Computational modeling and ligand-based design of some novel hypothetical compound as prominent inhibitors against *Mycobacterium tuberculosis*

Shola Elijah Adeniji and Olajumoke Bosede Adalumo

**Abstract**

**Background:** Time consumed and expenses in discovering and synthesizing new hypothetical drugs with improved biological activity have been a major challenge toward the treatment of multi-drug-resistant strain *Mycobacterium tuberculosis* (TB). To solve the above problem, quantitative structure activity relationship (QSAR) is a recent approach developed to discover novel agents with better biological activity against *M. tuberculosis*.

**Results:** A validated QSAR model was developed in this study to predict the biological activities of some anti-tubercular compounds and to design new hypothetical drugs is influenced with the molecular descriptors, AATS7s, VR1_Dzi, VR1_Dzs, SpMin7_Bhe, and TDB8e, which has been validated through internal and external validation test. Prior to high anti-tubercular activity of the lead compound, compound 17 served as a template structure to design compounds with improved activity. Among the compounds designed, compounds 17i, 17j, and 17n were observed with improved anti-tubercular activities which ranges from 8.8981 to 9.0377 pBA.

**Conclusion:** The outcome of this research is recommended for pharmaceutical and medicinal chemists to synthesis and carry out an in vivo and in vitro screening for the proposed designed compounds in order to substantiate the computational findings.

**Keywords:** Model, Quinoline, QSAR, Tuberculosis

**Background**

Multi-drug resistance strain (TB) has posed a challenge toward the treatment of tuberculosis in the global community. World Health Organization in (2018) has reported 9.0 million people infected with tuberculosis, 360,000 HIV patients who were living with tuberculosis, death of 230,000 children, and death of 1.6 million people worldwide [1]. Some of the notable commercially sold drugs administered to people infected with tuberculosis are isoniazid (INH), pyrazinamide (PZA), rifampicin (RMP), and para- amino salicylic acid (PAS). The emergence of multi-drug-resistant strains of *M. tuberculosis* toward the aforementioned drugs has led to advances in searching for new and better approach that is precise and fast in developing a novel compound with improved biological activity against *M. tuberculosis*.

The advance of computational chemistry has led to development of new drug. Computational methods which reduced the cost for effective evaluation of large virtual database of chemical compounds are currently employed in designing new drugs. Such method includes complex network theory, quantitative structure–activity relationships (QSAR) models, Machine Learning (ML), and Artificial Neural Networks (ANN) analysis. For the time being, QSAR is a theoretical approach with widely used computational method in predicting and designing new hypothetical drug candidate [2]. The application of QSAR...
technique to this problem has a potential to minimize the
effort and time required to discover new compounds or to
improve current compounds in terms of their efficiency.
Multi variant QSAR model is expressed mathematically to
relate the physical, chemical, biological, or environmental
activities of interest with measurable or computable pa-
rameters such as physicochemical, topological, stereochem-
ical, or electronic indices called molecular descriptors.
Meanwhile, some prominent researchers [3–6] have suc-
cessfully established QSAR models to show the relation-
ship between some derivatives such as triazole, chalcone,
quionolone, 7-methyijuglone, pyrrole, and their respective
biological activities using the QSAR approach. Hence, this
research was aimed to build a robust QSAR model with
high predictability and to design new potent hypothetical
compounds with proposed better anti-tubercular activity.

**Methods**

**Data collection**
The molecules of derivatives of 2,4-disubstituted quin-
oline derivatives reported as anti-Mycobacterium tuber-
culosis which were used in this study were obtained
from the literature [7]. The list of these compounds and
their biological activities were presented in Table 1.

**Biological activities**
The biological activities of 2,4-disubstituted quinoline
derivatives as potent anti-tubercular agents were initially
expressed in percentage (%) and then converted to loga-
rithm unit using Eq. 1 below in order to increase the lin-
earity and approach normal distribution of the activity
values. The observed structures and the biological activ-
ities of these compounds were presented in Table 1 [4].

\[
pBA = \log \left( \frac{\text{molecular weight (g/mol)}}{\text{Dose (g/ml)}} \right) \left( \frac{\text{percentage} \%}{100 - \text{percentage} \%} \right)
\]

(1)

**Molecular optimization**
The Spartan 14 software version 1.1.4 was used to
optimize all the inhibitory compounds in order for the
compounds to attain stable conformation at a minimal
energy. The strain energy from the molecules was re-
moved by employing molecular mechanics force field
(MMFF), and complete optimization was achieved with
the aid of density functional theory (DFT) by utilizing
the (B3LYP) basic set [4].

