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

QSAR studies on some C_{14}-urea tetrandrine compounds as potent anti-cancer agents against Leukemia cell line (K562)

Mustapha Abdullahi[a]*, Gideon Adamu Shallangwa[a], Muhammad Tukur Ibrahim[a], Abdullahi Umar Bello, David Ebuka Arthur[a], Adamu Uzairu[a], and Paul Mamza[a]

[a]Faculty of Physical Sciences, Chemistry Department, Ahmadu Bello University (ABU) Zaria, Kaduna State, Nigeria

Abstract: This research applied Quantitative Structure Activity Relationship (QSAR) technique in developing a Multiple-Linear Regression (MLR) model using Genetic Functional Approximation (GFA) method in selecting optimum molecular descriptors from the structures of 24 C_{14}-urea tetrandrine compounds. Firstly, the compounds were optimized at the Density Functional Theory (DFT) level using Becke's three-parameter Lee-Yang-Parr hybrid functional (B3LYP) with the 6-31G* basis set in the Spartan 14 Version 1.1.4 software. The descriptors of the compounds were computed using Padel-software, and data set was divided into training and test set. A model was built from the training set with internal validation parameter R^2_{train} as 0.9104. The external validation of the model was done using the test set compounds with validation parameter R^2_{test} as 0.6443 that passed the criteria for acceptability of a QSAR model globally. The coefficient of determination (cR^2_p) parameter was calculated as 0.8192 which is greater than 0.5, this affirms that the generated model is robust. Furthermore, AST4p, GATS8v, and MLFER are descriptors in the model with the positive mean effect of 0.0899, 0.9098 and 0.0002 respectively. This study depicts a route in designing and synthesizing new C_{14}-urea tetrandrine compounds with better inhibitory potentials.

Keywords: QSAR; Mean Effect; Validation; Descriptors; Model; Y-randomization

Submitted: September 06, 2018. **Accepted:** December 14, 2018.

Cite this: Abdullahi M, Shallangwa G, Ibrahim M, Bello A, Arthur D, Uzairu A, et al. QSAR studies on some C14-urea tetrandrine compounds as potent anti-cancer agents against Leukemia cell line (K562). JOTCSA. 2018;5(3):1391–402.

DOI: http://dx.doi.org/10.18596/jotcsa.457618.

***Corresponding author. E-mail:** mustychem19@gmail.com.

INTRODUCTION

Leukemia is one of the most fatal cancer type that affects tissue for blood formation in the bone marrow, lymphatic system, and spleen in the body (1). The K562 leukemic cell lines were the first human immortalized myelogenous leukemia cell line to be understood which was obtained from a 53-year-old female chronic myelogenous leukemia patient in blast crisis (2). The cells are non-adherent, rounded, positive for the BCR/ABL fusion gene, and bear some proteomic similarity to indistinguishable erythrocytes (2). In culture, they show much less clattering than many other suspension lines, perhaps due to the down-regulation of surface adhesion molecules by BCR/ABL. However, additional study lament that BCR/ABL

over-expression may actually increase cell adherence to cell culture plastic (3). The problem with K562 cells is that it undergoes excess of Aurora kinases which plays a role in the improvement of spindles, the partition of chromosomes, and cytokinesis (4). These functions are important in cells so as to split, redevelop tissues, and assume a support part in their homeostatic abilities. However, the excess of Aurora kinases takes uncontrolled cell division in to account and bringing about the tumor (4). Tetrandrines are compounds of dibenzyltetrahydroisoquinoline, derived from Chinese medicinal plant called *Stephania tetrandra* and it is reported to have anti-tumor activities, proliferation chemotherapeutic drugs and converses multidrug resistance (MDR) of tumor cell (5).

In recent decades, there was a significant number of studies that proved the success of the Quantitative Structure-Activity Relationship (QSAR) approach for prediction of various properties, such as solubility, lipophilicity, toxicity, mutagenicity, activities (6). By definition, a QSAR model is a mathematical linear equation involving molecular descriptors used in predicting the biological activity of a compound which is ought to be very useful in designing the new compound with better activity. Therefore, the main aim of this research was to develop a QSAR model of some C_{14}-urea tetrandrine compounds which can be used to predict the biological activities of compounds against the leukemia K562 cell line using Genetic Function Approximation–Multi-Linear Regression (GFA-MLR) method.

