Abdullahi *et al. Bulletin of the National Research Centre* (2020) 44:136
https://doi.org/10.1186/s42269-020-00389-7

Bulletin of the National Research Centre

RESEARCH Open Access

# Quantitative structure-activity relationship (QSAR) modelling study of some novel carboxamide series as new anti-tubercular agents

Mustapha Abdullahi[*], Shola Elijah Adeniji, David Ebuka Arthur and Shuaibu Musa

## Abstract

**Background:** QSAR modelling was performed on thirty-five (35) newly discovered compounds of N-(2-phenoxy) ethyl imidazo[1,2-a] pyridine-3-carboxamide (IPA) to predict their biological activities against *Mycobacterium tuberculosis* (MTB-H37Rv strain) by using some numerical data derived from structural and chemical features (descriptors) of the compounds.

**Results:** At first, the structure of the compounds was accurately drawn and optimized using the Spartan 14 software at DFT level of theory with B3LYP/6-31G** basis set in a vacuum. The diverse chemometric descriptors were computed from the optimized structures using the PaDEL descriptors software, and the division of the dataset into training and test sets was done based on Kennard-Stone's algorithm. Five (5) models were generated from the training set using genetic function approximation, and model 1 was chosen as the best due to its robust internal and external validation metrics ($R^2_{\text{train}}$ = 0.8563, $R^2_{\text{adjusted}}$ = 0.8185, PRESS = 3.5724, average $\overline{R}^2_m$ (LOO-train) = 0.6751, $Q^2_{\text{cv}}$ = 0.7534, $R^2_{\text{pred}}$ = 0.7543, $R^2_{\text{test}}$ = 0.6993) which passed the model criteria of acceptability. 6-Bromo-N-(2-(4-bromophenoxy) ethyl)-2-ethylimidazo[1,2-a] pyridine-3-carboxamide (compound 13) was used as the structural template for the in silico design due to its high pMIC, and it is within the model's chemical space.

**Conclusion:** Based on the information obtained from model 1, six (6) designed compounds with higher anti-tubercular activity were obtained. Furthermore, the ADME and drug-likeness prediction of the designed molecules showed good pharmacokinetic properties which indicate the application prospect of these compounds as novel MTB-H37Rv inhibitors. This research could help the medicinal chemists and pharmaceutical practitioners in future designing and development of more potent drug candidates.

**Keywords:** QSAR, Template, Docking, Hydrogen bonding, Genetic algorithm, Multi-linear regression, Model, Descriptors, Leave-one out

## Background

*Mycobacterium tuberculosis* (MTB) is the bacterium that causes one of the world's most deadly respiratory communicable diseases called tuberculosis (TB). It was among the ranked top 10 deadliest diseases caused by a single infectious agent (above HIV/AIDS) (Zhai et al. 2019). In recent times, the number of persons receiving life-saving treatment for TB in 2018 has tremendously increased due to enhanced detection and diagnosis (Mabhula and Singh 2019). Nigeria is among the top seven (7) countries that account for 64% of the total burden of tuberculosis worldwide (Ogbuabor and Onwujekwe 2019). The Philippines Department of Health (DOH) and the World Health Organization (WHO) jointly called for an all-out-war against tuberculosis (TB)

[*] Correspondence: mustychem19@gmail.com
Faculty of Physical sciences, Chemistry Department, Ahmadu Bello University, P.M.B. 1044, Zaria, Kaduna State, Federal Republic of Nigeria

in early 2019, because it is regarded as the number one infectious killer in the country. According to their report, TB claims the lives of over 70 Filipinos every day. There are also one million Filipinos who have active TB disease, the third-highest global prevalence rate next to South Africa and Lesotho (World Health Organization (WHO) 2019). Similarly, in the global tuberculosis report of WHO (2019), data were reported by 202 countries and territories that account for more than 99% of the world's population and estimated number of TB cases (World Health Organization (WHO) 2019). The existence of extensively drug-resistant (XDR) and the evolution of multidrug-resistant (MDR) TB have attracted the attention of drug scientists who are in search of novel anti-tubercular agents with better bioactivities (Wang et al. 2019). Researches have shown that imidazo[1,2-a] pyridine-3-carboxamides (IPA) as an anti-tubercular candidate is currently in the second phase of clinical trials, and it was reported to have resilient inhibitory potency or anti-mycobacterial activity (Wang et al. 2019). It was established that the development of more potent compounds with improved bioactivities is very costly and time-consuming (Adeniji et al. 2019). In recent decades, computational chemistry techniques such as computer-aided drug design (CADD) might save the time of discovering new compounds which also reduce the cost of synthesis (Abdullahi et al. n.d.). The quantitative structure-activity relationship (QSAR) technique provides a mathematical model containing some structural features represented as numerical data which predicts the response properties of the compound such as activity, toxicity, and so on (Ibrahim et al. 2018). The ultimate goal of this study was to derive a robust QSAR model from the structures of some synthesized IPAs compounds which predicts their biological activities against *Mycobacterium tuberculosis* (MTB-H37Rv strain), then utilized the model to design new compounds with improved activity.

## Methods

### Dataset collection and optimization

Thirty-five (35) compounds were selected from the newly discovered and synthesized series of N-(2-phenoxy) ethyl imidazo[1,2-a] pyridine-3-carboxamide (IPA) as anti-tubercular agents (Wang et al. 2019). The compound's response against the *Mycobacterium tuberculosis* (MTB-H37Rv) was measured as minimum inhibitory concentration (MIC) which is the lowest concentration affecting a decrease in fluorescence of greater than 90% relative to the mean of replicate bacterium-only controls in microgram per milliliter (Wang et al. 2019). These values were converted into logarithmic MIC values (pMIC) in order to reduce skewness using Eq. 1

$$pMIC = -\log\left(\frac{MIC(\mu g/mL)}{Mw\left(gmol^{-1}\right)} \times 10^{-3}\right) \tag{1}$$

where Mw is the molar weight of the compound in *grams* per *mole* and MIC is the minimum inhibitory concentration of the compound. The IPAs core structure and the substitution arrangement of the compounds based on $R_1$, $R_2$, and $R_3$ along with their anti-tubercular activities were presented in Table 1. The molecular structure of the IPAs showed above were accurately drawn using the ChemDraw Ultra level software V12.0.2, then saved in (*cdx) format. Consequently, the drawn compounds were exported to the Spartan 14 wave function program for equilibrium geometry optimization at ground state with density functional theory calculations (DFT/B3LYP/6-31G**) in a vacuum, starting from the initial molecular geometry (Adedirin et al. 2018a). In principle, geometry optimization is an iterative process whereby the energy and its first derivative with respect to all geometrical coordinates are calculated from a guess geometry, then used the information to project new geometry (Adedirin et al. 2018a). Thus, the process continues until the lowest energy or optimized structure of the molecule is achieved.

### Descriptors computation

The thirty-five (35) optimized structures of IPAs from Spartan 14 were accordingly saved as SD file format, then exported to the PaDEL descriptors software which is a product of Pharmaceutical Data Exploration Laboratory, created by Yap Chun Wai (Sanyal et al. 2019). This software allows QSAR users to compute diverse molecular descriptors and fingerprints of a molecule, including electrostatic, topological, spatial, autocorrelation, geometrical, constitutional, and thermodynamic descriptors (Abdullahi et al. 2018).