**Generation of molecular descriptor**
A descriptor is a mathematical logic that defines the
properties of a molecule in a numeral term based on the
connection between the biological activity of each mol-
ecule and its molecular structure. Descriptors for all the

| Table 1 Geometrical structures of inhibitory compounds as anti-tubercular agents |
|---|

Where superscript a represent the test set
inhibitory molecules were calculated with the aid of the PaDEL descriptor software version 2.20, and a total of 1879 molecular descriptors were generated.

**Normalization and pretreatment of data**

For each of the variable (descriptor) to have the same chance at the inception so as to influence the QSAR model, the descriptor values generated from the PaDEL descriptor software version 2.20 were subjected to normalization using Eq. 2 [8].

\[
D = \frac{d_i - d_{\text{min}}}{d_{\text{max}} - d_{\text{min}}} 
\]

where \(d_{\text{max}}\) and \(d_{\text{min}}\) are the maximum and minimum value for each descriptor column of \(D\). \(d_i\) is the descriptor value for each of the molecule. Immediately after the data have been normalized, the normalized data were then subjected to pretreatment so as to remove redundant descriptors.

**Generation training and test set**

The whole compounds that made up the data set were divided into training and test set in proportion of 70 to 30% using Kennard and Stone’s algorithm which was incorporated in DTC lab software. The development of the QSAR model and internal validation test were performed on the training set while the confirmation of the developed model was performed on test set.

**Building of QSAR models and internal validation test**

The QSAR models were built by adopting the Genetic Function Approximation (GFA) technique incorporated in the Material Studio software version 8.0 to select the optimum descriptors for the training set. Meanwhile, multi-linear regression approach (MLR) was used as a modeling tool to develop the multi-variant equations by placing the activity data in the last column of Microsoft Excel 2013 spread sheet which was later imported into the Material Studio software version 8.0 to generate the QSAR model. The internal validation test to affirm the built model is robust and also has a high predictability that was also performed in the Material Studio software version 8.0 and reported.

**Evaluation of leverage values (applicability domain)**

Influential and outlier molecule present in both the training and test set were determined by employing the applicability domain approach. The leverage \(h_i\) approach as defined in Eq. 3 was used to define applicability domain space ±3 for outlier molecule [9].

\[
h_i = M_i (M_i^T M_i)^{-1} M_i^T 
\]

where \(M_i\) represents the matrix of \(i\) for the training set. \(M\) represents the \(n \times d\) descriptor matrix for the training set, and \(M_i^T\) is the transpose of the training set \((M)\). \(M_i^T\) represents the transpose matrix \(M_i\). Meanwhile, the warning leverage \(h^*\) defined in Eq. 4 is the limit boundary to check for an influential molecule.

\[
h^* = 3 \frac{(d + 1)}{N} 
\]

where \(d\) is the total number of descriptors present in the built model, and \(N\) is the total number of compounds that made up the training set.

**Y-randomization validation test**

Y-randomization test is one of the external validation criteria which has to be considered in order to ascertain that the developed model is not built by chance [10]. Random shuffling of the data was performed on the training set following the principle laid by [11]. The activity data (dependent variable) were shuffled while the descriptors (independent variables) were kept unchanged in order to generate the multi-linear regression (MLR) model. For the developed QSAR to pass the Y-randomization test, the \(R^2\) and \(Q^2\) values for the model must be significantly low for numbers of trials while Y-randomization coefficient \(cR_p^2\) shown in Eq. 5 must be ≥0.5 in order to establish the robustness of the model.

\[
cR_p^2 = R \times [R^2 - (R_f)^2]^2 
\]

where \(cR_p^2\) is the Y-randomization coefficient, \(R\) is the correlation coefficient, and \(R_f\) is the average “R” of random models.

**Affirmation of the build model**

The internal and external validation criteria for both test and training set reported were compared with the generally accepted threshold value shown in Table 6 for any QSAR model [9–12] in order to affirm the reliability, fitting, stability, robustness, and predictability of the developed models.