MATERIALS AND METHODS

Data Set collection

A data set of twenty-four (24) C_{14}-urea tetrandrine compounds as potent anti-cancer agents for this study were sourced from the literature (7). The biological activities of the compounds against leukemia K562 cell line were measured in IC_{50} (μM) which is the concentration of compound required to reduce 50% of the cell viability. This is further transformed to logarithm scale (Eq. 1) so as to have linearity or normality in the concentration values. The 2D structures of the compounds were drawn using ChemDraw software version 12.0.2 as shown in "Fig 1", then aligned with their respective IC_{50} values as shown in Table 1.

$$pIC_{50} = -\log(IC_{50} \times 10^{-6}) \quad \text{(Eq. 1)}$$

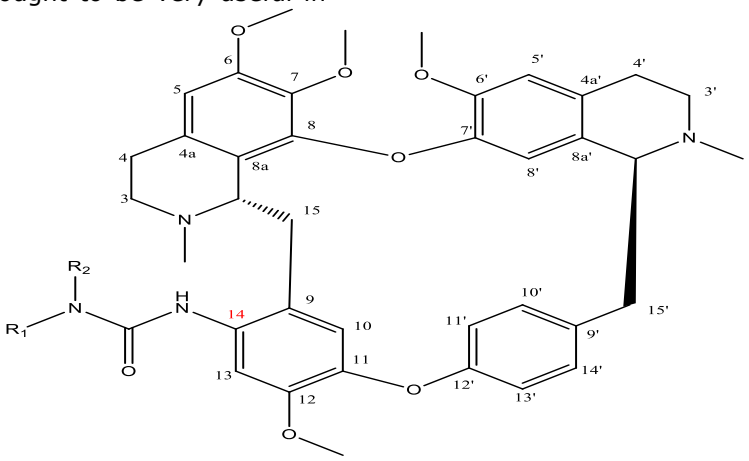
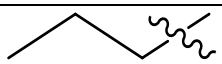
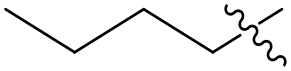
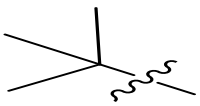
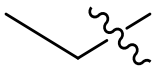
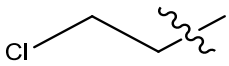
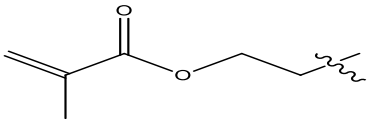
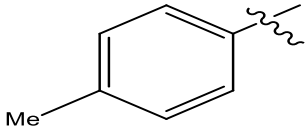
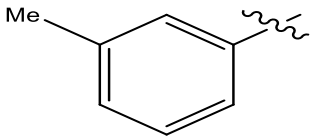
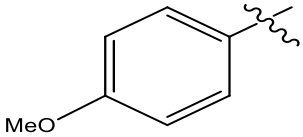
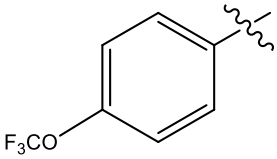
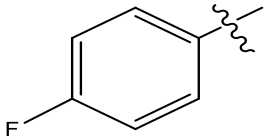
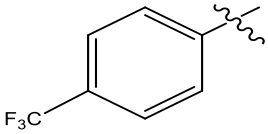
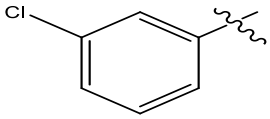
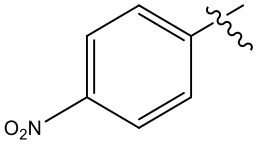
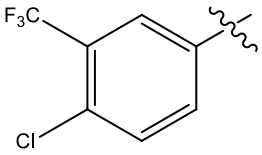
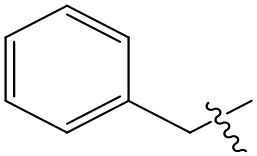
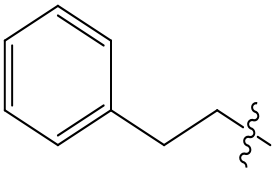
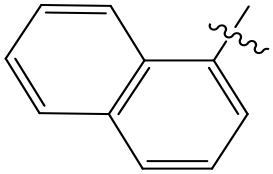
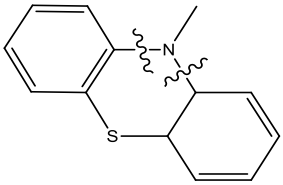
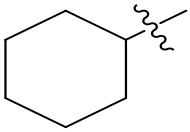
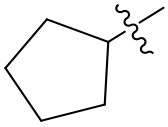
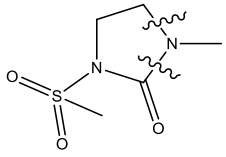
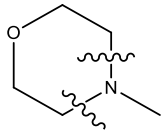


Figure 1: Main C_{14}-urea tetrandrine structure.

Table 1: Substitution pattern of C_{14}-urea tetrandrine compounds and their inhibitory concentrations (IC_{50}) against leukemia K562 cell line.

