



IN-SILICO STUDY FOR PREDICTING THE INHIBITION CONCENTRATION OF SOME HETEROCYCLIC AND PHENYLIC COMPOUNDS AS POTENT HERBICIDES USING THE MLR - GFA APPROACH

ESTUDO IN-SILICO PARA PREDIÇÃO DA CONCENTRAÇÃO DE INIBIÇÃO DE ALGUNS COMPOSTOS HETEROCÍCLICOS E FENÍLICOS COMO HERBICIDAS POTENTES UTILIZANDO A ABORDAGEM MLR-GFA

S. TUKUR^{1,*}, G.A. SHALLANGWA¹, A. UZAIRU¹ and A. IBRAHIM¹

Ahmadu Bello University, Department of Chemistry, Zaria, Kaduna State, Nigeria

*Corresponding author. Ahmadu Bello University, Department of Chemistry, Zaria, Kaduna State, Nigeria, Phone: +2348139475967 e-mail address: saidutukur@gmail.com (S. TUKUR).

ARTICLE INFO

Article history:

Received 2018-12-20

Accepted 2019-01-29

Available online 2019-03-08

Keywords

Herbicide

QSAR

Multiple Linear Regression (MLR)

Genetic Function Algorithm (GFA)

Applicability Domain

Y - Randomization

ABSTRACT

The study of the quantitative structure-activity relationship (QSAR) was used in a set of data from 43 heterocyclic and phenylic inhibitor compounds in order to establish a correlation between the inhibitory concentrations of the compounds in question and their structures. The optimization method of the density functional theory (DFT) was used to minimize the energy of the 3D structures using the Becke functional hybrid Exchange (B3) parameter with the Lee, Yang, and Parr Functional Correlation (LYP), commonly called the B3LYP functional Hybrid and 6-31G* Basis Set (B3LYP/6-31G*) method, to discover their molecular Quantum descriptors. Five models of QSAR were generated with the technique of genetic function algorithm (GFA). Among the five models generated, model 1 was selected as the best model because of its statistical significance (Friedman's LOF = 0.3008, $R^2 = 0.9784$, $R^2_{adj} = 0.9739$, $Q_{cv}^2 = 0.9675$ and $R^2_{pred} = 0.7348$). The meticulous model was evaluated by means of the Leave One out cross-validation (LOO-CV) approach, external validation of the compounds of the test set, Y-randomization test and applicability domain (Williams Plot). The proposed QSAR model was highly predictive and vigorous with good validation parameters. The molecular descriptors used in the model should be considered of great importance in improving the inhibitory concentrations of the herbicides and also in the conception of new herbicides with a higher concentration of inhibitor.

RESUMO

O estudo da relação quantitativa estrutura-atividade (QSAR) foi utilizado em um conjunto de dados de 43 compostos de inibidores heterocíclicos e fenílicos, a fim de estabelecer uma correlação entre as concentrações inibitórias dos compostos em questão e suas Estruturas. O método de otimização da teoria funcional da densidade (DFT) foi utilizado para minimizar a energia das estruturas 3D utilizando o parâmetro híbrido funcional de Becke (B3) com a correlação funcional de Lee, Yang e Parr (LYP), comumente denominada B3LYP funcional híbrido e 6-31G * base Set (B3LYP/6-31G *) método, para descobrir seus descritores Quantum molecular. Cinco modelos de QSAR foram gerados com a técnica de algoritmo de função genética (GFA). Entre os cinco modelos gerados, o modelo 1 foi selecionado como o melhor modelo por causa de sua significância estatística (LOF de Friedman = 0,3008, $R^2 = 0,9784$, $R^2_{adj} = 0,9739$, $Q_{cv}^2 = 0,9675$ e $R^2_{pred} = 0,7348$). O modelo meticuloso foi avaliado por meio da abordagem Leave One out Cross-Validation (LOO-CV), validação externa dos compostos do conjunto de teste, teste de randomização Y e domínio de aplicabilidade (Williams Plot). O modelo QSAR proposto foi altamente preditivo e vigoroso

com bons parâmetros de validação. Os descritores moleculares utilizados no modelo devem ser considerados de grande importância na melhoria das concentrações inibitórias dos herbicidas e também na concepção de novos herbicidas com maior concentração de inibidores.

1. INTRODUCTION

Herbicides are chemical compounds that inhibit different metabolic processes in plants. They are extensively used in agricultural purposes as an agent of destruction against undesirable crops (weeds). Weeds usually struggle with other species of crops for nutrients, sunlight, and water. In addition, weeds are capable of influencing plant growth, causes a decline in crop production and quality (Prasad et al., 2011). However, weeds are usually the causative agent of most pathogens, pests, and viruses, which can contribute extremely to the blowout of diseases in plant and insect pests. The herbicides being the safest and the most effective way to control weed can play a vital role in mechanized farming. The fortification of crops continually needs the discovery of new innovative herbicides. However, mechanized and large-scale farming is attainable with the help of herbicides. The herbicides play a reliable role in the unified management of weeds (Funar et al., 2017).

A large species of herbicide was developed early from the structure to the genre of actions (Troyer et al., 2001). The first commercial herbicide that was discovered at the beginning of 1948 was 2,4-dichlorophenoxyacetic acid (2,4-D) by Zimmerman (Zimmerman et al., 1948). In addition, the tendency in its structure and chemical activity was used in the designation of new inhibitors (Liu et al., 2013). The herbicides perform their function by influencing the electron transport system of the plants during the process of photosynthesis (Dayan et al., 2012). The slowdown of the electron flow to participate in the binding of the plastoquinone and the eccentricity of the electron in the photosystem I (PSI) was of great significance in photosynthesis. The relocation of electron from the PSI to its *QB* binding site of the D1 subunit of photosystem II (PSII) is succeeded with the help of the plastoquinone. A lot of classes of herbicides (e.g. urea, triazine pyridines etc.) can be destined to the undiscerning PSII *QB* binding site. PSII plays a vital role in photosynthesis and also in plant growth (Zhang et al. 2014). *QB* is a complex proteins sheath, which captures light from the sun and finishes the light oxidation of water to atmospheric oxygen. A sequence of light-inspired electron transfer reactions proceeds in the D1 and D2 subunit of PSII. The *QB* binding site of D1 protein is the finishing point of the electron transfer process in the PSII and is also the livelihood target of most herbicides. Herbicides always displaces the plastoquinone from its *QB* binding position (Pfister et al., 1979).

The quantitative structure-activity relationship (QSAR) is a technique used in the establishment of mathematical or computational models that strive to establish a good correlation between molecular structures (descriptors) and properties (concentrations) using a Chemometric method (Cruz et al., 2014). However, in the context of the molecular design, the word "structure" refers to molecular descriptor of the compounds (Verma et al., 2010). While, the Chemometric method employs artificial neural networks (ANN), principal component regression (PCR), Principal Component Analysis (PCA), multiple linear regression (MLR), partial least squares (PLS), genetic function algorithm (GFA) etc. However, several approaches to the QSAR studies have been established over a period of about a hundred years ago and have been abetted as a good predictive measure, particularly in the design of new

narcotics drugs and in the synthesis of new agrochemicals. These QSAR methods were established by Hansch and free-Wilson (Takač et al., 1991) starting from 1D, 2D linear free-energy correlation, 3D QSAR of Crammer, 4D. Hopfinger, 5D Vedani and 6D (Hansch et al., 1963). The 1D, 2D and other similar methods were generally referred to as "Classic Methods". The classical method was succeeded with the approximation of the genetic function algorithm (GFA) (Rasulev et al., 2005). The GFA technique does not only select the integer number of descriptors in the regression study but, also introduces multiple linear regression (MLR) on the use of straight-line, higher order polynomials, splines and Gaussian (Cho et al., 2001, 18). The algorithm technique was used to select the best descriptors in the regression equation. The genetic function algorithm could help as a valuable procedure for probing the large probability space with a large number of descriptors for a small number of molecules. The essence of this research was to develop a predictive QSAR model using 30 compounds as training set. However, if the developed QSAR models were reasonable enough, it may be use to predict the inhibition concentration of the unknown compounds. In addition, the prosperous QSAR models will certainly decrease the number of compounds to be synthesized more especially in pharmaceutical laboratories, thereby, making it an easy and affordable technique to reduce cost and time duration with regards to developing new drugs and agrochemicals.