**Results**

**Model 1**

\[
pBA = -7.836545646 \times AATS7s + 0.201962934 \times VR1_Dzi + 0.087893211 \times VR1_Dsz - 4.204663658 \times SpMin7_Bhe + 0.674915710 \times TDB8e + 29.11653208
\]

**Model 2**

\[
pBA = -4.790218643 \times AATS5e + 0.082643756 \times VR3_Dzv - 3.953009651 \times SpMin7_Bhe + 0.094784839 \times TDB7e + 0.024520722 \times RDF90i + 41.534742802
\]

EE is the standard error of estimation, \(w\) is the total number of terms present in the built model except the constant term, \(j\) is the number of descriptors confined in
the built model, \( q \) is a user-defined factor, and \( N \) is the number of compounds of training set. \( Y_{obs} \), \( Y_{training} \), and \( Y_{pred} \) are the observed activity, mean observed activity of the training compounds and the predicted activity respectively. \( r^2 \) is correlation coefficients of the plot of observed activity against predicted activity values, \( r_o^2 \) is the correlation coefficients of the plot of observed activity against predicted activity values at zero intercept, \( r_o^2 \) is the correlation coefficients of the plot of predicted activity against observed activity at zero intercept [7, 9, 10].

### Discussion on designed compounds

#### Discussion

**QSAR studies**

Optimum QSAR model for predicting the derivatives of 2,4-disubstituted quinoline against *M. tuberculosis* was successfully achieved by adopting the combination of computational and theoretical method. Data set comprises of 36 compounds was partitioned into 25 training set and 11 test set set using Kennard and Stone algorithm method. The 25 training set compounds were used to derive QSAR model using the multi-linear regression technique which also served as data set for internal validation test while the external validation test for the derived model was conducted on the test set.

The observed activities reported in literature, the calculated activities calculated for all the anti-tubercular compounds, the leverage values, and the residual values were presented in Table 1. The difference between the observed activities and calculated activities is the residual values which were observed to be significant low [13–15]. The low residual value indicates that the model built has a good predictive ability.

The optimum (2D and 3D) descriptors that efficiently describe the anti-tubercular compounds in relation to their biological activities were selected by the GFA approach. The characterization and relative information on the molecular structure of the anti-tubercular agent illustrated by the descriptors were reported in numerical value as shown in Table 2.

Various statistical analyses were conducted on the calculated descriptors in order to check the validity of the built model as reported in Table 3. Variance inflation factor (VIF) was evaluated for all the descriptors in order to determine the degree of correlation between each the descriptor. Generally, VIF value equals to 1 or falls with 1 and 5 signify non-existence of inter-correlation among the descriptors. However, if the VIF value is greater than 10, it signify that the model developed is unstable, hence the model should be re-checked if necessary. Regarding the VIF values for each of the descriptors which were found to be less than 5 as reported in Table 3 affirm that the descriptors were significantly orthogonal to each order since there is no inter-correlation between them. The degree of contribution that each descriptor plays in the built model was evaluated by determining the standard regression coefficient \( (b_j) \) and mean effect (ME). The magnitude and signs for \( b_j \) and ME values reported in Table 4 indicate strength and direction with which each descriptor influence the activity model. The relationship between the descriptors and biological activity of each compound was determined by one way analysis of variance (ANOVA). The probability value of each of the descriptor at 95% confidence level was found to be \( (p < 0.05) \) as presented in Table 3. Therefore, this signify that the alternative hypothesis that says there is a direct relationship between the biological activity of each compound and the descriptor swaying the built model is accepted; thus, null hypothesis proposing no direct relationship between biological activity of each compound and the descriptor swaying the built model is rejected. To further justify the validation of the

### Table 2: Descriptors name and class in model 1

| S/NO | Name of descriptor(s) | Descriptor symbols | Class |
|------|----------------------|-------------------|------|
| 1    | Average Broto-Moreau autocorrelation-lag 7/weighted by l-state | AATS7s | 2D |
| 2    | Randic-like eigenvector-based index from Barysz matrix/weighted by first ionization potential | VR1_Dzi | 2D |
| 3    | Randic-like eigenvector-based index from Barysz matrix/weighted by l-state | VR1_Dzs | 2D |
| 4    | Smallest absolute eigenvalue of Burden modified matrix-n 7/weighted by relative Sanderson electronegativities | SpMin7_Bhe | 2D |
| 5    | Topological distance based autocorrelation-lag 8/weighted by Sanderson electronegativities | TDB8e | 3D |

### Table 3: Statistical analysis and validation of descriptors model 1

| Descriptor | Standard regression coefficient \( (b_j) \) | Mean effect (ME) | \( p \) value (confidence interval) | VIF | Standard error |
|------------|------------------------------------------|------------------|----------------------------------|-----|----------------|
| AATS7s     | -0.2854                                  | -0.2651           | 2.23E-04                         | 3.0925 | 5.11E-05 |
| VR1_Dzi    | 0.4312                                   | 0.4176            | 0.00719                          | 1.6328 | 2.98E-06 |
| VR1_Dzs    | 0.1729                                   | 0.1623            | 4.34E-04                         | 2.7301 | 4.78E-05 |
| SpMin7_Bhe | -0.4287                                  | -0.4017           | 3.42E-04                         | 3.2001 | 3.09E-04 |
| TDB8e      | 0.6341                                   | 0.6287            | 2.84E-05                         | 1.0034 | 0.4302 |
descriptors in the activity model, Pearson’s correlation statistic was conducted to also check whether there is inter-correlation between each descriptor. The correlation coefficient between each descriptors reported in Table 4 was all < ± 0.8. Hence, this implies that all the descriptors were void of multicollinearity.

