The Journal of Engineering and Exact Sciences – JCEC, Vol. 05 N. 01 (2019)
journal homepage: https://periodicos.ufv.br/ojs/jcec
doi: https: 10.18540/jcecvl5iss1pp0125-0136
OPEN ACCESS – ISSN: 2527-1075

THEORETICAL MODELLING FOR INVESTIGATING SOME ACTIVE COMPOUNDS AS POTENT INHIBITORS AGAINST LUNG CANCER: A MULTI-LINEAR REGRESSION APPROACH

A. E. SHOLA[1,*], M.O. IDRIS[1] , S. TUKUR[1], A.U. SAVIOUR[1], G.A. SHALLANGWA[1] AND A. UZAIRU[1]

[1] Ahmadu Bello University, Department of Chemistry, Zaria, Kaduna State, Nigeria.

[*]*Corresponding author. Ahmadu Bello University, Department of Chemistry, Zaria, Kaduna State, Nigeria, Phone: +2347060511720*
e-mail addressl: shola4343@gmail.com (A. E. SHOLA).

ARTICLE INFO

Article history:
Received 2018-03-24
Accepted 2018-12-20
Available online 2019-03-08

keywords
Ciprofloxacin
Descriptor
Genetic Function
Approximation
Lung Cancer
QSAR

ABSTRACT

A Quantitative Structure Activity Relationship (QSAR) study has been attempted on ciprofloxacin derivatives as potent anti-lung cancer. QSAR models were derived with the aid of multi-linear regression (MLR) approach using topological, molecular shape, electronic and structural descriptors. The predictive ability of the QSAR models generated were validated and the best model selected has squared correlation coefficient (R2) of 0.954801, adjusted squared correlation coefficient (Radj) of 0.939265, Leave one out (LOO) cross validation coefficient (Q_cv^2) value of 0.907523. The external validation set used for confirming the predictive power of the model has its R2pred of 0.8387. The QSAR models point out that AATSC2m, VR3_Dzp and BIC2 are the important descriptors effectively describing the bioactivity of these compounds.

1.0 INTRODUCTION

Cancer incidence worldwide has been increasing over the year. Lung cancer (LC) is a disease with a poor prognosis once diagnosed. LC is the leading cause of death in men worldwide and the second cause of mortality in women.

Lung cncer arises from oncogenic alterations in tissues from the respiratory epithelium, namely in bronchi, bronchioles and alveoli (Longo *et al.*, 2012). This cancer results from multiple morphological, molecular and genetic changes, leading to an accumulation of malignant cells (Shahid *et al.*, 2016). LC is mainly classified into two categories, according to its histological characteristics: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC is categorized into three different categories, namely adenocarcinoma, squamous cell carcinoma and large cell carcinoma. NSCLC is the most common LC type (about 80% of total cases) and adenocarcinoma is the most common subtype (about 40%). Moreover, bronchioloalveolar carcinoma subtype is more associated with women and non-smokers. In contrast, squamous cell carcinoma is linked to tobacco

consumption (MacConaill, 2012; Raparia et al., 2013). On the other hand, SCLC tends to affect the neuroendocrine system and is related to smoking habit, being diagnosed in only 1% of non-smokers patients (Mendes *et al.*, 2015).

There are some mediators which may play a predominant role in the treatment of LC, such as Epidermal Growth Factor Receptor (EGFR), Vascular Endothelial Growth Factor (VEGF), (Gold *et al.*, 2012), Anaplastic Lymphoma Kinase (ALK), among others (Lamelas *et al.*, 2012). Ciprofloxacin (CP), an antibiotic has been shown to have anti-proliferative and apoptotic activities in several cancer cell lines (Azéma et al., 2009). Moreover, several reports have highlighted the interest of increasing the lipophilicity to improve the antitumor efficacy.

Synthesis of novel compounds are developed using a trial and error approach which is time consuming and expensive. The advent of computational chemistry led to challenges of drug discovery (Cramer *et al.*, 1988). QSAR establish a relationship

between various molecular properties of molecules and their experimentally known activities (Ibezim *et al.*, 2009). The application of Quantitative Structure Activity Relationship (QSAR) technique to this problem has potential to minimize effort and time required to discover new compounds or to improve current ones in terms of their efficiency. The aim of this research was to develop a QSAR model for predicting the activity of ciprofloxacin derivatives against lung cancer.

2. MATERIALS AND METHOD

2.1 Data set

Data set of ciprofloxacin derivatives as potential anti-lung cancer that were used in this study were obtained from the literature lines (Azéma et al., 2009).

