computed based on the resulting QSAR equation. This cycle is repeated until all the molecules of the training set have been deleted once, and the predicted activity data obtained for

all the training set compounds. The result is then used for the calculation of various internal validation parameters including predicted error sum of square (PRESS), standard deviation of error of prediction (SPRESS) and cross-validated determination coefficient designated R^2_{CV} or Q^2 which are defined by equations 6 to 8:

$$PRESS = \sum\left(Y_{obs(train)} - Y_{pred(train)}\right)^2 \quad (6)$$

$$S_{PRESS} = \sqrt{\frac{PRESS}{n}} \quad (7)$$

$$Q^2 = 1 - \frac{\sum\left(Y_{obs(train)} - Y_{pred(train)}\right)^2}{\sum\left(Y_{obs(train)} - \bar{Y}_{train}\right)^2} \equiv 1 - \frac{PRESS}{\sum\left(Y_{obs(train)} - \bar{Y}_{train}\right)^2} \quad (8)$$

In equations 6 to 8, $Yobs_{(train)}$ 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, $\bar{Y}_{train}$ 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.

For the Y-randomization test, process randomization was employed in the present study by permuting the observed activity data with respect to the models descriptor matrix which was kept constant. For each permutation, a new model was developed at the same confidence level as the original model. Then, the determination coefficients for the randomized models R^2_r were estimated (48). Also, the deviation in the value of the mean determination coefficient of the randomized models ($\bar{R}^2_r$) from the determination coefficient of the original (non-randomized) models (R^2) is reflected in the value of a parameter designated ${}^cR^2_p$ which was computed from equation 9:

$${}^cR^2_p = R \times \sqrt{R^2 - \bar{R}^2_r} \quad (9)$$

The threshold value of ${}^cR^2_p$ is 0.5. A QSAR model having ${}^cR^2_p$ value above the stated threshold may be considered not to be obtained by chance.

Validation through the external set was done by the evaluation of predictive correlation coefficient for the test set designated R^2_{pred}, which reflect the degree of correlation between the observed and predicted activity data for the test set. R^2_{pred} is defined by equation 10:

$$R^2_{pred} = 1 - \frac{\sum\left(Y_{obs(test)} - Y_{pred(test)}\right)^2}{\sum\left(Y_{obs(test)} - \bar{Y}_{training}\right)^2} \quad (10)$$

Here, $Y_{obs(test)}$ and $Y_{pred(test)}$ are the observed and predicted activity data for the test set compounds, while $\bar{Y}_{training}$ indicates the mean observed activity of the training set. The external predictive abilities

of the models for the test set were further judged using the Golbraikh and Tropsha criteria for a predictive model listed below:

a. $Q^2 > 0.5$
b. $R^2_{pred} > 0.6$
c. $r^2 - r^2_0/r^2 < 0.1$ and $0.85 \leq k \leq 1.15$ or $r^2 - r'^2/r^2 < 0.1$ and $0.85 \leq k' \leq 1.15$
d. $|r^2_0 - r'^2_0| < 0.3$

where Q^2 and R^2_{pred} were as discussed above, r^2 is the square correlation coefficients of the plot of observed against predicted activity values, r^2_0 is the square correlation coefficients of the plot of observed against predicted activity values at zero intercept, r'^2_0 the square correlation coefficients of the plot predicted against observed activity values at zero intercept, k is the slope of the plot of observed against predicted activity values at zero intercept and k' is the slope of the plot of predicted versus observed activity values at zero intercept. Finally modified square correlation coefficient designated R^2_m introduced by Roy *et al.* was also used to corroborate the other validation parameters and further verify the propinquity between the observed and predicted data. This was estimated for both the internal LOO cross validated training data ($R^2_{m(loo)}$) and the predicted test set data ($R^2_{m(test)}$). In general R^2_m is defined by equation 11:

$$R^2_m = r^2 \times \left(1 - \sqrt{(r^2 - r_0^2)}\right) \quad \text{(Eq. 11)}$$

where r^2 and r^2_0 are the square correlation coefficients of the plot of observed against predicted activity values of compounds (either for the training LOOCV or test set) with and at zero intercept respectively.

Relative importance of each descriptor in the model

Absolute value of the mean effect of each descriptor was used to evaluate the relative importance and contribution of the descriptor to the model. The mean effect is defined by Equation 12:

$$MF_j = \frac{\beta_j \sum_{i=1}^{i=n} d_{ij}}{\sum_j^m \beta_j \sum_i^n d_{ij}} \quad \text{(Eq. 12)}$$

where MF_j is the mean effect of a descriptor j in a model, β_j is the coefficient of the descriptor J in that model and d_{ij} is the value of the descriptor in the data matrix for each molecule in the training set, m is the number of descriptor that appear in the model and n is the number of molecules in the training set (49).

Models applicability domain

The extent of extrapolation method based on leverages value was employed to define the applicability domain (AD) of the QSAR models. The leverage (h_{ii}) value for each molecule was obtained has the diagonal elements of the hat matrix constructed for both training set and test set using the following consecutive steps. All calculations were done using Microsoft Excel software version 2007.

a. The descriptors for the training set were arranged in (n×m) matrix with the addition of a column vector of identity element (*i.e.* 1s) as the first column. This was designated $\mathbf{X_{tr}}$ (n×m) where n is the number of molecule that constitute the training set and m is the number of descriptors in the model
b. The transpose of the descriptor matrix was obtained designated $\mathbf{X^T_{tr}}$ (m× n).
c. The descriptor matrix was pre-multiply by its transpose resulting in symmetric matrix designated $\mathbf{X^T_{tr}X_{tr}}$ (m×m). Note that this multiplication is not commutative.
d. The inverse of the symmetric matrix in step c above designated $(\mathbf{X^T_{tr}X_{tr}})^{-1}$ (m×m) was evaluated. This was also a symmetric matrix and sometimes called "the clone"
e. The clone matrix was pre-multiply by the descriptor matrix and the result was designated $\mathbf{X_{tr}}$ $(\mathbf{X^T_{tr}X_{tr}})^{-1}$(n×m)
f. Finally the n×m matrix obtained in step e above was used to pre-multiply the transpose matrix obtained in step b above to give the hat matrix designated $\mathbf{H_{tr}}$ (n×n). The hat matrix for the training set was a symmetric matrix whose diagonal element represents the leverages for the training set. $\mathbf{H_{tr}}$ (n×n) = $\mathbf{X_{tr}}$ $(\mathbf{X^T_{tr}X_{tr}})^{-1}$(n×m)· $\mathbf{X^T_{tr}}$ (m× n)

In a similar manner but with slight modification the hat matrix for the test set was evaluated as follows:

a. The descriptors for the test set were arranged in (z×m) matrix with the addition of a column vector of identity element (i.e. 1s) as the first column. This was designated $\mathbf{X_{ext}}$ (z×m) where z is the number of molecule that constitute the test set and m is the number of descriptors in the model.
b. The transpose of the descriptor matrix was obtained designated $\mathbf{X^T_{ext}}$ (m× z).
c. The clone matrix obtained for the training test was used to post-multiply the descriptor matrix of the test set i.e $\mathbf{X_{ext}}$ (z×m)· $(\mathbf{X^T_{tr}X_{tr}})^{-1}$ (m×m) = $\mathbf{X_{ext}}(\mathbf{X^T_{tr}X_{tr}})^{-1}$(z×m). This step mapped the test set data into the training set data space.
d. Finally the z ×m matrix obtained in step c above was used to pre-multiply the transpose matrix obtained in step b above to give the hat matrix for the test set designated $\mathbf{H_{ext}}$ (z×z). The hat matrix for the test was also a symmetric matrix whose diagonal element represents the leverages for the test set. $\mathbf{H_{ext}}$ (z×z) = $\mathbf{X_{ext}}$ $(\mathbf{X^T_{tr}X_{tr}})^{-1}$(z×m)· $\mathbf{X^T_{tr}}$ (m×z).

Thereafter, a cut of leverage designated h* was evaluated using the equation 13 below.

$$h^* = \frac{3(m+1)}{n} \tag{13}$$

where m is the number of descriptors that appear in a model and n is the number of molecule in that make up the training set only. Any data point (activity value) whose leverage value exceeds h* is termed an influential point. Such influential data is not similar to the majority of the data used to train or build the model. However, such a data is not necessarily an outlier (50). A data is said to be an outlier to a given model if the standardized cross-validated residual of the data (activity

value) produced by the model is greater than ±3. Standardized cross-validated residual was calculated using equation 14 below.

$$SDR = \frac{\hat{y} - y}{\sqrt{\frac{\sum_{i=1}^{n}(\hat{y} - y)^2}{n}}} \quad (14)$$

where y is the observe activity value for either the training or the test set, $\hat{y}$ is the predicted activity value by the model and n is the number of molecules either in the training or test set. A graphical view of leverage values for each molecule in the entire data set is term William's plot which is a plot of SDR against the leverages.

RESULT AND DISCUSSION

The Kennard-Stone algorithm used in the present study partitioned the data of 62 derivatives into training set of 49 compounds and a test set of 13 compounds (see Table 1 note). Descriptive statistics of the activity values of the training and test set data showed that for test set values range (5.0533 to 3.3267) was within the training set value range (5.268 to 2.952). Also, the mean and standard deviation of the test set activity value (4.549 and 0.567) were approximately similar to that of the training set value (4.183 and 0.673). This indicated that the test set is interpolative within the training set and the spread or point distribution of the two set were comparable, which imply the Kennard and Stone algorithm employed was able to obtained a test set that is a good reflection of the training set data. After rigorous validation and inspection, the best QSAR model obtained by the GFA-MLR method employed in this study is presented in equation 15 together with its validation parameter.

$$pED50 = 1.968(\pm0.147) - 0.486(\pm0.135)\,\mu + 1.182(\pm0.206)\,\epsilon LUMO + 0.786(\pm0.135)\,PSA + 3.117(\pm0.189)\,WAA - 0.613(\pm0.193)\,QA + 0.457(\pm0.093)\,QF \quad \text{(Eq. 15)}$$

Internal validation parameters

N= 49, d = 6, R = 0.968, R^2 = 0.937, R^2_{adj} = 0.928, $F_{4, 42}(0.05)$ = 104.119, SEE = 0.181, Q^2 = 0.913, PRESS = 1.885, S_{PRESS} = 0.196, $^cR^2p$ = 0.890.