Validation results for both the external and internal assessment to assure that the built models are reliable and robust were presented in Table 5. These results were all in full agreement with the general validation criteria presented in Table 5 to truly indorse that the stability and robustness of the model is valid. Reference to these validation results obtained, model one was selected and established to be the prime model which was used to predict the biological activities of 2, 4-disubstituted quinoline against *M. tuberculosis*.

The built QSAR model and results obtained in this research were compared with recent model developed in the literature [3, 6] as shown below:

| Table 4 | Coefficient of Pearson’s correlation for descriptor in model |
|---------|-------------------------------------------------------------|
| Inter-correlation | AATS7s | VR1_Dzi | VR1_Dzs | SpMin7_Bhe | TDB8e |
| AATS7s | 1 | | | | |
| VR1_Dzi | 0.4318 | 1 | | | |
| VR1_Dzs | 0.5324 | 0.4091 | 1 | | |
| SpMin7_Bhe | 0.3912 | −0.0123 | −0.0291 | 1 | |
| TDB8e | 0.1008 | −0.0189 | 0.0183 | 0.0298 | 1 |

Table 4 was all < ± 0.8. Hence, this implies that all the descriptors were void of multicollinearity.

| Table 5 | Internal and external validation parameters for each model |
|---------|-------------------------------------------------------------|
| S/NO | Validation parameters | Formula | Threshold | Model 1 | Model 2 |
|---------|-------------------------------------------------------------|---------------------------------|-------------|--------|--------|
| Internal validation | | | | | |
| 1 | Friedman’s lack of fit (LOF) | \[ \frac{\hat{\sigma}_{Y_{\text{pred}}}}{\sigma_{Y}} \] \[ \frac{1}{N} \sum_{i=1}^{N} (Y_{i} - \hat{Y}_{i})^2 \] | \[ R^2 > 0.6 \] | 0.9183 | 0.8429 |
| 2 | R-squared | \[ \frac{1}{N} \sum_{i=1}^{N} (Y_{i} - \hat{Y}_{i})^2 \] | \[ R^2_{\text{adj}} > 0.6 \] | 0.9183 | 0.8429 |
| 3 | Adjusted R-squared | \[ \frac{1}{N-1} \sum_{i=1}^{N} (Y_{i} - \hat{Y}_{i})^2 \] | \[ Q^2 > 0.6 \] | 0.8202 | 0.7400 |
| 4 | Cross validated R-squared \( (Q^2_{\text{cv}}) \) | \[ \frac{1}{N} \sum_{i=1}^{N} (Y_{i} - \hat{Y}_{i})^2 \] | \[ Q^2_{\text{cv}} > 0.6 \] | 0.8202 | 0.7400 |
| 5 | Significant regression | | Yes | Yes | |
| 6 | Critical SOR F-value (95%) | \[ F_{0.05, p-1} > 2.09 \] | | 4.3892 | 4.3892 |
| 7 | Replicate points | | 0 | 0 | |
| 8 | Computed observed error | | 0 | 0 | |
| 9 | Min exp. error for non-significant LOF (95%) | | 0.0278 | 0.0278 | |
| Model randomization | | | | | |
| 10 | Average of the correlation coefficient for randomized data \( (\bar{R}) \) | | \[ \bar{R} < 0.5 \] | 0.3371 | 0.3983 |
| 11 | Average of determination coefficient for randomized data \( (\bar{R}^2) \) | | \[ \bar{R}^2 < 0.5 \] | 0.1521 | 0.1763 |
| 12 | Average of leave-one-out cross-validated determination coefficient for randomized data \( (\bar{Q}^2) \) | | \[ \bar{Q}^2 < 0.5 \] | −1.3198 | −1.3719 |
| 13 | Y-randomization coefficient \( (c^2_{R^2}) \) | | \[ c^2_{R^2} > 0.6 \] | 0.7362 | 0.7058 |
| External validation | | | | | |
| 14 | Slope of the plot of observed activity against predicted activity values at zero intercept \( (K) \) | \[ \frac{\Delta Y_{\text{obs}}}{\Delta Y_{\text{pred}}} \] | \[ 0.85 < k < 1.15 \] | 1.0013 | 1.0582 |
| 15 | Slope of the plot of predicted against observed activity at zero intercept \( (k^\prime) \) | \[ \frac{\Delta Y_{\text{pred}}}{\Delta Y_{\text{obs}}} \] | \[ 0.85 < k < 1.15 \] | 0.9290 | 0.9016 |
| 16 | \[ \frac{\Delta Y_{\text{obs}}}{\Delta Y_{\text{pred}}} - 1.0 \] | | < 0.3 | 0.0834 | 0.0610 |
| 17 | \[ \frac{\Delta Y_{\text{pred}}}{\Delta Y_{\text{obs}}} - 0.0 \] | | < 0.1 | 0.0028 | 0.0042 |
| 18 | \[ \sqrt{\frac{\Delta Y_{\text{obs}}}{\Delta Y_{\text{pred}}}} \] | | < 0.1 | 0.0610 | 0.0582 |
| 19 | \[ R^2_{\text{test}} = 1 - \frac{\sum (Y_{\text{obs, test}} - \hat{Y}_{\text{pred, test}})^2}{\sum (Y_{\text{obs, test}} - \bar{Y}_{\text{test}})^2} \] | | \[ R^2_{\text{test}} > 0.6 \] | 0.8052 | 0.7281 |
pBA = -0.307001458(MATS2s) + 1.528715398(nHBint3) + 3.976720227(maxtsC) + 0.016199645(TDB9e) + 0.089381479(RDF90i) - 0.107407822(RDF110s) + 4.057082751,