S/No.	R_1	R_2	IC_{50} (μM)	pIC_{50}
1		H	5.09	5.2932
2		H	6.88	5.1624
3		H	4.89	5.3106

4	-Me	Methyl	9.21	5.0357
5		Methyl	3.20	5.4948
6		H	6.24	5.2048
7		H	8.01	5.0963
8		H	2.89	5.5391
9		H	2.15	5.6675
10		H	3.22	5.4921
11		H	1.25	5.9030
12		H	1.81	5.7423
13		H	1.74	5.7594
14		H	2.02	5.6946
15		H	1.84	5.7351
16		H	2.05	5.6882

17		H	4.59	5.3381
18		H	4.94	5.3062
19		H	1.97	5.7055
20		-	2.49	5.6038
21		H	5.56	5.2549
22		H	6.42	5.1924
23		-	3.06	5.5142
24		-	4.33	5.3635

Equilibrium Geometry

The equilibrium geometries of all the compounds were obtained by engaging Spartan 14 version software at the density functional theory (DFT) level using Becke's three-parameter Lee-Yang-Parr hybrid functional (B3LYP) with the 6-31G* basis set (8, 9). The geometry optimization is an atomic arrangement process which gives the most stable state of the starting molecular structure.

Molecular descriptor calculation

The optimized twenty-four (24) molecules were subjected to PaDEL-Descriptor software V2.20 to calculate a total of 1875 molecular descriptors including electronic, spatial, structural, thermodynamic, and topological descriptor (10). The data generated from the PADEL- software in MS Excel (.csv) format were observed to contain redundant data, zero columns or non-informative descriptors.

Data pretreatment and Division

The data was subjected to a pretreatment process using Data Pretreatment software downloaded from Drug Theoretical and Cheminformatics Laboratory so as to curate the results (11). Consequently, the pretreated data were divided into training and test sets using Data Division software also gotten from Drug

Theoretical and Cheminformatics Laboratory (DTC Lab) using Kennard and Stone's algorithm method (12).

Model Building and Validation

The training set was used in developing the model from Material studio software version 8 by engaging the Genetic Function Approximation (GFA) method in which the dependent variable is the inhibitory concentration (IC_{50}) and the independent variables are the molecular descriptors. The model generated was evaluated using Friedman formula (Eq. 2) which determines the finest fitness score defined as; (13).

$$LOF = \frac{SEE}{\left(1 - \frac{m + s \times d}{T}\right)^2} \quad \text{(Eq. 2)}$$

m is the number of the terms in the model, **s** is a user-defined smoothing parameter, ***d*** is the total number of descriptors in the model and ***T*** is the number of data in the training set (14).

$$SEE = \sqrt{\frac{(Y_{exp} - Y_{pred})^2}{N - P - 1}} \quad \text{(Eq. 3)}$$

Where **SEE** is the Standard Error of Estimation or Sum of Squares of Errors **(SSE)**. It gives an idea about the quality of a model, low SEE value signifies better model and vice versa. It was defined by the expression (Eq. 3);

Internal Validation

The established QSAR model was validated so as to check the predictive capability and reliability of the models. The internal validation of the models was examined using the leave-one-out (LOO) cross-validation method. The cross-validation regression coefficient, R^2 (Q^2_{cv}) were also calculated using Eq. 4:

$$R^2 = 1 - \left[\frac{\sum\left(y_{exp} - y_{pred}\right)^2}{\sum\left(y_{exp} - \overline{y}_{training}\right)^2}\right] \quad \text{(Eq. 4)}$$

Where $\overline{y}_{training}$ is the mean of experimental activities, y_{exp} is the experimental activities, and y_{pred} is the predicted activity in the training set respectively (15).

External Validation

The R^2 value are directly proportional to the number of descriptors. However, the R^2 values is not consistent for evaluating the strength of the model. Thus, R^2 is adjusted with the mandate to refurbish and stabilize the model. The adjusted R^2 is defined as like in Eq. 5:

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

Where **p** is the number of descriptors in the model, **n** is the number of compounds that made up the training set (15).

The model developed was further subjected to external validation in order to measure its prediction competency using the test set and the coefficient of determination(R^2_{test}) value is given in Equation 6;

$$R^2_{test} = 1 - \frac{\sum\left(Ypred_{test} - Y_{exp_{test}}\right)^2}{\sum\left(Ypred_{test} - \overline{Y}_{training}\right)^2} \quad \text{(Eq. 6)}$$

Where; $Ypred_{test}$ and $Y_{exp_{test}}$ are the predicted and experimental activity test set respectively. $\overline{Y}_{training}$ is mean values of experimental activity of the training set (15).

Y-Randomization

In order to have confidence in the model built, Y-Randomization test was executed on the training set descriptors matrix (16). This is done by randomly shuffling the inhibitory concentrations (dependent variable) while keeping the descriptors (independent variables) constant resulting in the generation of random MLR models. The new QSAR models are anticipated to have significantly low R^2 and Q^2 values for 10 trials, which certify that the models are robust and cR^2_p is also calculated which should be more than 0.5 defined as:

$$cR^2_p = R \times [R^2 - (R_r)^2]^{1/2} \quad \text{(Eq. 7)}$$

Where cR^2_p is coefficient of determination, R is the coefficient of regression and R_r is average 'R' of random models.