### Data pretreatment and division

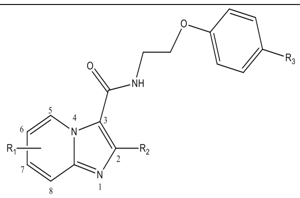
PaDEL descriptors output in MS Excel sheet was subjected to a variable reduction method so as to eliminate constant and highly inter-correlated descriptors based on user-specified variance and correlation coefficient cut-off value using Data Pre-treatment GUI 1.2, downloaded from Drug Theoretics and Cheminformatics (DTC) Laboratory. In order to have a rational selection of training set and test set, the Dataset Division GUI 1.0 software was used by engaging Kennard-Stone's algorithm division technique (Adeniji et al. 2019).

### QSAR Model generation and Validation

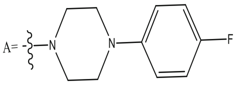
#### *Internal validation*

The training set compounds were used to build the five (5) multi-linear regression (MLR) models using Material

**Table 1** Substitution arrangement of the imidazo[1,2-a] pyridine-3-carboxamide (IPA) core and their anti-tubercular properties



IPAs core structure



| Compd | R1 | R2 | R3 | MIC($\mu g/mL$) | Experimental pMIC | Predicted pMIC | Residual | Leverage |
|---|---|---|---|---|---|---|---|---|
| 1[a] | 6-$NO_2$ | Me | Br | 1.28 | 5.515035 | 5.58658 | -0.07155 | 0.6776 |
| 2[a] | 6-F | Me | Br | 1.208 | 5.511263 | 5.592481 | -0.08122 | 0.1378 |
| 3[b] | 6-Cl | Me | Br | 0.488 | 5.922251 | 6.884464 | -0.96221 | 0.5507 |
| 4[b] | 6-Br | Me | Br | 0.358 | 6.101217 | 7.89805 | -1.79683 | 0.5785 |
| 5[a] | 6-OMe | Me | Br | 0.028 | 7.159288 | 7.523572 | -0.36428 | 0.4339 |
| 6[a] | 7-Me | Me | Br | 0.463 | 5.923318 | 6.146127 | -0.22281 | 0.3545 |
| 7[a] | 7-Cl | Me | Br | 2.682 | 5.182212 | 4.833061 | 0.349151 | 0.2951 |
| 8[a] | 8-Me | Me | Br | 13.925 | 4.445104 | 4.796104 | -0.351 | 0.3931 |
| 9[a] | 8-Cl | Me | Br | 0.505 | 5.907379 | 5.129843 | 0.777536 | 0.2190 |
| 10[a] | 5-Cl | Me | Br | 2.524 | 5.208581 | 5.089772 | 0.118809 | 0.1455 |
| 11[a] | 6-Cl | Et | Br | 0.023 | 7.263605 | 6.78672 | 0.476885 | 0.0987 |
| 12[a] | 6-F | Et | Br | 0.202 | 6.303228 | 5.687492 | 0.615736 | 0.1458 |
| 13[a] | 6-Br | Et | Br | 0.025 | 7.270418 | 7.799587 | -0.52917 | 0.2905 |
| 14[a] | 7-Cl | Et | Br | 6.354 | 4.822286 | 4.645328 | 0.176957 | 0.2719 |
| 15[a] | 8-Cl | Et | Br | 17.524 | 4.3817 | 5.052435 | -0.67074 | 0.1511 |
| 16[a] | 6-Me | Et | Br | 0.275 | 6.16498 | 6.298099 | -0.13312 | 0.2338 |
| 17[a] | 6-Me | n-Pr | Br | 0.415 | 6.001139 | 6.179424 | -0.17828 | 0.1484 |
| 18[a] | 6-Me | c-Pr | Br | 0.052 | 6.901081 | 6.485117 | 0.415964 | 0.2572 |
| 19[a] | 6-F | n-Pr | Br | 2.682 | 5.194863 | 5.545951 | -0.35109 | 0.1552 |
| 20[a] | 6-F | c-Pr | Br | 0.928 | 5.65368 | 5.798172 | -0.14449 | 0.2077 |
| 21[a] | 6-Cl | n-Pr | Br | 0.054 | 6.907133 | 6.497305 | 0.409827 | 0.1404 |
| 22[b] | 6-Cl | c-Pr | Br | 0.098 | 6.646284 | 6.777038 | -0.13075 | 0.9051 |
| 23[a] | 7-Cl | c-Pr | Br | 32 | 4.13236 | 4.633385 | -0.50103 | 0.3444 |
| 24[b] | 7-Me | c-Pr | Br | 5.004 | 4.917767 | 6.395343 | -1.47758 | 0.5851 |
| 25[b] | 6-Me | n-Pr | Cl | 0.631 | 5.770677 | 5.977592 | -0.20692 | 0.8379 |
| 26[a] | 6-F | Me | Cl | 3.841 | 4.957247 | 4.997813 | -0.04057 | 0.2087 |
| 27[a] | 7-Cl | Et | Cl | 5.862 | 4.809526 | 4.997813 | -0.18829 | 0.2087 |
| 28[a] | 6-Cl | Et | Cl | 0.029 | 7.115174 | 6.641123 | 0.474051 | 0.0841 |
| 29[b] | 6-Cl | n-Pr | Cl | 0.631 | 5.793356 | 6.511716 | -0.71836 | 0.1656 |
| 30[b] | 6-F | n-Pr | Cl | 1.141 | 5.518041 | 5.198363 | 0.319678 | 0.6873 |
| 31[b] | 7-Cl | c-Pr | Cl | 32 | 4.085993 | 4.538986 | -0.45299 | 0.6816 |
| 32[b] | 6-Cl | Et | OMe | 0.056 | 6.824823 | 7.918572 | -1.09375 | 0.4347 |
| 33[a] | 6-Me | c-Pr | OMe | 0.027 | 7.132331 | 7.348759 | -0.21643 | 0.1946 |
| 34[a] | 6-Me | Et | OMe | 0.023 | 7.187496 | 6.958363 | 0.229133 | 0.2011 |
| 35[b] | 6-Br | Et | A | 0.0625 | 6.921957 | 8.636852 | -1.7149 | 0.5731 |

Superscript “a” = training set
Superscript “b” = test set

[a] Training set
[b] Test set

Studio Software (Version 8.0) based on genetic function approximation (GFA) as the variable selection technique where the dependent variable is the logarithmic values of the minimum inhibitory concentration (pMIC) and the independent variables are the descriptors generated from PaDEL program (Tropsha 2010). Numerous internal validation metrics of the models were also generated using MLR-plus validation program such as:

a. Friedman lack-of-fit parameter (LOF) from the material studio is defined as

$$\mathrm{LOF} = \frac{\mathrm{SEE}}{j\left[1 - \beta\left(\frac{a+b\times c}{j}\right)\right]^2} \tag{2}$$

where $a$ = number of the terms in the model, $b$ = scaled smoothing factor, $c$ = corresponds to the entire number of descriptors in the model, $j$ = total number of compounds in the training set, and $\beta$ = a safety factor with a value of 0.99 which guarantee that the denominator of the equation can never be equal to zero.

b. Cross-validated parameter ($Q^2{}_{\mathrm{cv}}$)

$$\begin{aligned}\mathrm{Q2cv} = {} & 1 - \left[\frac{\sum\left(Y_{-\,Y_{\mathrm{pred}}}\right)^2}{\sum\left(Y - \overline{Y}_{tr}\right)^2}\right] \\ & = 1 - \frac{\mathrm{PRESS}}{\sum\left(Y - \overline{\mathrm{Y}}_{\mathrm{tr}}\right)^2} =\end{aligned} \tag{3}$$

$\overline{Y}_{tr}$ = average observed concentrations of the training set, $Y$ = observed concentration, and $Y_{\mathrm{pred}}$ = predicted concentration in the training set, respectively.