2.2 BIOLOGICAL ACTIVITIES (PIC50)

The Biological activities of ciprofloxacin derivatives against lung cancer measured in IC50 (μM) were converted to logarithm unit (pIC50) using the equation (1) below in order to increase the linearity activities values and approach normal distribution.

$$pIC50 = -\log(IC50) \quad (1)$$

The observed structures and the biological activities of these compounds were presented in Figure 1 and Table 1.

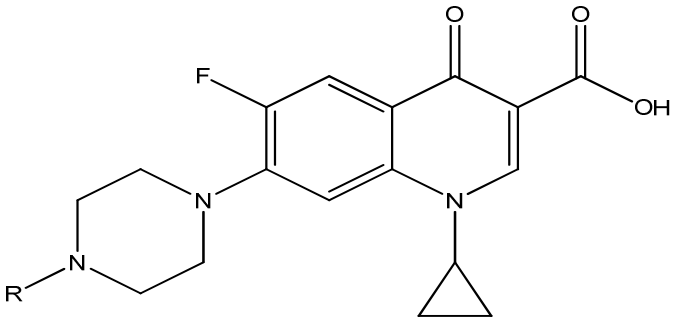

Figure 1: General structure of ciprofloxacin derivatives

Table 1: Molecular structure of ciprofloxacin derivatives as potent anti-prostate cancer

S/N	R	Experimental Activity (pBA)
1[a]	H	280
2	$COCH_2Cl$	10
3	$C(O)OC(CH_3)_3$	18
4	$COCH_2OCOCH_3$	373
5	$COCH_2OCO(CH_2)2CH_3$	456
6[a]	$COCH_2OCO(CH_2)4CH_3$	20
7	$COCH_2OCO(CH_2)6CH_3$	16
8[a]	$COCH_2OCO(CH_2)7CH_3$	216
9[a]	$COCH_2OCO(CH_2)8CH_3$	273
10	$COCH_3$	584
11	$COCH_2\ CH_3$	402
12	$CO(CH_2)_2CH_3$	697

13	$CO(CH_2)_3\ CH_3$	64
14	$COC(CH_3)_3$	509
15	$CO(CH_2)_5CH_3$	808
16	$CO(CH_2)_7CH_3$	6
17	$CO(CH_2)_8CH_3$	3
18 [a]	$CO(CH_2)_{10}CH_3$	7
19	$CO(CH_2)_{12}CH_3$	56
20	$CO(CH_2)_{14}CH_3$	65
21 [a]	$COCH_2C_6H_5$	29
22 [a]	$COCH_2OH$	456

Where superscript **a** represent the test set

2.3 OPTIMIZATION

The 2D structures of the compounds presented in the Table 1 were drawn utilizing chemdraw programming (Adeniji et al., 2018a). The spatial conformations of the compounds were exported from 2D structure to 3D format using the Spartan 14 V1.1.4 Wave Function programming package. All 3D structures were geometrically optimized by minimizing energy. The chemical structures were initially minimized by Molecular Mechanics Force Field (MMFF) count to remove strain energy before subjecting it to quantum chemical estimations. Density Functional Theory (DFT) method was later employed by utilizing the Becke's three parameter exchange functional (B3) hybrid with Lee, Yang and Parr correlation functional (LYP) which is termed (B3LYP) hybrid functional for complete geometric optimization of the structures (Adeniji et al., 2018a).

2.4 MOLECULAR DESCRIPTOR CALCULATION

Molecular descriptors are mathematical values that describe the properties of a molecule. Descriptors calculation for all the inhibitory were calculated using PaDEL-Descriptor software V2.20. A total of 1876 molecular descriptors were calculated (Adeniji et al., 2018a).

2.5 NORMALIZATION AND DATA PRETREATMENT

The descriptors' value were normalized using Equation 2 in order to give each variable the same opportunity at the onset to influence the model (Adeniji et al., 2018b; Singh, 2013)

$$X = \frac{X_1 - X_{min}}{X_{max} - X_{min}} \qquad (2)$$

Where Xi is the value of each descriptor for a given molecule, Xmax and Xmin are the maximum and minimum value for **each** column of descriptors X. The normalized data were subjected to pretreatment using Data

Pretreatment software obtained from Drug Theoretical and Cheminformatics Laboratory (DTC Lab) in order to remove noise and redundant data.

2.6 DATA DIVISION INTO TRAINING AND TEST SET

Kennard and Stone's algorithm approach was employed in this study to divide the data set into a training set and a test compounds in proportion of 70 to 30%. The training set was used to develop the QSAR model while the test was used to confirm the developed model.