Statistical analysis of the descriptors

Mean Effect

The mean effect values of each descriptor were used to evaluate their relative significances in the model and it is defined as:

$$Mean\ Effect = \frac{\beta_j \sum_i^n D_j}{\sum_j^m (\beta_j \sum_i^n D_j)} \quad \text{(Eq. 8)}$$

Where β_j is the coefficient of the descriptor *j* in that model, *Dj* is the value of each descriptor in the data matrix for each molecule in the training set and m is the number of the descriptor that appears in the model and n is the number of molecules in the training set (17).

Varian Inflation Factor (VIF)
The Variance Inflation Factor is a measure of the multi-collinearity among the descriptors, usually expressed as:

$$VIF = \frac{1}{(1-R^2)} \quad \text{(Eq. 9)}$$

Where R^2 is the correlation coefficient of the multiple regression between the variables within the model. If VIF equals to 1, no inter-correlation exists for each variable, if VIF falls into the range of 1–5, the related model is acceptable; and if VIF is larger than 10, the related model is unstable and unacceptable (18).

Applicability Domain
A QSAR model applicability domain is usually tasked to explore the area where the compound predictions can be dependably useful. As such, chemical compounds that fall outside the applicability domain cannot make a very good prediction (19, 20). Consequently, the prediction that is interpolated in the chemical space is acceptable while extrapolated predictions in the chemical space are rejected as well. The leverage method was engaged in evaluating the applicability domain of the established QSAR model and it is defined as the leverage values for the *ith* compound (Eq. 10) (21):

$$hi = X_i(X^T X)^{-1} X_i^T \quad \text{(Eq. 10)}$$

Where; *Xi* is training compounds matrix of I, X is the n × k descriptor matrix of the training set compound and X^T is the transpose matrix of X used in developing the model. The warning leverage (h^*) is the borderline of normal values for X outliers and is defined as follows (Eq. 11):

$$\boldsymbol{h}^* = 3\frac{(r+1)}{n} \quad \text{(Eq. 11)}$$

Where ***n*** is the number of training compounds and ***r*** is the number of descriptors in the model.

The leverages of the test compounds with $\boldsymbol{hi} < \boldsymbol{h}^*$ are measured to be consistently predicted by the model. A plot of standardized residuals versus leverage values (Williams plot) is utilized to interpret the relevance area of the model in terms of chemical space. The area of unfailing predictions for the external test compounds, defined as compounds whose leverage values are within the threshold and standardized residuals is not greater than 2σ (2 standard deviation units). Therefore, the test compound ($\boldsymbol{hi} < \boldsymbol{h}^*$) are accepted as Y outlier. Similarly, the test set compounds having ($\boldsymbol{hi} > \boldsymbol{h}^*$) are variably projected by the model since they are extrapolated (21)

RESULT AND DISCUSSION

Descriptor Calculations
The QSAR studies were performed to generate a model that relates the structure activity relationship of twenty-four C_{14}-urea tetrandrine compounds as a potential anticancer agent against leukemia (K562) cell lines. Initially, the 32 quantum chemical descriptors for all the drawn compounds were obtained from Spartan 14 software via optimization process. These were pooled with the 1875 molecular descriptor calculated by PaDEL-Descriptor software V2.20 to give 1907.

Data Pretreatment and Division
The descriptors result in MS Excel (.csv) were subjected to data pretreatment which removed non-informative constant data and a pair of variables with a correlation coefficient greater than 0.7 using the Data Pretreatment software. The data set results from the pretreatment process was divided by using Kennard-Stone algorithm method where 16 compounds (70% of the total compounds) are considered as training set and 8 compounds (30% of the total compounds) are the test set. The division was successfully done using the Dataset Division GUI 1.2 software.

Model Building and Validation
In building the QSAR model, three (3) descriptors were used to build the model by the Genetic Function Approximation (GFA) of Material studio software and the model generated is illustrated below:

$$\begin{aligned} \mathrm{pIC50} = & -0.064954009 * \mathbf{ATSC4p} + 6.794973156 \\ & * \mathbf{GATS8v} - 0.626117779 \\ & * \mathbf{MLFER\_A} \\ & - 2.008205026 \end{aligned} \quad \text{(Eq. 12)}$$

The validation parameters of the model were presented in Table 2 which clearly shows that the model passed the criteria of acceptability. In addition, the coefficients of regression (R-squares) are 0.9104 and 0.6443 for both the training and test set compounds respectively. This is an indication of a good relationship between the predicted and experimental activities. The Centered Broto-Moreau autocorrelation-lag 4 per weighted by

polarizabilities (**ATSC4p**) descriptor is an autocorrelation of a topological structure defined as the most recognized spatial autocorrelation on a molecular graph which is given as;

$$ATS_k = \frac{1}{2}\sum_{i=1}^{A}\sum_{j=1}^{A} w_i \cdot w_j \delta(d_{ij};k) = \frac{1}{2}\cdot(w^T \cdot {}^K B \cdot w) \quad \text{(Eq. 13)}$$

where ***w*** is any atomic property, ***A*** is the number of atoms, ***k*** is the interval, and δ_{ij} is the topological distance between ***ith*** and ***jth*** atoms; δ (d_{ij}; k) is a Kronecker delta function which is equivalent to1 if d_{ij}=k, but if d_{ij} is not equal k, the function is said to be zero. ^{k}B is the k^{th} order corresponding to the geodesic matrix, whose elements are equal to 1 only for vertices v_i and v_j at topological distance k, and zero otherwise; w is the dimensional vector of atomic properties (22).