External validation parameters

R^2_{pred} = 0.929, r^2 = 0.913, r^2_0 = 0.906, r'^2_0 = 0.877, $R^2_{m(test)}$ = 0.864, $|r^2_0 - r'^2_0|$= 0.029, k = 1.012, $r^2 - r^2_0/r^2$ = 0.008, k' = 0.987 and $r^2 - r'^2/r^2$ = 0.039

The values in the parenthesis in model 15 are the standard deviation of the regression coefficients. There were d = 6 descriptors in the model and N= 49 compounds in the training set therefore, the QSAR 'rule of thumb' was achieved because there were at least five compounds for every

independent variable present in the equation and as a result the risk of chance correlation was acceptably low (51). Thus, the GFA-MLR performed on this data was justified. Model 15 had coefficient of determination R^2 = 0.937 which indicated that it was able to explain 93.7 % of the variance in the anticonvulsant activity of the data set against MES induced seizure, which was also confirmed it low standard deviation (standard error of estimation) value (SEE = 0.181). In addition, Model 15 possesses excellent adjustment level because it had high correlation coefficient R = 0.968 and low SEE (52). Furthermore, the SEE value represents about 4.33 % of the mean response (mean of the ant-MES value of the training set) variable and this is far less than 15 % which is an acceptable value for biological measurement such as ED50 (4). More also, the models coefficient of determination R^2_{adj} = 0.928, leave one out cross validation correlation coefficient Q^2 = 0.913 were greater than 0.6 which indicated that the model had good internal predictive ability (46, 53). This was also evident in the value of the root of the mean square error in the cross validated test (standard deviation of PRESS) S_{PRESS} = 0.196 which was only 0.35 % higher than SEE (4). For the statistics and inter-correlation matrix of the six variables employed in the QSAR Model 15 (Table 3), descriptors included in the model were significant were significant at 95 % level as evident in the model F-test ($F_{6, 42}(0.05)$ = 104.119) and the descriptors t-statistics greater than 2 (see Table 3). The estimated VIF values for all the descriptors were less than 4 (see Table 3) corroborating the conclusion that Model15 was statistically significant and the descriptors were orthogonal *i.e.* no problem of multicolinearity among them (53-54).

Table 3: Model 15 descriptors inter-correlation matrix and statistics

	Inter-correlation						statistics		
Descriptors	μ	ϵLUMO	PSA	WAA	QA	QF	t-sat	VIF	MF
μ	1						-3.611	2.085	-0.116
ϵLUMO	-0.287	1					5.752	1.506	0.183
PSA	0.576	-0.474	1				5.846	2.084	0.219
WAA	0.408	-0.475	0.481	1			16.43	2.652	0.776
QA	0.513	-0.335	0.586	0.742	1		-3.176	2.812	-0.113
QF	-0.310	-0.193	0.059	0.214	0.098	1	4.951	1.389	0.050

The auto-scaled descriptor, experimental and predicted activity, and residual values for the training data by the Model 15 are presented in Table 4. The plot of experimental versus predicted $-LogED_{50}$ for the training set data by the model (Fig. 2) showed the existence of linearity the two variables as evident in the plots regression coefficient r^2 value. Furthermore, the plot of standardized residual against predicted activity value (Fig. 3) showed a symmetric distribution or random scattering of data points above and below the line standardizes residual = 0 with nearly all the data points being within a boundary defined by standardized residual = ±2.5 corroborating the conclusion that the model has a good predictive ability for the training data and with no outlier in site (55-57). To

further check the robustness of the model ten y-randomization runs was performed on the training set data (Table 5) with resultant randomization modified determination coefficient ($^cR^2p$ = 0.890) greater than 0.5, averages of randomized correlation coefficients $\overline{R}^2_{rand}$ and cross-validated correlation coefficient $\overline{Q}^2_{rand}$ were less than 0.2. Thus, confirming the conclusion that Model 15 is not a product of chance correlation (58).

As a further and stronger predictive criterion, the obtained model was evaluated for its ability to predict the anti-MES activity of external test set data against MES and the results were included in Table 4. The plot of experimental versus predicted $-LogED_{50}$ for the test set data by the model (Fig. 4) showed the existence of linearity between the two variables as evident by its r^2 value. Also, the predictive ability of the model was further confirmed by the results of each of the following statistics: (a) R^2_{pred} = 0.929 > 0.6 (b) $R^2_{m(test)}$ = 0.864 > 0.5 (c) $r^2 - r^2_0/r^2$ = 0.008 < 0.1 (d) $r^2 - r'^2/r^2$ =0.039 < 0.1 (e) $|r^2_0 - r'^2_0|$ = 0.029 < 0.3 (f) k = 1.012 where 0.85 ≤ k ≤ 1.15 and (g) k′ = 0.987 where 0.85 ≤ k′ ≤ 1.15. From the above discussion, the model reported in the study was judged to be predictive, because, it passed all the necessary criteria and can therefore be used as tool for evaluating the anti-MES activity of novel compounds.

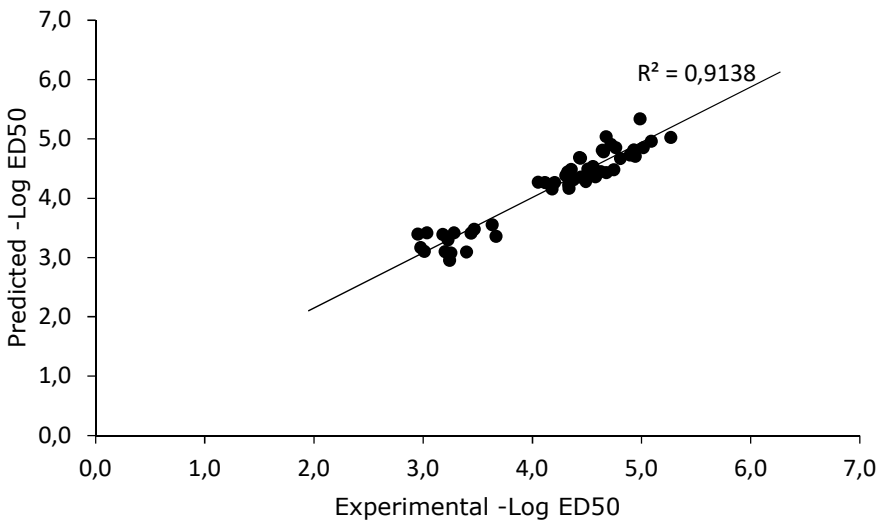


Figure 2: Training data predicted versus experimental anti-MES activity by Model 15.

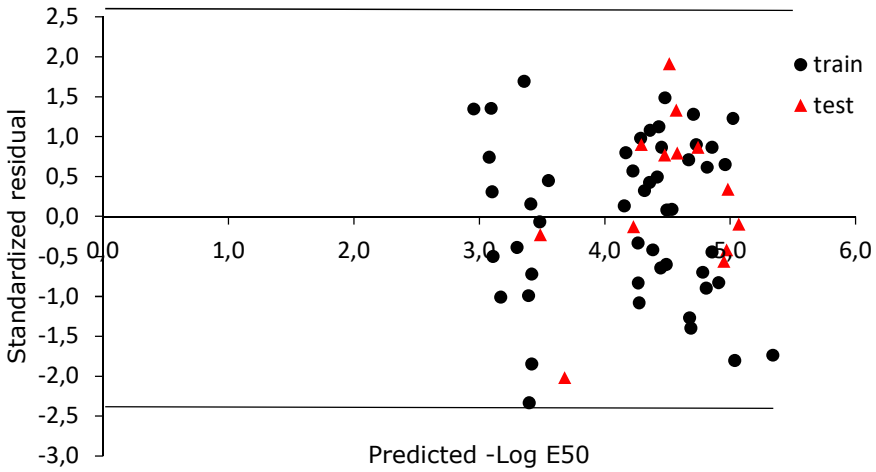


Figure 3: Standardized residual versus predicted anti-MES activity by Model 15.

Table 4: Normalized descriptors, experimental, predicted and residual of anti-MES seizure by Model 15.

No.	μ	ϵLumo	PSA	WAA	QA	QF	Y_{exp}	Y_{pred}	Res.	Stres.	h_{ii}
1[a]	0.352	0.606	0.000	0.353	0.208	0.000	3.447	3.486	-0.039	-0.226	0.151
2	0.811	0.176	0.533	0.280	0.098	0.000	3.242	3.014	0.228	1.349	0.211
3	0.758	0.297	0.523	0.476	0.471	0.000	3.633	3.557	0.076	0.452	0.083
4	0.871	0.000	0.563	0.431	0.332	0.000	3.253	3.128	0.125	0.741	0.280
5	0.590	0.341	0.739	0.349	0.662	0.000	3.181	3.348	-0.167	-0.988	0.203
6	0.710	0.326	0.744	0.232	0.361	0.000	3.011	3.095	-0.084	-0.498	0.131
7	0.548	0.215	0.760	0.322	0.345	0.000	2.952	3.346	-0.394	-2.333	0.115
8	0.361	0.244	0.452	0.275	0.239	0.000	2.976	3.147	-0.171	-1.011	0.118
9	0.690	0.315	0.600	0.417	0.486	0.000	3.468	3.479	-0.011	-0.066	0.079
10	0.749	0.441	0.507	0.354	0.364	0.000	3.283	3.404	-0.122	-0.721	0.083
11	0.410	0.297	0.454	0.326	0.184	0.000	3.667	3.380	0.287	1.697	0.072
12	0.179	0.810	0.284	0.043	0.048	0.000	3.396	3.167	0.229	1.358	0.239
13*	0.434	1.000	0.391	0.000	0.163	0.000	3.199	3.147	0.052	0.309	0.463
14	0.450	0.663	0.368	0.215	0.130	0.000	3.440	3.413	0.027	0.157	0.126
15[a]	0.512	0.720	0.383	0.305	0.234	0.000	3.327	3.679	-0.352	-2.018	0.153
16	0.153	0.462	0.189	0.241	0.076	0.000	3.228	3.293	-0.065	-0.386	0.1
17	0.552	0.527	0.076	0.309	0.000	0.000	3.034	3.346	-0.312	-1.845	0.185
18	0.197	0.444	0.591	0.489	0.396	0.512	4.307	4.377	-0.070	-0.417	0.076
19	0.173	0.290	0.433	0.569	0.299	0.000	4.180	4.158	0.023	0.134	0.128
20	0.000	0.251	0.759	0.458	0.322	0.505	4.488	4.323	0.166	0.980	0.186
21	0.251	0.269	0.431	0.572	0.428	0.505	4.114	4.254	-0.140	-0.828	0.073
22	0.304	0.244	0.454	0.565	0.508	1.000	4.444	4.373	0.072	0.425	0.183
23	0.197	0.251	0.757	0.462	0.386	0.932	4.577	4.394	0.183	1.084	0.159
24	0.495	0.444	0.281	0.588	0.451	0.449	4.331	4.235	0.096	0.571	0.100
25	0.171	0.459	0.005	0.681	0.545	0.515	4.354	4.456	-0.101	-0.598	0.252
26	0.319	0.405	0.328	0.593	0.228	0.000	4.203	4.258	-0.056	-0.328	0.103
27	0.350	0.391	0.328	0.600	0.302	0.482	4.507	4.424	0.084	0.496	0.070