2.7 MODEL DEVELOPMENT

MLR is a strategy, utilized for displaying direct relationship between a dependent variable Y (pMIC) and independent variable X (atomic descriptors). The model is fit such that sum-of square difference between the experimental and predicted values of set biological activity is minimized. In regression analysis, contingent mean of dependent variable (pMIC) Y relies on (Descriptors) X. MLR examination extends this thought to incorporate more than one autonomous variable and regression equation takes the form.

$$Y = b_1x_1 + b_2x_2 + b_3x_3 + C \quad (3)$$

Where Y is dependent variable, 'b's are regression coefficients for corresponding 'x's (independent variable), and 'C' is a regression constant or intercept.

2.8 VALIDATION OF MODEL

Validation of the model was carried out using Material studio software version 8 using Genetic Function Approximation (GFA) method. The numbers of descriptors in the regression equation were three; population and generation were set to 10000 and 10000, respectively. The numbers of top equations returned were four. Mutation probability was 0.1, and the smoothing parameter was 0.5. The models were scored based on Friedman's LOF. In GFA algorithm, an individual or model was represented as one-dimensional string of bits. It was a distinctive characteristic of GFA that it could create a population of models rather than a single model.

The models were estimated using the LOF, which was measured using a slight variation of the original Friedman formula, so that the best fitness score can be received. In materials studio version 8, LOF is measured using a slight variation of the original Friedman formula. The revised formula is:

$$LOF = \frac{SEE}{\left(1 - \frac{C + d \times p}{M}\right)^2} \quad (4)$$

Where:

SEE is the Standard Error of Estimation which is equivalent to the model's 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 below;

$$SEE = \sqrt{\frac{\left(Y_{exp} - Y_{pred}\right)^2}{N - P - 1}} \quad (5)$$

c is the number of terms in the model, other than the constant term, d is a user-defined smoothing parameter, p is the total number of descriptors contained in the model and M is the number of data in the training set (Adeniji et al., 2018a)

The square of the correlation coefficient (R^2) describes the fraction of the total variation attributed to the model. The closer the value of R^2 is to 1.0, the better the regression equation explains the Y variable. R^2 is the most commonly used internal validation indicator and is expressed as follows:

$$R^2 = 1 - \left[\frac{\Sigma\left(Y_{exp} - Y_{pred}\right)^2}{\Sigma\left(Y_{exp} - \overline{Y}_{training}\right)^2}\right] \quad (6)$$

Where:

Y_{exp}, Y_{pred} and $\overline{Y}_{training}$ training are the experimental activity, the predicted activity and the mean experimental activity of the samples in the training set, respectively.

R^2 value varies directly with the increase in number of repressors i.e. descriptors, thus, R^2 cannot be a useful measure for the stability of model. Therefore, R^2 is adjusted for the number of explanatory variables in the model. The adjusted R^2 is defined as:

$$R^2_{adj} = \frac{R^2 - P\,(n-1)}{n - p + 1} \quad (7)$$

Where p = number of independent variables in the model.

The capability of the QSAR equation to predict bioactivity of new compounds was determined using the leave-one-out cross validation method. The cross-validation regression coefficient (Q^2_{cv}) was calculated with the equation below:

$$Q^2_{cv} = 1 - \left[\frac{\Sigma\left(Y_{pred} - Y_{exp}\right)^2}{\Sigma\left(Y_{exp} - \overline{Y}_{training}\right)^2}\right] \quad (8)$$

Where
Y_{pred}, Y_{exp}, and $\overline{Y}_{training}$ are the predicted, experimental and mean values of experimental activity of the training set.
The coefficient of determination for the test set R^2_{test} was calculated with the equation below;

$$R^2_{test} = 1 - \frac{\Sigma\left(Ypred_{test} - Y_{exp_{test}}\right)^2}{\Sigma\left(Ypred_{test} - \overline{Y}_{training}\right)^2} \quad (9)$$

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

2.8 EVALUATION OF THE APPLICABILITY DOMAIN OF THE MODEL

The built QSAR model was evaluated based on applicability domain approach in order establish that the model is robust and reliable to predict the activities of the inhibitor compounds. The leverage approach was employed in defining and describing the applicability domain of the built QSAR models. Leverage of a given chemical compound hi, is defined as follows:

$$hi = X_i\,(X^T X)^{-1}\,X_i^T \qquad (10)$$

Where Xi is training compounds matrix of i. X is the m× k descriptor matrix of the training set compound and X^T is the transpose matrix of X used to build the model. The warning leverage (h^*) is the boundary of values for X outliers and is defined as:

$$h^* = 3\frac{(d+1)}{m} \qquad (11)$$

Where m is the descriptors and d is the compound that made up the training set. (Adeniji et al., 2018a)

