Moroccan Journal of Chemistry
ISSN: 2351-812X
http://revues.imist.ma/?journal=morjchem&page=login
Elijah Adeniji & al. / Mor. J. Chem. 7 N°4 (2019) 697-713

Computational investigation and molecular docking simulation of some bioactive compounds as potent inhibitors against tuberculosis receptor

Shola Elijah Adeniji, Sani Uba and Adamu Uzairu

Department of Chemistry, Ahmadu Bello University, Zaria-Nigeria

** Corresponding author:*
shola4343@gmail.com
Received 09 April 2019,
Revised 13 March 2019,
Accepted 12 Dec 2019

Abstract

Multi-drug resistant strains of *Mycobacterium Tuberculosis* still remain a major challenge toward the first hand drugs for treating tuberculosis. Development and syntheses of novel compounds with more potent anti-tubercular agents are usually by trial approach with lots of errors which is time consuming and expensive. QSAR is a theoretical approach which has the potential to reduce the aforementioned problem in discovering new potent drugs against *M. Tuberculosis*. This approach was employed to develop multivariate QSAR model to correlate the chemical structures of the 1, 2, 4-Triazole analogues with their observed activities using a theoretical approach. In order to build the robust QSAR model, Genetic Function Approximation (GFA) was employed as a tool for selecting the best descriptors that could efficiently predict the activities of the inhibitory agents. The developed model was validated through internal and external validation test. Molecular docking studies was as well carried for all the studied compounds in order to show the interactions and binding modes between the ligand and the receptor (DNA gyrase). The lead compound (compound 3) with higher anti-tubercular activity was observed with prominent binding affinity of -11.8 kcal/mol. Therefore, compound 3 could serve as a template structure to designed compounds with more efficient activities. The outcome of this research is recommended for pharmaceutical and medicinal chemists to design and synthesis more potent compounds with prominent anti-tubercular activities using the model designed in this study.

Keywords: Anti-tuberculosis, Binding Affinity, DNA gyrase, Molecular Docking, QSAR

1.Introduction

Tuberculosis (TB) is the most deadly bacterial disease caused by specie of bacteria known as *Mycobacterium Tuberculosis*. In (2013), World Health Organization (WHO) estimated death of 1.5 million people, 9.0 million people living with tuberculosis and 360,000 people whom were HIV positive [1]. At present, pyrazinamide (PZA), para-amino salicylic acid (PAS), isoniazide (INH) and rifampicin (RMP) are the current drugs administered to patient suffering from tuberculosis. The resistances of the *M. tuberculosis* toward the current drugs led to development of new approach that is fast and precise which could able to predict the biological activity for the new compounds against *M. tuberculosis.* Meanwhile, a theoretical approach; quantitative structure activity relationships (QSARs) is one of the most widely used computational method which helps in drug designing and prediction of drugs activities [2]. QSAR model is a mathematical linear equation which relates the molecular structures of the compounds and their biological activities. In this research, a data set of , 2, 4-Triazole derivatives which had been synthesized and evaluated as anti-*Mycobacterium tuberculosis* [3] have been selected for QSAR study. Few researchers [4–7] have established relationship between some anti-tubercular inhibitor's like quinolone, chalcone, pyrrole and 7-methyijuglone using QSAR approach. However, QSAR alongside with molecular docking simulation study have not been fully established to relate the structures and activities of the inhibitory compounds as well as the interaction mode with the receptor (DNA gyrase. Therefore, this study aimed to establish a valid QSAR model that could correlate the structures of 1, 2, 4-Triazole derivatives, molecular docking simulation and to design new potent compounds with better anti-tubercular activities against *Mycobacterium tuberculosis*. The aim of this research was to build a QSAR model that will predict the activities for 1, 2, 4-Triazole derivatives against M. tuberculosis and to carry out molecular docking studies to elucidate the kind of interaction existing between the inhibitory compounds and the target site (DNA gyrase) of M. tuberculosis.

2.0 Materials and method

2.1 Data collection

Thirty (30) molecules of 1,2,4-Triazole derivatives as potent anti-*mycobacterium tuberculosis* that were used in this study were obtained from the literature [3]. The observed structures and the biological activities of these compounds were presented in Table 1.

2.2 Structure optimization

In order for the molecules to attain a stable conformer at a minimal energy, all the molecules were geometrically optimized with the aid of Spartan 14 V1.1.4 by employing Molecular Mechanics Force Field (MMFF) count to remove strain energy and later subjected to Density Functional Theory (DFT) by utilizing the (B3LYP) basic set [5].

2.3 Molecular descriptor calculation

Descriptor is a mathematical logic that describes the properties of a molecule based on the connection between the structure of the compound and its biological activity. Descriptors calculation for all the inhibitory compounds was achieved using PaDEL-Descriptor software V2.20. A total of 1876 molecular descriptors were predicted.

2.4 Normalization of data and pretreatment

The values for the predicted descriptors' were normalized using Equation 1 below so that each variable will have the same prospect at the inception so as to sway the model [8].

$$Y = \frac{Y_1 - Y_{min}}{Y_{max} - Y_{min}} \quad \text{(Eq.1)}$$

Where Y_1 is the descriptor value for each molecule, Ymin and Ymax are the minimum and maximum value for each descriptors column of Y. After successful normalization of the data, the data were further subjected to pretreatment using in order to remove noise and redundant data.