No.	μ	ϵLumo	PSA	WAA	QA	QF	Y_{exp}	Y_{pred}	Res.	Stres.	h_{ii}
28[a]	0.246	0.369	0.643	0.492	0.155	0.000	4.207	4.229	-0.022	-0.128	0.139
29	0.155	0.240	0.614	0.593	0.380	0.482	4.509	4.495	0.014	0.083	0.091
30	0.266	0.341	0.765	0.472	0.132	0.000	4.052	4.234	-0.182	-1.078	0.179
31	0.220	0.366	0.453	0.577	0.203	0.000	4.379	4.324	0.055	0.324	0.125
32	0.328	0.337	0.767	0.491	0.203	0.475	4.325	4.433	-0.108	-0.641	0.099
33	0.339	0.362	0.455	0.596	0.272	0.472	4.747	4.496	0.252	1.488	0.058
34	0.226	0.362	0.447	0.656	0.393	0.498	4.433	4.669	-0.236	-1.395	0.069
35	0.373	0.351	0.453	0.722	0.413	0.472	4.653	4.771	-0.118	-0.697	0.068
36[a]	0.253	0.358	0.183	0.659	0.352	0.498	4.613	4.478	0.135	0.775	0.120
37[a]	0.419	0.323	0.768	0.484	0.262	0.912	4.849	4.515	0.334	1.914	0.155
38[a]	0.393	0.341	0.456	0.585	0.333	0.915	4.716	4.576	0.139	0.799	0.132
39	0.149	0.308	0.455	0.601	0.394	0.957	4.807	4.687	0.120	0.710	0.133
40	0.743	0.269	0.999	0.704	0.518	0.858	4.675	4.980	-0.304	-1.801	0.159
41	0.980	0.283	0.999	0.698	0.529	0.870	4.721	4.861	-0.140	-0.827	0.251
42	0.647	0.215	0.999	0.713	0.570	0.902	5.089	4.979	0.111	0.654	0.143
43	0.767	0.287	0.982	0.768	1.000	0.000	4.678	4.488	0.190	1.127	0.235
44	0.878	0.290	1.000	0.775	0.587	0.000	4.943	4.727	0.217	1.282	0.088
45	0.869	0.290	1.000	0.779	0.588	0.000	4.895	4.743	0.152	0.902	0.089
46	0.874	0.258	0.991	0.836	0.721	0.000	4.640	4.792	-0.151	-0.895	0.091
47[a]	0.891	0.294	1.000	0.925	0.788	0.000	5.053	5.069	-0.016	-0.092	0.122
48	0.821	0.297	1.000	1.000	0.888	0.000	4.986	5.279	-0.293	-1.734	0.168
49	0.867	0.287	0.999	0.947	0.919	0.000	5.268	5.060	0.208	1.231	0.156
50	0.865	0.333	0.955	0.757	0.645	0.000	4.442	4.657	-0.214	1.269	0.079
51	0.831	0.319	0.954	0.762	0.767	0.502	4.930	4.826	0.104	0.617	0.104
52	0.821	0.330	0.952	0.815	0.863	0.502	4.851	4.949	-0.098	0.904	0.175
53	0.876	0.276	0.953	0.846	0.627	0.000	5.018	4.871	0.147	0.868	0.092
54[a]	0.872	0.287	0.951	0.896	0.724	0.000	5.041	4.981	0.060	-0.560	0.131
55[a]	0.867	0.276	0.953	0.878	0.631	0.000	4.900	4.972	-0.073	0.347	0.103
56[a]	0.956	0.186	0.849	0.737	0.195	0.000	4.802	4.569	0.233	-0.416	0103
57	0.805	0.176	0.855	0.750	0.289	0.485	4.764	4.840	-0.075	-0.445	0.174
58[a]	1.000	0.219	0.856	0.820	0.372	0.000	4.893	4.742	0.151	1.334	0.263
59	0.705	0.258	0.972	0.512	0.150	0.000	4.335	4.199	0.136	0.803	0.193
60[a]	0.869	0.308	0.379	0.741	0.369	0.000	4.449	4.291	0.158	0.866	0.186
61	0.809	0.380	0.358	0.793	0.485	0.000	4.627	4.480	0.147	0.867	0.161
62	0.823	0.376	0.358	0.817	0.493	0.000	4.554	4.539	0.015	0.091	0.175

[a] stand for test set compound, Y_{exp} = experimental -Log ED_{50}, Y_{pred} = predicted –Log ED_{50}

Res. = Y_{exp} - Y_{pred} , Stres = standardized residual, h_{ii} = leverage values

Table 5: Result of ten randomization runs of the models training data.

Model	R	R^2	Q^2	$^cR_p^2$
Original	0.968	0.937	0.913	
Random 1	0.303	0.092	-0.271	
Random 2	0.169	0.028	-0.377	
Random 3	0.343	0.118	-0.186	
Random 4	0.394	0.155	-0.143	
Random 5	0.279	0.078	-0.231	
Random 6	0.169	0.029	-0.275	
Random 7	0.249	0.062	-0.245	
Random 8	0.316	0.100	-0.228	
Random 9	0.538	0.289	0.044	
Random 10	0.257	0.066	-0.305	
Average	0.302	0.102	-0.222	0.890

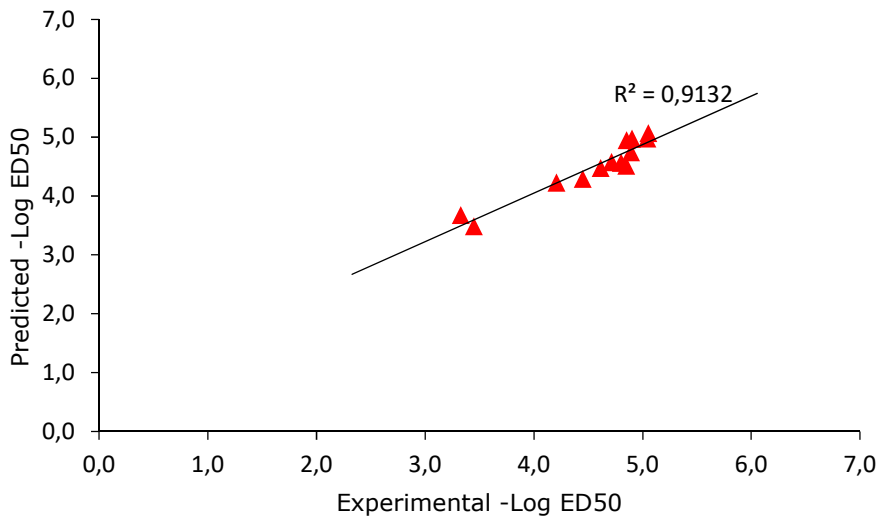


Figure 4: Test set data predicted versus experimental anti-MES activity by Model 15.

However, no matter how robust, significant and thoroughly validated a QSAR model may be, it cannot be expected to reliably predict the anti-MES activity for the entire universe of chemicals, rather, only those that are within its applicability domain (AD). The extent of extrapolation method employed in the present study to define the chemical space where the model makes reliable prediction gave a pictorial representation of the training and the test set data within the applicability domain of the model (Fig.5) with a cut off leverage h* = 0.43. And it was observed that all compounds of both set are within the AD of the model except for a structurally influential molecule 13 with leverage value h_{ii} = 0.46 > h* = 0.43*(Table 4).

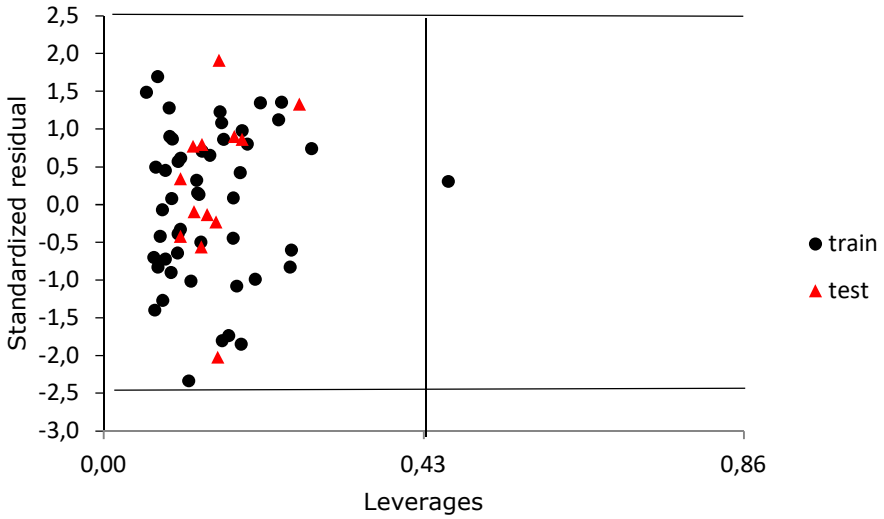


Figure 5: Williams plot for Model 15.

Interpretation of descriptors contained in the model

The model reported in the present study contained six descriptors which are the molecular dipole moment (μ), energy of the lowest unoccupied molecular orbital (ϵLUMO), polar surface area (PSA), accessible area of the molecule obtained from wave function calculation (WAA), sum of the absolute value of the charges on all atom in a molecule (QA) and sum of the square of the charges on all fluorine atom in a molecule (QF). Among the descriptors ϵLUMO, PSA, WAA and QF bear positive coefficients, while μ and QA bears negative coefficients. As usually, the positive sign of the descriptor means that it has a positive effect on the dependent variable and vice versa. Also, it should be noted that increasing the value of pED_{50} (*i.e.* -Log ED_{50}) means reducing the value of ED_{50} and this implies that increase in the values of descriptors that bears positive coefficient increases the anti-MES activities of studied compounds and *vice versa*. The relative importance and contribution of each descriptor was described and evaluated using absolute mean effect (MF) value and the order of their absolute MF value is: WAA > PSA > ϵLUMO > μ > QA > QF (see Table 3). This showed that WAA, PSA, ϵLUMO played major positive role on anti-MES activities of studied compounds, while μ and QA had similar influence and QF had the least influence on the studied bioactivity. Detailed explanations of the six descriptors were given in the following.