The Geary autocorrelation-interval 8 per weighted by the Vander Waals volumes (**GATS8v**) is a 2D autocorrelation descriptor, which is obtained from molecular graphs by summing the products of atom weights of the terminal atoms of all the paths of the considered path length (the lag 8) (22). Whereas, the **MLFER_A** descriptor is a linear free energy relation (LFERs) descriptor whose coefficient measures the acidity of hydrogen bond due to the interaction of basic solutes with acidic phase (22). The positive mean effect of these three (3) descriptors in this study inferred that there will be a positive influence on the inhibitory concentrations when each descriptor value increases in the same direction.

Table 2: Validation parameters of the model.

Validation Parameters	Model	QSAR Validation Standard
Friedman LOF	0.0280	-
R-squared (Training set)	0.9104	≥ 0.6
Adjusted R-squared	0.8880	-
Cross validated R-squared	0.8172	≥ 0.5
Significant Regression	Yes	-
Significance-of-regression F-value	40.6445	-
Critical SOR F-value (95%)	3.6506	-
Replicate points	0	-
Computed experimental error	0	-
Lack-of-fit points	12	-
Min expt. error for non-significant LOF (95%)	0.0601	-
R-square (test set)	0.6443	≥ 0.6

Univariate analysis were conducted on the inhibitory concentration values of the two set (i.e. training set and test set) as presented in Table 3. These clearly show that the training set range values are within the test set range values. Furthermore, the mean activities and standard deviation of both the training set were almost alike when compared to the test set value. This inferred that test set compounds activities were interpolative within the activities of the training set.

The experimental, predicted inhibitory concentration (pIC_{50}) and the residual values generated from the compounds were shown in Table 4. The residual value is defined as the differences between experimental and predicted activity, and lower residual values signify that the model has a high predictive ability.

Table 3: Univariate analysis for the inhibitory concentrations (IC_{50}).

	All	Training Set	Test Set
Number of sample points	24	16	8
Range	0.8673	0.7065	0.7106
Maximum	5.9030	5.7423	5.9030
Minimum	5.0357	5.0357	5.1924
Mean	5.4624	5.4179	5.5513
Median	5.4934	5.4027	5.5798
Variance	0.0567	0.0532	0.0517
Standard deviation	0.2432	0.2383	0.2432
Mean absolute deviation	0.2106	0.2099	0.2048
Skewness	-0.0291	-0.024	-0.0499
Kurtosis	-1.3179	-1.557	-1.6534

Table 4: Experimental, predicted and residual values of tetrandrine derivatives

Training Set				Test Set			
Compound	Experimental	Predicted	Residual	Compound	Experimental	Predicted	Residual
1	5.2932	5.2300	0.0632	9	5.6675	5.7165	-0.0489
2	5.1624	5.1855	-0.0235	10	5.4921	5.7516	-0.2594
3	5.3106	5.2425	0.0680	11	5.9030	5.9345	-0.0314
4	5.0354	5.1089	-0.0732	13	5.7594	5.6038	0.1555
5	5.4945	5.4227	0.0720	14	5.6946	5.5557	0.1388
6	5.2048	5.2385	-0.0337	17	5.3381	5.2026	0.1355
7	5.0963	5.1650	-0.0687	22	5.1924	5.2129	-0.0205
8	5.5391	5.6657	-0.1266	24	5.3635	5.5647	-0.2012
12	5.7423	5.6612	0.0811	-			
15	5.7358	5.7416	-0.0064	-			
16	5.6882	5.7723	-0.0840	-			
18	5.3062	5.2613	0.0448	-			
19	5.7055	5.6388	0.0666	-			
20	5.6038	5.4989	0.1048	-			
21	5.2549	5.2974	-0.0425	-			
23	5.5142	5.5562	-0.0419	-			

Statistical Analysis of the Descriptors

In order to assess the relationships between each descriptor used in the model, the values of the three (3) descriptors were extracted from the training set, then subjected to Pearson's correlation analysis and the results were described in Table 5. These show that there is no significant inter-correlation between the descriptors used in the model because the correlation coefficients between all pairs are less than 0.5. The Variance Inflation Factor (VIF) values for all the three (3) descriptors are not greater than 2 which signifies that the descriptors are and the model is said to be stably acceptable.