c. Regression coefficient squared ($R^2{}_{\text{train and test}}$)

$$R^2 = \frac{\left[\sum\left\{\left(Y - \overline{Y}_{\mathrm{tr}}\right)\left(Y_{\mathrm{pred}} - \overline{Y}_{\mathrm{pred}}\right)\right\}\right]^2}{\sum\left(Y - \overline{Y}_{\mathrm{tr}}\right)^2 \sum\left(Y_{\mathrm{pred}} - \overline{Y}_{\mathrm{pred}}\right)^2} \tag{4}$$

$Y_{\mathrm{pred}}$ and $Y$ were predicted and observed training set concentration (experimental), respectively. $\overline{Y}_{\mathrm{tr}}$ and $\overline{Y}_{\mathrm{pred}}$ were the average observed (experimental) and predicted training set response, respectively

d. Coefficient of determination adjusted ($R^2{}_{\mathrm{adjusted}}$)

$$R^2_{\mathrm{adj}} = \frac{R^2 - p(n-1)}{n-1-p} \tag{5}$$

where $\boldsymbol{p}$ = number of the descriptor in the model, $\boldsymbol{n}$ = number of compounds in the training set, $R^2$ is the correlation coefficient, and $n - 1 - p$ is the degree of freedom.

e. Variance ratio $F$ (Fischer's value)

$$F = \frac{\dfrac{\sum\left(Y_{\mathrm{pred}} - \overline{Y}_{\mathrm{tr}}\right)^2}{p}}{\dfrac{\sum\left(Y - Y_{\mathrm{pred}}\right)^2}{n-p-1}} \tag{6}$$

f. Standard errors of estimate (SEE)

$$\mathrm{SEE} = \sqrt{\frac{\left(Y - Y_{\mathrm{pred}}\right)^2}{n-p-1}} \tag{7}$$

g. Average modified square of correlation coefficient ( $\overline{R}^2_m$)

$$\overline{R}^2_m = \frac{r'^2_{\mathrm{m}} + r^2_{\mathrm{m}}}{2} \tag{8}$$

where $r'^2_{\mathrm{m}}$ and $r^2_{\mathrm{m}}$ represent the reverse and modified square of correlation coefficient computed according to the expressions below:

$$r'^2_{\mathrm{m}} = r^2 \times \left(1 - \sqrt{r^2 - \mathrm{r}'^2_0}\right) \tag{9}$$

$$r^2_{\mathrm{m}} = r^2 \times \left(1 - \sqrt{r^2 - r^2_o}\right) \tag{10}$$

where $r^2$ and $r^2_o$ represent the correlation coefficients of the plot of observed against predicted training set concentrations with and without intercept, respectively. $r_0^{'2}$ is a squared correlation coefficient of the plot of predicted versus observed training set response without intercept (Veerasamy et al. 2011).

h. Delta modified square of correlation coefficient ($\Delta r_\mathrm{m}^2$)

$$\Delta r_\mathrm{m}^2 = \left| r_\mathrm{m}^2 - r_\mathrm{m}^{'2} \right| \tag{11}$$

i. $Y$ randomization parameters ($^{C}R_\mathrm{p}^2$)

$$^{C}R_p^2 = R^2 \times \left( R^2 - (\mathrm{Average}R_r)^2 \right)^{1/2} \tag{12}$$

where c$R_\mathrm{p}^2$ = coefficient of determination, $R$ = correlation of coefficient, and $R$r = average "$R$" of random models.

### *External validation*

The QSAR models predictive competency were examined by using independent test set compounds for external validation, and the metrics proposed by Golbraikh and Tropsha (Tropsha 2010) were also computed using MLRplusValidation 1.3 program as follows:

i. Predicted determination coefficient for test set data $R_\mathrm{Pred}^2$ expressed as:

$$R_\mathrm{pred}^2 = 1 - \frac{\sum \left( Y\mathrm{pred}_\mathrm{test} - Y_\mathrm{test} \right)^2}{\sum \left( Y_\mathrm{test} - \bar{Y}_\mathrm{tr} \right)^2} \tag{13}$$

where $Y\mathrm{pred}_\mathrm{test}$ and $Y_\mathrm{test}$ are the predicted and observed concentration of test set compounds respectively. $\bar{Y}_{tr}$ = average values of observed concentration of the training set compounds.

ii. $\frac{r^2 - r_0^{'2}}{r^2} < 0.1$ or $\frac{r^2 - r_0^2}{r^2} < 0.1$ and $\left| r_0^2 - r_0^{'2} \right| < 0.3$

where $r^2$ = squared correlation coefficient between the observed and predicted activities with intercept and $r_0^2$ = squared correlation coefficient between the predicted and observed concentration without intercept (Tropsha 2010).

iii. Root mean square error of prediction RMSEP is defined as:

$$\mathrm{RMSEP} = \sqrt{\frac{1}{n} * \sum \left( Y_\mathrm{test} - Y_\mathrm{pred(test)} \right)^2} \tag{14}$$

where $Y_\mathrm{(test)}$ and $Y_\mathrm{pred(test)}$ are the observed and predicted test set concentrations, respectively.

iv. The slope of the plot of observed against the predicted concentration of test set compounds without intercept ($k$) and plot of predicted against the observed concentration of test set compounds without intercept ($k^{'}$) are expressed as:

$$k = \frac{\sum \left( Y_\mathrm{(test)} \times Y_\mathrm{pred(test)} \right)}{\sum \left( Y_\mathrm{pred(test)} \right)^2} \tag{15}$$

$$k^{'} = \frac{\sum \left( Y_\mathrm{obs(test)} \times Y_\mathrm{pred(test)} \right)}{\sum \left( Y_\mathrm{obs(test)} \right)^2} \tag{16}$$

However, an acceptable and predictive QSAR model should have $0.85 < k < 1.15$ or $0.85 < k' < 1.15$ (Tropsha 2010)

### *Applicability domain (AD)*

The applicability domain (AD) of the developed model is defined as the chemical space of compound structure and response where the model predictions are highly reliable (Veerasamy et al. 2011). This technique is used to detect the presence of response and structural outliers in the test set and training set compounds, respectively (Tropsha 2010; Veerasamy et al. 2011; Gramatica 2007). The leverage approach of evaluating the model's applicability domain based on models distance measure can be utilized by plotting a scatter plot of standardized residual response and leverage values ($f$) of both training set and test set (Williams plot). The leverage of compound ($f$) can be determined as follows:

$$\boldsymbol{f} = x\left( x^T x \right)^{-1} x^T \tag{17}$$

where $x$ is the model's descriptors matrix, $x^T$ represents the transpose matrix $x$, and $F$ is the diagonal element of the hat matrix. In this study, AD of the QSAR model was evaluated as the square area with vertical boundary $0 < \boldsymbol{f}_\mathbf{i} < f^*$ and horizontal boundary $-3 <$ standardized residual $< 3$, where $\boldsymbol{f}_\mathrm{i}$ is leverage values of compounds and $f^*$ is the threshold leverage expressed as:

$$f* = \frac{3 \cdot (m+1)}{p} \tag{18}$$

where $\boldsymbol{p}$ is the number of molecules in the training set and $\boldsymbol{m}$ is the number of molecular descriptors used in