WAA is the wave function (quantum mechanical) calculated accessible area based on electron density as well as electrostatic potential map (59). It represents electrostatic potential surface area, a valuable parameter in computer-aided drug design that helps in optimization of electrostatic interactions between ligand and protein. Also, it helps in predicting the behavior of complex molecules and shows the total surface area of the molecule accessible to electrophilic and nucleophilic attack as well as depicts the overall molecular size and shape (59). 3D representation

of the parameter illustrates the charge distributions in a molecules and by color convention, colors toward red depict negative potential *i.e.* electron-rich region and subject to attack by electrophiles, while colors toward blue depict positive potential *i.e.* electron poor region and subject to attack by nucleophiles and colors in between (orange, yellow, green) depict intermediate values of the potential (59). In the regression model the WAA is positively correlated to the anti-MES activity as indicated by it positive mean effect value (see Table 3), this shows that electronegativity of atoms has prominent positive influence on anti-MES activity of studied. Higher values of pED_{50} observed in compounds 1 and 3 when compare to similar compounds 2 and 4 could be attributed to the presence of more electronegative Cl atom in compounds 1 and 3. Similar trend was observed when compounds 60, 61 and 62 are compared (Table 1)

PSA represents molecular polar surface area and is defined as the area due to nitrogen and oxygen and any attached hydrogen *i.e.* the molecular surface sum, usually van der Waals, over all polar atoms (59). It is commonly used in for the optimization of a drug's ability to permeate cells. Molecules with a polar surface area greater than 140 $Å^2$ tend to be poor at permeating cell membranes (60). Rather, a PSA less than 90 $Å^2$ is usually needed (61). In the regression model the PSA is positively correlated to the anti-MES activity as indicated by it positive mean effect value (see Table 3), this shows that nitrogen and oxygen atoms have prominent positive influence on anti-MES activity of studied. The little increase in the pED_{50} value of compound 3 may be attributed to the presence of additional O atom when compare to similar compound 1 (Table 1).

εLUMO represents the energy of the lowest unoccupied molecular orbital. It is a very popular quantum chemical descriptor with great importance in several chemicals and pharmacological processes. It has been shown to play prominent role in the formation of many charge transfer complexes (62). According to the frontier molecular orbital theory (FMO) of chemical reactivity, the formation of a transition state is due to an interaction between the frontier orbitals (HOMO and LUMO) of reacting species (63). Also, the energy of the LUMO has been directly related to the electron affinity and characterizes the susceptibility of the molecule toward attack by nucleophiles (64). In the regression model the εLUMO is positively correlated to the anti-MES activity as indicated by it positive mean effect value (Table 3), this shows that electron donating groups impart the positive influence on anti-MES activities of studied compounds. pED_{50} value of compounds 32 < 33 and that of 34 < 36 may be attributed to electron donating alkyl group attached to the amine nitrogen of the 1H-[1,2,3]triazolo[4,5-c]pyridine-4-amine presents in these compounds (Table1).

μ represents the total molecular dipole moment, a 3D electronic descriptor that indicates the strength and orientation behavior of a molecule in an electrostatic field. It explains the charge distribution in the molecule and often considered as the direct characteristic of the global polarity

of a molecule because it is obtained from partial charges defined on the atoms of the molecule and if no partial charges is defined, the molecular dipole moment will be zero (65). Dipole moment may be related to information about site-specific receptor binding ability of a molecule, because drug size and charge distributions are essential factors to bind active site of receptor molecule (66, 67). In the regression model the dipole moment is negatively correlated to the anti-MES activity, this indicates that decreasing the polarity of the molecule by substituting such groups (large substituent) that decrease the polarity of a molecule as a whole (36) will account for increase in anti-MES activity. The order of pED_{50} value of similar compounds 35 > 36 > 34 (Table 1) may be attributed to bulky group attached to the amine nitrogen of the 1H-[1,2,3]triazolo[4,5-c]pyridine-4-amine presents in these compounds.

QA and QF are the sum of the absolute value of the charges on all atoms in a molecule and sum of the square of the charges on all fluorine atoms in a molecule respectively. They are atomic charge based descriptors and it has been reported that atomic partial charges have been used as static chemical reactivity indices (62). Also, various sums of absolute or squared values of partial charges have been also used to describe intermolecular interactions (68-70). In fact, according to classical chemical theory, all chemical interactions are by nature either electrostatic (polar) or orbital (covalent). Electrical charges in the molecule are obviously the driving force of electrostatic interactions. Indeed, it has been proven that local electron densities or charges are important in many chemical reactions and physico-chemical properties of compounds and atomic charges are also used for the description of the molecular polarity of molecules (71, 72). In the regression model the QA is negatively correlated, while QF is positively correlated to the anti-MES, indicating the number of F atom impart positive influence on the anti-MES activities of studied compounds. The order of pED50 value of similar compounds 30 < 32 < 37 may be attributed to the number of F atom in the compounds.

***In silico* screening**

The main goal of *in silico* (virtual) screening is to identify compounds that are promising synthetic targets for the studied biological activity. Accomplishing this, is to determine whether the developed QSAR model could predict structures as more or less active, as those used for the training and validation sets and to identify which structural modifications could be allowed using the domain of applicability. In the present study a template based new compound design method was employed starting with compound 61 with $-\text{Log } ED_{50}$ value of 4.627(see Table 1) as the template since it has the lowest ED50 value among the 1H-pyrazole-5-carboxylic derivatives present in the data set and therefore offers a good anti-MES activity and looked promising as a useful scaffold.

The chosen scaffold was divided into some structural fragments: 1H-pyrrole, ethyl formate, morpholine, and chlorobenzene (Fig. 6) upon which simple modifications were made in order to obtain the new compounds for the virtual screening process. Since the chemistry of five membered pyrrole has been established and well understood, introducing structural modifications around the 1H-pyrole ring was considered synthetically viable and as a result the *in silico* screening employed in the present study proceeds from this angle. A total of 101 compounds were designed in the study, their structures were subsequently optimized and their molecular descriptors were calculated using the combination of MMFF, AMI and DFT B3LYP/6-31G (d, p) methods as used for the data set used in building the model. Thereafter, the model was used to predict the activity values for the designed compounds and the result showed that the model tolerated most of the structural modification made, as the leverage values of most of the designed compounds were less than the warning leverage (h* = 0.43) (see Table 6). However, compounds obtained from the modification of the ethyl formate fragments, *i.e.* compounds 12 m to 14 m and 65 m to 71 m had leverage values greater than 0.43, therefore, they are not within the applicability domain of the model. Also, substitution of the morpholine ring with lesser ring systems like five- to three-membered rings led to compounds that were not within the activity domain of the model i.e. compounds 88m, 89m and 94m to 99m (Table 6).

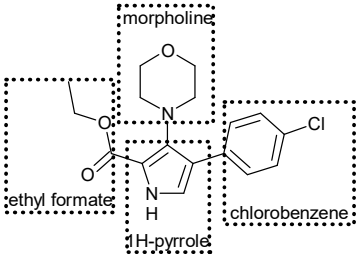


Figure 6: Structural fragment of the chosen scaffold.

Among the designed compounds tolerated by the model, compounds designated with asterisk in Table 6 *i.e.* 11 m, 16 m, 22 m, 23 m, 24 m, 45 m, 57 m, 58 m, 59 m, 81 m and 101 m with pED50 (leverage) values of 4.638(0.395), 4.632(0.248), 4.805(0.316), 4.629(0.237), 4.826(0.315), 5.406(0.445), 4.851(0.268), 4.649(0.237), 4.664(0.269), 4.634(0.340) and 5.243(0.421) respectively had improve anti-MES activity when compared with the chosen scaffold whose pED50 (leverage) value was 4.627 (0.161) (Table 4). Substitution of the morpholine ring fragment in chosen scaffold (Fig.6) with methylpiperazine (compounds 11 m), cyclopropylpiperazine (compounds 101 m, 45 m), methyl-1,4-dihydropyrazine (compounds 24 m, 81 m), methylpiperidine (compound 22 m) and 4,4-dimethylpiperidine (compound 23 m) ring systems accounted for increase in the number of nitrogen atoms in the molecule, increase in the molecular size and introduction of more electron donating groups, *e.g.* methyl and cyclopropyl, into the molecular system thereby leading to increase in the values of WAA, PSA, εLUMO and relative decrease in the

value of dipole moment μ of the corresponding designed compounds which in turn leads to higher pED_{50} values for the compounds. Also, substitution of the morpholine with seven membered ring systems of 4,5-dihydro-1H-1,4-diazepine (compound 57 m), 4,5-dihydro-1H-1,4-oxazepine (compound 58 m) and 4,5-dihydro-1H-1,4-thaizepine (compound 59 m) led to similar observation. The higher pED_{50} values predicted for these compounds only shows which structures should be targeted for synthesis on the basis that they approach the optimal values for the chosen descriptors in the model developed in the present study. Also, the *in silico* screen based on the developed QSAR model clearly achieved its objective in identifying derivatives of 1H-pyrazole-5-carboxylic acid with improved predicted activity while simultaneously identifying structural modifications that were out of the models domain of applicability and therefore the scope of the models reliability. This study thus demonstrates the usefulness of constructing QSAR models which can aid in identifying new synthetic targets for drug discovery.

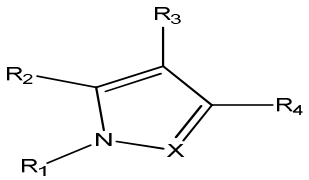


Table 6: Structural modification around the 1H-pyrole ring and predicted activities.