2.9 Y-RANDOMIZATION TEST

To guarantee the created QSAR model is strong and not inferred by chance, the Y-randomization test was performed on the training set data as suggested by (Tropsha *et al*., 2003). Random MLR models are generated by randomly shuffling the dependent variable (activity data) while keeping the independent variables (descriptors) unaltered. The new QSAR models are expected to have significantly low R^2 and Q^2 values for several trials which confirm that the developed QSAR models are robust. Another parameter, cR_p^2 is also calculated which should be more than 0.5 for passing this test.

$$cR_p^2 = R \times [R^2 - (R_r)^2]^2 \qquad (12)$$

Where,
cR_p^2 is Coefficient of determination for Y-randomization, R is coefficient of determination for Y-randomization and Rr is an average 'R' of random models.

2.10 QUALITY ASSURANCE OF THE MODEL

The fitting ability, stability, reliability and predictive ability of the developed models were evaluated by internal and external validation parameters. The validation parameters were compared with the minimum recommended value for a generally acceptable QSAR model (Adeniji et al., 2018b; Veerasamy et al., 2011) showed in Table 3.

3.0 RESULTS AND DISCUSSION

A QSAR examination was performed to investigate the structure activity relationship of 22 compounds as potent anti-lung cancer agents. The nature of models in a QSAR study is expressed by its fitting and forecast capacity. In order to assemble a decent QSAR model for anti-lung cancer with good predictive power for the selected test set. Kennard-Stone algorithm was used to divide the dataset of 22 compounds into a training set of 15 compounds which was used to developed the model and a test set of 7 compounds which was applied to assess the predictive ability-built model.

Experimental and Predicted activity for ciprofloxacin derivatives as a potent anti-lung cancer and the residual values were presented in Table 2. The low residual value between Experimental and Predicted activity indicates that the model is of high predictability.

The Genetic Algorithm- Multi Linear Regression (GA–MLR) investigation led to the selection of three descriptors which were used to assemble a linear model for calculating predictive activity on lung cancer. Four QSAR models were built using Genetic Function Algorithm (GFA), but due to the statistical significance, model 1 was selected, reported and its s parameters were as well calculated.

Model 1
Y = - 0.260777641* AATSC2m - 1.673908378 * VR3_Dzp + 0.431577310 * BIC2 + 0.174310823

Model 2
Y = - 0.301042455 * AATSC2m + 0.354110306 * CIC4 - 0.007605618 * GGI7 - 5.484268669

Model 3
Y = - 0.301042455 * AATSC2m + 0.354110306 * CIC4 - 0.007605618 * CIC4 - 1.943165607

Model 4
Y = - 0.260777641 * AATSC2m - 1.673908378 * VR3_Dzp + 0.431577310 * BIC2 + 4.490083927

All the validation parameters for the optimum model were reported in Table 4 and were all in agreement with parameters presented in Table 3 which actually confirmed the robustness of the model.

Table 2: Experimental, Predicted and Residual values of ciprofloxacin derivatives

Molecule	Experimental Activity (pIC_{50})	Predicted activity	Residual
1 [a]	3.552842	3.54632	0.006522
2	5.000000	5.09629	-0.09629
3	4.744727	4.754117	-0.00939
4	3.428291	3.247843	0.180448
5	3.341035	3.310391	0.030644
6 [a]	4.69897	4.7533	-0.05433
7	4.79588	4.6833	0.11258
8 [a]	3.665546	3.889306	-0.22376
9 [a]	3.563837	3.567258	-0.003421
10	3.233587	3.330907	-0.09732
11	3.395774	3.315635	0.080139
12	3.156767	3.119305	0.037462
13	4.19382	4.128542	0.065278
14	3.293282	3.371872	-0.07859
15	3.092589	3.156149	-0.06356
16	5.221849	5.132853	0.088996
17	5.522879	5.672869	-0.14999
18 [a]	5.154902	5.146353	0.008549
19	4.251812	4.243278	0.008534
20	4.187087	4.222457	-0.03537
21 [a]	4.537602	4.614322	-0.07672
22 [a]	3.341035	3.264104	0.076931

Where superscript **a** represent the test set

Table 3: Minimum recommended value of Validation Parameters for a generally acceptable QSAR model