$R^2 = 0.92024$, $R_{adj}^2 = 0.9102$, $Q^2_{cv} = 0.8954$, and $R_{pred}^2 = 0.8842$ [3]

pIC50 = -2.040810634(nCl) -19.024890361(MATS2m) + 1.855704759(RDF140s) + 6.739013671 = 27, $R^2 = 0.9480$, $R_{adj}^2 = 0.9350$, $Q^2_{cv} = 0.8799$, and $R_{pred}^2 = 0.7690$ [6]

The validation factors reported in this work and those reported in the literature were all in agreement with the validation parameters presented in Table 5 which really inveterate that the model generated is predictive and robust.

The coefficient of Y-randomization ($c R_p^2$) with significant value of 0.7362 greater than threshold value of 0.5 reported in Table 5 provides a reasonable support that the model built is robust and not just by chance.

The graphical representation to show the degree of correlation between the calculated activities and observed activities of the training and test set were shown in Figs. 1 and 2. The correlation coefficient ($R^2$) value of 0.9183 and 0.8052 for both the training set and test set shows that there is a high correlation existing between the calculated activities and observed activities of the training and test set which were also in agreement with the accepted QSAR threshold values reported in Table 6.

The residual plot shown in Fig. 3 signify that there is no indication of computational incompetency and inaccuracy in the QSAR model derived as all the standard residual values for both training and test set were found within the defined boundary of ±2 on the standard residual activity axis.