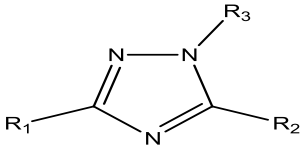
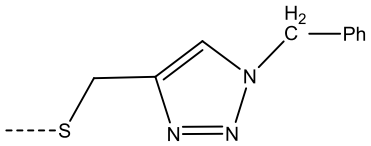
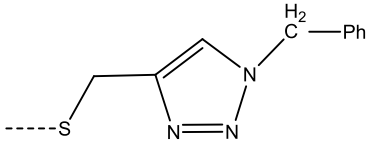
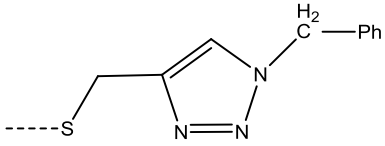
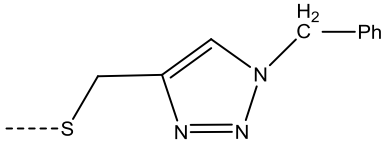
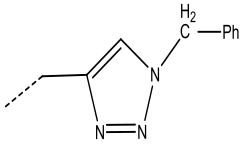
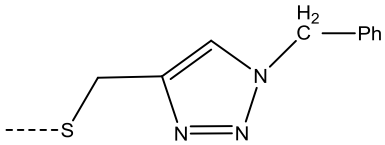
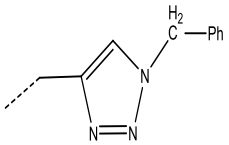
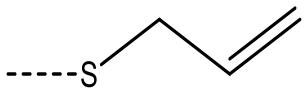
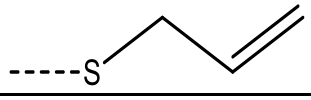
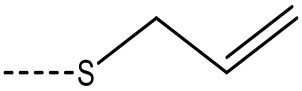
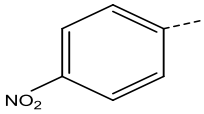
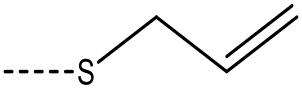
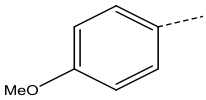
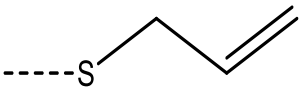
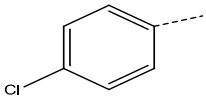
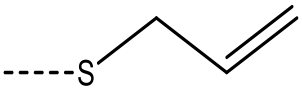
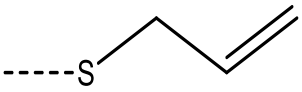
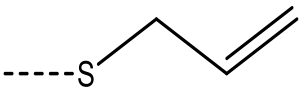
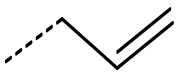
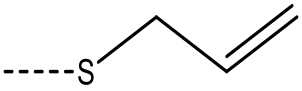
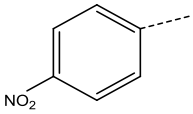
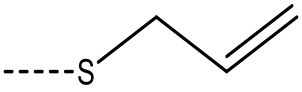
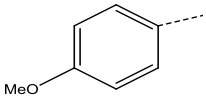
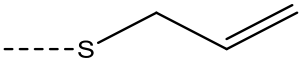
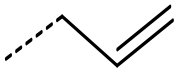
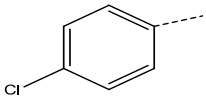
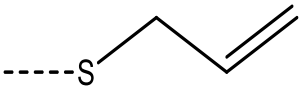
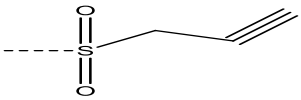
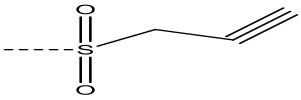
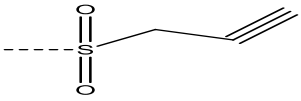
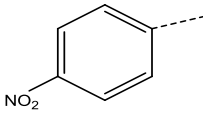
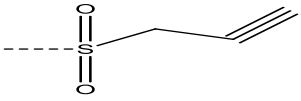
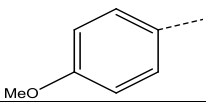
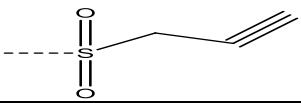
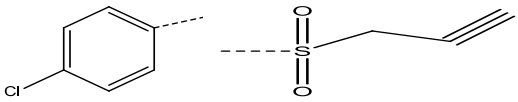
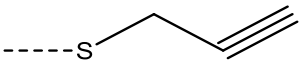
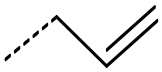
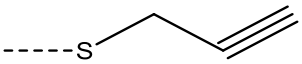
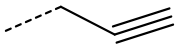
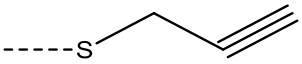
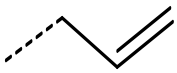
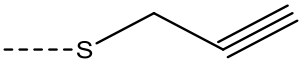
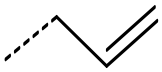
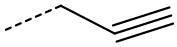
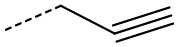


Figure 1: General structure of 1, 2, 4-Triazole derivatives

Table 1: Molecular structures of 1, 2, 4-Triazole derivatives and their activities as potent anti- mycobacterium tuberculosis.

Molecule	R_1	R_2	R_3	Activity (pBA)
1	H		H	8.0250
2	methyl		H	8.0345
3	isobutyl		H	8.7064
4	Methyl			5.7441
5	isobutyl			5.9258
6	H		H	6.1667
7	methyl		H	5.8765

8	isobutyl		H	6.4171
9			H	5.9413
10			H	6.6397
11			H	8.0899
12	H			6.5267
13	Methyl			5.7405
14	isobutyl			5.6533
15				6.1923
16				7.3233
17				6.0097
18	H		H	6.0928
19	methyl		H	7.3279
20	isobutyl		H	6.8568
21			H	6.2234
22			H	7.0079

23			H	7.314
24	isobutyl			7.0854
25	isobutyl			7.2615
26	H			5.2346
27	Methyl			6.4218
28	Br	Br		5.1016
29	Br	Br		6.1213
30	Br	Br		5.4406

2.5 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.

2.6 Development of the model

Multi-linear regression approach (MLR) is a strategy used to develop the QSAR. MLR display a direct relationship between the dependent variable Y (activity) and independent variable X (descriptors). In MLR analysis, the mean of the dependent variable Y relies on X . MLR equation below is used to incorporate more than one independent variable (Descriptors) with a single response variable (activity). Where Y represent the dependent variable, represent the independent variables, 'k's are regression coefficients for each '*x*'s and 'C' is a regression intercept.

$$Y = k_1x_1 + k_2x_2 + k_3x_3 + C \qquad \text{(Eq.2)}$$

2.7 Generation of QSAR Model and validation

The combinations of the optimum descriptors for the training set were obtained from the descriptor pool using the Genetic Function Approximation technique. Their anti-lung cancer activities were placed as the last column in their respective spread sheets in Microsoft Excel 2010 which were later imported into the Material Studio software version 8.0 to generate the QSAR model by employing Multi-linear regression Approach (MLR) and to evaluate the internal validation parameters.