Symbol Value	Name	Value
$\mathbf{R^2}$	Coefficient of determination	≥ 0.6
$\mathbf{P_{(95\%)}}$	Confidence interval at 95% confidence level	< 0.05
$\boldsymbol{Q_{cv}^2}$	Cross validation coefficient	> 0.5
$\mathbf{R^2 - Q_{cv}^2}$	Difference between R^2 and Q_{cv}^2	≤ 0.3
$\mathbf{N_{ext.\ test\ set}}$	Minimum number of external test set	≥ 5
$\mathbf{cR_p^2}$	Coefficient of determination for Y-randomization	> 0.5

Table 4: Validation parameters of the Genetic Function Approximation from material studio

S/N	Validation Parameters	Model 1	Model 2	Model 3	Model 4
1	Friedman LOF	0.02864	0.034616	0.040668	0.046369
2	R-squared	0.954801	0.920137	0.853168	0.843168
3	Adjusted R-squared	0.939265	0.919265	0.830396	0.810396
4	Cross validated R-squared	0.907523	0.897523	0.80322	0.76344
5	Significant Regression	Yes	Yes	Yes	Yes
6	Significance-of-regression F-value	88.314223	83.65342	74.627565	71.457543
7	Critical SOR F-value (95%)	3.748716	3.748716	3.748716	3.748716
8	Replicate points	0	0	0	0
9	Computed experimental error	0	0	0	0
10	Lack-of-fit points	11	11	11	11
11	Min expt. error for non-significant LOF (95%)	0.065834	0.06867	0.07043	0.071357
12	R^2 test	0.8387	0.7467	0.7198	0.6216

The name and symbol of the descriptors used in the QSAR optimization model was reported in Table 5. The presence of the three 2D descriptors in the model suggests that these types of descriptors are able to characterize better anti-lung cancer activities of the compounds.

Pearson's correlation matrix and statistics of the three descriptors employed in the QSAR Model were reported in Table 6. Which shows clearly that the correlation coefficients between each pair of descriptors is very low thus, it can be

inferred that there exists no significant inter-correlation among the descriptors used in building the model. The Mean Effect (ME) reported in Table 6 provides important information on the effect of the molecular descriptors and the degree of contribution in the developed model. The signs and the magnitude of these descriptors combined with their mean effects indicate their individual strength and direction in influencing the activity of a compound. The estimated Variance Inflation Factor (VIF) values for all the descriptors were less than 4 which imply that the model generated was statistically significant and the descriptors were orthogonal. The p-value is a probability that measures the evidence against the null hypothesis. Lower probabilities provide stronger evidence against the null hypothesis. The null hypothesis implies that there is no association between the descriptors and the activities of the molecules. The P-values of all the descriptors in the model at 95% confidence level shown in Table 6 are less than 0.05. This implies that the alternative hypothesis is accepted. Hence there is a relationship between the descriptors used in the model and the activities molecules which take preference over the null hypothesis.

Y- Randomization parameter test was reported in Table 7. The low R^2 and Q^2 values for several trials confirm that the developed QSAR model is robust. While the cR_p^2 value greater than 0.5 affirms that the created model is powerful and not inferred by chance.

Table 5: List of some descriptors used in the QSAR optimization model

S/NO	Descriptors symbols	Name of descriptor(s)	Class
1	AATSC2m	Average Broto-Moreau autocorrelation - lag 2 / weighted by mass	2D
2	VR3_Dzp	Logarithmic Randic-like eigenvector-based index from Barysz matrix / weighted by polarizabilities	2D
3	BIC2	Bond information content index (neighborhood symmetry of 2-order)	2D

Table 6: Pearson's correlation matrix and statistics for descriptor used in the QSAR optimization model

Inter-correlation				Statistics		
Descriptors	*AATSC2m*	VR3_Dzp	BIC2	ME	VIF	P- value
AATSC2m	1			-0.6346	1.43322	4.342E-04
VR3_Dzp	-0.13208	1		0.2455	2.32121	2.566E-07
BIC2	-0.11093	0.362855	1	-0.5428	1.24554	2.345E-05

Table 7: Y- Randomization Parameters test

Model	R	R^2	Q^2
Original	0.954869	0.936082	0.900868
Random 1	0.557607	0.310925	-0.23015
Random 2	0.399318	0.159455	-0.54789
Random 3	0.366365	0.134224	-0.58322
Random 4	0.823539	0.678216	0.368781
Random 5	0.511228	0.261355	-0.42421
Random 6	0.346023	0.119732	-0.53935
Random 7	0.243492	0.059288	-1.17941
Random 8	0.671342	0.4507	-0.17635
Random 9	0.552076	0.304788	-0.47477
Random 10	0.345377	0.119285	-0.63643