The Williams plot to show the applicability domain space (AD) is shown in Fig. 4. However, compound (number 30) is found to have a leverage value greater than the predicted warning leverage ($h^* = 0.60$). Therefore, it can be infer that compound (number 30) is an influential molecule. Moreover, it is also
| Ligand | Binding affinity (BA) Kcal/mol | Target | Hydrogen bond Amino acid Bond length (Å) | Hydrophobic interaction Amino acid |
|--------|-------------------------------|--------|----------------------------------------|----------------------------------|
| 1      | − 7.2                         | DNA gyrase | PRO124 2.251 | VAL278, TRP103, HIS220, GLN277 |
| 2      | − 7.5                         | DNA gyrase | ARG98 2.9399 | GLN277, PRO285, HIS220, VAL78 |
| 3      | − 7.7                         | DNA gyrase | ASP94 TRP182 2.3878 | PRO124, VAL138, GLN101, CYS112 |
| 4      | − 7.7                         | DNA gyrase | ARG98 1.4999 | PRO124, VAL97, HIS220 |
| 5      | − 7.9                         | DNA gyrase | ASP94 2.1801 | VAL278, PRO119, GLN101, ASP122 |
| 6      | − 8.3                         | DNA gyrase | SER102 2.529 | ASP122, ALA167, TRP182, SER247 |
| 7      | − 8.2                         | DNA gyrase | ARG98 GLY120 4.287 | TYR276, ASP94, VAL97, PRO124 |
| 8      | − 8.8                         | DNA gyrase | HIS220 2.4765 | PHE228, ALA173, PRO119, TRP182, SER247 |
| 9      | − 8.4                         | DNA gyrase | LEU213 ARG184 1.461 | MET99, VAL78, TRP182, SER118, ASP122 |
| 10     | − 8.2                         | DNA gyrase | PRO119 GLY120 2.1738 | VAL77, SER247, ARG98, ASP94, VAL182 |
| 11     | − 9.7                         | DNA gyrase | ASP94 TRP103 1.383 | GLY120, GLY120, SER118, PHE168, PRO285, VAL78 |
| 12     | − 8.6                         | DNA gyrase | SER104 VAL77 2.023 | TRP162, CYS145, ASP122, VAL78, PRO126, ARG98 |
| 13     | − 8.1                         | DNA gyrase | VAL169 ARG134 PRO285 2.6021 | MET99, ASP122, PHE232 |
| 14     | − 8.4                         | DNA gyrase | GLY145 SER205 2.4909 | VAL98, ALA223, MET145, MET99, LEU164 |
| 15     | − 9.1                         | DNA gyrase | ARG98 SER118 GLY120 3.3701 | PRO124, VAL97, VAL97, ASP94, PRO123, ASP122 |
| 16     | − 8.1                         | DNA gyrase | ARG98 SER118 GLY120 2.8704 | 1.9128, 3.2821 |
| 17     | − 18.8                        | DNA gyrase | ARG98 1.99395 | CYS174, ALA67, ASN74, GLY120, MET99 |
| 18     | − 9.1                         | DNA gyrase | LEU114 ALA78 2.3983 | LEU164, VAL228, PHE168, GLY232, TYR276 |
| 19     | − 9.7                         | DNA gyrase | ALA167 ARG94 1.3965 | ALA233, LYS136, MET99, VAL228 |
| 20     | − 11.6                        | DNA gyrase | MET99 2.3975 | TRP142, PHE88, PRO169, VAL78, LEU156 |
| 21     | − 9.9                         | DNA gyrase | GLN223 TYR276 2.5093 | ARG98, LEU103, ALA167, PHE168, MET234 |
| 22     | − 6.8                         | DNA gyrase | PHE212 TRP182 1.8408 | VAL78, LEU123, SER119, ALA233, TYR276 |
| 23     | − 10.7                        | DNA gyrase | LSY146 TRP143 2.1665 | PHE168, CYS254, TRP182, ALA167, VAL78, VAL82 |
| 24     | − 7.9                         | DNA gyrase | ARG98 CYS156 1.5984 | ALA167, LEU 103, TRP112, ARG386 |
| 25     | − 7.1                         | DNA gyrase | TRP182 2.3663 | ARG72, ALA143, VAL78, GLN154 |
| 26     | − 8                           | DNA gyrase | PHE256 ARG143 1.287 | TRP182, CYS345, ALA176, PHE 168 |
| 27     | − 9.2                         | DNA gyrase | – – | PRO285, MET 232, SER108, ALA137 |
| 28     | − 8.5                         | DNA gyrase | – – | LEU164, VAL178, PRO169, PHE98, VAL228 |
| 29     | − 8.1                         | DNA gyrase | ARG145 1.9217 | LEU234, VAL228, CYS 144, ALA233, VAL78 |
| 30     | − 8.2                         | DNA gyrase | TRP182 MET99 2.3896 | PRO34, PRO94, PHE93, VAL178, PHE241, PRO169 |
observed that all the compounds fall within the defined space of ±3 which indicates that no compound is said to be outlier.

**In silico design for new derivatives based on Lead compound 17**

Ligand-based approach was employed to design new compounds with better anti-tubercular activities via modification of the template by deletion, insertion, and substitution of active substituent(s) into the template structure. The choice of template used in this study was (E)-N-benzyl-2-(2-benzylidenehydrazinyl)-quinoline-4-carboxamide (i.e., molecule 17 in Table 1) due to its relative high anti-tubercular activity which also falls within the model applicability domain (AD) space shown in Fig. 4. The modification was easily made around N-ethylacetamide and 2-methylhydrazine moiety of the template at positions 16 and 23 shown in Fig. 5. The QSAR model built indicated that increase in the values of descriptors, VR3_Dzp, VR1_Dzi, and influences the activity positively. This implies that increase in the value of these descriptors also augment the values of the activity in the same direction. Variation of the substituent at positions 16 and 23 of the template structure with alkyl group, benzene derivatives, and substituted alky amines lead to generation of fourteen compounds with better anti-tubercular activities reported in Table 7. The leverage values predicted for the designed compounds were used to screen and confirm whether these compounds were within their model AD. Based on the leverage value predicted for each compound in