Table 5: Pearson's correlation analysis for descriptor used in the QSAR model.

	ATSC4p	GATS8v	MLFER_A	VIF
ATSC4p	1			1.2954
GATS8v	0.3439	1		1.2183
MLFER_A	0.4354	0.3719	1	1.3255

**VIF is the variance inflation factor*

The results in Table 6 illustrate some statistical parameters of descriptors in the developed model. From results, the absolute t-statistics values for each descriptor are greater than 2, this also inferred that the selected descriptors were good (23). The p-values of all descriptors in the model are less than 0.05 which means that there is a relationship between the descriptors and the inhibitory concentration of the compounds

Table 6: Statistical parameters.

	Coefficients	***Standard Error***	***t Stat***	***P-value***	***Mean Effect***
ATSC4p	-0.0649	0.0086	-7.5382	6.87E-06	0.0899
GATS8v	6.7949	1.2599	5.3931	0.000162	0.9098
MLFER_A	-0.6261	0.1293	-4.8403	0.000405	0.0002

The output of Y-Randomization test was presented in Table 7. The cR^2_p value was calculated as 0.8192 which is greater than 0.5, this affirms that the generated model is robust.

A Plot of standardized residual against experimental activity in "Fig 3" illustrated a random scattering around the baseline of data at the standardized residual equal to zero. Hence, there was no systematic error in the model built.

Table 7: Y-randomization test

Model	**R**	**R^2**	**Q^2**
Original	0.9541	0.9104	0.8172
Random 1	0.4285	0.1836	-0.2395
Random 2	0.5090	0.2591	-0.1873
Random 3	0.3475	0.1208	-1.0654
Random 4	0.2729	0.0744	-0.5785
Random 5	0.3166	0.1002	-0.8467
Random 6	0.3393	0.1151	-0.5946
Random 7	0.5352	0.2865	-0.0898
Random 8	0.4387	0.1924	-0.7466
Random 9	0.2490	0.0620	-0.5746
Random 10	0.7233	0.5232	0.2392
Random models parameters			
Average R :	0.4160		
Average R^2 :	0.1917		
Average Q^2 :	-0.4684		
cRp^2 :	0.8192		

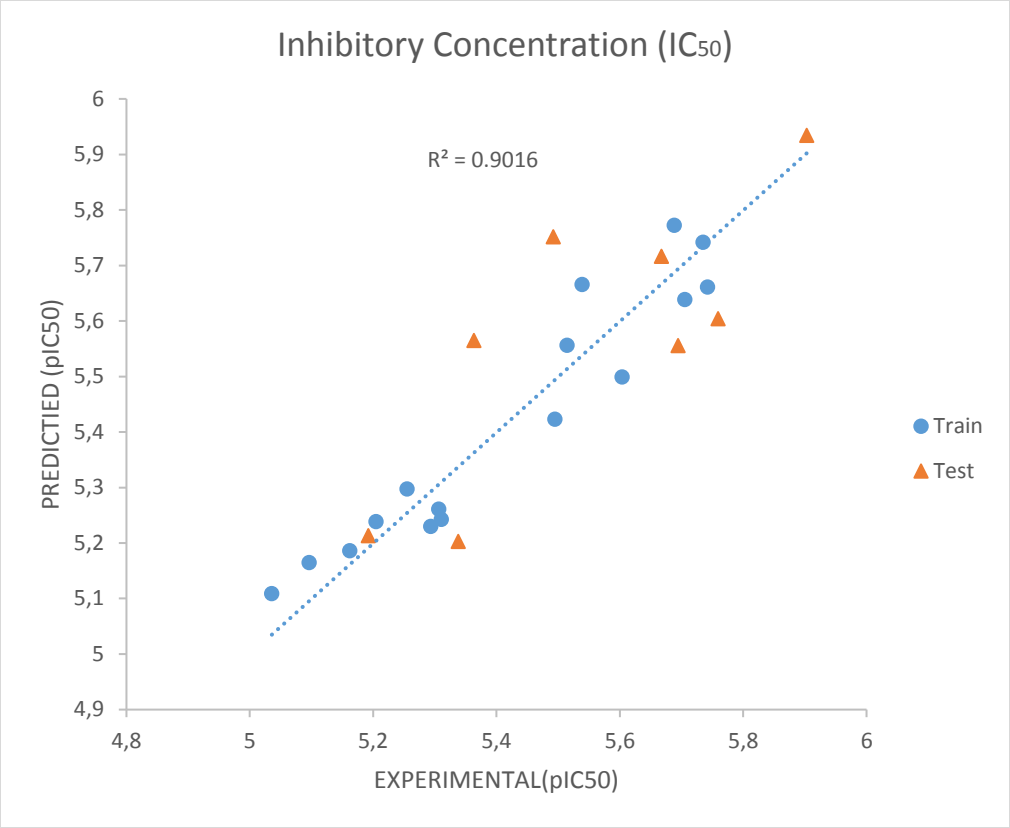


Figure 2: Plot of predicted against experimental activities (pIC_{50}).