2. MATERIALS AND METHOD

2.1 Experimental Dataset

43 compounds were used in this study as data set, comprising a mixture of Sulfonyl Urea, Triazines, Benzoate, Acetamides, Pyridines, etc. The inhibition concentration of these compounds has been reported in the literature (Gand et al., 2015). The dataset, along with its chemical abstract service record number (CASRN), IUPAC names, common names, and various physicochemical properties has been selected from a simplified molecular input line String file system Commonly abbreviated as SMILES. The result of the inhibitory concentration was expressed in pLC_{50} ($pLC_{50} = -\log(LC_{50})$) which is the concentration at 50% maximum inhibition.

2.2 Geometry Optimization

The procedure of calculating the equilibrium geometry or lowest energy conformation of a molecule is called molecular optimization (Abdulfatai et al., 2015). The 2D structures of the compounds were drawn with the help of ChemDraw Ultra version 12.0 software and the drawn compounds were saved in a format called CDX file format (Ibrahim et al., 2018). The 3D conformations of the compounds were optimize using a programmed software package called Spartan 14.0 VI. 1.4 wave function (Arthur et al., 2016). The molecular structures were in the first place pre-optimized using the Semiempirical method purposely to reduce the tension in the molecule before submitting the whole molecules to DFT (density functional theory) exploring the Becke three parameters exchange functional (B3) hybrid with Lee, Yang and Parr Functional Correlation (LYP), usually abbreviated as B3LYP (Lee et al.,

1988) and by the use of 6-31G* basis set for complete optimization of the 3D structures. The optimized spartan files of all the molecules were then saved in a format called the sdf file format, which is the only input entry format that can be recognized by the Padel-Descriptor V 2.20 software (Adeniji et al. 2018).

2.3 Molecular Descriptor Calculations

Molecular descriptor is a mathematical value or figure that describe properties of molecules acquired from an accurate algorithm or experimental procedure (Olasupo et al., 2017). Quantum chemical descriptors were calculated using the optimization software of Quantum chemistry known as Spartan '14 version 1.1.2 (Abdulfatai et al., 2017). The descriptors of the various dimensions (1D, 2D, and 3D) were calculated with the help of Padel software version 2.18. The total molecular descriptors of 1875 (1444 1D, 2D, and 431 3D) were generated from the Padel descriptor software and the Spartan "14" Software (Arthur et al., 2016).

2.4 Data Normalization

The descriptors obtained from the Padel software were unvarying by a procedure that makes use of range; Maximum and minimum with the dispersion of the molecular descriptors using standard deviation or variance. Thus, the sprincling of the descriptors was achieved before the data were transformed over into an N distribution (0, 1, 2... N). This procedure enables the correlation between descriptors much less redundant (Panchal et al., 2013).

2.5 Data Pretreatment

The descriptors generated from the PaDEL software were pre-treated using software popularly known by the pharmacist as DTC LAB, meaning; Drug Theoretical Cheminformatics Laboratory. However, since a large number of descriptors have been generated, it is therefore important to employ an applicable technique to eliminate the less important descriptors. Molecular descriptors showing almost equal values were detached from all ensembles and the molecular descriptors having a very low variance were also ignored in order to decrease the Loch descriptor. The descriptors showing a small level of intercorrelation were retained (Roy et al., 2015).

2.6 Generation of Training and Test Set

In the development of a validated QSAR model, the data set of 43 compounds were divided according to Kennard and Stones into 70% of training and 30% of test set using information division software obtained from the Drug Theoretical Cheminformatics Laboratory (DTC LAB.) (Kennard and Stone 1969). The Kennard and Stone algorithm existed with a prodigious success in the recent QSAR studies and remained a valuable tool to build the both training and test sets compounds. 70% of the entire data set were used in the development of the QSAR model. The remaining 30% of the data set were used to validate the developed QSAR model externally.

2.7 Relative Importance of Each Descriptor to the Model

The definitive meaning of the mean effect (ME) of each descriptor was used as a good measure to evaluate the absolute prominence and impact of each descriptor to the model. The mean effect can be calculated using Equation 1, given below as:

$$ME = \frac{\beta_j \sum_i^n D_j}{\sum_j^m (\beta_j \sum_i^n D_j)} \quad (1)$$

Where, ME is the mean effect of descriptor J in the developed QSAR model, while β_j , is the coefficient of descriptor J , D_j is the value of each descriptor in the data set for each molecule in the training compounds, m is the number of descriptors that appear in the model and n is the number of training set compounds (Minovski et al., 2013)

2.8 Degree of Contribution of Selected Descriptors

The significant impact of each designated descriptor is to take into account the standardized regression coefficients (b_j^s) as the factor that must be considered in measuring the degree of impact of each descriptor to the model. The standardized regression coefficients b_j^s can be calculated using Equation 2, given by the expression below:

$$b_j^s = \frac{S_j b_j}{S_y} \quad (2)$$

Where b_j is the regression coefficient of descriptor j . S_j and S_y are the standard deviations of each descriptor and concentration respectively. Emphasis was placed on certain descriptors that reveal a larger standardized coefficient value.

2.9 Model development

The multiple linear regression (MLR) study was performed with the aid of the genetic function algorithm (GFA) that was incorporated in the material studio software purposely to discover the ideal number of the descriptor in the model. Genetic function algorithm is a computer-aided descriptor calculation method that has been stirred from engineering and computer scientist (Saidi et al., 2016). GFA not only inevitably selects the prime number of descriptors in the regression analysis but, also creates a multiple linear regression models (MLR) by the use of linear, higher order polynomials, splines and Gaussians (Cho et al., 2001). The MLR showed an orthodox correlation between the concentrations (pLC₅₀) and the molecular descriptors. The universal expression for MLR is given by equation 3, below:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + \dots a_n X_n \quad (3)$$

Where, Y is the dependent variable (inhibition concentration), X_1, X_2, \dots, X_n are the independent variables (descriptors) existing in the model with their equivalent regression coefficients of a_1, a_2, \dots, a_n , respectively, and a_0 is the constant term of the model. The explanation of the impact of

the individual descriptors X_1, X_2, \dots, X_n depends on the corresponding coefficient value of each descriptor in the regression equation and its numerical sign (Roy et al., 2015). However, each regression coefficient should be significant at 95% probability level ($P < 0.05$), otherwise called 95% confidence limit. This can be confirmed from a Student 't' test. The number of compounds and the number of descriptors must be in a ratio of at least 5:1. MLR model that fits well the normalized data will give a scatter plot of experimental inhibition concentration versus predicted inhibition concentration which eventually shows a lower deviation of the contour points from the line of fit (figure 1).