S/N	R_1	R_2	R_3	R_4	X	$-LogED_{50}$ (predicted)	leverages
1 m	H	$C_3H_5O_2$	C_4H_8NO	C_6H_4Cl	N	3.652	0.049
2 m	H	$C_3H_5O_2$	C_4H_8NO	C_6H_4Cl	NCH_3	3.950	0.104
3 m	H	$C_3H_5O_2$	C_4H_8NO	C_6H_4Cl	NOH	3.961	0.194
4 m	H	$C_3H_5O_2$	C_4H_8NO	C_6H_4Cl	$NOCH_3$	4.493	0.097
5 m	H	$C_3H_5O_2$	C_4H_8NO	C_6H_4Cl	NNH_2	4.036	0.189
6 m	H	$C_3H_5O_2$	C_4H_8NO	C_6H_4Cl	$NNHCH_3$	4.105	0.110
7 m	H	$C_3H_5O_2$	C_4H_8NO	C_6H_4Cl	$NN(CH_3)_2$	4.585	0.112
8 m	H	C_3H_5OS	C_4H_8NO	C_6H_4Cl	CH	3.710	0.493
9 m	H	C_3H_5ONH	C_4H_8NO	C_6H_4Cl	CH	4.247	0.369
10 m	H	$C_3H_5ONCH_3$	C_4H_8NO	C_6H_4Cl	CH	4.619	0.431
11 m*	H	C_3H_5ONOH	C_4H_8NO	C_6H_4Cl	CH	4.638	0.395
12 m	H	$C_3H_5ONOCH_3$	C_4H_8NO	C_6H_4Cl	CH	5.183	0.604
13 m	H	$C_3H_5ONNHCH_3$	C_4H_8NO	C_6H_4Cl	CH	5.304	0.609
14 m	H	$C_3H_5ONN(CH_3)_2$	C_4H_8NO	C_6H_4Cl	CH	5.074	0.803
15 m	H	$C_3H_5O_2$	C_4H_8NNH	C_6H_4Cl	CH	4.113	0.145
16 m*	H	$C_3H_5O_2$	$C_4H_8NNCH_3$	C_6H_4Cl	CH	4.632	0.248
17 m	H	$C_3H_5O_2$	C_4H_8NS	C_6H_4Cl	CH	4.140	0.148
18 m	H	$C_3H_5O_2$	$C_4H_5N_2$	C_6H_4Cl	CH	4.532	0.220
19 m	H	$C_3H_5O_2$	C_4H_4NO	C_6H_4Cl	CH	4.367	0.175
20 m	H	$C_3H_5O_2$	C_4H_4NS	C_6H_4Cl	CH	4.141	0.242
21 m	H	$C_3H_5O_2$	$C_5H_{10}N$	C_6H_4Cl	CH	4.084	0.178
22 m*	H	$C_3H_5O_2$	$CH_3C_5H_9N$	C_6H_4Cl	CH	4.805	0.316
23 m*	H	$C_3H_5O_2$	$(CH_3)_2C_5H_8N$	C_6H_4Cl	CH	4.629	0.237
24 m*	H	$C_3H_5O_2$	$C_4H_4NNCH_3$	C_6H_4Cl	CH	4.826	0.315
25 m	H	$C_3H_5O_2$	C_4H_8N	C_6H_4Cl	CH	3.964	0.180
26 m	H	$C_3H_5O_2$	C_4H_4N	C_6H_4Cl	CH	3.883	0.159
27 m	H	$C_3H_5O_2$	C_3H_6NO	C_6H_4Cl	CH	4.058	0.199
28 m	H	$C_3H_5O_2$	C_3H_4NO	C_6H_4Cl	CH	3.858	0.104
29 m	H	$C_3H_5O_2$	C_3H_4NS	C_6H_4Cl	CH	4.008	0.153
30 m	H	$C_3H_5O_2$	C_3H_6NS	C_6H_4Cl	CH	4.273	0.203
31 m	H	$C_3H_5O_2$	C_3H_6NNH	C_6H_4Cl	CH	3.788	0.079

S/N	R_1	R_2	R_3	R_4	X	-LogED$_{50}$ (predicted)	leverages
32 m	H	$C_3H_5O_2$	$C_3H_6NNCH_3$	C_6H_4Cl	CH	4.550	0.290
33 m	H	$C_3H_5O_2$	$C_3H_4NNCH_3$	C_6H_4Cl	CH	4.491	0.253
34 m	H	$C_3H_5O_2$	C_3H_4NNH	C_6H_4Cl	CH	4.602	0.345
35 m	H	$C_3H_5O_2$	C_3H_6N	C_6H_4Cl	CH	3.824	0.248
36 m	H	$C_3H_5O_2$	C_2H_4NO	C_6H_4Cl	CH	3.445	0.122
37 m	H	$C_3H_5O_2$	C_2H_4NS	C_6H_4Cl	CH	3.848	0.152
38 m	H	$C_3H_5O_2$	C_2H_4NNH	C_6H_4Cl	CH	3.519	0.083
39 m	H	$C_3H_5O_2$	$C_2H_4NNCH_3$	C_6H_4Cl	CH	4.254	0.244
40 m	H	$C_3H_5O_2$	C_2H_4N	C_6H_4Cl	CH	3.403	0.163
41 m	H	$C_3H_5O_2$	CH_2NO	C_6H_4Cl	CH	3.007	0.174
42 m	H	$C_3H_5O_2$	CH_2NS	C_6H_4Cl	CH	2.989	0.197
43 m	H	$C_3H_5O_2$	CH_2NNH	C_6H_4Cl	CH	3.241	0.134
44 m	H	$C_3H_5O_2$	CH_2NNCH_3	C_6H_4Cl	CH	3.621	0.137
45 m*	H	$C_3H_5O_2$	$C_4H_8NNC_3H_5$	C_6H_4Cl	CH	5.406	0.445
46 m	H	$C_3H_5O_2$	C_2H_2N	C_6H_4Cl	CH	3.520	0.182
47 m	H	$C_3H_5O_2$	$C_6H_{12}N$	C_6H_4Cl	CH	4.448	0.271
48 m	H	$C_3H_5O_2$	$C_5H_{10}N$-20-O	C_6H_4Cl	CH	4.363	0.320
49 m	H	$C_3H_5O_2$	$C_5H_{10}N$-19-O	C_6H_4Cl	CH	4.124	0.094
50 m	H	$C_3H_5O_2$	$C_5H_{10}N$-18-O	C_6H_4Cl	CH	4.370	0.134
51 m	H	$C_3H_5O_2$	$C_5H_{10}N$-18-S	C_6H_4Cl	CH	4.192	0.157
52 m	H	$C_3H_5O_2$	$C_5H_{10}N$-19-S	C_6H_4Cl	CH	3.960	0.146
53 m	H	$C_3H_5O_2$	$C_5H_{10}N$-20-S	C_6H_4Cl	CH	4.385	0.434
54 m	H	$C_3H_5O_2$	$C_5H_{10}N$-20-NH	C_6H_4Cl	CH	4.477	0.151
55 m	H	$C_3H_5O_2$	$C_5H_{10}N$-19-NH	C_6H_4Cl	CH	4.606	0.266
56 m	H	$C_3H_5O_2$	$C_5H_{10}N$-18-NH	C_6H_4Cl	CH	4.424	0.149
57 m*	H	$C_3H_5O_2$	C_5H_6N-20-NH	C_6H_4Cl	CH	4.851	0.268
58 m*	H	$C_3H_5O_2$	C_5H_6N-20-O	C_6H_4Cl	CH	4.649	0.237
59 m*	H	$C_3H_5O_2$	C_5H_6N-20-S	C_6H_4Cl	CH	4.664	0.269
60 m	H	$C_3H_5O_2$	C_5H_6N-20-S	C_6H_4F	CH	4.457	0.343
61 m	H	$C_3H_5O_2$	C_4H_8NO	C_6H_4F	N	3.421	0.304
62 m	H	$C_3H_5O_2$	C_4H_8NO	C_6H_4F	NCH_3	3.546	0.360
63 m	H	$C_3H_5O_2$	C_4H_8NO	C_6H_4F	$NOCH_3$	4.202	0.249
64 m	H	$C_3H_5O_2$	C_4H_8NO	C_6H_4F	$NNHCH_3$	4.013	0.353
65 m	H	C_3H_5OS	C_4H_8NO	C_6H_4F	CH	3.448	0.487
66 m	H	C_3H_5ONH	C_4H_8NO	C_6H_4F	CH	4.006	0.685
67 m	H	$C_3H_5ONCH_3$	C_4H_8NO	C_6H_4F	CH	4.538	0.809
68 m	H	C_3H_5ONOH	C_4H_8NO	C_6H_4F	CH	4.440	0.677
69 m	H	$C_3H_5ONOCH_3$	C_4H_8NO	C_6H_4F	CH	5.030	0.732
70 m	H	$C_3H_5ONNHCH_3$	C_4H_8NO	C_6H_4F	CH	5.093	0.757
71 m	H	$C_3H_5ONN(CH_3)_2$	C_4H_8NO	C_6H_4F	CH	5.206	0.857
72 m	H	$C_3H_5O_2$	C_4H_8NNH	C_6H_4F	CH	3.908	0.403
73 m	H	$C_3H_5O_2$	$C_4H_8NNCH_3$	C_6H_4F	CH	4.434	0.356
74 m	H	$C_3H_5O_2$	C_4H_8NS	C_6H_4F	CH	3.923	0.349
75 m	H	$C_3H_5O_2$	C_4H_4NNH	C_6H_4F	CH	4.329	0.337
76 m	H	$C_3H_5O_2$	C_4H_4NO	C_6H_4F	CH	4.154	0.260
77 m	H	$C_3H_5O_2$	C_4H_4NS	C_6H_4F	CH	3.937	0.365
78 m	H	$C_3H_5O_2$	$C_5H_{10}N$	C_6H_4F	CH	3.854	0.411
79 m	H	$C_3H_5O_2$	$CH_3C_5H_9N$	C_6H_4F	CH	4.353	0.392
80 m	H	$C_3H_5O_2$	$(CH_3)_2C_5H_8N$	C_6H_4F	CH	4.578	0.445
81 m*	H	$C_3H_5O_2$	$C_4H_4NNCH_3$	C_6H_4F	CH	4.634	0.340
82 m	H	$C_3H_5O_2$	C_4H_8N	C_6H_4F	CH	3.601	0.387
83 m	H	$C_3H_5O_2$	C_4H_4N	C_6H_4F	CH	3.655	0.285
84 m	H	$C_3H_5O_2$	C_3H_6NO	C_6H_4F	CH	3.821	0.403
85 m	H	$C_3H_5O_2$	C_3H_4NO	C_6H_4F	CH	3.633	0.303
86 m	H	$C_3H_5O_2$	C_3H_4NS	C_6H_4F	CH	3.790	0.305
87 m	H	$C_3H_5O_2$	C_3H_6NS	C_6H_4F	CH	3.828	0.328
88 m	H	$C_3H_5O_2$	C_3H_6NNH	C_6H_4F	CH	4.040	0.464
89 m	H	$C_3H_5O_2$	$C_3H_6NNCH_3$	C_6H_4F	CH	4.358	0.451
90 m	H	$C_3H_5O_2$	$C_3H_4NNCH_3$	C_6H_4F	CH	4.281	0.376