| Ligand     | Binding affinity (BA) Kcal/mol | Target   | Hydrogen bond Amino acid | Hydrophobic interaction Amino acid |
|------------|-------------------------------|----------|--------------------------|-----------------------------------|
| 31         | − 9.3                         | DNA gyrase | ARG98                    | VR1_Dzi 1.3896                     |
| 32         | − 11.4                         | DNA gyrase | TYR276                   | VR1_Dzi 3.1345                     |
| 33         | − 7.1                          | DNA gyrase | GLN277                   | VR1_Dzi 2.5007                     |
| 34         | − 9.6                          | DNA gyrase | HIS220, SER104, MET99    | VR2_Dzi 3.2896, VR1_Dzi 1.3904    |
| 35         | − 9.4                          | DNA gyrase | TYR276                   | VR1_Dzi 2.5007                     |
| 36         | − 8.2                          | DNA gyrase | ALA167, LEU137           | VR1_Dzi 1.3907                     |
| Ethambutol | − 5.8                          | DNA gyrase | ALA337                   | VR1_Dzi 2.59739                    |
| Isoniazid  | − 14.6                         | DNA gyrase | SER279, ALA337           | VR1_Dzi 2.29943, VR1_Dzi 2.52954, |

| Fig. 3 | Plot of standardized residual activity versus observed activity |
Table 7 assured that each compound have a low leverage value compared with the warning leverage $h^* = 0.64$ shown in Fig. 5. Hence, this signified that all the designed compounds fall within their model AD space. Among the compounds designed, compounds 17i, 17j, and 17n were observed with better antitubercular activities. This was as a result of substitution at positions 16 and 23 of the template structure with N-substituted alky amine which act as electron releasing group via positive inductive effect (+I). Due to the positive +I effect of the alkyl group attached to the template structure, the nitrogen becomes strongly electronegative, so the lone pair of electron on N-atom is easily available. The steric hindrance of the bulky alkyl group (3⁰ amine) observed in the compound 17j accounts for the decrease in its reactivity when compared to compound 17i (1⁰ amine) and 17n (2⁰ amine). Based on the decreasing order of amine, $(\text{CH}_3)_2\text{NH} > \text{CH}_3\text{NH}_2 > (\text{CH}_3)_3\text{N} > \text{NH}_3$, suggests why compound 17n was observed with prominent activity.

**Conclusion**

This work addresses the quantitative structure activity relationship (QSAR) between quinoline derivatives and their biological activities against *Mycobacterium tuberculosis*. The QSAR model was established to predict the reported experimental activities of 2,4-disubstituted quinoline derivatives against *M. tuberculosis* via computational modeling approach under the influence of optimum descriptors: AATS7s, VR1_Dzi, VR1_Dzs, SpMin7_Bhe, and TDB8e. The lead compound (compound 17) with higher antitubercular activity was used as a structural template to designed new hypothetical drug candidates. Among the...
Table 7 Designed molecule, predicted descriptors, and calculated activities for template 17 of 2,4-quino line derivative

| Compound | $R_1$ | $R_2$ | AATS | VR3_D | VR1_D | VR2_D | SpMin7 | Predicted Activity (pBA) |
|----------|------|------|------|-------|-------|-------|--------|--------------------------|
| 17a      | CH₃  |      | 7.339| 17.429| 451.48| 418.73| 0.6986 | 8.2004 | 0.1327 |
| 17b      |      |      | 7.743| 19.261| 512.23| 463.40| 0.8252 | 8.2614 | 0.3711 |
| 17c      |      |      | 7.637| 19.672| 530.40| 477.50| 0.9564 | 8.3264 | 0.4193 |
| 17d      |      |      | 7.516| 26.591| 577.81| 604.65| 1.1100 | 8.5258 | 0.2586 |
| 17e      |      |      | 7.428| 25.823| 623.03| 687.69| 1.1891 | 8.6570 | 0.2674 |
| 17f      |      |      | 7.416| 25.047| 638.78| 637.42| 1.1966 | 8.6670 | 0.3510 |
| 17g      | CH₃  |      | 7.439| 14.500| 257.17| 246.88| 1.0736 | 7.9161 | 0.4394 |
| 17h      |      |      | 7.480| 24.753| 583.47| 501.81| 1.1357 | 8.4934 | 0.3877 |
| 17i      |      |      | 7.430| 27.626| 877.38| 688.65| 1.2192 | 9.0377 | 0.2115 |
| 17j      |      |      | 7.375| 26.656| 770.48| 650.33| 1.2577 | 8.8981 | 0.3711 |
| 17k      | CH₃  |      | 7.520| 20.229| 550.27| 466.01| 1.0563 | 8.3886 | 0.4542 |
| 17l      | CH₃CH₂ |     | 7.503| 23.096| 562.85| 490.70| 1.1001 | 8.4280 | 0.0906 |
| 17m      |      |      | 7.487| 26.505| 569.84| 496.54| 1.1467 | 8.4484 | 0.2126 |
| 17n      |      |      | 7.440| 29.626| 977.38| 746.65| 1.2324 | 9.2007 | 0.0332 |
compounds designed, compounds 17i, 17j, and 17n were observed with improved anti-tubercular activities which ranges from 8.8981 to 9.0377 pBA. The outcome of this research is recommended for pharmaceutical and medicinal chemists to synthesis and carry out an in vivo and in vitro screening for the proposed designed compounds in order to substantiate the computational findings.