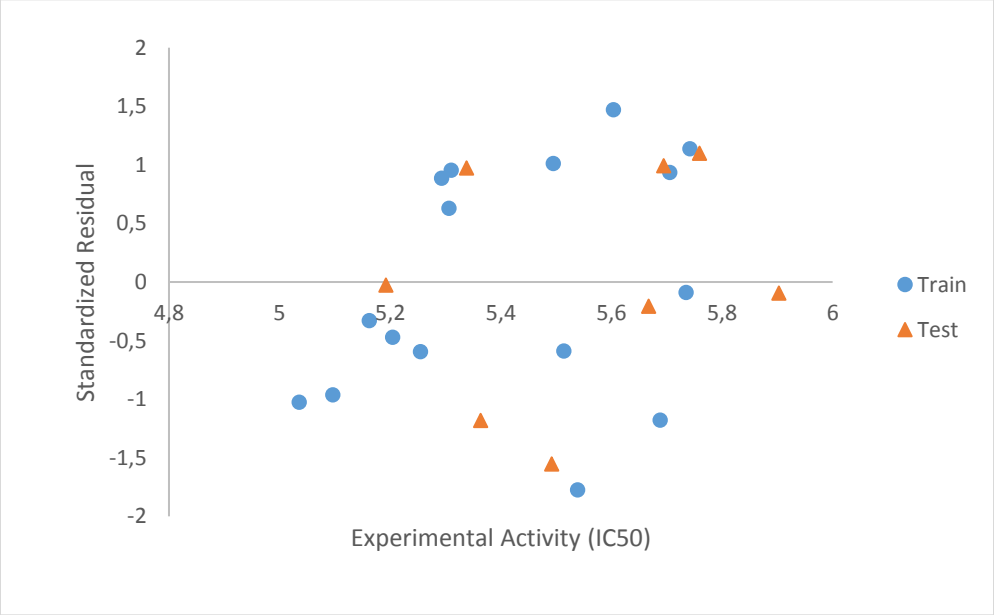


Figure 3: Plot of standardized residual against experimental activity (pIC_{50}).

A scatter plot for standardized residuals against the leverages termed as Williams Plot was presented in "Fig 4" so as to detect the presence of outliers and influencing compounds in the models. Our results revealed that all the compounds are within the square area ±2 of standardized deviation unit which means there is no outlier. However, the calculated warning leverage (h*) is 0.75. The plot also revealed that two (2) test set compounds (i.e.,

compound 17 and 22) are considered to be the influencing compounds because their leverages are more than the warning leverage. The reason may be attributed to the differences in the substitution pattern of the chemical structure in the data set.

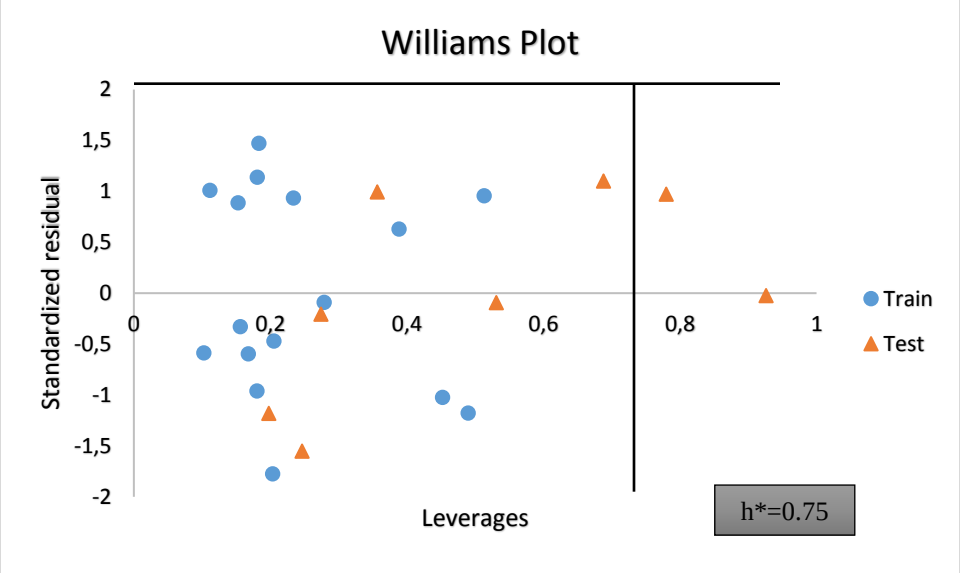


Figure 4: The williams plot (Standardized residuals vs the leverage values)

CONCLUSION

In conclusion, this research has successfully achieved its aim of constructing a QSAR model for the tetrandrine compounds which predicts the inhibitory concentration against leukemia K562 cell line using Genetic functional algorithm method. Our research findings revealed the molecular descriptors AST4p, GATS8v and MLFER with a mean effect of 0.0899, 0.9098 and 0.0002 respectively, were found to positively influence the inhibitory concentrations. This knowledge could be of vital importance in designing and synthesizing new C_{14}-urea tetrandrine compound with excellent inhibitory potentials.