S/N	R_1	R_2	R_3	R_4	X	-LogED$_{50}$ (predicted)	leverages
91 m	H	$C_3H_5O_2$	C_3H_4NNH	C_6H_4F	CH	3.808	0.159
92 m	H	$C_3H_5O_2$	C_2H_4NO	C_6H_4F	CH	4.284	0.374
93 m	H	$C_3H_5O_2$	C_2H_4NS	C_6H_4F	CH	3.628	0.374
94 m	H	$C_3H_5O_2$	C_2H_4NNH	C_6H_4F	CH	3.526	0.505
95 m	H	$C_3H_5O_2$	C_2H_4NNCH3	C_6H_4F	CH	3.997	0.433
96 m	H	$C_3H_5O_2$	C_2H_4N	C_6H_4F	CH	3.164	0.503
97 m	H	$C_3H_5O_2$	CH_2NO	C_6H_4F	CH	2.803	0.511
98 m	H	$C_3H_5O_2$	CH_2NS	C_6H_4F	CH	2.646	0.547
99 m	H	$C_3H_5O_2$	CH_2NNH	C_6H_4F	CH	3.034	0.484
100 m	H	$C_3H_5O_2$	CH_2NNCH_3	C_6H_4F	CH	3.671	0.369
101 m*	H	$C_3H_5O_2$	$C_4H_8NNC_3H_5$	C_6H_4F	CH	5.243	0.421

'*' represent compounds with improved activity when compared to the chosen scaffold

Docking studies

Compounds with activity in MES test were reported to prevent convulsion by acting as voltage-gated sodium channel blockers. Therefore, molecular docking study was carried out in order to elucidate which of the designed compounds have good affinity against solution structure of the human-voltage gated sodium channel (VGSC), brain isoform ($N_aV1.2$) reported by Milousher *et al* (73). The structure of the VGSC used in the study was downloaded from protein data bank with PDB code: 2 KaV (Fig 7a). The optimized structure of compounds 11 m, 16 m, 22 m, 23 m, 24 m, 45 m, 57 m, 58 m, 59 m, 81 m, and 101 m (Table 6) and phenyltion (a standard molecule known for its selective activity in MES test) saved as SDF files were converted to PDB files using Discovery studio software. These compounds were docked with prepared structure of 2KaV using Autodock vina (74) incorporated in pyrx software. The grid box was set to maximum (X = 55.53, Y = 52.05 and Z = 31.58) to cover entire 2 KaV and ligands structures. The docking results were compiled and analyzed using Autodock Tools-1.5.6 (incorporated in the pyrx) and discovery studio and reported in Table 7.

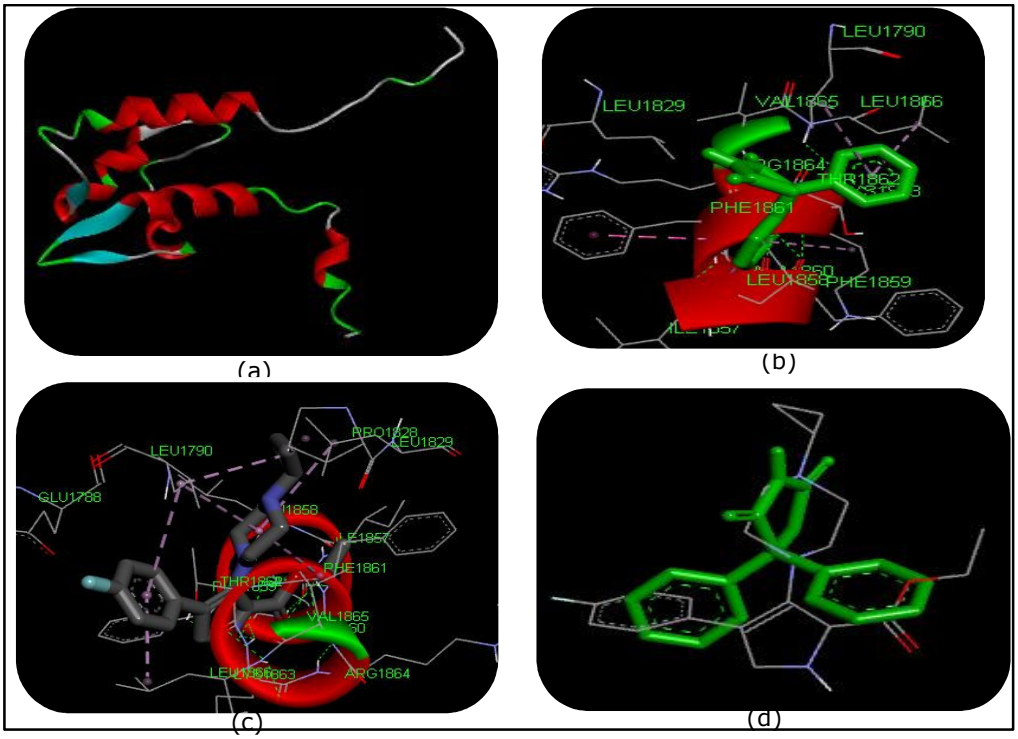


Figure 7: Structure and interaction of ligand with voltage-gated sodium channel (a) structure of 2KaV model 1 (b) H-bond and hydrophobic interaction between phenyltoin and 2KaV (c) H-bond and hydrophobic interaction between compound 101 m and 2 KaV (d) compound 101 m superimposed on phenyltoin structure.

From the table, it was observed that all the designed compounds had binding affinity lower than that of the template compound 61 (-4.70 Kcal/mol). These corroborate the claim that the designed compounds had improved anti-MES activity. However, when compared to phenytoin (a molecule known for its activity against MES induced seizure), their binding affinities were slightly higher than that of phenytoin (-6.60 Kcal/mol) but comparable. Furthermore, most of the compounds formed hydrogen bond with LEU1858, ASP1856, ILE1857 and PRO1828 amino acids of the 2KaV coil structure while phenytoin formed hydrogen bond with VAL1865 suggesting these compounds may be acting as voltage-gated sodium channel blocker with different mechanism. However, all the compounds including phenytoin had Pi-Alkyl and Alkyl-Alkyl hydrophobic interactions with the target. The green-colored broken lines in Figure 7 a and b represent hydrogen bonds between the target (2 KaV) and the ligands *i.e.* phenytoin and compounds 101 m respectively while, the purple-colored broken lines represent the hydrophobic interactions. Figure 7d represented an attempt to superimpose compounds 101m on phenytoin in the docked configuration, however, it was noticed that the two compounds cannot be perfectly superimposed.

Table 7: Binding affinity of designed compounds with voltage-gated sodium channel.

Compounds	BA(Kcal/mol)	Hydrogen bond	Hydrophobic interaction	
		Amino acid	Amino acid	Interaction
11 m	-6.00	ASP1856 and PRO1828	LEU1790, VAL1865 and LEU1866	Alkyl-Alkyl and Pi-Alkyl
16 m	-5.80	ILE1857	LEU1790, LEU1866,VAL1865, ALA1860 and LYS1863	Pi-Alkyl and Alkyl-Alkyl
22 m	-6.00	LEU1858 and ILE1857	LEU1790, LEU1866,VAL1865, ALA1860 and LYS1863	Pi-Alkyl and Alkyl-Alkyl
23 m	-6.00	LEU1858 and ILE1857	LEU1790, LEU1866,VAL1865, ALA1860 and LYS1863	Pi-Alkyl and Alkyl-Alkyl
24 m	-5.70	LEU1858 and ILE1857	LEU1790, LEU1866,VAL1865, ALA1860 and LYS1863	Pi-Alkyl and Alkyl-Alkyl
45 m	-6.20	LEU1858	LEU1790, LEU1829, LEU1790, PRO1828, LEU1866, VAL1865 and LYS1863	Pi-Alkyl and Alkyl-Alkyl
57 m	-5.70	ILE1857	LEU1790, VAL1865 and LEU1866	Pi-Alkyl and Alkyl-Alkyl
58 m	-5.60	LEU1858 and ILE1857	LEU1790, VAL1865 and LEU1866	Pi-Alkyl and Alkyl-Alkyl
59 m	-5.90	LEU1858 and ILE1857	LEU1790, VAL1865 and LEU1866	Pi-Alkyl and Alkyl-Alkyl
81 m	-5.50	LEU1858 and ILE1857	LEU1790, LEU1866 and LYS1863	Pi-Alkyl
101 m	-6.00	LEU1858 and ASP1856	LEU170, PRO1828, LEU1829, VAL1865 and LEU1866	Pi-Alkyl and Alkyl-Alkyl
Compound 61	-4.70	PHE1861	LEU1829, VAL1865 and LEU1866	Pi-Alkyl and Alkyl-Alkyl
Phenyltoin	-6.60	VAL1865	PHE1861, LEU1790, LEU1866, ALA1860 and LYS1863	Pi-Alkyl and Alkyl-Alkyl

BA is the binding affinity of the ligands to the receptor

CONCLUSION

The activity of 1H-pyrazolo [3, 4-d]pyrimidine, 1H-pyrazole-5-carboxylic and hydrazine carboxamide derivatives against maximal electroshock-induced seizure has been quantitatively analyzed in terms of chemometric descriptors. The statistically validated QSAR model obtained provided rationales to explain the anticonvulsant activities of these classes of compounds. The descriptors identified through GFA-MLR analysis have highlighted the role of the molecular dipole moment (μ), energy of the lowest unoccupied molecular orbital (ϵLUMO), polar surface area (PSA), accessible area of the molecule obtained from wave function calculation (WAA), sum of the absolute value of the charges on all atom in a molecule (QA) and sum of the square of the charges on all fluorine atom in a molecule (QF) in explaining the variation in anti-MES activities of the used compounds. And it was observed that for a compound to be more potent, the higher values of descriptors ϵLUMO, PSA, WAA and QF and lower values of descriptors μ and QA are conducive. The statistics that emerged from the test-set validated the obtained model. Applicability domain analysis revealed that the obtained model have acceptable predictability except one for influential point (compound 13, Table 4), all the compounds remained within the applicability domain of the

proposed models and were evaluated correctly. Few new compounds having better anti-MES activity than highest active 1H-pyrazole-5-carboxylic derivative (compound 61), have been suggested for further exploration. The binding affinity of these newly suggested compounds to voltage-gated sodium channel (PBD code: 2 KaV) were found to be better than that of compounds 61 and comparable to that of phenytoin.