Acknowledgements
None

Authors’ contributions
SE and OB did the conception and design of the work. SE and OB did the acquisition and analysis of the data. SE interpreted the data. SE drafted the manuscript. SE and OB substantively revised the manuscript. All authors read and approved the final manuscript.

Funding
None

Availability of data and materials
It has been reported and cited in the methodology part of the manuscript.

Ethics approval and consent to participate
Not applicable for that section

Consent for publication
Not applicable for that section

Competing interests
Not applicable for that section

Author details
1Chemistry department, Ahmadu Bello University, Zaria, Nigeria Kaduna State 810107, Nigeria. 2Department of Science Education, Ajasin University, Akungba Akoko, Ondo, State 342211, Nigeria.

Received: 17 December 2019 Accepted: 8 April 2020
Published online: 25 May 2020

References
1. W.H.O (2018) http://www.who.int/news-room/fact-sheets/detail/tuberculosis
2. Hansch C, Kurup A, Garg R, Gao H (2001) Chem-bioinformatics and QSAR: a review of QSAR lacking positive hydrophobic terms. Chem Rev 101:619–672
3. Ogadimma AI, Adamu U (2016) Analysis of selected chalcone derivatives as Mycobacterium tuberculosis inhibitors. Open Access Library J 3:1–13
4. Adeniji SE, Uba S, Uzairu A (2018) QSAR Modeling and molecular docking analysis of some active compounds against Mycobacterium tuberculosis receptor (Mtb CYP121). J Pathog 2018
5. Adeniji SE, Uba S, Uzairu A (2018) 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 6:479–492
6. Adeniji SE, Uba S, Uzairu A (2019) A derived QSAR model for predicting some compounds as potent antagonist against M. tuberculosis: a theoretical approach. Hindawi Adv Prev Med5173786
7. Nayyar A, Jain R (2008) Synthesis and anti-tuberculosis activity of 2,4- disubstituted quinolines. Ind J Chemist Sect B Org Med Chem 47:117–128
8. Singh P (2013) Quantitative structure-activity relationship study of substituted-(1, 2, 4) oxadiazoles as S1P agonists. J Curr Chem Pharm Sci 3: 64–79
9. Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK (2011) Validation of QSAR models-strategies and importance. Int J Drug Des Discov 3:511–519
10. Tropscha A, Gramatica P, Gombar VK (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSRR models. Mol Inform 22:69–77
11. Adeniji SE, Uba S, Uzairu A (2018) 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 4:284–295
12. Roy K, Chakraborty P, Mitra I, Ojha PK, Kar S, Das RN (2013) 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 34:1071–1082
13. Ibrahim MT, Uzairu A, Shallangwa GA, Uba S (2020) In-silico activity prediction and docking studies of some 2, 9-disubstituted 8-phenylthio/ phenylsulfinyl-9 h-purine derivatives as Anti-proliferative agents. Helicon 6: e03158
14. Adeniji SE, Uba S, Uzairu A (2020) Multi-linear regression model, molecular binding interactions and ligand-based design of some prominent compounds against M. tuberculosis. Network Model Anal Health Inform Bioinform 9(8):1–18
15. Abdullahi M, Shallangwa GA, Uzairu A (2020) In silico QSAR and molecular docking simulation of some novel aryl sulfonamide derivatives as inhibitors of HSN1 influenza A virus subtype. Beni-Suef Univ J Basic Appl Sci 9(2):1–13

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.