2.8 Determination of outlier and influential molecule (applicability domain)

The applicability domain approach was employed to determination of outlier and influential molecule. Any compound outside the applicability domain space of ±3 is said to be an outlier. To define and describe the applicability domain of the built QSAR models, the leverage h_i approach was employed and defined as; [9].

$$h_i = M_i \left(M^T M\right)^{-1} M_i^T \quad \text{(Eq.3)}$$

M_i is training set matrix of *i*. *M* is the n× k descriptor matrix of the training set compound and M^T is the transpose of the training set (*M*). M_i^T is the transpose matrix M_i used to build the mode. The warning leverage h^* is the limit values to check for influential molecule. The warning leverage h^* is defined as;

$$h^* = 3\frac{(j+1)}{N} \quad \text{(Eq.4)}$$

Where *j* is the number of descriptors in the build model and *N* is the number of compounds that made up the training set.

2.9 Assessment of Y-Randomization

Y-Randomization test is a confirmatory test to show that the developed QSAR model created is reliable, strong, robust and not gotten by chance. This test was performed on the training set data as described by [10]. Multi-linear regression (MLR) models were generated by randomly shuffling of the dependent variable (activity data) while keeping the independent variables (descriptors) unaltered. It is expected that the developed QSAR model should have significantly low R^2 and Q^2 values for numbers of trials in order to ascertain that the developed QSAR models is robust. For the built model to be valid, the value for Y-randomization Coefficient (cR_p^2) must be equal or greater than 0.5.

$$cR_p^2 = R \times \left[R^2 - (R_r)^2\right]^2 \quad \text{(Eq.5)}$$

Where,

cR_p^2 is Y-randomization Coefficient, R is correlation coefficient for *Y*-Randomization and Rr is average 'R' of random models.

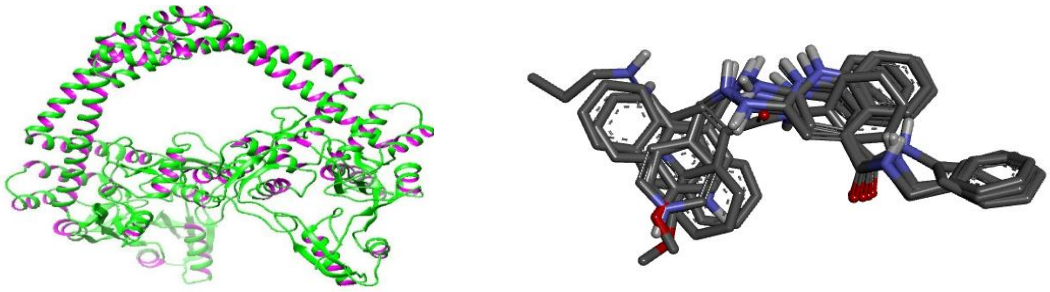
2.10 Affirmation of the build model

The fitting ability, stability, reliability, predictive and robustness of the developed models were evaluated by internal and external validation parameters. The validation parameters were compared with the accepted threshold value for any QSAR model [9–12] shown in Table 2.

2.11 Docking studies

2.11.1 Preparation of the receptor (DNA gyrase)

The DNA gyrase receptor was gotten from protein data bank with PDB code 31FZ in form crystal structure shown in Figure 1. Foreign bound substances such as ligands, solvent molecules and cofactors associated with the receptor were removed with the aid of Discovery Studio Visualizer software. The prepared receptor was saved in an input format (PDB) which is recognized by Discovery Studio Visualizer and Pyrx software. In the Pyrx software, the prepared receptor was transported in order to become a macro molecule.



A B

Figure 1: (A) Crystal structure of DNA gyrase. (B) prepared ligands

2.11 2 Preparation of the ligands

All the studied compounds (1, 2, 4 Triazole derivatives.) were optimized in order to have a stable conformer at a minima energy with the aid of Spartan 14 software at Density Functional Theory (DFT) level. After optimization, the ligand was saved as a PDB format, imported into the Pyrx software in order to become micro molecules (ligands).

2.11.3 Receptor-ligand complex docking

Molecular docking of the receptor with the ligands were carried out utilizing the PyRx virtual screening software. The software comprises the combination of several software's such as AutoDock Vina 4.2, Open Babel and Mayavi, etc. In order to perform receptor-ligand docking, the ligands and the receptor (DNA gyrase) were converted from pdb format to pdbqt (protein data bank, partial charge and atom type) format. The conversion of pdb format to pdbqt format (Vina input format) was done by launching the PyRx virtual screening software in order compute the Binding Affinity (kcal/mol). The more the negative the binding affinity, the better the orientation of the ligand in the binding site of DNA gyrase. The docked results were visualized, analyzed and compiled with the aid of Discovery Studio Visualizer software.

3.0 Results and discussion

A theoretical approach was employed to derive a QSAR model for predicting the activities of 1, 2, 4 Triazole analogues against *Mycobacterium tuberculosis*. Kennard-Stone algorithm approach employed in this research was able to divide the studied compounds which comprises of 30 compounds into a training set of 21 compounds while the remaining 9 compound serve as the test set. The model generated was built on the basis of the training set while validation of the model was accessed by the test set. The combination of the utmost descriptors that could better predict the activities of the inhibitory compounds were selected with the approach of Genetic Function Algorithm (GFA) while multi-linear Regression (MLR) method was used as modeling technique in generating the QSAR model. GFA-MLR led to selection of four descriptors and four QSAR models as shown below. However, based on the internal and external validation parameters reported in Table 5 for a valid QSAR model to robust and have a optimum predictive ability, model was selected as best model. These parameters were in agreement with the threshold value reported in Table 2 which actually confirmed the robustness and stability of the model.