REFERENCES

1. Lookward W. Leukemia (AML, CML, ALL and CLL). www.RN.ORG; ©RN.ORG®, S.A., RN.ORG®, LL; 2015. ISBN: 678-693-1-32.

2. Lozzio CB, Lozzio BB. Human Chronic Myelogenous Leukemia Cell-Line with Positive Philadelphia Chromosome. National Institute of Health. 1975; 45(1): 321–340.

3. Karimiani EG, Marriage F, Merritt AJ, Burthem J, Byers RJ, Day PJ. Single-cell analysis of K562 cells: an imatinib-resistant subpopulation is adherent and has upregulated expression of BCR-ABL mRNA and protein. Experimental Hematology. 2014; 42(3): 183-191

4. Fan Y, Lu H, An L, Wang C, Zhou Z, Feng F, Zhao Q. Effect of active fraction of Eriocaulon sieboldianum on human leukemia K562 cells via proliferation inhibition, cell cycle arrest and apoptosis induction. Environmental Toxicology and Pharmacology. 2016; 4(3):13-20.

5. Liu T, Liu X, Li WH. Tetrandrine, a Chinese Plant-Derived Alkaloid, Is a Potential Candidate for Cancer Chemotherapy, On Co-Target. 2016; 2(7):480100–480115.

6. Perkins R, Fang H, Tong W, Welsh WJ. Quantitative Structure-Activity Relationship Methods: Perspectives on Drug Discovery and Toxicology. 2003; 22(1): 1666–1679.

7. Lan J, Huang L, Lou H, Chen C, Liu T, Hu S, Yao Y, Song J, Luo J, Liu Y, Xia B, Xia L, Zeng X, Ben-David Y, Pan W. Design and Synthesis of Novel Tetrandrine Derivatives as Potential Anti-Tumor Agents against Human Hepatocellular Carcinoma. European Journal of Medicinal Chemistry. 2017 Sep; 3-4.

8. Becke AD. Becke's three-parameter hybrid method using the LYP correlation functional. Journal of Chemical Physics. 1993; 98: 5648–5652.

9. Lee C, Yang W, Parr RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Physical Review B. 1988; 37:785

10. Alisi IO, Uzairu A, Abechi SE, Idris SO. Quantitative Structure activity relationship analysis of coumarins as free radical scavengers by genetic function algorithm. Iranian Chemical Society. 2018 Mar; 2(6):208–222

11. Ambure P, Rahul BA, Agnieszka G, Tomasz P, Kunal R. "NanoBRIDGES" Software: Open Access Tools to Perform QSAR and Nano-QSAR Modeling. Chemical Intelligent Laboratory Systems. 2015; 147: 1–13.

12. Kennard RW, Stone LA. Computer Aided Design of Experiments. Technometrics. 1969 Feb; 11(1):137–48.

13. Friedman JH, Multivariate Adaptive Regression Splines. The Annals of Statistics. 1991: 1–67.

14. Khaled KF, Abdel-Shafi NS. Quantitative structure and activity relationship modeling study of corrosion inhibitors: Genetic function approximation and molecular dynamics simulation methods. International Journal of Electrochemical Science. 2011; 6:4077-4094

15. Brand V, Orr KA, Comprehensive R archive network (CRAN):http://CRAN.Rproject.org. retrieved; 2015

16. Tropsha A. Best Practices for QSAR Model Development, Validation, and Exploitation. Molecular Informatics. 2010 Jul 6; 29(6–7):476–88.

17. 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. 2013; 759:28–42.

18. Myers RH. Classical and modern regression application. 2nd edition. Duxbury Press. CA. 1990

19. Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM, Gramatica P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-baes QSARs. Environmental Health Perspectives 2003; 111:1361-1375

20. Nandi S, Monesi A, Drgan V, Merzel F, Novič M. Quantitative structure-activation barrier relationship modeling for Diels-Alder ligations utilizing quantum chemical structural descriptors. Chemistry Central Journal. 2013; 7:1-13.

21. Gramatica P, Giani E., Papa E., Statistical external validation and consensus modeling: A QSPR case study for KOC prediction. Journal of Molecular Graphics and Modelling. 2007; 25:755-66.

22. Todeschini R, Consonni V. Molecular descriptors for chemo-informatics. Weinheim: Wiley- VCH; 2009. (Methods and principles in medicinal chemistry). ISBN: 9783527318520

23. Adeniji SE, Sani U, Uzairu A, QSAR Modeling and Molecular Docking Analysis of Some Active Compounds against Mycobacterium Tuberculosis Receptor (Mtb CYP121). Journal of Pathogens Hindawi. 2018; 24-64.