**Table 2** Internal validation of the five (5) QSAR models

| Parameters | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Threshold |
|---|---|---|---|---|---|---|
| Friedman (LOF) | 0.8669 | 0.9431 | 0.9408 | 0.9376 | 1.4731 | – |
| *R*-squared (training set) | 0.8563 | 0.8436 | 0.8440 | 0.8446 | 0.7234 | > 0.6 |
| Adjusted *R*-squared | 0.8185 | 0.8025 | 0.8030 | 0.8037 | 0.6542 | > 0.6 |
| Cross validated ($Q^2_{cv}$) | 0.7543 | 0.7315 | 0.7423 | 0.6620 | 0.5781 | > 0.5 |
| Standard error of estimation (SEE) | 0.4336 | 0.4522 | 0.4517 | 0.4509 | 0.5671 | – |
| Variance ratio (*F* value) | 22.648 | 20.5100 | 20.5700 | 20.6538 | 10.4625 | – |
| PRESS | 3.5724 | 3.8866 | 3.8775 | 3.8638 | 6.4323 | > 0.5 |
| Average $\overline{R}^2_m$ (LOO-train) | 0.6751 | 0.6501 | 0.6564 | 0.5721 | 0.4771 | > 0.5 |
| Delta $\overline{R}^2_m$ (LOO-train) | 0.0898 | 0.0680 | 0.1184 | 0.0971 | 0.0755 | – |
| Computed experimental error | 0.1044 | 0.1044 | 0.1044 | 0.1044 | 0.1 | – |
| RMSEP | 1.0657 | 1.1101 | 1.1320 | 0.9995 | 0.3388 | – |
| *Y* randomization ($^{C}R^2_p$) | 0.7609 | 0.7392 | 0.7468 | 0.7516 | 0.6296 | > 0.5 |

the model. In addition, compounds with higher leverage scores which are greater than threshold leverage ($f_i > f^*$) tend to have unreliable predictions. However, compounds whose leverage scores are less than the threshold score ($f_i < f^*$) and the standardized residuals are not greater than ± 3α (3 standard deviation units) are said to fall within the applicability domain (Adedirin et al. 2018b). Similarly, the Euclidean approach of the applicability domain was also determined based on mean distance scores computed by the euclidean distance. As such, the Uzairu plot was determined by plotting the standardized residuals against normalized mean distance scores whose ranges are from 0 to 1 (Arthur et al. 2018). The normalized mean distance score for training set ranges from 0 which is for least diverse and 1 which is for the most diverse training set. However, the normalized mean distance scores for test compounds with scores outside 0 to 1 are regarded as outliers which are outside the applicability domain (Arthur et al. 2018)

### In silico ADME prediction

The designed hypothetical molecules in SMILES format (Simplified Molecular Input Line Entry System) were exported to SwissADME online webserver to predict their absorption, distribution, metabolism, and excretion (ADME) properties (Pan et al. 2019).

## Results

Model 1

$$\begin{aligned} pMIC = {} & 0.549089725*AATS1i - 17.773108385*ATSC4c \\ & - 0.002133080*ATSC3v + 28.701878113*MATS5p \\ & + 4.867881082*GATS6c - 79.733701170 \end{aligned} \tag{19}$$

Model 2

**Table 3** External validation of the models generated

| Parameter | Threshold score | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|---|
| $R^2_{\text{pred}}$ | $R^2_{\text{pred}} > 0.5$ | 0.7543 | 0.7315 | 0.7423 | 0.6620 | 0.5781 |
| $R^2_{\text{test}}$ | $R^2_{\text{test}} > 0.6$ | 0.6993 | 0.6858 | 0.6656 | 0.2745 | 0.8572 |
| Average $\overline{R}^2_m$ (test) | > 0.5 | 0.5954 | 0.6501 | 0.5530 | 0.0776 | 0.7284 |
| $\Delta\overline{R}^2_m$ (test) | < 0.5 | 0.0244 | 0.0680 | 0.1295 | 0.3724 | 0.1310 |
| RMSEP | – | 1.0657 | 1.1101 | 1.132 | 0.9995 | 0.3388 |
| $r^2_0$ | > 0.5 | 0.5246 | 0.4725 | 0.3121 | 0.26839 | 0.8556 |
| $r'^2_0$ | > 0.5 | 0.6975 | 0.6825 | 0.6531 | − 0.98759 | 0.8168 |
| $\lvert r^2_0 - r'_0 \rvert$ | $\lvert r^2_0 - r'_0 \rvert < 0.3$ | 0.1728 | 0.2100 | 0.3409 | 1.25598 | 0.0388 |
| $k$ | $0.85 < k < 1.15$ | 0.8674 | 0.8631 | 0.8671 | 0.89299 | 0.98011 |
| $k'$ | $0.85 < k' < 1.15$ | 1.1419 | 1.1464 | 1.1374 | 1.10363 | 1.0164 |
| $(r^2 - r_0^2)/r^2$ | $(r^2 - r_0^2)/r^2 < 0.1$ | 0.2497 | 0.3109 | 0.5308 | 0.0225 | 0.0018 |
| $(r^2 - r'^2_0)/r^2$ | $(r^2 - r'^2_0)/r^2 < 0.1$ | 0.0026 | 0.0047 | 0.0187 | 4.59689 | 0.0471 |

**Table 4** Predicted descriptors score values for training set compounds (model 1)

| Compd ID | AATS1i | ATSC4c | ATSC3v | MATS5p | GATS6c |
|---|---|---|---|---|---|
| 1 | 149.9963 | − 0.13352 | − 1.64514 | − 0.15433 | 1.029607 |
| 2 | 148.4884 | − 0.23029 | 6.144138 | − 0.15996 | 0.884163 |
| 5 | 147.9648 | − 0.27276 | − 24.5695 | − 0.09900 | 0.811955 |
| 6 | 147.2553 | − 0.07817 | − 252.455 | − 0.07509 | 1.078683 |
| 7 | 147.294 | − 0.04072 | − 9.00586 | − 0.08356 | 1.097935 |
| 8 | 147.2553 | − 0.19434 | − 176.800 | − 0.17680 | 1.010072 |
| 9 | 147.294 | − 0.15973 | 76.07726 | − 0.11515 | 0.947924 |
| 10 | 147.294 | − 0.22815 | 244.1761 | − 0.17113 | 1.093596 |
| 11 | 147.0975 | − 0.2269 | 305.2975 | − 0.11890 | 1.187731 |
| 12 | 148.2123 | − 0.25337 | 387.6355 | − 0.17959 | 1.133481 |
| 13 | 146.8087 | − 0.21656 | 266.3534 | − 0.07808 | 1.208384 |
| 14 | 147.0975 | − 0.06399 | 389.3469 | − 0.11307 | 1.345093 |
| 15 | 147.0975 | − 0.18486 | 474.4300 | − 0.14108 | 1.189836 |
| 16 | 147.0735 | − 0.19866 | 42.37038 | − 0.15855 | 1.311779 |
| 17 | 146.9131 | − 0.17938 | 128.5620 | − 0.14213 | 1.316857 |
| 18 | 146.2623 | − 0.21964 | 227.8612 | − 0.15131 | 1.403673 |
| 19 | 147.9707 | − 0.23392 | 477.2540 | − 0.16200 | 1.138185 |
| 20 | 147.3009 | − 0.27328 | 589.0215 | − 0.17161 | 1.227496 |
| 21 | 146.9255 | − 0.20751 | 406.1632 | − 0.10696 | 1.192311 |
| 23 | 146.2335 | − 0.08422 | 592.4304 | − 0.11019 | 1.438241 |
| 27 | 148.7978 | − 0.24243 | 27.51959 | − 0.19152 | 0.878249 |
| 26 | 148.7978 | − 0.24243 | 27.51959 | − 0.19152 | 0.878249 |
| 28 | 147.3862 | − 0.23904 | 322.5570 | − 0.13408 | 1.178039 |
| 33 | 147.1508 | − 0.28249 | 267.0596 | − 0.15284 | 1.27758 |
| 34 | 147.9338 | − 0.26151 | 96.44824 | − 0.16756 | 1.197733 |