Model 1

pBA = - 1.907001458 (AATS7s) + 1.310767463 (nHBint3) - 1.483029601 (minssCH2) + 5.089381479 (RDF90i) + 3.442395171

Model 2

pBA = 2.133597608 (nHBint3) - 1.313467537 (nHBint7) + 6.924567876 (E1i) + 0.972490084 (Vi) + 2.469905087

Model 3

pBA = -1.935644981 (AATS7s) + 2.133204345 (nHBint3) - 1.484532032 (minssCH2) + 0.964401389 (Vi) + 5.463456786

Model 4

pBA = - 1.913434667 (AATS7s) + 2.131200932 (nHBint3) + 3.430925453 (GATS4v) + 5.083434501 (RDF90i) + 3.230945311

The QSAR model generated in this research was compared with the models obtained in the literature [4,5] as shown below;

pBA = - 0.307001458(**AATS7s**)
+ 1.528715398(**nHBint3**)
+ 3.976720227(**minHCsatu**)
+ 0.016199645(**TDB9e**)
+ 0.089381479 (R**DF90i**)
- 0.107407822(**RDF110s**) + 4.057082751

N_{train} = 35, R^2 = 0.92023900, R_{adj} = 0.91017400, Q^2_{cv} = 0.89538600 and the external validation for the test set was found to be R^2pred = 0.8842 [5]

pIC50 = −2.040810634 * nCl
− 19.024890361 * MATS2m
+ 1.855704759 * RDF140s + 6.739013671

N_{train} = 27, R^2 = 0.9480, R_{adj} = 0.9350, Q^2_{cv} = 0.87994 and R^2pred = 0.7690 [4]

From the above models the validation parameters reported in this work and those reported in the literature were all in agreement with parameters presented in Table 2 which actually confirmed the robustness of the model generated. In order to validate the Kennard and Stone algorithm employed in this studies, the activity values of the training and test set data were subjected descriptive analysis which was reported in Table 3. It could be observed that the range value of the training set (3.1365) was almost equivalent to range value of the test set (3.1448). Moreover, the standard deviation and mean of the activity values of test set (1.2336626.and 336856) were approximately the same to that of the training set value (0.930345 and 6.660548). Based on this observations, it can be infer that the algorithm used was able to obtain a test set that is a good reflection of the training set since the test set is interpolative within the training. The observed activities, predicted activities of the inhibitors, and the residual values for each compound were reported in Table 4. The low residual values between observed activities and predicted activities indicate that the model generated has a high predictive ability. The names and symbols of each descriptors selected by GFA approach were presented in Table 6. The combination of the selected descriptors (2D and 3D) reported in model 1 indicates that these types of descriptors are able characterize and give better information on the structure of the anti-tubercular molecules. The Person correlation coefficients calculated for the descriptors in the model was reported in Table 7. The low correlation coefficients that exist between each descriptor in the model imply that there exists no significant inter-correlation between each descriptor. Statistical parameters calculated for the selected descriptors reported in Model 1 were repented in Table 7. The descriptors

were subjected to Variance Inflation Factor (VIF) in order to check for orthogonality. Meanwhile, the VIF values for each descriptor shown in Table 4 were less than 4 which confirm that the descriptors were statistically significant and orthogonal. The mean effect (ME) reported in Table 7 give a vital information on the effect of each descriptor and the degree of contribution in the developed model. The signs and the magnitude on the mean effects values indicate direction and strength with which each descriptor is influencing the activity of each compound. Table 7 represents the P-values of each of the descriptors in the model at 95% confidence level. Therefore the null hypothesis that says there is no association between the descriptors influencing the model and the activities of the molecules is rejected thus; the alternative hypothesis that says there is a relationship between the descriptors used in generating the model and the activities of the compounds at $p < 0.05$ is accepted. Y- Randomization coefficient (cR_p^2) was also conducted and has a significant value of 0.78232 greater than 0.5 which was reported in Table 8 supports the claim that the model generated is powerful and not inferred by chance.

Table 2: Recommended values value for the validation parameters for a given QSAR model

Validation Parameter	Formula	Threshold	comment	Reference
Internal validation				
$\mathbf{R^2}$	$\frac{\left[\sum\left\{(Y-\bar{Y})\times\left(\hat{Y}-\bar{\hat{Y}}\right)\right\}\right]^2}{\sum(Y-\hat{Y})^2\times\sum\left(\hat{Y}-\bar{\hat{Y}}\right)^2}$	$R^2 > 0.6$	passed	[13,14]
$\mathbf{R^2_{adj}}$	$\frac{(N-1)\times R^2-p}{N-1-p}$	$R^2_{adj} > 0.6$	passed	
$\mathbf{Q^2}$	$1-\frac{\sum(Y-\hat{Y}_{loo})^2}{\sum(Y-\bar{Y})^2}$	$Q^2 > 0.6$	Passed	
$\mathbf{F_{(4,15)}}$	$\frac{\sum(Y-\bar{Y})^2}{p}\Big/\frac{\sum(Y-\hat{Y})^2}{N-p-1}$	$F_{(test)} > 2.09$	Passed	
Rando model				
$\bar{R}_r$	an average of the correlation coefficient for randomized data	$\bar{R} < 0.5$	passed	[13,14]
$\bar{R}_r^2$	an average of determination coefficient for randomized data	$\bar{R}_r^2 < 0.5$	Passed	
$\bar{Q}_r^2$	an average of leave one out cross-validated determination coefficient for randomized data	$\bar{Q}_r^2 < 0.5$	Passed	
$^c\mathbf{R_p^2}$	$R^2\times\left(1-\sqrt{\lvert R^2-\bar{R}_r^2\rvert}\right)$	$^cR_p^2 > 0.6$	Passed	[15]
External validation				
$\mathbf{R^2_{Pred}}$	$1-\frac{\sum(Y_{ext}-\hat{Y}_{ext})^2}{\sum(Y_{ext}-\bar{Y})^2}$	$R^2_{pred} > 0.6$	Passed	[13,14]

Table 3: Descriptive statistics of the inhibition data

Statistical parameters	Observed activity	
	Training set	Test set
Mean	6.660548	6.336856
Median	6.2234	6.1667
Standard Deviation	0.930345	1.233662
Sample Variance	0.865542	1.521922
Kurtosis	-0.95391	-1.44935
Skewness	0.556104	0.265368
Range	3.1448	3.1365
Minimum	5.4406	4.925
Maximum	8.5854	8.0615
Number of sample points	21	9