REFERENCES

1. McCormick D A, Contreras D. On the cellular and network bases of epileptic seizures Annu. Rev. Physiol. 2001 March; 63: 815-846. DOI: 10.1146/annurev.physiol.63.1.815

2. Duncan JS, Sander JW, Sisodiya SM, Walker MC. Adult epilepsy. Lancet. 2006 Apil; **367**: 1087–100. doi:10.1016/S0140-6736(06)68477-8

3. Gavernet L, Barrios IA, Sella CM, Bruno-Blanch LE. Design, synthesis, and anticonvulsant activity of some sulfamides. Bioorganic & Medicinal Chemistry. 2007; 15:5604–5614. http://dx.doi.org/10.1016/j.bmc.2007.05.024.

4. Estrada E, Pena A. In Silico studies for the rational discovery of anticonvulsant compounds. Bioorg. Med.Chem. 2000 December; 8:2755- 2770. DOI: 10.1016/S0968-0896(00)00204-2.

5. Reijs R, Aldenkamp AP, De Krom M. Mood effects of antiepilepsy drug. Epilepsy Behav. 2004 Febuary; 5(1):66-76. DOI: http://dx.doi.org/10.1016/j.yebeh.2003.11.009

6. Bialer M, Johannessen SI, Kupferberg HJ, Levy RH, Perucca E, Tomson T. Progress report on new antiepileptic drugs: a summary of the Seventh Eilat Conference (EILAT VII). Epilepsy Res. 2004 September-October; 61(1-3) 1-48. DOI: 10.1016/j.eplepsyres.2004.07.010

7. Edafiogho IO, Ananthalakshmi KV, Kombian SB. Anticonvulsant evaluation and mechanism of action of benzylamino enaminones. Bioorg. Med. Chem. 2006 March 14(150):15566-5272. Dio:10.1016/j.bmc.2006.03.049

8. Eddington ND, Cox DS, Khurana M, Salama NN, Stables JP, Harrison SJ, Negussie A, Taylor RS, Tran UQ, Moore JA. Synthesis and anticonvulsant activity of enaminones. Part 7. Synthesis and anticonvulsant evaluation of ethyl 4-[(substituted phenyl)amino]-6-methyl-2-oxocyclohex-3-ene-1-carboxylates and their corresponding 5-methylcyclohex-2-enone derivatives. Eur J Med Chem. 2003 March; 38: 49–64. Doi:10.1016/S0223-5234(02)00006-5

9. Malawska B, Kulig K, S'Piewaka A, Stables J. Investigation into new anticonvulsant derivatives of α-substituted N-benzylamides of γ-hydroxy- and γ-acetoxybutyric acid. Part5. Search for new anticonvulsant compounds. Bioorg Med Chem. 2004 Febuary; 12: 625-632. DOI: 10.1016/j.bmc.2003.10.036

10. White HS. Comparative anticonvulsant and mechanistic profile of the established and newer antiepileptic drugs. Epilepsia. 1999; 40 (Suppl.): S2-10. https://www.ncbi.nlm.nih.gov/pubmed/ 10530688

11. Rogawski MA, Porter RJ. Antiepileptic drugs: pharmacological mechanisms and clinical efficacy with consideration of promising developmental stage compounds. Pharmacologica Reviews. 1990 September; 42(3): 223-286.

12. Leppik IE. Antiepileptic drugs in development: prospects for the near future. Epilepsia 1994 August, 35, 29-40. DOI: 10.1111/j.1528-1157.1994.tb05953.

13. Perucca E. The new generation of antiepileptic drugs: advantages and disadvantages Br. J. Clin. Pharmacol. 1996 November; 42: 531-543. DOI: 10.1111/j.1365-2125.1996.tb00107.x

14. Loscher W, Schmidt D. Strategies in antiepileptic drug development: is rational drug design superior to random screening and structural variation? Epilepsy Res. 1994, 17(2): 95-134. https://www.ncbi.nlm.nih.gov/pubmed/8194514

15. Ibezim EC, Duchowicz PR, Ibezim NE, Mullen LMA, Onyishi IV, Brown SA, Castro EA. Computer - aided linear modeling employing QSAR for drug discovery. Scientific Research and Essay. 2009 December;4(13):1559-1564. http://www.academicjournals.org/SRE

16. Cox B, Nobbs MS, Shah GP, Edney DD, Loft MS. Pyrazine compounds- Patent WO9838174. Exp.Opin.Ther.Patents. 1999; 9(10):1431-1436 http://dx.doi.org/10.1517/13543776.9.10.1431

17. Bhutoria S, Ghoshal N. A Novel Approach for the Identification of Selective Anticonvulsants Based on Differential Molecular Properties for TBPS Displacement and Anticonvulsant Activity: An Integrated QSAR Modelling, QSAR Comb. Sci. 2008 May; 27(7): 876 – 889. DOI: 10.1002/qsar.20086000

18. Macchiarulo A, De luca L, Costantino et al. QSAR study of anticonvulsant negative allosteric modulator of the AMPA receptor. J med. chem. 2004 March; 47:1860-1863. doi/abs/10.1021/jm0310838

19. Abdulfatai U, Uzairu A, Uba S. Quantitative structure activity relationship study of anticonvulsant activity of α_substituted acetamido-N-benzylacetamide derivatives, Cogent Chemistry. 2016 April ;(2):1-12. http://dx.doi.org/10.1080/23312009.2016.1166538

20. Lesk AM. Introduction to bioinformatics. Oxford University Press; 2002. 207p. ISBN: (pbk)0199251967

21. Pedro JB, John BOM. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. Bioinformatics. 2010 May; 26(9): 1169-1175. doi:10.1093/bioinformatics/btq112

22. Kornet MJ, Chu JYR. Synthesis and anticonvulsant testing of 4-phenylsemicarbazides J. Pharm. Sci. 1983 October; 72:1213-1215. DOI: 10.1002/chin.198414170

23. Pajouhesh H, Parson R, Popp FD. Potential Anticonvulsants VI: Condensation of Isatins with Cyclohexanone and other Cyclic Ketones. J. Pharm. Sci. 1983 March; 72:318-321. doi: 10.1002/jps.2600720330

24. Popp, F. D.; Pajouhesh, H. Potential anticonvulsants IV: Condensation of isatin with benzoylacetone and isopropyl methyl ketone. J. Pharm. Sci. 1982 September;71:1052-1054. DOI: http://dx.doi.org/10.1002/jps.2600710924

25. Haj-Yehia A, Bialer M. Structure-pharmacokinetic relationships in a series of short fatty acid amides that possess anticonvulsant activity. J. Pharm. Sci. 1990 August, 79, 719-724.DOI: http://dx.doi.org/10.1002/jps.2600790814

26. Hernandez-Gallegos Z, Lebmann PA. Partition Coefficients of three new anticonvulsants. J. Pharm. Sci. 1990 November; 79:1032-1033.DOI : http://dx.doi.org/10.1002/jps.2600791118

27. Crider AM, Kolczynski TM, Miskell DL. Synthesis and anticonvulsant activity of racemic 2-amino-N-substituted succinimide derivatives. J. Pharm. Sci. 1981 Febuary, 70:192-195. DOI: http://dx.doi.org/10.1002/jps.2600700220

28. Darling CM, Pryor P. Anticonvulsant activity of alkyl-substituted *N*-benzylcyanoacetamides J. Pharm. Sci. 1979 January; 68, 108-110. DOI:10.1002/jps.2600680137

29. Kelley JL, Linn JA, Bankston DD, Burchall CJ, Soroko FE, Cooper BR. 8-Amino-3-benzyl-1,2,4-triazolo[4,3-a]pyrazines. Synthesis and anticonvulsant activity. J. Med. Chem. 1995 September; 38: 3676-3679.DOI: 10.1021/jm00018a029

30. Kelley JL, Davis RG, McLean EW, Glen RC, Soroko FE, Cooper BR. Synthesis and anticonvulsant activity of N-Benzylpyrrolo[2,3-d]-, -pyrazolo[3,4-d]-, and -triazolo[4,5-d]pyrimidines: imidazole ring-modified analogs

of 9-(2-Fluorobenzyl)-6-(methylamino)-9H-purine. *J. Med. Chem.*1995 September; *38* (19): 3884–3888.DOI: 10.1021/jm00019a019

31. Kelley JL, Koble CS, Davis RG, McLean EW, Soroko FE, Cooper BR. l-(Fluorobenzyl)-4-amino-1H 1,2,3-triazolo[4,5-c]pyridines:Synthesis and Anticonvulsant ActivityJ. Med. Chem. *1995 September*;38 (20): 4131–4134. DOI: 10.1021/jm00020a030

32. Dimmock JR, Puthucode RN, Smith JM, Hetherington M, Quail JW, Pugazhenthi U, Lechier T, Stables JP.(Aryloxy)aryl semicarbazones and related compounds: a novel class of anticonvulsant agents possessing high activity in the maximal electroshock screen. J Med Chem. 1996 September; 39(20):3984-97. DOI:10.1021/jm9603025

33. Puthucode RN, Pugazhenthi U, Quail JW, Stables JP, Dimmock JR. Anticonvulsant activity of various aryl, arylidene and aryloxyaryl semicarbazones Eur. J. Med. Chem. 1998 July-August; 33: 595-607. http://dx.doi.org/10.1016/S0223-5234(98)80018-4

34. Unverferth K, Engel J, Hofgen N, Rostock A, Gunther R, Lankau HJ, Menzer M, Rolfs A, Liebscher J, Mu¨ ller B, et al. Synthesis, anticonvulsant activity, and structure-activity relationships of sodium channel blocking 3-aminopyrroles. J Med Chem.1998 January; 41:63–73. DOI: 10.1021/jm970327j