**Table 5** Predicted descriptors score values for external test set compounds (model 1)

| Test set | AATS1i | ATSC4c | ATSC3v | MATS5p | GATS6c |
|---|---|---|---|---|---|
| 22 | 146.2335 | − 0.24714 | 508.3810 | − 0.11640 | 1.283596 |
| 24 | 146.2623 | − 0.12293 | 302.5973 | − 0.09294 | 1.426912 |
| 25 | 147.1679 | − 0.19151 | 140.3329 | − 0.15875 | 1.305459 |
| 35 | 148.1172 | − 0.36160 | 396.3149 | − 0.12156 | 1.016583 |
| 29 | 147.1962 | − 0.21965 | 418.3325 | − 0.1166 | 1.182565 |
| 3 | 147.2940 | − 0.20363 | − 93.0553 | − 0.09084 | 0.930660 |
| 30 | 148.2413 | − 0.24606 | 486.5724 | − 0.18462 | 1.129426 |
| 31 | 146.5099 | − 0.09636 | 608.9053 | − 0.12294 | 1.425727 |
| 32 | 148.0084 | − 0.28975 | 324.8536 | − 0.11810 | 1.091907 |

$$\begin{aligned} p\mathrm{MIC} = {} & -14.169744187 \ast AATS4i \\ & + 3.834886578 \ast ATSC1v \\ & + 31.519004609 \ast MATS1c \\ & - 5.211789063 \ast MATS1s \\ & + 0.978391331 \ast SM1Dzi + 12.391321643 \end{aligned} \tag{20}$$

Model 3

$$\begin{aligned} pMIC = {} & 0.394587981 \ast AATS6p \\ & - 18.951952930 \ast AATS4i \\ & - 0.002471214 \ast ATSC3c \\ & + 29.595804745 \ast MATS1c \\ & + 5.833470377 \ast MATS2e - 64.610615353 \end{aligned} \tag{21}$$

Model 4

$$\begin{aligned} pMIC = {} & 12.944502696 \ast AATS2i \\ & + 0.713555545 \ast ATSC5s \\ & - 8.188486787 \ast AATSC4p \\ & - 21.641364078 \ast AATSC0s \\ & - 1.244294688 \ast GATS6i + 9.131873810 \end{aligned} \tag{22}$$

Model 5

$$\begin{aligned} pMIC = {} & 0.004102352 \ast ATS4e \\ & + 14.864143471 \ast AATSC6v \\ & - 19.476022072 \ast AATSC3e \\ & + 13.552890779 \ast MATS7s \\ & + 6.735482707 \ast GATS5c - 19.644784827 \end{aligned} \tag{23}$$

## Discussion

### QSAR modelling analysis

The logic behind the development of a QSAR model is to arrive at relevant molecular descriptors that describe changes in the structural features of a compound. Molecular descriptors of all compounds in this study were generated using the PaDEL software as mentioned earlier. A total sum of 625 diverse descriptors was generated in MS Excel (.csv) format, and the result was exported to the DTC lab software for the pretreatment and division. In the data pretreatment process, non-informative and highly inter-correlated descriptors with correlation cutoff greater than 0.8 were removed, which reduces about 24.32% of the total descriptors computed by the PaDEL program amounting to 152 descriptors. The pretreated data were divided into the training set and test set based on Kennard-Stone permutation, where 70% of the dataset (25 compounds) are the training set and the remaining 30% (10 compounds) are the test set. Based on the genetic function approximation of the descriptors from the Material Studio software, five (5) multilinear regression models (Eqs.19, 20, 21, 22 and 23) were

**Table 6** Definition of the descriptors in the QSAR model 1

| Descriptor java class | Descriptor | Description | Class | Contribution |
|---|---|---|---|---|
| Autocorrelation descriptor | AATS1i | Average Broto-Moreau autocorrelation–lag 1/weighted by first ionization potential | 2D | Positive |
| Autocorrelation descriptor | ATSC4c | Centered Broto-Moreau autocorrelation–lag 4/weighted by charges | 2D | Negative |
| Autocorrelation descriptor | ATSC3v | Centered Broto-Moreau autocorrelation–lag 3/weighted by van der Waals volumes | 2D | Negative |
| Autocorrelation descriptor | MATS5p | Moran autocorrelation–lag 5/weighted by polarizabilities | 2D | Positive |
| Autocorrelation descriptor | GATS6c | Geary autocorrelation–lag 6/weighted by charges | 2D | Positive |

developed containing five (5) optimum descriptors, and model 1 was selected as the best model due to its statistical significance of the internal and external validation metrics. The experimental pMIC reported in the literature, the predicted pMIC computed by model 1 for all the 35 anti-tubercular agents, the residual scores, and the leverage values are shown in Table 1. The residual score is the difference between the experimental pMIC and predicted pMIC, and the lower residual score signifies that the developed model has good predictive potentials. The internal validation parameters of the models generated are presented in Table 2, and model 1 revealed the most significant descriptors. The external validations of the models generated are shown in Table 3, and model 1 has also passed the external validation metrics proposed by Golbraikh and Tropsha including the error-based judgment of test set compounds (Tropsha 2010). In addition, the validation metrics reported in this study are in agreement with the metrics from literature and for the purpose of reproducibility; all the computed descriptors for both the training and test set in model 1 are reported in Tables 4 and 5, respectively.

Table 6 provides a comprehensive description of the molecular descriptors in the model 1. Furthermore, the model showed a positive contribution of MATS5p, GATS6c, and AATS1i descriptors, while a negative contribution for descriptors ATSC4c and ATSC3v, respectively. This means that the increment in the magnitude of MATS5p, GATS6c, and AATS1i descriptors will positively influence the prediction of pMIC with the negative influence of ATSC4c and ATSC3v descriptors. However, the MATS5p descriptor has the highest contribution which is the most significant descriptor to be considered for designing new hypothetical compounds. The regression coefficient squared ($R^2_{\text{train}}$ = 0.8536 and $R^2_{\text{test}}$ = 0.7543) indicates good extrapolation between the training set and test set. In addition, the models generated are robust due to the small differences in $R^2$ and $Q^2_{\text{cv}}$ (< 0.3).

The regression statistics (Table 7) show *P* value and *t* values of the model which suggests that the coefficients of the descriptors are statistically significant at a 95% confidence level. Furthermore, inter-correlated descriptors in model 1 were assessed based on their multicollinearity computed as the variation inflation factor (VIF):

$$\text{VIF} = \left(1-R^2\right)^{-1} \tag{24}$$

where $R^2$ represents the correlation coefficient. VIF values corresponding to unity depict no inter-correlation among each variable; if the VIF scores ranging from 1 to 5, as such the model is acceptable and stable. But if the VIF scores larger than 10, it means that the model in question is unstable and unacceptable (Driouche and Messadi 2019). Table 8 shows the correlation analysis and VIF values of the descriptors in model 1. The non-existence of inter-correlation among the descriptors could be observed between descriptors pair, and the VIF values of each descriptor do not exceed 4 depicting that the descriptors in the model are stable. The plot of predicted pMIC against experimental pMIC values is shown in Fig. 1. It could be seen that the values of the test sets are in close agreement with the training set values. The scatter plot of standardized residual against experimental pMIC values (Fig. 2) showed a random scattering of data