Table 4: Observed, Predicted and Residual values for the inhibitory compounds

Molecule	Observed Activity	Predicted Activity	Residual	Leverage
1[a]	8.0250	8.021	0.0004	0.3435
2[a]	8.0345	8.3453	-0.3108	0.0849
3[a]	8.7064	8.0002	0.7062	0.0965
4	5.7441	5.902932	-0.15883	0.0900
5	5.9258	5.689154	0.236646	0.0675
6[a]	6.1667	6.4566	-0.2899	0.1013
7	5.8765	5.891691	-0.01519	0.2189
8[a]	6.4171	6.9833	-0.5662	0.0909
9	5.9413	5.784999	0.156301	0.0799
10	6.6397	6.773987	-0.13429	0.0755
11	8.0899	7.791209	0.298691	0.1547
12	6.5267	6.321249	0.205451	0.0423
13	5.7405	5.615602	0.124898	0.0598
14	5.6533	5.906997	-0.2537	0.3572
15	6.1923	6.332476	-0.14018	0.2146
16[a]	7.3233	7.1234	0.1999	0.2008
17	6.0097	5.857651	0.152049	0.4327
18	6.0928	6.1928	-0.1	0.2637
19	7.3279	7.213392	0.114508	0.2553
20	6.8568	6.916443	-0.05964	0.0623
21	6.2234	6.5234	-0.3	0.2798
22	7.0079	7.391271	-0.38337	0.4100
23	7.314	7.259008	0.054992	0.2571

24	7.0854	7.606443	-0.52104	0.0559
25 [a]	7.2615	7.0023	0.2592	0.5172
26	5.2346	5.726447	-0.4916	0.2496
27 [a]	6.4218	6.2343	0.1875	0.3414
28	5.1016	5.799815	-0.6982	0.2571
29 [a]	6.1213	6.3244	-0.2031	0.0559
30	5.4406	5.774536	-0.33394	0.5172

Where superscript **a** represent the test set

Table 5: QSAR internal and external validation parameters for each model using Genetic Function Approximation (GFA)

S/NO		Model 1	Model 2	Model 3	Model 4
1	Friedman LOF	0.17565	0.18237	0.18437	0.19034
2	R-squared	0.96484	0.95434	0.95183	0.93446
3	Adjusted R-squared	0.95223	0.93235	0.91435	0.90328
4	Cross validated (R-squared (Q^2_{cv})	0.91325	0.90239	0.88322	0.88232
5	Significant Regression	Yes	Yes	Yes	Yes
6	Significance of regression F-value	82.41933	80.23228	78.33212	72.43429
7	Critical SOR F-value (95%)	2.92126	2.92126	2.92126	2.92126
8	Replicate points	0	0	0	0
9	Computed observed error	0	0	0	0
10	Lack-of-fit points	15	15	15	15
11	Min expt. error for non-significant LOF (95%)	0.155757	0.153431	0.135759	0.115753
12	R2 test	0.8681	0.7213	0.7021	0.5213

Table 6: List of some descriptors used in the QSAR optimization model

S/NO	Descriptors symbols	Name of descriptor(s)	Class
1	AATS7s	Average Broto-Moreau autocorrelation - lag 7 / weighted by I-state	2D
2	nHBint3	Count of E-State descriptors of strength for potential Hydrogen Bonds of path length 3	2D
3	minHCsatu	Minimum atom-type H E-State: H on C sp3 bonded to unsaturated C	2D
4	RDF90i	Radial distribution function - 090 / weighted by relative first ionization potential	3D

Table 7: Pearson's correlation and statistics for descriptor used in the QSAR model

Inter- correlation				Statistics		
AATS7s	nHBint3	minssCH2	RDF90i	P- Value (Confidence	VIF	Mean Effect (ME)

				interval)		
AATS7s	1			0.00087	2.3321	-0.7456
nHBint3	-0.125	1		0.00006	2.2872	0.3635
minHCsatu	0.03233	-0.37024	1	4.4E-09	1.2353	-0.5432
RDF90i	0.13017	0.2933	0.283317	6.51E-07	1.6453	0.5443

Table 8: Y- Randomization Parameters test

Model	R	R^2	Q^2
Original	0.96484	0.952230	0.91325
Random 1	0.545618	0.234729	-1.0723
Random 2	0.656406	0.226963	-0.28460
Random 3	0.621344	0.483220	0.014324
Random 4	0.564345	0.345343	-0.08742
Random 5	0.345456	0.290871	-0.89655
Random 6	0.547372	0.345677	0.021234
Random 7	0.655542	0.143334	-0.86654
Random 8	0.678965	0.245443	-0.58765
Random 9	0.543466	0.256783	-0.60112
Random 10	0.597654	0.367888	-0.00324

Random Models Parameters	
Average r :	0.538767
Average r^2 :	0.287544
Average Q^2 :	-0.44307
cRp^2 :	0.78232

The graph of predicted activities plotted against observed activities of the training and test set are presented in Figure 2 and 3. The correlation coefficient (R^2) value of 0.9649 for the training set and (R^2) value of 0.8681 for the test set recorded in this work was found to in line with accepted QSAR threshold values reported in Table 2. This affirms the stability, reliability and predictive power of the built model. The plot of residual activity against observed activities shown in Figure 4 designates that there exist no computational inaccuracy in the derived QSAR model as the range of residuals values fall within an accepted limit of $\pm$2 on residual activity axis.