2.10 Evaluation of the QSAR Models

The developed QSAR models were evaluated using the following statistical parameters, N (number of compounds in regression), P (number of descriptors), R^2 (squared correlation coefficient), F - test (Fischer's value), Q^2_{CV} (Cross-validation correlation coefficient) R^2_{pred} . (square correlation coefficient for external testset). The R^2 regression coefficient and the Q^2_{CV} cross-validation correlation coefficient are the two most important determining factors that should be taken into consideration in accepting any validated QSAR model (Arthur et al., 2016). A model is alleged to be significant only if it has fulfilled the following conditions: $R^2 > 0.6$, $Q^2_{CV} > 0.6$ and $R^2_{pred} > 0.5$. Taking into account the statistical parameters, we opted for a high-value model Q^2_{CV} (cross-validation correlation coefficient) and R^2 (correlation coefficient for the training set) as the best model.

2.11 Internal Validation of the QSAR Model

Internal validation is the first step in validating the QSAR model. The expected results of internal validation indicate that, the model permeates a higher level of stability and reliability (Abdulfatai et al., 2016, 1). The Square correlation coefficient (R^2) designates the division of the total variant accredited to the model. The closer the value of R^2 to 1.0; the better the model generated. R^2 is one of the most commonly used internal validation parameters and can be calculated using Equation 4, expressed below as:

$$R^2 = 1 - \frac{\sum(Y_{obs.} - Y_{pred.})^2}{\sum(Y_{obs.} - Y_{training})^2} \quad (4)$$

Where, $Y_{obs.}$, $Y_{pred.}$, and $Y_{training}$ are the observed, predicted a mean inhibition concentration of the compounds in the training set (Alho et al., 2010). The R^2 value is directly proportional to the number of descriptors in the model. Therefore, we cannot solely rely on the R^2 value in the development of the model. However, R^2 needs to be adjusted to the number of descriptors used in the model. The adjusted R^2 can be calculated using equation 5 given below as:

$$R^2_{adj.} = 1 - (1 - R^2) \frac{n - 1}{n - p - 1} = \frac{(n - 1)R^2 - p}{n - p + 1} \quad (5)$$

Where n is the number of training set compounds and p is the number of descriptors used in the model.

However, the predictive power of the QSAR model was also determined by the use of Friedman's lack of Fit (LOF), which is one among other criteria of internal validation assessment. The lack of Friedman's Fit (LOF) was calculated using Equation 6, given below as:

$$LOF = \frac{SEE}{\left(1 - \frac{c + dp}{N}\right)^2} \quad (6)$$

Where SEE is the standard estimation error, also known as standard deviation (SD), p is the number of independent variables in the model, d is a user-defined smoothing parameter, c is the number of terms in the model, and N is the number of the compound in the training set. However, a model is alleged to be vigorous if it has a smaller SEE value. The SEE can be calculated using Equation 7, expressed below as:

$$SEE = \sqrt{\frac{(Y_{exp.} - Y_{pred.})^2}{N - P - 1}} \quad (7)$$

Where, $Y_{exp.}$ and $Y_{pred.}$ are the experimental and predicted inhibition concentration of the compounds in the training set. N is the number of training compounds, P is the number of descriptors in the model (Jalali et al., 2004). In addition, another factor that should be considered highly important when accessing the internal validation of a QSAR model is the Leave-One-out cross-validation coefficient. The cross-validation regression coefficient (Q^2_{CV}) can be calculated using Equation 8, given by:

$$q^2_{cv} = 1 - \frac{\sum_{i=1}^n (Y_{exp.} - Y_{pred.})^2}{\sum_{i=1}^n (Y_{exp.} - \bar{Y})^2} \quad (8)$$

Where $Y_{exp.}$, $Y_{pred.}$, and \bar{Y} are experimental, predicted and the mean inhibition concentration values of the training set compounds (Jalali et al., 2004).

2.12 External validation of the QSAR Model

The developed QSAR model was externally validated to confirm its robustness. Therefore, the external validation of the model was evaluated centered on its R^2 value for the compounds in the test set. Therefore, the external predictive strength and the extrapolation of the models were calculated using the regression coefficient expression given by equation 9, below:

$$R^2_{Test} = 1 - \frac{\sum(Y_{pred.test} - Y_{exp.test})^2}{\sum(Y_{pred.test} - \bar{Y}_{Training})^2} \quad (9)$$

Where, $Y_{pred.test}$, $Y_{exp.test}$, are the predicted, experimental inhibition concentration of the compounds in the test set and $\bar{Y}_{Training}$ is the mean experimental inhibition concentration of the training set compounds (Tropsha et al., 2003). In addition, the predictive capacity of the QSAR model was also calculated using the excluded compounds in the test set with the aid of the root mean square error (rm^2) given by equation 10, below:

$$rm^2 = r_{test}^2 \left(1 - \sqrt{r_{test}^2 - r_{test_0}^2} \right) \quad (10)$$

Where, r_{test}^2 is the coefficient of determination for test set compounds and $r_{test_0}^2$ is the square correlation coefficient when the plot of the experimental inhibition concentration against the predicted inhibition concentration of the test set was plotted at zero intercepts (Roy et al., 2012). The values of k and k' are the slopes of the regression equation of the predicted inhibition concentration versus experimental inhibition concentration and vice versa (Shen et al., 2002). The k and k' can be calculated using equation 11, given below as:

$$k = \frac{\sum Y_i \bar{Y}_i}{\sum \bar{Y}_i^2} \quad \text{and} \quad k' = \frac{\sum Y_i \bar{Y}_i}{Y_i^2} \quad (11)$$

Where Y_i and \bar{Y}_i are the predicted, mean experimental concentrations of the training set compounds respectively. Other computable relationships between the inhibition concentration and the descriptors were investigated by means of the randomization test, notably known as the Y - randomization test, which is one among other QSAR external validation methods. The areas of the Y -segment were mixed and new QSAR models were developed using the same arrangement of variables as they existed in the unrandomized model.

However, we employed the use of a parameter denoted by R_p^2 , which incriminate the model R^2 for the difference between the mean square correlation coefficient ($R^2_{rand.}$) of the randomized models and the square correlation coefficient (R^2) of the un-randomized model. The R_p^2 parameter was calculated by the mathematical expression given by equation 12, below:

$$R_p^2 = R^2 \times \sqrt{(R^2 - R_{and.}^2)} \quad (12)$$

Where the parameter R_p^2 ensure that, the generated models were not acquired by coincidence. Meanwhile, we projected that, the estimation of R_p^2 should be more prominent than 0.5 for a credible model.

2.13 Y-randomization test

The Y- randomization test is an important criterion for the external validation of the developed QSAR model. To safeguard this, the QSAR model produced was predictive and not acquired by fortuitous, the Y-randomization test was employed using the training set compounds, as suggested by Tropsha (Tropsha et al., 2003). However, for a model to pass the Y randomization test, cR_p^2 should be greater than 0.5 ($cR_p^2 > 0.5$). The cR_p^2 can be calculated using equation 13, expressed below as:

$$cR_p^2 = R[R^2 - (R_r)^2]^2 \quad (13)$$

Where, cR_p^2 is the coefficient of determination for Y randomization test, R is the correlation coefficient for Y-randomization test and R_r is average R of the random model (Tropsha et al., 2003).