**Table 7** Regression statistics of the descriptors in model 1

| | Coefficients | Standard error | *t* stat | *P* value | Lower 95% | Upper 95% |
|---|---|---|---|---|---|---|
| Intercept | − 79.7337 | 26.01779 | − 3.06458 | 0.006379 | − 134.19 | − 25.2778 |
| AATS1i | 0.54909 | 0.172825 | 3.177139 | 0.004962 | 0.187362 | 0.910817 |
| ATSC4c | − 17.7731 | 1.752082 | − 10.144 | 4.18E− 09 | − 21.4403 | − 14.106 |
| ATSC3v | − 0.00213 | 0.000515 | − 4.14019 | 0.000556 | − 0.00321 | − 0.00105 |
| MATS5p | 28.70188 | 3.637264 | 7.891062 | 2.05E− 07 | 21.089 | 36.31476 |
| GATS6c | 4.867881 | 0.924533 | 5.26523 | 4.41E− 05 | 2.93281 | 6.802952 |

**Table 8** Correlation matrix of the descriptors from the built model 1

| Descriptors | AATS1i | ATSC4c | ATSC3v | MATS5p | GATS6c | VIF |
|---|---|---|---|---|---|---|
| AATS1i | 1 | − 0.20028 | − 0.36769 | − 0.45573 | − 0.69559 | 2.713443 |
| ATSC4c | | 1 | − 0.15354 | 0.569614 | 0.222739 | 1.832864 |
| ATSC3v | | | 1 | − 0.0374 | 0.58439 | 1.754699 |
| MATS5p | | | | 1 | 0.172072 | 2.068945 |
| GATS6c | | | | | 1 | 3.130259 |

above and below the baseline of the standardized residual of zero which signified the non-existence of systematic error. The Williams plot (Fig. 3) which is the scatter plot of standardized residuals versus leverages revealed two (2) response outliers (compounds 22 and 25). This is because their leverage scores are greater than the threshold ($h^*$) of 0.72 which may be due to the changes in substitution arrangement of the substituents on the parent structure. However, the remaining compounds whose leverage score less than the threshold score are within the applicability domain of square area of ± 2.5. Also, the Uzairu plot (Fig. 4) showed that all compounds fall within the chemical space of the model which confirmed its predictive capabilities.

### In silico design of new compounds

In order to explore newly hypothetical compounds with improved anti-tubercular activity, the QSAR model 1 was utilized due to its robust statistical metrics as mentioned earlier. In silico screening reduced cost and time of identifying new hits or lead compounds. This is done by substitution, deletion, insertions, or addition of substituents to the scaffold (template) or lead compound. Compound 13 was chosen as the template due to its relatively higher pMIC of 7.2704, a low absolute residual score of 0.5291, and it is within the chemical space or applicability domain of the model with leverage score of 0.2905. The alteration was successfully done around its benzene ring and imidazo[1,2-a] pyridine moiety as shown in Table 9. Subsequently, modifications were done by inserting different amino-N analogs such as $-N(CH_3)_2$, $-NO_2$, $-NHCH_3$, and sulfur (S) containing substituents, and $-OCH_3$ which enhances the molecular polarizability as suggested by the MATS5p descriptor. On this note, six (6) newly designed compounds were obtained with excellent predicted MIC value greater than the experimental MIC value for the template (7.2704) and also having leverage scores less than the threshold leverage (0.72) which indicated that the designed compounds are within the chemical space of the model used for predicting the anti-tubercular response.

### In silico ADME prediction of designed compounds

The ADME and drug-likeness analysis are very important in the drug discovery which helps to make a rational decision on whether inhibitors can be administered to a biological system or not (Attique et al. 2019). Furthermore, the inhibitors with poor ADME properties and high toxicity effects on the biological systems are often the major cause of most failed medicines in the clinical phase of

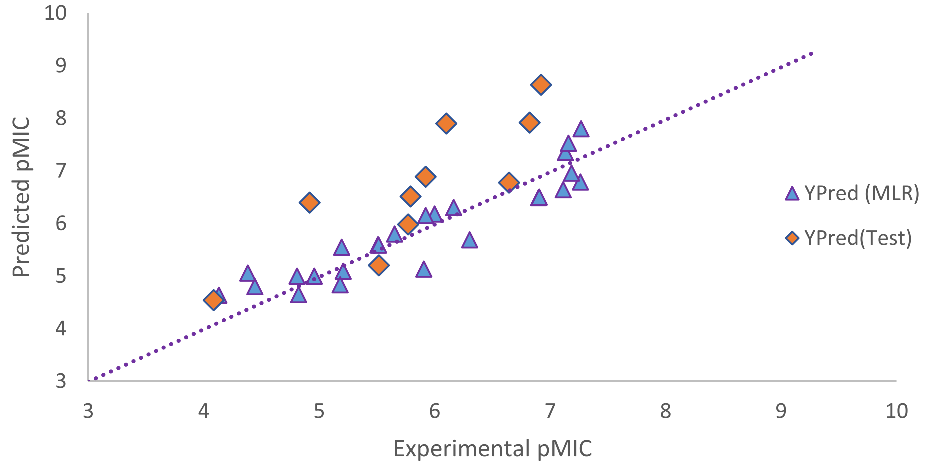


**Fig. 1** The plot of predicted against experimental pMIC for training and test set

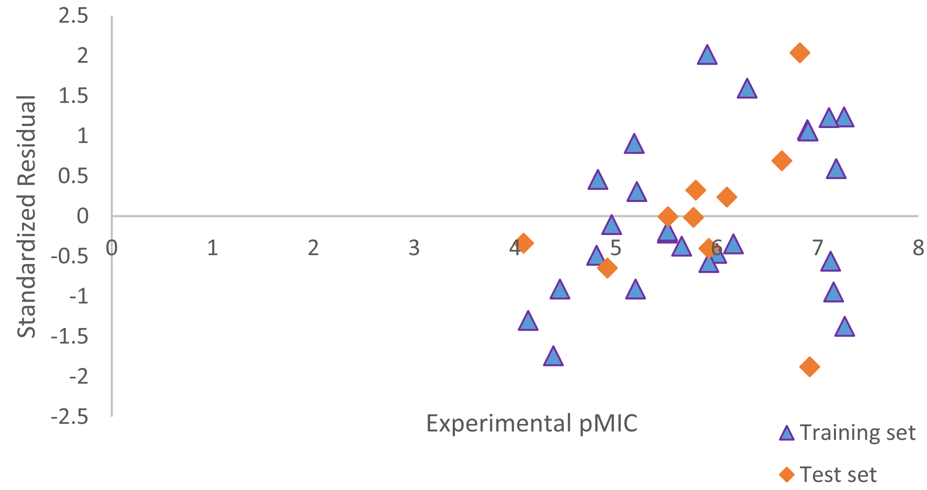


**Fig. 2** The scatter plot of standardized residual versus experimental pMIC

experiments (Attique et al. 2019). In an effort to improve the success rate of our QSAR modelling analysis, the pharmacokinetics properties of the designed compounds were predicted and assessed using the SwissADME online software as mentioned earlier, revealing that D1, D2, D3, D4, D5, and D6 obey Linpinski's rule of five (5), which indicates the claim prospect of these compounds as novel MTB inhibitors. The Lipinski's rule of five (5) is a thumb-rule for evaluating drug-likeness and to decide if an inhibitor with a certain pharmacological or biological properties would be an orally active drug in the human body (Daina et al. 2017). The rule states that a molecule or an inhibitor can be orally absorbed/active if two (2) or more of these thresholds, molecular weight (Mw) of molecule ˂ 500, octanol/water partition coefficient (iLOGP) ≤ 5, number of hydrogen bond acceptors (nHBA) ≤ 10, number of hydrogen bond donors (nHBD) ≤ 5, and topological polar surface area (TPSA ˂ 140 Å$^2$), are not violated (Daina et al. 2017). From the output of some ADME and drug-likeness properties shown in Table 10, it was observed that only D1 molecule has zero violation of the Lipinski's rule, but D2, D3, D4, D5, and D6 respectfully violated molecular weight rule.