The standardized residuals activities plotted against the leverage value known as The Williams plot is shown in Figure 5. The plotted graph clearly shows that all the compounds falls within limit boundary $\pm$3 of standardized cross-validated residual. Hence, it can be infer that no outlier is observed in the data set. Moreover, all the compounds were as well found to be lower than the warning leverage (h* = 0.86). Therefore no influential molecules were observed. Molecular docking was carried out between the targets (DNA gyrase) of M. tuberculosis and 1, 2, 4-Triazole derivatives. Ten (10) inhibitor ligands (compounds 1, 2, 3, 11, 16, 19, 22, 23, 24 and 25) with better activity were selected and docked with the DNA gyrase in order to elucidate the interaction and the binding mode. The binding affinity values for these ligands ranges

from (-6.2 to -11.8 kcal/mol) as reported in Table 9. All these ligands have higher binding score greater than the binding affinity of enthambutol (-5.8 kcal/mol) and isoniazid (-5.3 kcal/mol) which are the standard anti-tuberculosis drug. The ligand (compound 3) with best activity was selected for visualization purpose utilizing Discovery Studio Visualizer as shown in Figure 7 and 8 below. Ligand 3 formed four hydrogen bonds (2.82662, 2.35700, 3.07242 and 2.91737) with SER104, ASN176, GLN101 and GLN101of the target. In addition, it also formed hydrophobic bond with LEU105, ARG98, HIS52 and LEU105 of the target site

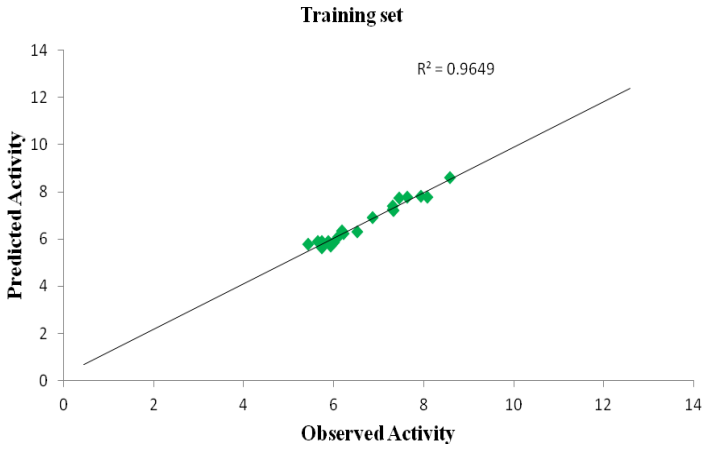


Figure 2: Plot of predicted activity against observed activity of training set

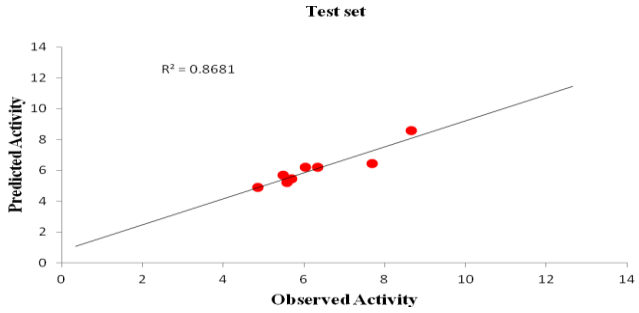


Figure 3: Plot of predicted activity against observed activity of test set

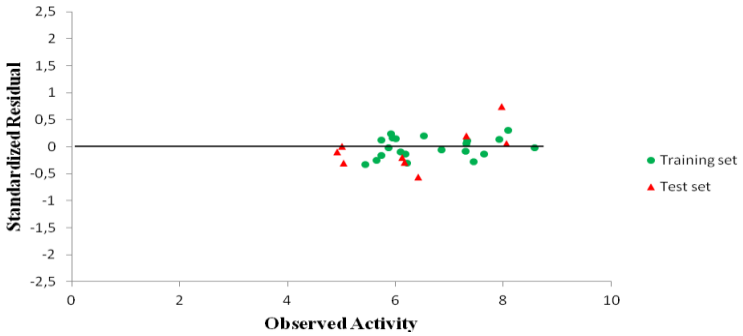


Figure 4: Plot of standardized residual versus observed activity

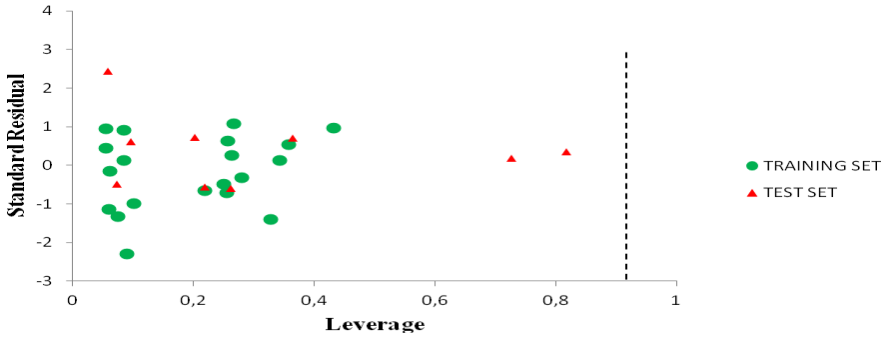
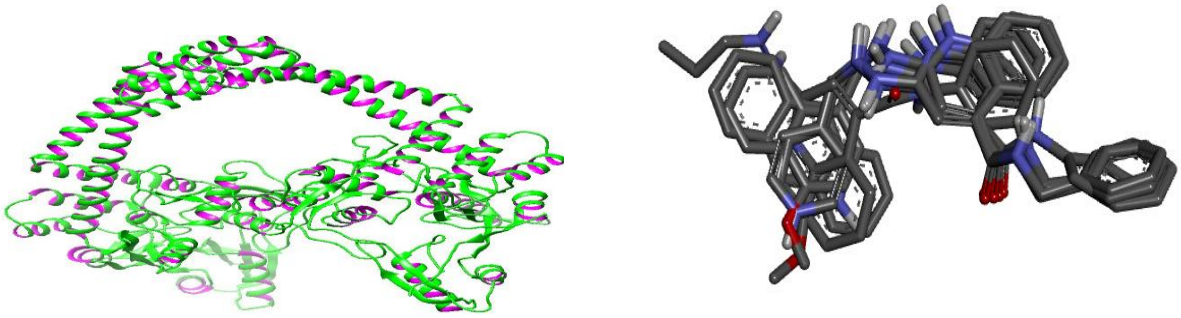


Figure 5: Plot of the standardized residuals versus the leverage value

Docking studies



A B

Figure 6: (A) Prepared structure of DNA gyrase (B) 3D structures of the prepared ligands.