In the randomization test, the Multiple Linear Regression models were completed by scrambling the inhibition concentration (pLC50), while maintaining the descriptors unchanged. The models that come after the first model usually have a fundamentally low R^2 and cross-validation Q^2_{cv} values for a few numbers of trials, which confirms that the generated models remained predictive and reliable. 10 Y-randomization tests were performed and it was perceived from the result that, about four of the models have an estimation of R^2 and $Q^2_{cv} < 0.5$. This evaluation confirms that, the model generated was powerful enough and not acquired by chance occurrence (Arthur et al., 2016).

2.14 Evaluation of the Applicability Domain of the Model

The developed QSAR model was further evaluated by plotting the applicability domain (Williams plot), which is the plot of standardized residual versus leverages. The applicability domain of a QSAR model is used purposely to examine the presence of outliers and influential compounds if there is and to confess the consistency and robustness of the generated model (Ibrahim et al., 2018). Leverage is used to plot the applicability domain of a QSAR model. The leverage can be calculated using equation 14, expressed below as:

$$h_i = X_i(X^T X)^{-1} X_i^T, \quad (i = k, \dots, p) \quad (14)$$

Where X_i is the training set matrix I , X is a $n \times k$ descriptor matrix of the training set, and X^T is the transpose matrix X used to generate the model. However, leverage being a prognostic tool; therefore, compounds in the test set or training set with leverage values $h_i < h^*$ are considered to be firmly predictive by the model. The domain with reliable predictive ability is that, with has compounds (both training and test set) with leverage values within the threshold ($h_i < h^*$) and standardized residual values not $> \pm 3$. Therefore, the compounds in the test set or training set with ($h_i > h^*$) are believed to be influential as projected by the model. For the training set, the Williams plot is purposely exploited to identify compounds with the best structural impact ($h_i < h^*$) in the development of the model (Arthur et al., 2016). The warning leverage (h^*) that is the boundary for the X values can be expressed using equation 15, given below as:

$$h^* = \frac{3(k+1)}{n} \quad (15)$$

Where n is the number of compounds in the training set, and k is the number of descriptors used in the model.

2.15 Quality assurance of the model

The methods of internal and external validation of a QSAR model were the two most significant techniques used in assessing the stability, robustness, reliability and predictive capacity of the QSAR model. The validation parameters were compared with the recommendation standard (Veerasamy et al., 2011). Table 1.0; outline the Standard General recommendation values for the internal and external validation parameters that guarantee whether to accept or to reject a model.

3. RESULTS AND DISCUSSION

Multiple linear regression and genetic function algorithms that was incorporated in a material Studio software were used to spawn five models. Among the five models generated, model 1 was selected as the best model because of its statistical significance (Friedman's LOF = 0.3008, $R^2 = 0.9784$, $R^2_{adj} = 0.9739$, $Q^2_{cv} = 0.9675$ and $R^2_{pred} = 0.7348$). However, all the validation parameters that best endorse the reliability and robustness of the model were reported in table 1. This QSAR model is considered highly predictive because it has fulfilled all the prerequisite necessary for accepting any validated QSAR model. The IUPAC names of the compounds, the experimental concentrations, the predicted concentrations and the residual values of the compounds were reported in table 2. The low residual values (the difference between the predicted and experimental inhibition concentrations) means that, the model has an extraordinary predictive capacity. Five molecular descriptors were selected to build a linear model which shows an improved inhibition concentrations result of the compounds from genetic function algorithm.

Model 1

$$\text{pLC50} = 0.037617197 * (\text{AMR}) + 0.948665366 * (\text{SpMax8_Bhp}) - 2.448219792 * (\text{FPSA-2}) + 0.133743691 * (\text{MOMI-YZ}) + 0.114650757 * (\text{RDF50m}) - 0.342918766$$

$$\text{Friedman's LOF} = 0.3008, R^2_{\text{training}} = 0.9784, R^2_{\text{adjusted}} = 0.9739, Q^2_{\text{cv}} = 0.9675, N_{\text{training}} = 30, R^2_{\text{predicted}} = 0.7348, N_{\text{test}} = 13$$

The validation parameters (internal and external) developed for each model using the genetic function approximation were reported in table 3. The molecular descriptors used for the training and test compounds with their experimental and predicted inhibition concentration were reported in table 4 and 5, respectively.

The Pearson correlation matrix of the five descriptors used in this computational study was reported in table 6. The correlation coefficient between each descriptor in the model was significantly low, which means that, there is no much correlation between the descriptors used in the development of the model.

Table 1- General minimum recommended value for an acceptable QSAR model.

Symbol	Name	Acceptable Value
R^2	Coefficient of determination	≥ 0.6
$P_{95\%}$	Confidence interval at 95% confidence level	< 0.05
Q^2_{cv}	Cross-validation coefficient	≥ 0.5
$R^2 - Q^2_{cv}$	Difference between R^2 and Q^2_{cv}	< 0.3
$N_{\text{ext. test set}}$	Minimum number of external test set	≥ 5
cR^2_p	The coefficient of determination for Y-randomization	> 0.5

Table 2- Showing Herbicide IUPAC Name, Experimental pIC50, Predicted pIC50 values and Residual Values of the generated MLR model.

S/N	IUPAC Name	pLC ₅₀ Exp	pLC ₅₀ Pred.	Residual
1		2.96	2.81	-0.15
2		2.51	2.03	-0.48
3		3.67	3.45	-0.22
4		3.04	3.01	-0.03
5		5.63	5.50	-0.13
6		4.41	4.14	-0.27
7		3.26	3.07	-0.19
8		0.95	1.29	0.34
9		3.03	2.77	-0.26
10		3.04	2.76	-0.28