## Conclusion

In this research, chemometric modelling analysis has been thoroughly used on 35 IPA molecules as potential anti-tubercular agents. As such, a regression-dependent quantitative structure-activity relationship (QSAR) model was fabricated and defended with multiple statistical

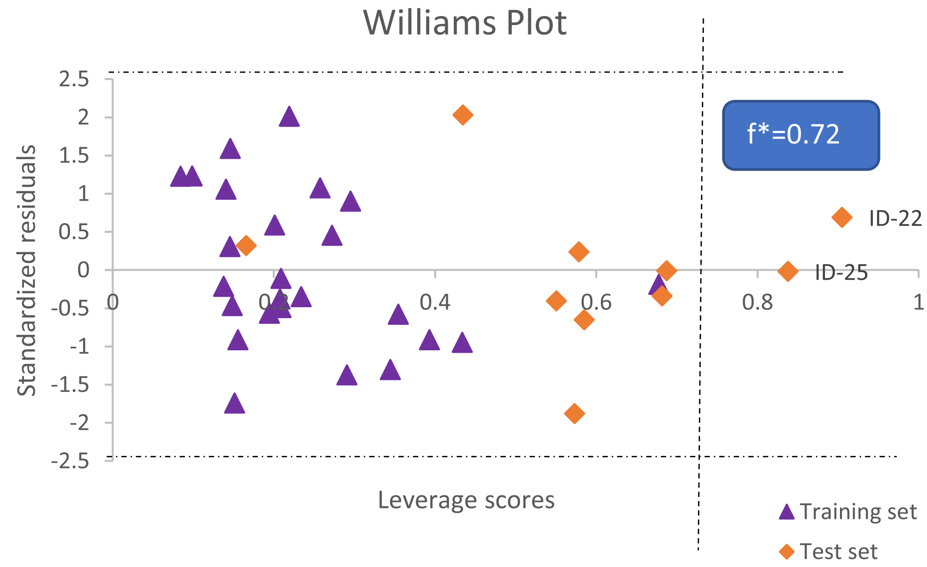


**Fig. 3** The scatter plot of standardized residuals and leverages (Williams plot)

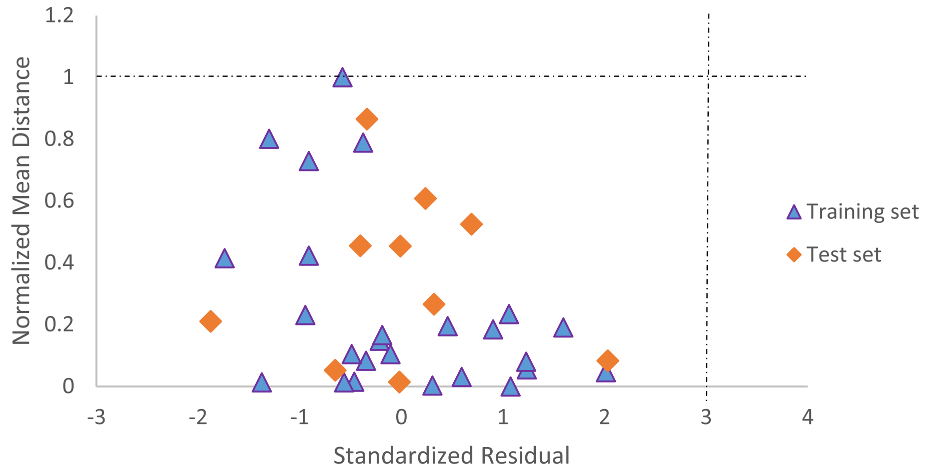
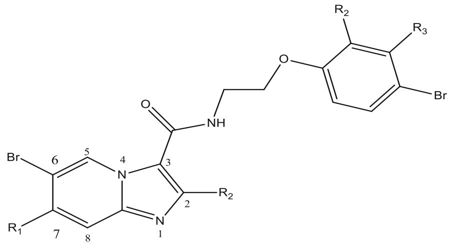


**Fig. 4** The scatter plot of standardized residuals and normalized mean distance (Uzairu's plot)

parameters according to Golbraikh and Tropsha (Tropsha 2010). The internal and external validation confirmed the robustness and reliability of the built QSAR model. Molecular descriptors, MATS5p, GATS6c, AATS1i, ATSC4c, and ATSC3v, from the results (model 1) are the optimum descriptors needed to predict the bioactivities of the compounds. Based on the information obtained from model 1, compound 13 was used as a template for the in silico design due to its high pMIC, and it is within the chemical space of the model. Thereafter, six (6) newly designed compounds with better anti-tubercular activity and good ADME/drug-likeness properties were obtained. According to the above work, the designed compounds have shown substantial prospective therapy against *Mycobacterium tuberculosis*. However, the research encouraged further experimental validation of the designed compounds against *Mycobacterium tuberculosis* through in vivo and in vitro considerations.

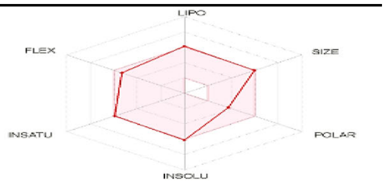
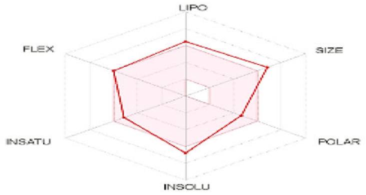
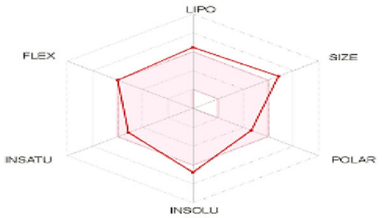
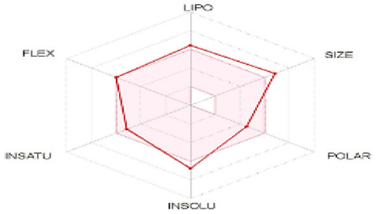
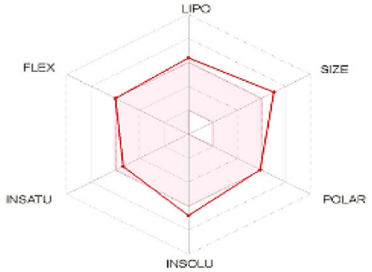
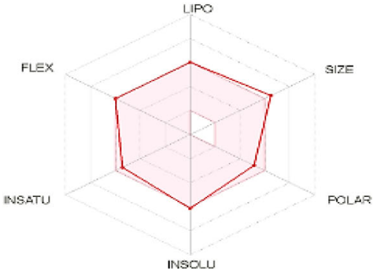
**Table 9** In silico designed IPAs and their predicted anti-tubercular activities



Template

| Compd | $R_1$ | $R_2$ | $R_3$ | AATS1i | ATSC4c | ATSC3v | MATS5p | GATS6c | Predicted. p MIC | Leverage Score |
|---|---|---|---|---|---|---|---|---|---|---|
| D1 | H | NH(CH3) | H | 148.7519 | -0.22086 | 377.9094 | -0.08605 | 1.04392 | 7.6757 | 0.3896 |
| D2 | $S(CH_3)_3$ | NH(CH3) | H | 147.2452 | -0.15159 | -603.332 | 0.053355 | 0.922259 | 11.1192 | 0.3658 |
| D3 | NH(CH3) | $S(CH_3)_3$ | H | 147.2452 | -0.17844 | -567.192 | -0.02876 | 0.903645 | 9.0717 | 0.2947 |
| D4 | $OCH_3$ | $S(CH_3)_3$ | H | 146.0809 | -0.14582 | -676.493 | -0.0034 | 0.883605 | 8.7161 | 0.4310 |
| D5 | $S(CH_3)_3$ | H | $NO_2$ | 148.4825 | -0.05314 | -353.443 | -0.13867 | 1.043886 | 11.2328 | 0.6642 |
| D6 | NH(CH3) | $SCH_3$ | H | 146.8828 | -0.12767 | -592.93 | 0.085611 | 0.888175 | 9.6034 | 0.4324 |