Random Models Parameters

Average r :	0.481637
Average r^2 :	0.259797
Average Q^2 :	-0.4423
cRp^2 :	0.852239

Plot of predicted activity against experimental activity of training and test set where shown in Figure 2 and Figure 3 respectively. The R^2 value of 0.9548 for training set and R^2 value of 0.8387 for test set reported in this study was in agreement with Genetic Function Approbation (GFA) derived R^2 value reported in Table 3 which confirms the robustness and reliability of the model. Plot of standardized residual versus experimental activity shown in Figure 4 indicates that there was no systematic error in the model built as the spread of standardized residual values were on both sides of zero (Adeniji et al., 2018b). The leverage values for the entire compounds in the dataset were plotted against their standardized residual values leading to discovery of outliers and influential compound in the models. The Williams plot of the standardized residuals versus the leverage value is shown in Figure 5. From our result it is an evident that all the compounds were within the square area ± 3 of standardized cross-validated residual produced by the model. Therefore, no compound is said to be an outlier. However, only one compound (molecule number 15) is said to be an influential compound since its leverage value is greater than the warning leverage ($h^* = 0.80$). This was attributed to difference in its molecular structure compared to other compounds in the dataset.

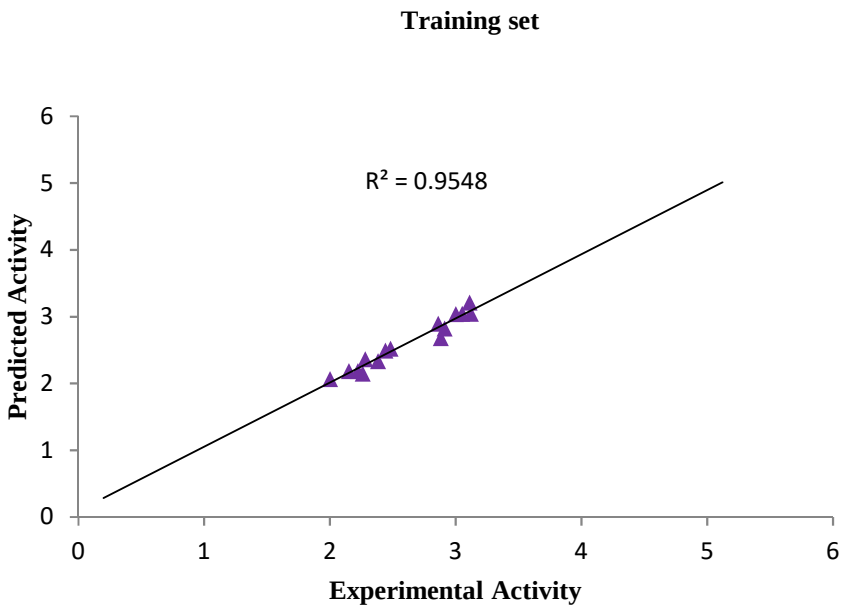

Figure 2: Plot of predicted activity against experimental activity of training set