11	1-(3,4-dichlorophenyl)-3-methoxy-3-methylurea	2.18	2.30	0.12
12	1-(3-chloro-4-methoxyphenyl)-3,3-dimethylurea	3.32	3.17	-0.15
13	1-(3-chloro-4-methoxyphenyl)-3,3-dimethylurea	1.8	1.64	-0.16
14	1-(3-methylphenyl)-5-phenyl-1H-1,2,4-triazole-3-carboxamide	2.02	2.29	0.26
15	1-(4,6-dimethoxypyrimidin-2-yl)-3-[(2-ethoxyphenoxy)sulfonyl]urea	0.52	0.46	-0.06
16	1-(4,6-dimethoxypyrimidin-2-yl)-3-[(N-methylmethanesulfonamido)sulfonyl]urea	5.51	5.34	-0.17
17	1-(4,6-dimethoxypyrimidin-2-yl)-3-[[1-methyl-4-(2-methyl-2H-1,2,3,4-tetrazol-5-yl)-1H-pyrazol-5-yl]sulfonyl]urea	1.66	1.88	0.22
18	1-(4,6-dimethoxypyrimidin-2-yl)-3-[[2-(ethanesulfonyl)imidazo[1,2-a]pyridin-3-yl]sulfonyl]urea	3.55	2.37	-1.18
19	1-(4,6-dimethoxypyrimidin-2-yl)-3-[[3-(ethanesulfonyl)pyridin-2-yl]sulfonyl]urea	4.14	4.43	0.29
20	1-(4-bromo-3-chlorophenyl)-3-methoxy-3-methylurea	2.61	2.66	0.05
21	1-(4-bromophenyl)-3-methoxy-3-methylurea	4.98	4.19	-0.79
22	1-(4-chlorophenyl)-3,3-dimethylurea	2.61	2.30	-0.31
23	1-(4-chlorophenyl)-3-methoxy-3-methylurea	3.18	3.30	0.12
24	1-(5-tert-butyl-1,2-oxazol-3-yl)-3,3-dimethylurea	5.53	5.56	0.03
25	1-(5-tert-butyl-1,3,4-thiadiazol-2-yl)-1,3-dimethylurea	3.21	3.45	0.24
26	1-(dimethoxy-1,3,5-triazin-2-yl)-3-[2-(2-methoxyethoxy)benzenesulfonyl]urea	4.16	3.17	-0.99
27	1,2-dimethyl-3,5-diphenyl-1H-pyrazol-2-ium methyl sulfate	4.2	5.15	0.95
28	1,3-dimethyl-1-[5-(trifluoromethyl)-1,3,4-thiadiazol-2-yl]urea	4.16	4.26	0.10
29	1,4-dimethyl 2,3,5,6-tetrachlorobenzene-1,4-dicarboxylate	6.4	6.19	-0.21
30	1-[(2-cyclopropanecarbonylphenyl)sulfamoyl]-3-(4,6-dimethoxypyrimidin-2-yl)urea	5.71	8.08	2.37
31	1-[2-(2-chloroethoxy)benzenesulfonyl]-3-(4-methoxy-6-methyl-1,3,5-triazin-2-yl)urea	5.95	6.22	0.27
32	1-[3-({[(4,6-dimethoxypyrimidin-2-yl)carbamoyl]amino}sulfonyl)pyridin-2-yl]-2-fluoropropyl 2-methoxyacetate	3.6	3.78	0.18
33	1-[4-(4-chlorophenoxy)phenyl]-3,3-dimethylurea	5.85	5.45	-0.40
34	1-[4-(5-tert-butyl-2-oxo-2,3-dihydro-1,3,4-oxadiazol-3-yl)-3-chlorophenyl]-3,3-dimethylurea	2.04	1.46	-0.58
35	1-[5-(ethanesulfonyl)-1,3,4-thiadiazol-2-yl]-1,3-dimethylurea	3.11	3.35	0.24
36	1-{3-chloro-4H,5H,6H,7H-pyrazolo[1,5-a]pyridin-2-yl}-5-[methyl(prop-2-yn-1-yl)amino]-1H-pyrazole-4-carbonitrile	1.91	2.42	0.51
37	1-{4-chloro-3-[(2,2,3,3,3-pentafluoropropoxy)methyl]phenyl}-5-phenyl-	4.98	4.84	-0.14
38	1-3-tert-butylphenoxy-N-(6-methoxypyridin-2-yl)-N-methylmethanethioamide	5.34	5.65	0.31
39	1-chloro-N-(2-chloro-4-fluoro-5-{6-fluoro-1,3-dioxo-hexahydro-1H-pyrrolo[1,2-]imidazolidin-2-yl}phenyl)methanesulfonamide	4.19	3.98	-0.21
40	1-methyl-1-phenyl-3-(2-phenylpropan-2-yl)urea	6.04	5.98	-0.06
41	1-methyl-2-[(2-methylphenyl)methoxy]-4-(propan-2-yl)-7-oxabicyclo[2.2.1]heptane	4.33	4.03	-0.30
42	1-methyl-3-phenyl-5-[3-(trifluoromethyl)phenyl]-1,4-dihydropyridin-4-one	4.44	4.85	0.41

Table 3- Validation parameters for each model using Genetic Function Approximation (GFA)

S/N		Model 1	Model 2	Model 3	Model 4	Model 5	Threshold value
1	Friedman LOF	0.300795	0.305711	0.324264	0.326245	0.327771	0.5
2	R-squared	0.9784	0.978047	0.976714	0.976572	0.976463	≥ 0.6
3	Adjusted R-squared	0.9739	0.973473	0.971863	0.971691	0.971559	> 0.6
4	Cross validated R-squared	0.967515	0.967247	0.964517	0.964853	0.960061	> 0.5
5	Significant Regression	Yes	Yes	Yes	Yes	Yes	
6	Significance-of-regression F-value	217.4193	213.846	201.3358	200.0839	199.1301	
7	Critical SOR F-value (95%)	2.64405	2.64405	2.64405	2.64405	2.64405	
8	Replicate points	0	0	0	0	0	
9	Computed experimental error	0	0	0	0	0	
10	Lack-of-fit points	24	24	24	24	24	
11	Min expt. error for non-significant LOF (95%)	0.21014	0.21185	0.218184	0.218849	0.219361	

Table 4- Calculated descriptor values for the training sets with their experimental and predicted concentrations.

Molecules	AMR	SpMax8_Bhp	FPSA-2	MOMI-YZ	RDF50m	Experimental pLC ₅₀	Predicted pLC ₅₀
4	77.1585	2.192375	0.971161	1.368516	4.938247	3.04	3.01
5	81.3337	2.246656	0.597721	6.505179	10.84811	5.63	5.50
6	73.6698	1.855351	0.62989	3.441717	9.041087	4.41	4.14
8	77.7131	2.368499	0.990879	1.027995	4.619998	3.26	3.07
9	47.3379	1.147726	0.663697	2.028204	1.060182	0.95	1.29
10	85.9427	2.45754	1.622312	1.648164	11.35794	3.03	2.77
11	61.6751	1.918317	1.000945	2.015574	9.995276	3.04	2.76
12	63.455	2.303188	1.210624	2.76899	5.756426	2.18	2.30
13	75.9645	2.413115	1.603034	3.894527	15.41641	3.32	3.17
14	49.6439	1.400916	0.854616	1.769071	5.600613	1.8	1.64
15	46.9027	0.907671	0.557558	1.631555	10.00544	2.02	2.28
16	18.5507	0.817568	0.413431	2.539778	0.001403	0.52	0.46
18	56.9644	1.303324	0.87168	2.500567	5.616158	1.66	1.88
20	84.3649	2.276436	0.832989	2.848138	9.616859	4.14	4.43
21	98.324	2.699828	1.846134	2.136735	8.547054	2.61	2.66
23	61.3448	1.467804	0.612111	1.200149	2.45737	2.61	2.300
25	71.256	2.257071	0.735835	1.7323	23.1274	5.53	5.56
26	68.3294	2.31343	0.87362	4.708859	4.723874	3.21	3.45
30	133.5519	2.811904	1.9543	3.782769	27.18359	6.4	6.19
32	122.1024	2.768123	1.194735	3.399947	15.81433	5.95	6.22
33	97.112	2.650921	1.659118	1.221423	16.1747	3.6	3.78
34	85.6045	2.349084	0.610816	5.913709	9.17447	5.85	5.45
36	108.0014	2.75524	2.319869	3.399239	19.56198	3.11	3.35
39	91.0298	2.247382	0.490221	4.857742	8.623428	5.34	5.65
42	111.577	2.671396	1.600715	8.795199	20.34854	6.04	5.98
43	92.7797	2.502606	1.116026	2.007745	8.517236	4.33	4.03

44	76.0325	2.45754	0.797488	2.146059	14.51517	4.44	4.85
45	114.8064	2.882648	1.452049	5.385797	11.07858	5	5.15
1	57.8325	1.25614	0.476029	2.03909	5.959689	2.96	2.81
17	107.7082	2.665925	1.421546	1.768612	20.44443	5.51	5.34

Table 5- Calculated descriptors for the test set with their experimental and predicted Activity.