**Table 10** ADME and drug-likeness parameters of the designed IPA molecules

| Properties | BS | MW | nHBA | nHBD | TPSA | iLOGP | nLV | Bio |
|---|---|---|---|---|---|---|---|---|
| D1 | 0.55 | 496.2 | 3 | 2 | 67.66 | 3.85 | 0 |  |
| D2 | 0.55 | 572.36 | 3 | 2 | 92.96 | 0 | 1 |  |
| D3 | 0.55 | 572.36 | 3 | 2 | 92.96 | 0 | 1 |  |
| D4 | 0.55 | 573.34 | 4 | 1 | 90.16 | 0 | 1 |  |
| D5 | 0.55 | 588.31 | 5 | 1 | 126.75 | 0 | 1 |  |
| D6 | 0.55 | 543.28 | 4 | 2 | 105.85 | 4.44 | 1 |  |

Key: Synthetic Accessibility (SA), Molecular weight (MW), Number of hydrogen bond donor (nHBD), Number of hydrogen bond acceptor (nHBA), Topological polar surface area (TPSA), octanol/water partition coefficient (iLOGP), Number of Lipinski violation (nLV)

Key: *SA* synthetic accessibility, *MW* molecular weight, *nHBD* number of hydrogen bond donor, *nHBA* number of hydrogen bond acceptor, *TPSA* topological polar surface area, *iLOGP* octanol/water partition coefficient, *nLV* number of Lipinski violation

### Abbreviations

MLR: Multi-linear regression; GFA: Genetic function approximation; QSAR: Quantitative structure-activity relationship; MTB: *Mycobacterium tuberculosis*

### Acknowledgements

We are thankful to the members of staff in the physical chemistry unit, chemistry department of Ahmadu Bello University for providing the essential facilities to carry out this research work.

### Authors' contributions

MA handled the computational chemistry softwares and drafted the manuscript. SEA contributed with technical advice during revision process. DEA designed the research work. SM helped in proofreading and sentence corrections. All authors have read and approved the final manuscript.

### Funding

Not applicable.

### Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

Received: 7 January 2020 Accepted: 28 July 2020
Published online: 18 August 2020

### References

Abdullahi M, Shallangwa GA, Ibrahim MT et al (2018) QSAR studies on some C14-urea tetrandrine compounds as potent anti-cancer agents against leukemia cell line (K562). Journal of the Turkish Chemical Society, Section A: Chemistry 5(3):1387–1398. https://doi.org/10.18596/jotcsa.457618

Abdullahi M, Uzairu A, Shallangwa GA, Mamza P, Arthur DE, Ibrahim MT. In-silico modelling studies on some C14-urea-tetrandrine derivatives as potent anti-cancer agents against prostate (PC3) cell line. Journal of King Saud University - Science https://doi.org/10.1016/j.jksus.2019.01.008. Published 2019.

Adedirin O, Uzairu A, Shallangwa GA, Abechi SE (2018a) Computational studies on α-aminoacetamide derivatives with anticonvulsant activities. Beni-Suef University Journal of Basic and Applied Sciences 7(4):709–718. https://doi.org/10.1016/j.bjbas.2018.08.005

Adedirin O, Uzairu A, Shallangwa GA, Abechi SE (2018b) Optimization of the anticonvulsant activity of 2-acetamido-N-benzyl-2-(5-methylfuran-2-yl) acetamide using QSAR modeling and molecular docking techniques. Beni-Suef University Journal of Basic and Applied Sciences 7(4):430–440. https://doi.org/10.1016/j.bjbas.2018.03.010

Adeniji SE, Uba S, Uzairu A, Arthur DE (2019) A derived QSAR model for predicting some compounds as potent antagonist against Mycobacterium tuberculosis : a theoretical approach. Adv Prev Med 2019:1–18. https://doi.org/10.1155/2019/5173786

Arthur DE, Uzairu A, Mamza P, Abechi SE, Shallangwa G (2018) Activity and toxicity modelling of some NCI selected compounds against leukemia P388ADR cell line using genetic algorithm-multiple linear regressions. Journal of King Saud University - Science. https://doi.org/10.1016/j.jksus.2018.05.023

Attique SA, Hassan M, Usman M et al (2019) A molecular docking approach to evaluate the pharmacological properties of natural and synthetic treatment candidates for use against hypertension. Int J Environ Res Public Health 16(923):1–17. https://doi.org/10.3390/ijerph16060923

Daina A, Michielin O, Zoete V. SwissADME : a free web tool to evaluate pharmacokinetics , drug- likeness and medicinal chemistry friendliness of small molecules. Nature Publishing Group. 2017;(March):1-13. doi:https://doi.org/10.1038/srep42717

Driouche Y, Messadi D (2019) Quantitative structure-retention relationship model for predicting retention indices of constituents of essential oils of Thymus vulgaris (Lamiaceae). Journal of the Serbian Chemical Society 84(4):405–416. https://doi.org/10.2298/jsc180817010d

Gramatica P (2007) Principles of QSAR models validation: internal and external. QSAR Comb Sci 26(5):694–701. https://doi.org/10.1002/qsar.200610151

Ibrahim MT, Uzairu A, Shallangwa GA, Ibrahim A. In-silico studies of some oxadiazoles derivatives as anti-diabetic compounds Journal of King Saud University – Science In-silico studies of some oxadiazoles derivatives as anti-diabetic compounds. Journal of King Saud University - Science. 2018;(June). doi:10.1016/j.jksus.2018.06.006

Mabhula A, Singh V (2019) Drug-resistance in Mycobacterium tuberculosis : where we stand. MedChemComm. 10(8):1342–1360. https://doi.org/10.1039/c9md00057g

Ogbuabor DC, Onwujekwe OE (2019) Governance of tuberculosis control programme in Nigeria. Infectious Diseases of Poverty 8(1):1–11. https://doi.org/10.1186/s40249-019-0556-2

Pan Z, Wang Y, Gu X, Wang J, Cheng M (2019) Refined homology model of cytochrome Bcc complex B subunit for virtual screening of potential anti-tuberculosis agents. J Biomol Struct Dyn 1102. https://doi.org/10.1080/07391102.2019.1688196

Sanyal S, Amin SA, Adhikari N, Jha T (2019) QSAR modelling on a series of arylsulfonamide-based hydroxamates as potent MMP-2 inhibitors. SAR QSAR Environ Res 30(4):247–263. https://doi.org/10.1080/1062936X.2019.1588159

Tropsha A (2010) Best practices for QSAR model development, validation, and exploitation. Molecular Informatics 29(6-7):476–488. https://doi.org/10.1002/minf.201000061

Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK (2011) Validation of QSAR models - strategies and importance. International Journal of Drug Design and Disocovery 2(3):511–519

Wang A, Lv K, Li L et al (2019) Eur J Med Chem 178:715–725. https://doi.org/10.1016/j.ejmech.2019.06.038

World Health Organization (WHO). Global tuberculosis report-executive summary, Geneva, 2019

Zhai W, Wu F, Zhang Y, Fu Y, Liu Z (2019) The immune escape mechanisms of Mycobacterium tuberculosis. Int J Mol Sci 20:2. https://doi.org/10.3390/ijms20020340

## Publisher's Note

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