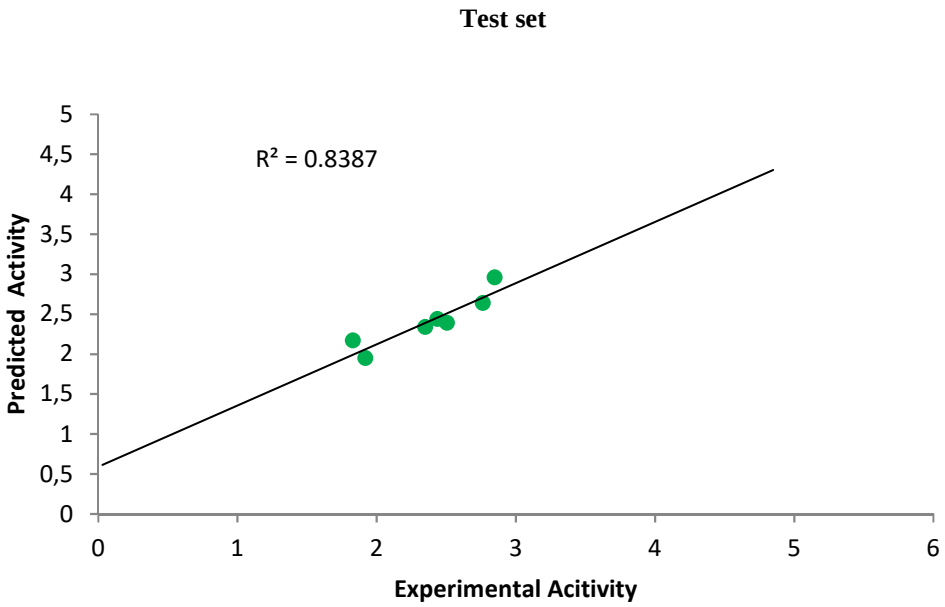

Figure 3: Plot of predicted activity against experimental activity of test set

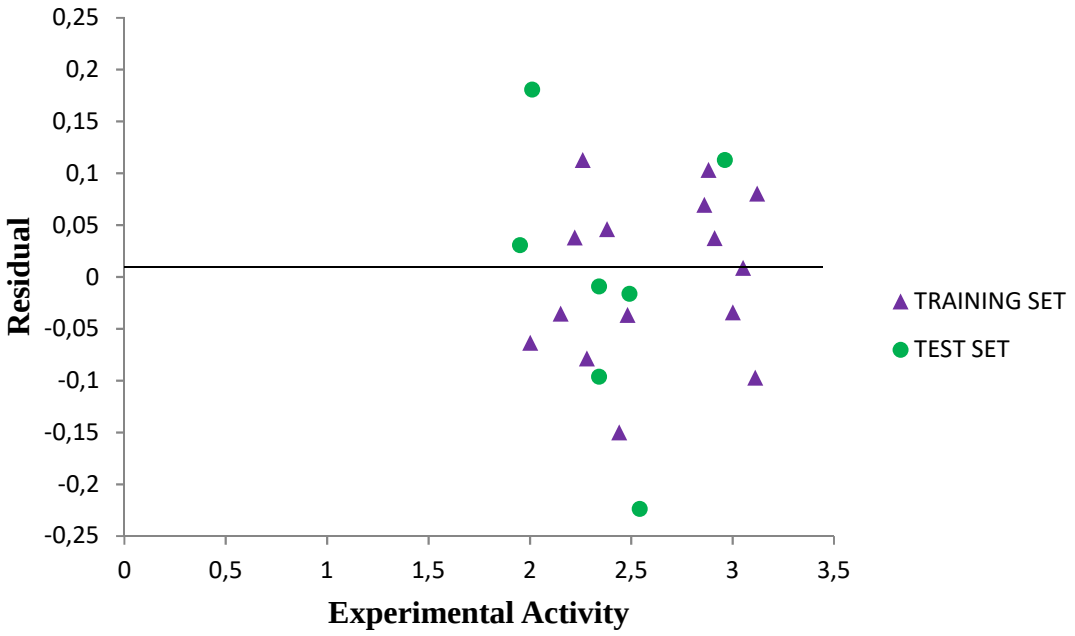

Figure 4: Plot of residual values versus experimental activity

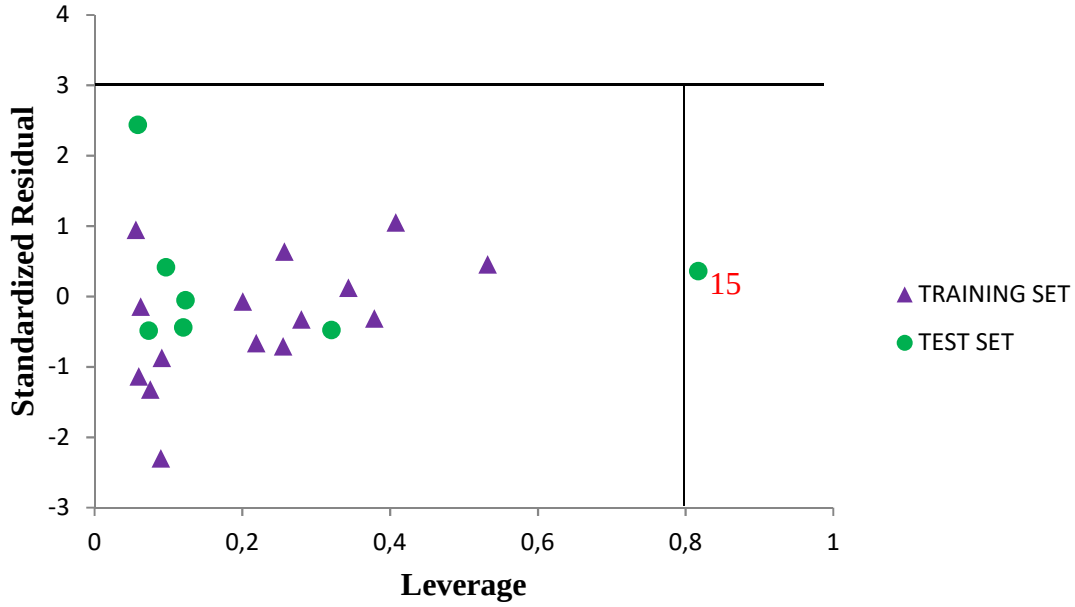

Figure 5: Plot of Standardized residual activity versus experimental activity

4. CONCLUSION

QSAR analysis on a series of ciprofloxacin derivatives was carried out using the GFA technique. The best model was selected based on the statistical parameters. The internal and external validation test for the QSAR model generated agreed with recommended value of validation parameters for a generally acceptable QSAR model. Thus, the descriptors, AATSC2m, VR3_Dzp and BIC2 in the built model are important descriptors to determine the activity of the compounds to function as effective lung cancer inhibitors. This knowledge can be used for designing more effective chemical entities and may also provide important insights into structural variants leading to the development of novel lung cancer inhibitors.

REFERENCES

ADENIJI, S.E.; ARTHUR, D.E.; OLUWASEYE, A.. Computational modeling of 4-Phenoxynicotinamide and 4-Phenoxypyrimidine-5-carboxamide derivatives as potent anti-diabetic agent against TGR5 receptor. *Journal King Saud University-Science*, 2018a

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*, v6, p.479–492, 2018b

AZÉMA, J.; GUIDETTI, B.; DEWELLE, J.; LE CALVE, B.; MIJATOVIC, T.; KOROLYOV, A.; VAYSSE, J.; MALET-MARTINO, M.; MARTINO, R.; KISS, R. 7-((4-Substituted) piperazin-1-yl) derivatives of ciprofloxacin: synthesis and in vitro biological evaluation as potential antitumor agents. *Bioorganic. Medical Chemistry*, v17, p.5396–5407, 2009.