Molecules	AMR	SpMax8_Bhp	FPSA-2	MOMI-YZ	RDF50m	Activity	Predict
2	53.1163	1.246209	0.547225	2.763915	1.450453	2.51	2.03
3	60.7273	1.479485	0.592166	8.461216	3.651049	3.67	3.45
19	54.3212	1.66715	0.850731	2.593675	7.153605	3.55	2.37
22	91.2674	2.665925	1.074292	2.856846	7.112731	4.98	4.19
24	70.4077	1.840996	0.723742	1.335393	7.309965	3.18	3.30
27	98.6457	2.684169	1.884564	5.179453	10.28315	4.16	3.17
28	115.3564	2.665925	1.619995	2.890615	19.21301	4.2	5.15
29	74.3812	1.977655	0.678168	9.250508	3.035912	4.16	4.26
31	125.5065	2.745364	1.025938	5.845965	24.63727	5.71	8.08
35	63.5926	2.315266	1.754143	4.465842	7.913456	2.04	1.46
37	55.904	1.718534	0.723825	3.15644	3.283551	1.91	2.42
38	81.9939	2.45754	0.880807	2.441991	13.95963	4.98	4.84
41	88.4505	2.476088	1.183455	2.169109	10.96011	4.19	3.98

Table 6- Pearson's correlation matrix of the descriptors in Selected QSAR model.

	AMR	SpMax8_Bhp	FPSA-2	MOMI-YZ	RDF50m
AMR	1				
SpMax8_Bhp	0.890651	1			
FPSA-2	0.694943	0.699339	1		
MOMI-YZ	0.378794	0.333113	0.089971	1	
RDF50m	0.719901	0.639811	0.605965	0.256842	1

Table 7- Y-Randomization Test Parameters

Model	R	R ²	Q ²
Original	0.989141	0.9784	0.967515
Random 1	0.522795	0.273315	-0.16266
Random 2	0.500953	0.250954	-0.13068
Random 3	0.46971	0.220627	-0.16702
Random 4	0.579175	0.335444	-0.09836
Random 5	0.354832	0.125906	-0.29534
Random 6	0.314135	0.098681	-0.47183
Random 7	0.559705	0.31327	-0.13038
Random 8	0.306126	0.093713	-0.41786
Random 9	0.262277	0.068789	-0.70639
Random 10	0.360066	0.129647	-0.54853
Random Models Parameters			
Average r :	0.422977		
Average r ² :	0.191035		
Average Q ² :	-0.31291		
cRp ² :	0.884432		

The Y-randomization test being one among other essential methods used in the validation of the QSAR model externally. The Y- randomization test was reported in Table 7. It was observed from the result that, the low values of R^2 and Q_{cv}^2 for the number of trials allow us to say that, the constructed model was reliable, stable and robust. While, the cR_p^2 value was much higher than 0.5 which signifies that, the built model was powerful and not inferred by Trial and error.

The description and other related statistical parameters that may have a greater impact on the selected descriptors were reported in table 8. The appearance of 2D and 3D descriptors in the model shows that, these types of descriptors will have a greater affinity (increase the inhibition concentration of the compounds). The values of the variance inflation factor (VIF) for all the five descriptors of the model were lower than 7, which indicates that, the descriptors were orthogonal and the model validity was highly significant. The null hypothesis suggesting that, there is no significant relationship between the inhibition concentration and the descriptor used in the construction of the model at $P < 0.05$. The P- values of the descriptors in the

model at 95% confidence limit were reported in table 8, which are all less than 0.05. This implies that, the null hypothesis was actually rejected. Thus, we accepted the alternative hypothesis which predicts that, there exists a significant relationship between the concentration of the inhibitor molecule and the descriptor. Thus, we can conclude that, there is a significant relationship between the concentration of the inhibitor compounds and the descriptors used in the construction of the model at 95% probability level.

The plot of the predicted inhibition concentration against the experimental concentration of inhibition of the training and test compounds was reported in Figure 1 and 2, respectively. The R^2 value of 0.9784 for the training set and the R^2 value of 0.7348 for the test compounds as reported in this study, was in agreement with the genetic function approximation recommendations report reported in table 1. This confirms the robustness and reliability of the model. The plot of the standardized residual versus experimental inhibition concentration shown in Figure 3 indicates that, there was no reasonable error in the generated model since the spread of the

standardized residual values was virtually on both sides of the zero (Jalali et Al., 2004).

The leverage values for the complete set of compounds in the data set were plotted against their standardized residual values. The Williams plot, that is, a plot of standardized residual versus leverage value was shown in Figure 4. This plot really helps us to inspect outliers and influential compounds in the model. From our result, it is clear that, all compounds (training and testing) were within the square discrimination of ± 3 of standardized residual. Therefore, no outlier compound is found in this study. However, only two compounds from the test set with ID-31 and ID-35 are said to be an influential compound. This is because their leverage values are greater than the warning leverage ($h > h^*$). This behavior was attributed due to differences in their molecular structures as compared to other compounds in the entire data set.

Table 8 - List of the descriptors, their description, classes, and their statistical significance.

S/N	Descriptor Symbol	Description	Class	VIF	Statistics	
					ME	P-Value
1	AMR	Molar refractivity	2D	6.31	0.74	8.54E-08
2	SpMax8_Bhp	The largest absolute eigenvalue of Burden modified matrix - n 8 / weighted by relative polarizabilities	2D	5.18	0.50	4.15E-05
3	FPSA-2	PPSA-2 / total molecular surface area	3D	2.33	-0.64	4.03E-14
4	MOMI-YZ	Y/Z	3D	1.26	0.10	0.00024
5	RDF50m	Radial distribution function - 050 / weighted by relative mass	3D	2.18	0.30	4.19E-10

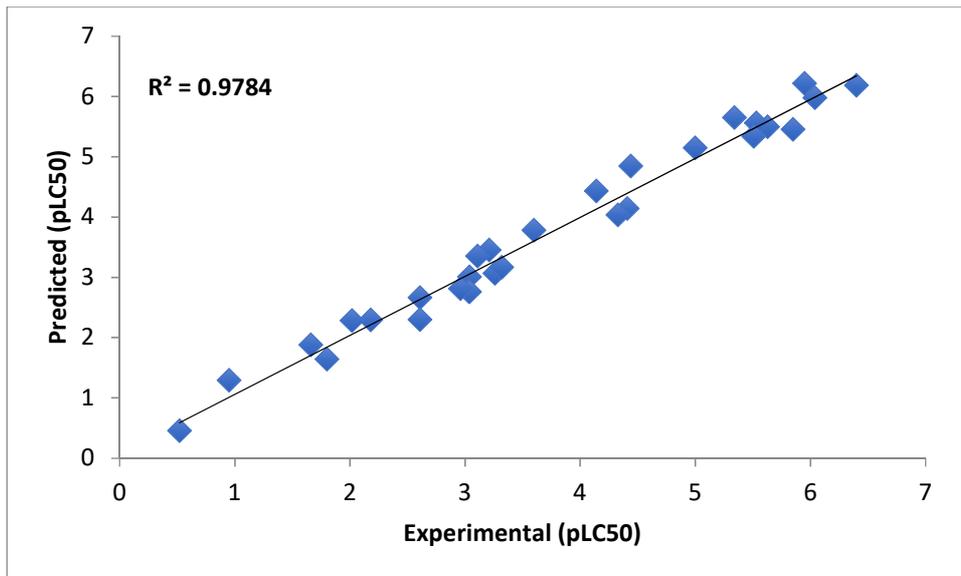


Figure 1- A plot of Predicted (pLC50) versus Experimental (pLC50) of the training set