35. Shao Y, Molnar LF, Jung Y, Kussmann J, Ochsenfeld C, Brown ST, Gilbert ATB, Slipchenko LV, Levchenko SV, O'Neill DP, DiStasio RA, Lochan RC, Wang T, Beran GJO, Besley NA, Herbert JM, Lin CY, Van Voorhis T, Chien SH, Sodt A, Steele RP, Rassolov VA, Maslen PE, Korambath PP, Adamson RD, Austin B, Baker J, Byrd EFC, Dachsel H, Doerksen RH, Dreuw A, Dunietz BD, Dutoi AD, Furlani TR, Gwaltney SR, Heyden A, Hirata S, Hsu CP, Kedziora G, Khalliulin RZ, Klunzinger P, Lee AM, Lee MS, Liang WZ, Lotan I, Nair N, Peters B, Proynov EI, Pieniazek PA, Rhee YM, Ritchie J, Rosta E, Sherril CD, Simmonett AC, Subotnik JE, Woodcock III HL, Zhang W, Bell AT, Chakraborty AK, Chipman DM, Keil FJ, Warshel A, Hehre WJ, Schaefer HF, Kong J, Krylov AI, Gill PMW, Head-Gordon M. Advances in methods and algorithms in modern quantum chemistry program package. Phys. Chem. Chem. Phys., 2006 June; 8, 3172. DOI: 10.1039/B517914A

36. Singh P. Quantitative structure-activity relationship study of substituted-[1,2,4] oxadiazoles as $S1P_1$ agonists. Journal of Current Chemical and Pharmaceutical Sciences. 2013 Febuary; 3(1):64-67. www.sadgurupublications.com

37. Ambur P, Aher RB, Gajewicz A, Pyzyt WRK, "NanoBRIDGES" software:open acess tools to perform QSAR and nano-QSAR modeling. Chem. Intel. Lab. Syst, 2015 Octomber; 147, 1-13. http://dx.doi.org/10.1016/j.chemolab.2015.07.007

38. Kennard RW, Stone LA. Computer aided design of experiments. Technometrics. 1969 Febuary; 11:137–148. doi:10.2307/1266770

39. Melagraki G, Afantitis A, Makridima K, Sarimveis H, Igglessi- Markopoulou O. Prediction of toxicity using a novel RBF neural network training methodology. J Mol Model. 2006 Febuary; 12:297–305. doi:10.1007/s00894-005-0032-8

40. Afantitis A, Melagraki G, Sarimveis H, Koutentis PA, Markopoulos J, Igglessi-Markopoulou O. A novel QSAR model forpredicting induction of apoptosis by 4-aryl-4H-chromenes. Bioorg Med Chem. 2006 May; 14:6686–6694. doi:10.1016/j.bmc.2006.05.061

41. Chakraborti AK, Gopalakrishnan B, Sobhia ME, Malde A. 3D-QSAR studies of indole derivatives as phosphodiesterase IV inhibitors. Eur J Med Chem. 2003 September; 38:975–982. doi:10.1016/j.ejmech. 2003.09.001

42. Wu W, Walczak B, Massart DL,Heuerding S, Erni F, Last IR, Prebble KA. Artificial neural networks in classification of NIR spectral data: design of the training set. Chemom Intell Lab Syst. 1996 May; 33:35–46. doi:10.1016/0169-7439(95)00077-1

43. Rogers D, Hopfinger AJ. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. J.Chem. Inf. Comp. Sci.1994 July; 34, 854-866. DOI: 10.1021/ci00020a020

44. Allison PD. Multiple regression: a primer. London: Pine Forge Press; 1999. 10-147p. *ISBN-10*: 0761985336

45. Larose TD. Data mining methods and models. Wiley Interscience; 2002.114p. *ISBN*-13 978-0-471-66656-1.

46. Golbraikh A, Tropsha A. Beware of q2!. J Mol Graphics Mod 2002 January; 20, 269–276. http://dx.doi.org/10.1016/S1093-3263(01)00123-1

47. Roy K, Mitra I, Kar S, Ojha PK, Das RN, Kabir H. Comparative studies on some metrics for external validation of QSPR models. J Chem Inf Model. 2012 Febuary; 52:396–408. DOI:10.1021/ci200520g

48. Tropsha A, Gramatica P, Gombar VK. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb Sci. 2003 Apil; 22:69–77. doi:10.1002/qsar.200390007

49. Minovski N, Zuperl S, Drgan V, Novi M. Assessment of applicability domain for multivariate counter-propagation artificial neural network predictive models by minimum Euclidean distance space analysis: A case study. Analytica Chimica Acta. 2013 January;759:28–42. DOI: 10.1016/j.aca.2012.11.002

50. Netzeva TI, Worth AP, Aldenberg T, Benigni R, Cronin MTD, Gramatica P, Jaworska JS, Kahn S, Klopman G, Marchant CA, Myatt G, Nikolova-Jeliazkova N, Patlewicz GY, Perkins R, Roberts DW, Schultz TW, Stanton DT, Van De Sandt JJM, Tong W, Veith G, Yang C. Current status of methods for defining the applicability domain of (quantitative) structure–activity relationships. Alternative to Laboratory Animal. 2005 April; 33:155–173. http://www.atla.org.uk

51. Topliss JG, Costello RJ. Chance correlations in structure– activity studies using multiple regression analysis. J Med Chem. 1972 Octomber; 15, 1066–1068 DOI: 10.1021/jm00280a017

52. Motta LF, Gaudio AC, Takahata Y. Quantitative Structure–Activity Relationships of a Series of Chalcone Derivatives (1,3–Diphenyl–2–propen–1–one) as Anti *Plasmodium falciparum* Agents (Anti Malaria Agents), *Internet Electron. J. Mol. Des.* 2006, *5*, 555–569, http://www.biochempress.com.

53. David AB, Edwin K, Roy EW. Regression diagnostics: indentifying influential data and source of collinearity. Wiley-interscience: John Wiley & Sons, lnc., Hoboken, New Jersey, 2004. 85-93p. ISBN 0-47 1-69 1 17-8

54. Belsley DA. Regression diagnostics: identifying influential data and sources of collinearity. Hoboken, N.J: Wiley-Interscience; 2004. 292 p. (Wiley series in probability and statistics).

55. Shapiro S, Guggenheim B. Inhibition of oral bacteria by phenolic compounds. Part 1 QSAR analysis using molecular connectivity. Quant. Struct.-Act. Relat. 1998 August; 17, 327–33. DOI: 10.1002/(SICI)1521-3838(199808)17:04<327::AID-QSAR327>3.0.CO;2-O

56. Noolvi MN, Patel HM, Bhardwaj V. 2D qsar studies on a series of 4-anilino quinazoline derivatives as tyrosine kinase (egfr) inhibitor: an approach to design anticancer agents. Digest Journal of Nonmaterial and Biostructures. 2010 April; 5(20):387–401. http/:www.researchgate.net/262187015

57. Habibi-Yangjeh A, Danandeh-Jenagharad M. Application of a genetic algorithm and an artificial neural network for global prediction of the toxicity of phenols to Tetrahymena pyriformis. Monatsh Chemistry 2009 October; 140, 1279–1288. doi: 10.1007/s00706-009-0185-8

58. Tropsha A. Best practices for QSAR model development, validation, and exploitation. Mol Inf. 2010 July; 29, 476–488. DOI: 10.1002/minf.201000061

59. Warren H, Sean O. Spartan 14 for windows, machintosh and linux tutorial and user's Guide. Wavefunction Incoorperated: Von Karman Avenue. 2014.289-291p. ISBN: 978-1-890661-42-2

60. Stenberg P, Luthman K, Ellens H, Lee CP, Smith PL, Lago A, Elliot JD, Artursson P. Prediction of the intestinal absorption of endothelin receptor antagonists using three theoretical methods of increasing complexity. Pharm. Res. 1999 Octomber; 16:1520–1526. doi:10.1023/A:1015092201811

61. Clark, DE. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption. J. Pharm. Sci. 1999 August; 88:807–816. DOI:10.1021/js980402t

62. Franke, R. Theoretical Drug Design Methods. Elsevier: Amsterdam. 1984.115-123p.

63. Osmialowski K, Halkiewicz J, Radecki A, Kaliszan, R. Quantum chemical parameters in correlation analysis of gas liquid chromatographic retention indices of amines *J. Chromatography.* 1985 December, *346*, 53-60. DOI: 10.1016/S0021-9673(00)90493-X

64. Fleming, I. *Frontier Orbitals and Organic Chemical Reactions*. Somerset New Jersey: John Wiley and Sons; 1976. 308p. *ISBN* 10: 0471018198

65. Todeschini R, Consonni V, Mannhold R. Handbook of molecular descriptors. Wiley-VCH, Weinheim; 2000. 1-688p. ISBN 3-52-29913-0

66. Cartier A, Rivail, JL. Electronic Descriptors in Quantitative Structure-Activity Relationship. *Chemom. Intell. Lab. Sys.* 1987 October, *1*, 335-347. DOI: 10.1016/0169-7439(87)80039-4

67. Bodor N, Gabanyi Z, Wong CK. A new method for the estimation of partition coefficient *J. Am. Chem. Soc.* 1989 May; *111*: 3783-3786. DOI: 10.1021/ja00193a003

68. Buydens L, Massart D, Geerlings P. Prediction of gas chromatographic retention indexes with topological, physicochemical, and quantum chemical parameters Anal. Chem. 1983 April, 55, 738-744. DOI: 10.1021/ac00255a034

69. Klopman G, Iroff LD. Calculation of partition coefficients by the charge density method. J. *Comput. Chem.* 1981 June; *2*: 157-160 DOI: 10.1002/jcc.540020204

70. Ruslin N, Syamsudin. Quantitative Structure-Activity Relationship Analysis of Antimalarial Compound of Mangostin Derivatives Using Regression Linear Approach. Asian Journal of Chemistry. 2013 September; 25(11): 6136-6140. https://www.researchgate.net/.../260600323

71. Clare, B. W.; Supuran, C. T. Carbonic anhydrase activators. 3: structure-activity correlations for a series of isozyme II activators. *J. Pharm. Sci.* 1994 June; *83*:768. DOI: http://dx.doi.org/10.1002/jps.2600830603

72. Ordorica MAV, Velázquez MLM, Ordorica JGV, Escobar JLV, Lehmann PAF. A Principal Component and Cluster Significance Analysis of the Antiparasitic Potency of Praziquantel and some Analogues. Quant. Struct.-Act. Relat. 1993; 12: 246–250. doi:10.1002/qsar.19930120305

73. Milouser VZ, Leven JA, Arbing MA, Hunt JF, Put GS, Palmer AG. Solution structure of the NaV1.2 C-terminal EF-hand domain. J boil chemistry. 2009; 284: 6646-6654. DOI: 10.1074/jbc.M807401200

74. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, Journal of Computational Chemistry. 2010 January; 31: 455-461 Doi 10.1002/jcc-21334.