CRAMER, R. D., PATTERSON, D. E., & BUNCE, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society*, v. 110, p. 5959–5967, 1998.

GOLD, K.A.; WISTUBA, I.I.; KIM, E.S. New strategies in squamous cell carcinoma of the lung: identification of tumor drivers to personalize therapy. *Clinical Cancer Research*, v18, p. 3002–3007, 2012.

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

LAMELAS, I.P.; ARCA, J.A.; PÉREZ, J.L.F.. Directed therapies in lung cancer: new hope? *Bronconeumol. Engl. Ed*. v 48, p. 367–371, 2012.

LONGO, D.L.; FAUCI, A.S.; KASPER, D.L.; HAUSER, S.L.; JAMESON, J.L.; LOSCALZO, J. Harrison's Principles of Internal Medicine 18E Vol 2 EB. McGraw Hill Professional. 2012

MACCONAILL, L.E. Advancing personalized cancer medicine in lung cancer. *Pathology Laboratory Medica*l, v136, p. 1210–1216, 2012.

MENDES, F.; ANTUNES, C.; ABRANTES, A.; GONCALVES, A.; NOBRE-GOIS, I.; SARMENTO, A.; BOTELHO, M.; ROSA, M. Lung cancer: the immune system and radiation. *Journal of Biomedical Scence*, v72, p. 78–84, 2015.

RAPARIA, K.; VILLA, C.; DECAMP, M.M.; PATEL, J.D.; MEHTA, M.P. Molecular Profiling in Non–Small Cell Lung Cancer: A Step Toward Personalized Medicine. *Pathology Laboratory Medica*l, v137, p. 481–491., 2013.

SHAHID, M.; CHOI, T.G.; NGUYEN, M.N.; MATONDO, A.; JO, Y.H.; YOO, J.Y.; NGUYEN, N.N.Y.; YUN, H.R.; KIM, J.; AKTER, S. An 8-gene signature for prediction of prognosis and chemoresponse in non-small cell lung cancer. Oncotarget 7, 8656, 2016.

SINGH, P. Quantitative Structure-Activity Relationship Study of Substituted-[1, 2, 4] Oxadiazoles as S1P1 Agonists. Journal of *Current Chemistry and Pharmaceutical Science*. v3, 2013

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. *Molecular Informatics*, v.22, p.69–77, 2003.

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*. v3, p. 511–519, 2011.