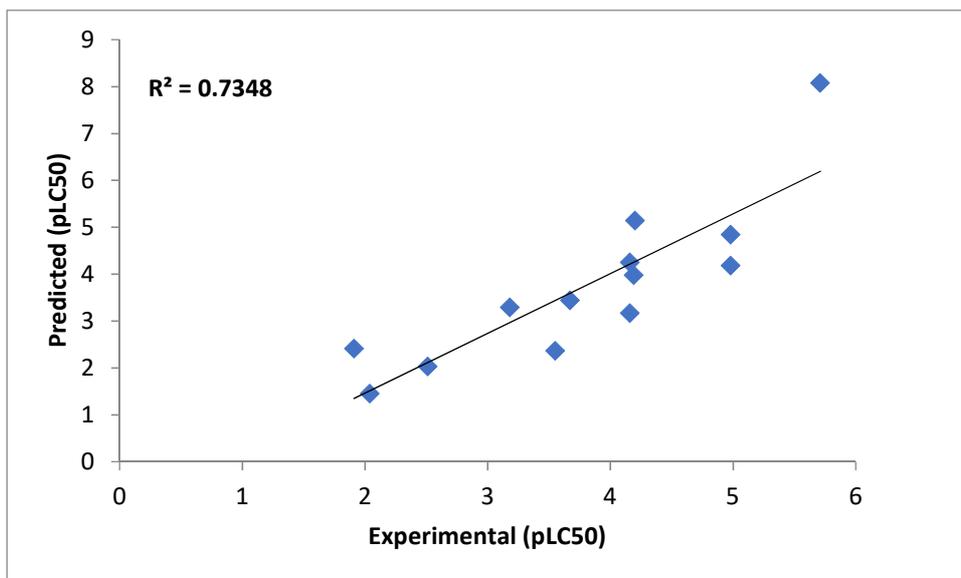


Figure 2- A plot of Predicted (pLC50) versus Experimental (pLC50) of the test set.

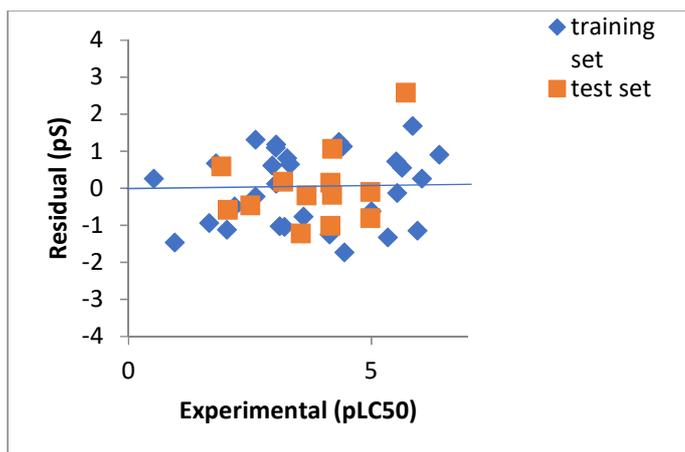


Figure 3- A plot of Residual versus Experimental (pLC50) of the training and test set

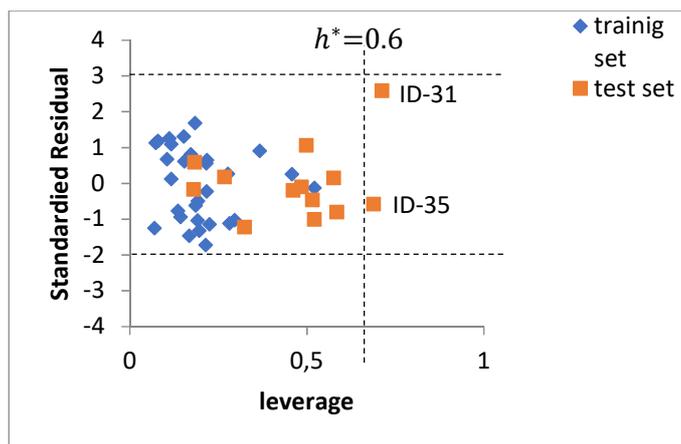


Figure 4- Williams Plot, A plot of standardized residual versus Leverage

4. CONCLUSION

QSAR studies were carried out in a set of data from 43 heterocyclic and phenylic compounds used as potent herbicides. The Density Functional Theory (DFT) optimization method was used in minimizing the energy of the 3D structures. The Genetic function algorithm (GFA) that was incorporated into the material studio software was used to developed five QSAR models. Among the five models generated, model 1 was selected as the best model because of its statistical significance (Friedman's $LOF = 0.3008$, $R^2 = 0.9784$, $R^2_{adj} = 0.9739$, $Q_{cv}^2 = 0.9675$ and $R^2_{pred} = 0.7348$). The relative impact of each descriptor on the model was accessed using the mean effect (ME). However, it is evident that, a decrease in the molecular surface area PPSA-2/total of the descriptor FPSA-2 with negative mean effect of -0.6429 will increase the inhibition concentration of the compounds. Whereas, an increase in AMR (molar refractivity), SPMAX8_BHP (Largest absolute eigenvalue of Burden modified matrix - n 8 / weighted by relative polarizabilities), Momi-YZ (Y/Z) and RDF50m (Radial distribution function - 050 / weighted by relative mass) of the descriptors with positive mean effects of 0.7402 , 0.5006 , 0.9992 and 0.3022 will increase the inhibition concentration of these compounds. The QSAR result developed in this study actually correlates with the standard recommendation values reported earlier in table 1. The positive mean effects values of some descriptors actually provide a good idea in designing new herbicides with improve activity.

REFERENCES

- ABDULFATAI, U.; UZAIRU, A.; UBA, S. Investigation of some acetamido-N-benzylacetamide derivatives as potent anti-convulsant inhibitors. **Journal of Computational Methods in Molecular Design**, v.5, n. 4, p. 77-83, 2015.
- ABDULFATAI, U.; UZAIRU, A.; UBA, S. Quantitative structure-activity relationship and molecular docking studies of a series of quinazolinonyl analogs as inhibitors of gamma-aminobutyric acid

aminotransferase. **Journal of Advanced Research**, v.8, n. 1, p. 33-43, 2017.