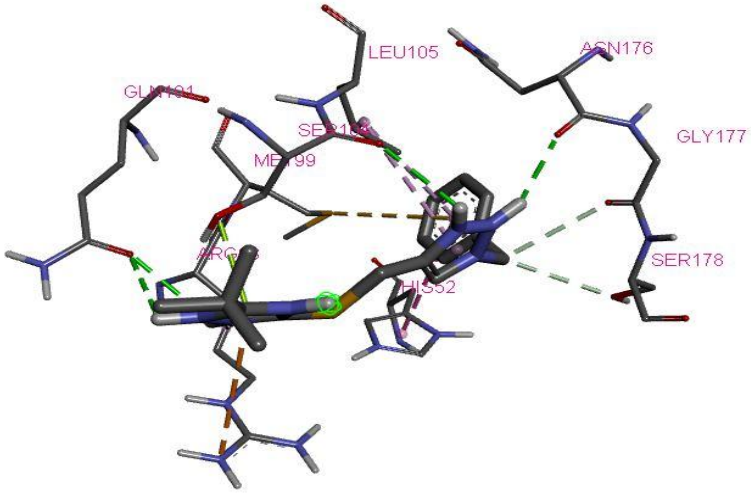


Figure 7: 3D interactions between DNA gyrase and Ligand 3.

.

Table 9: Binding Affinity, Hydrogen bond interaction and hydrophobic interaction formed between the ligands and the active site of DNA gyrase.

Ligand	Binding Affinity (BA) Kcal/mol	Target	Hydrogen bond		Hydrophobic
			Amino acid	Bond length (A^o)	
1	-9.5	DNA gyrase	SER118 SER118 GLY120	2.41232	PRO124
				2.47934	
				2.18555	
2	-9.3	DNA gyrase	GLN101 SER118 SER118	2.6091	TRP103, GLN277, VAL278, PRO124
				2.33129	
				2.69125	
3	-11.8	DNA gyrase	SER104	2.82662	LEU105, ARG98, HIS52, LEU105
			ASN176	2.35700	
			GLN101	3.07242	
			GLN101	2.91737	
11	-9.1	DNA gyrase	PRO119 SER118 GLY120	2.0904	TRP103, TRP103, HIS280 , VAL97, PRO124, PRO119 A, PRO124
				2.20816	
				2.58818	
16	-84	DNA gyrase	GLY112 PRO108	2.4735	HIS490 , PRO108, PRO102, PRO108
				2.96424	
19	-8.3	DNA gyrase	GLY17 ASN176	2.05316	LEU105
				2.79531	
22	-6.2	DNA gyrase	ALA100	2.66283	PRO102, PRO108
23	-8.1	DNA gyrase	TRP103 SER118	2.40711	TYR93, PRO124, VAL97, PRO124
				2.48976	
24	-6.6	DNA gyrase	ASP94	2.62857	PRO124, PRO124
25	-7.8	DNA gyrase	GLY177 SER178	2.82609	HIS52, HIS52
				2.47154	