- ADENIJI, S. E.; UBA, S.; UZAIRU, A. A Novel QSAR Model for the Evaluation and Prediction of (E)-N'-Benzylideneisonicotinohydrazide Derivatives as the Potent Anti-mycobacterium Tuberculosis Antibodies Using Genetic Function Approach. **Physical Chemistry Research**, v.6, n. 3, p. 479-492, 2018.
- ALHO, J. S.; VÄLIMÄKI, K.; MERILÄ, J. Rhh: an R extension for estimating multilocus heterozygosity and heterozygosity-heterozygosity correlation. **Molecular Ecology Resources**, v.10, n. 4, p. 720-722, 2010.
- ARTHUR, D. E.; UZAIRU, A.; MAMZA, P.; ABECHI, E.; SHALLANGWA, G. In Silico Modelling of Cytotoxic Behaviour of Anti-Leukemic Compounds on HL-60 Cell Line. **Journal of the Turkish Chemical Society, Section A: Chemistry**, v.3, n. 2, p. 147-158, 2016.
- ARTHUR, D. E.; UZAIRU, A.; MAMZA, P.; ABECHI, S. Quantitative structure-activity relationship study on potent anticancer compounds against MOLT-4 and P388 leukemia cell lines. **Journal of Advanced Research**, v.7, n. 5, p. 823-837, 2016.
- ARTHUR, D. E.; UZAIRU, A.; MAMZA, P.; ABECHI, E.; SHALLANGWA, G. Insilco study on the toxicity of anti-cancer compounds tested against MOLT-4 and p388 cell lines using GA-MLR technique. **Beni-Suef University Journal of Basic and Applied Sciences**, v.5, n. 4, p. 320-333, 2016.
- BECKE, A. D. Becke's three-parameter hybrid method using the LYP correlation functional. **Journal of Chemical Physics**, v. 98, p. 5648-5652, 1993.
- CHO, D. H.; LEE, S. K.; KIM, B. T.; NO, K. T. Quantitative structure-activity relationship (QSAR) study of new fluorovinylloxyacetamides. **Bulletin of the Korean Chemical Society**, v.22, n. 4, p. 388-394, 2001.
- CRUZ, V. L.; MARTINEZ, S.; RAMOS, J.; MARTINEZ-SALAZAR, J. 3D-QSAR as a tool for understanding and improving single-site polymerization catalysts. a review. **Organometallics**, v.33, n. 12, p. 2944-2959, 2014.

- DAYAN, F. E.; ZACCARO, M.; LETICIA DE, M. Chlorophyll fluorescence as a marker for herbicide mechanisms of action. **Pesticide Biochemistry and Physiology**, v.102, n. 3, p. 189-197, 2012.
- FUNAR-TIMOFEI, S.; BOROTA, A.; CRISAN, L. Combined molecular docking and QSAR study of fused heterocyclic herbicide inhibitors of D1 protein in photosystem II of plants. **Molecular diversity**, v.21, n. 2, p. 437-454, 2017.
- GANDY, M. N.; CORRAL, M. G.; MYLNE, J. S.; STUBBS, K. A. An interactive database to explore herbicide physicochemical properties. **Organic & biomolecular chemistry**, v.13, n. 20, p. 5586-5590, 2015.
- HANSCH, C.; MUIR, R. M.; FUJITA, T.; MALONEY, P. P.; GEIGER, F.; STREICH, M. The correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammett constants and partition coefficients. **Journal of the American Chemical Society**, v.85, n. 18, p. 2817-2824, 1963.
- IBRAHIM, M. T.; UZAIRU, A.; SHALLANGWA, G. A.; IBRAHIM, A. In-silico studies of some oxadiazoles derivatives as anti-diabetic compounds. **Journal of King Saud University-Science**, p. 2018.
- JALALI-HERAVI, M.; KYANI, A. "Use of computer-assisted methods for the modeling of the retention time of a variety of volatile organic compounds: a PCA-MLR-ANN approach. **Journal of chemical information and computer sciences**, v.44 n. 4, p. 1328-1335, 2004.
- KENNARD, R. W.; STONE, LARRY, A. Computer-aided design of experiments. **Technometrics**, v.11, n. 1, p. 137-148, 1969.
- LEE, C.; Yang, W.; PARR, R.G. Becke's three-parameter hybrid method using the LYP. **Phys. Rev. B**, v.37, p. 785, 1988.
- LIU, Y.; Zhao, H.; WANG, Z.; Li, Y.; Song, H.; RICHES, H.; BEATTIE, D.; Gu, Y.; WANG, Q. The discovery of 3-(1-aminoethylidene) quinoline-2, 4 (1H, 3H)-dione derivatives as novel PSII electron transport inhibitors. **Molecular diversity**, v.17, n. 4, p. 701-710, 2013.
- MINOVSKI, N.; ŽUPERL, Š.; DRGAN, V.; NOVIČ, 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**, v.759, p. 28-42, 2013.
- OLASUPO, S. B.; UZAIRU, A.; SAGA GIS, B. S. Density Functional Theory (B3LYP/6-31G*) Study of Toxicity of Polychlorinated Dibenzofurans., 2017.
- PANCHAL, J. H.; KALIDINDI, S. R.; MCDOWELL, D. L. Key computational modeling issues in integrated computational materials engineering. **Computer-Aided Design**, v.45, n. 1, 4-25, 2013.
- PFISTER, K.; ARNTZEN, C. J. The mode of action of photosystem II-specific inhibitors in herbicide-resistant weed biotypes. **Zeitschrift für Naturforschung C**, v.34, n. 11, p. 996-1009, 1979.
- PRASAD, R. K.; SHARMA, R. 2D QSAR Analysis of pyrazine carboxamide derivatives as an herbicidal agent. **Journal of Computational Method & Molecular Design**, v.1, p. 7-13, 2011.
- RASULEV, B. F.; ABDULLAEV, N. D.; SYROV, V. N.; LESZCZYNSKI, J. A Quantitative Structure-Activity Relationship (QSAR) Study of the Antioxidant Activity of Flavonoids. **QSAR & Combinatorial Science**, v.24, n. 9, p. 1056-1065, 2005.
- ROY, K.; KAR, S.; Das, R. N. Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment. **Academic Press**, 2015.
- SAIDI, A.; MIRZAEI, M. Prediction of AHAS inhibition by sulfonylurea herbicides using a genetic algorithm and artificial neural network. 2016.
- SHEN, M.; LET IRAN, A.; XIAO, Y.; GALBRAITH, A.; KOHN, H.; TROPSHA, A. Quantitative structure-activity relationship analysis of functionalized amino acid anticonvulsant agents using k nearest neighbor and simulated annealing PLS methods. **Journal of medicinal chemistry**, v.45, n. 13, p. 2811-2823, 2002.
- TAKAČ, M. J.; MEDIĆ-ŠARIČ, M. QSPR, and QSAR in Pharmacy. I. Classic QSAR models. Hansch and Fred Wilson's model. **Farmaceutski glasnik: glacial Hrvatskog farmaceutskog društva**, v.47, n. 6, p. 161-178, 1991.
- TROPSHA, A.; GRAMATICA, P.; GOMBAR, V. K. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. **QSAR & Combinatorial Science**, v.22, n. 1, p. 69-77, 2003.
- TROYER, J. R. In the beginning: the multiple discoveries of the first hormone herbicides. **Weed Science**, v.49, n. 2, p. 290-297, 2001.
- VEERASAMY, R.; RAJAK, H.; JAIN, A.; SIVADASAN, S.; VARGHESE, C. P.; AGRAWAL, R. K. Validation of QSAR models-strategies and importance. **International Journal of Drug Design & Discovery**, v.3, p. 511-519, 2011.
- VERMA, J.; KHEDKAR, V. M.; COUTINHO, E. C. "3D-QSAR in drug design-a review. **Current topics in medicinal chemistry**, v.10, n. 1, p. 95-115, 2010.
- ZHANG, C.; CHANG, S.; TIAN, X.; Tian, Y. 3D-QSAR and docking modeling study of 1, 3, 5-triazine derivatives as PSII electron transport inhibitor. **Asian Journal of Chemistry**, v.26, n. 1, p. 264, 2014.
- ZIMMERMAN, P. W.; HITCHCOCK, A. E. Plant hormones. **Annual review of biochemistry**, v.17, n. 1, p. 601-626, 1948.