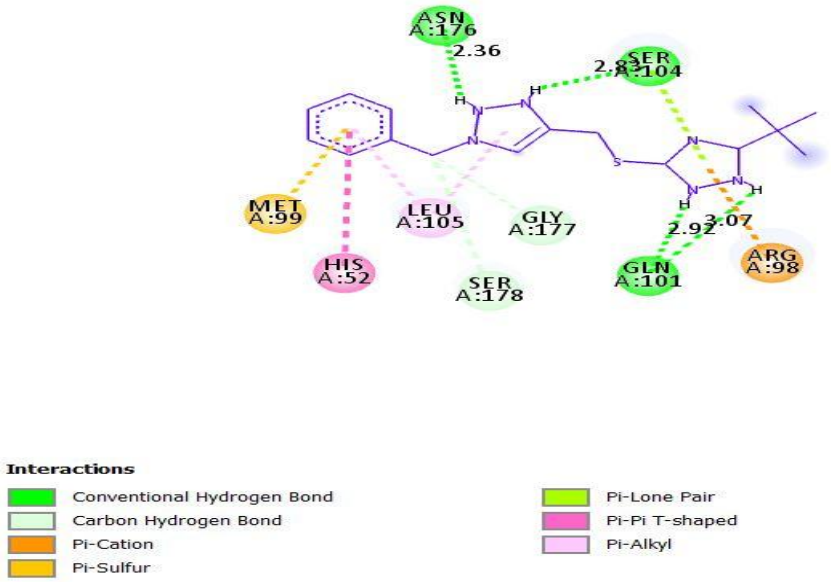


Figure 8: 2D interactions between DNA gyrase and Ligand 3

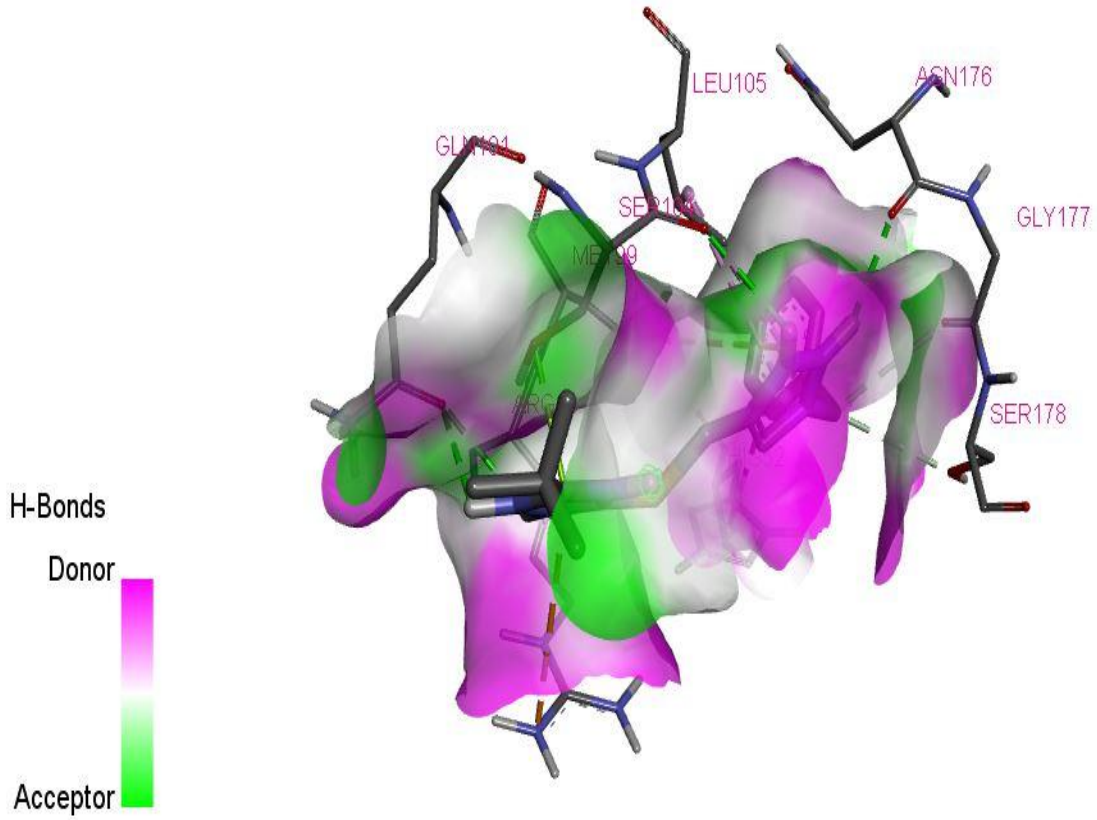


Figure 9: H-bond interaction between the ligand 3 and M. tuberculosis target (DNA gyrase)

Hydrogen bond interaction between the ligand 3 and DNA gyrase target of M. uberculosis is shown in Figure 9. A total of four hydrogen bonds were formed. The N-H group of triazole ligand as hydrogen donor and formed two hydrogen bonds with SER104 and ASN176 of the target. While the N-H group triazolidine of the ligand also acts as hydrogen donor and formed two hydrogen bonds with GLN101 of the target. The hydrogen bond formation alongside with the hydrophobic interaction provides an evidence that ligand 3 of the inhibitor compound is potent against DNA gyrase receptor.

4.0 Conclusion

A theoretical approach was employed in this study to selected molecular descriptors to derive a model that could be used to correlate the structure of 1, 2, 4-Triazole derivatives as potent inhibitors against *Mycobacterium tuberculosis* and their respective biological activities. The model derived was subjected to internal and external validation test to confirm that the built QSAR model is significant, robust, and reliable. From the results, it is concluded that 1, 2, 4-Triazole derivatives can be modeled using molecular descriptors; AATS7s, nHBint3, minssCH2 and RDF90i. Molecular docking simulation revealed that compounds; 1, 2, 3, 11, 16, 19, 22, 23, 24 and 25 with better activity have higher bind affinity ranging from (-6.2 and -11.8kcal/mol). However, the lead compound (compound 3) with higher anti-tubercular activity have prominent higher binding affinity of -11.8kcal/mol which indicates that the compounds could serve as a template structure to design compounds with more efficient activities. The outcome of this research and the propose QSAR model develop can be recommended for pharmaceutical and medicinal chemists to design, synthesis and also carry out an in-*vivo* and in-*vitro* screening in order to substantiate the computational findings.

References

[1] W.H. Organization, others, Tuberculosis Fact Sheet (No. 104) 2000, Site Accessed Www Who Intmediacentrefactsheetswho104enindex Html. (2016).

[2] C. Hansch, A. Kurup, R. Garg, H. Gao, Chem-bioinformatics and QSAR: a review of QSAR lacking positive hydrophobic terms, Chem. Rev. 101 (2001) 619–672.

[3] D. Sarkar, S.R. Deshpande, S.P. Maybhate, A.P. Likhite, S. Sarkar, A. Khan, P.M. Chaudhary, S.R. Chavan, 1, 2, 4-triazole derivatives and their anti-microbial activity, (2016).

[4] A.I. Ogadimma, U. Adamu, Analysis of Selected Chalcone Derivatives as Mycobacterium tuberculosis Inhibitors, (2016).

[5] S.E. Adeniji, S. Uba, A. Uzairu, QSAR Modeling and Molecular Docking Analysis of Some Active Compounds against Mycobacterium tuberculosis Receptor (Mtb CYP121), J. Pathog. 2018 (2018).

[6] S.E. Adeniji, S. Uba, A. Uzairu, 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, Phys. Chem. Res. 6 (2018) 479–492.

[7] G.M. Eric, A. Uzairu, P.A. Mamza, A Quantitative Structure-activity Relationship (QSAR) Study of the Anti-tuberculosis Activity of Some Quinolones, (n.d.).

[8] P. Singh, Quantitative Structure-Activity Relationship Study of Substituted-[1, 2, 4] Oxadiazoles as S1P1 Agonists, J. Curr. Chem. Pharm. Sci. 3 (2013).

[9] R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C.P. Varghese, R.K. Agrawal, Validation of QSAR models-strategies and importance, Int. J. Drug Des. Discov. 3 (2011) 511–519.

[10] A. Tropsha, P. Gramatica, V.K. Gombar, The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models, Mol. Inform. 22 (2003) 69–77.

[11] S.E. Adeniji, S. Uba, A. Uzairu, Theoretical modeling and molecular docking simulation for investigating and evaluating some active compounds as potent anti-tubercular agents against MTB CYP121 receptor, Future J. Pharm. Sci. 4 (2018) 284–295.

[12] K. Roy, P. Chakraborty, I. Mitra, P.K. Ojha, S. Kar, R.N. Das, Some case studies on application of “rm2” metrics for judging quality of quantitative structure–activity relationship predictions: emphasis on scaling of response data, J. Comput. Chem. 34 (2013) 1071–1082.
[13] S.E. Adeniji, S. Uba, A. Uzairu, In Silico Study for investigating and predicting the activities of 1, 2, 4-Triazole derivaties as potent anti-tubercular agents, J. Eng. Exact Sci. 4 (2018) 0246–0254.
[14] A. Tropsha, Best practices for QSAR model development, validation, and exploitation, Mol. Inform. 29 (2010) 476–488.
[15] K. Roy, On some aspects of validation of predictive quantitative structure–activity relationship models, Expert Opin. Drug Discov. 2 (2007) 1567–1577.