**Bulletin of Environment, Pharmacology and Life Sciences**
Bull. Env. Pharmacol. Life Sci., Vol 12 [4] March 2023 : 58-71

Online ISSN 2277-1808
Journal's URL:http://www.bepls.com
CODEN: BEPLAD

**ORIGINAL ARTICLE** **OPEN ACCESS**

# Molecular leads generation for 4-aminoquinolines having Mtb Gyrase B inhibitor activity: QSAR and Molecular dynamics simulation studies

**Gopinath Papichettypalle[1], P. Aravanan[2], Ramu Samineni[3], Bidhu Bhusan Karkara[4], Swathi Thumula[5], Guru Prasad Muppala[6], S. Subramanyam[7]**

[1,4]Department of Pharmaceutical Chemistry, GITAM School of Pharmacy, GITAM University, Hyderabad, Telangana, India- 502329

[2]Professor and HOD, Department of Pharmaceutical Chemistry, Sri Venkateswara College of Pharmacy, Chittoor - 517127

[3]Department of Pharmaceutical Sciences, Vignan's Foundation for Science, Technology and Research, (Deemed to be University), Vadlamudi, Guntur, A.P. India-522213

[4]Department of Chemistry, Sreenidhi Institute of Science and Technology, Yamnampet, Ghatkesar, Hyderabad-501301

[6]Assistant General Manager (R&D), Vaishnavi Microbial Pvt. ltd, Hyderabad, Telangana - 500033

[7]Department of Pharmaceutical Sciences, Vignan's Foundation for Science, Technology and Research, (Deemed to be University), Vadlamudi, Guntur, A.P. India-522213

***Corresponding author's Email:** gpappich@gitam.edu, gopi.pharma@gmail.com

**ABSTRACT**

*Mycobacterial DNA gyrase B subunit is one of the potential drug targets in the field of antitubercular drug discovery. Fluoroquinolones targeting gyrase A subunit have been facing a major hurdle of their resistance developed by Mtb which makes gyrase B subunit a druggable target for discovery of potent anti-tubercular agents. DNA gyrase B subunit is involved in the process of ATP hydrolysis which in turn provides energy to gyrase A subunit for maintaining the DNA topological state. In the present study, we employed structural optimization of the reported GyrB inhibitor possessing quinoline nucleus. QSAR studies were carried out by QSARINS software on 4-aminoquinoline derivatives for Mtb gyrase B inhibitory activity. The best model had four variables L1i, MoRSEN26, RDFM5 and RDFE25 with statistical values* $R^2$ *= 0.7430, LOF=0.0608,* $CCC_{tr}$ *= 0.8525,* $Q^2_{LOO}$ *= 0.6461,* $Q2_{LMO}$ *= 0.6189,* $CCC_{cv}$ *= 0.7972,* $R^2_{ext}$ *= 0.8294, and* $CCC_{ext}$ *= 0.8898. The developed qsar model suggests that the 3D WHIM, MoRSE and RDF descriptors play key roles and were extremely helpful in predicting bioactivity. Molecular docking studies were performed using Autodock v 4.2.6 and the residues of active site region involving both hydrophilic and hydrophobic parts interacted with best experimentally active compounds and designed compounds. Based on predicted docking interactions in the active site of enzyme, molecular dynamics simulations were carried out for 30ns using GROMACS 2020 to predict complex stability. RMSD, RMSF and Rg analysis were performed on 3zkd protein in complex with compound 16 and best designed compound 8e. The MDS results were found to be satisfactory with compound 8e exhibiting better conformational stability in the active site pocket of 3zkd protein.*

***Keywords****: Mycobacterium tuberculosis, DNA gyrase B, QSARINS, Autodock, GROMACS*

Received 14.01.2023 Revised 10.02.2023 Accepted 12.03.2023

## INTRODUCTION

It is estimated that tuberculosis will rank as the subsequent leading cause of death from a single infectious agent, next to COVID-19. Worldwide, tuberculosis (TB) caused by Mycobacterium tuberculosis, continues to be one of the deadliest infectious diseases with mortality rate of around 1.3 million lives in 2020 as reported by World Health Organization (WHO) [1]. Although TB can be treated with the commonly used drugs such as isoniazid, rifampicin, ethambutol, and pyrazinamide, unfortunately, it has been reported that TB treatment is hampered by resistance of M. tuberculosis to available therapies. The noncompliance to treatment regimens has led to the development of drug resistance in a self-sustaining process that has resulted in some forms of TB being untreatable by currently available anti-tuberculosis drugs. The rapid emergence of multi-drug-resistant (MDR) and extremely drug-resistant (XDR) Mycobacterium tuberculosis (*M. tuberculosis*) strains have now been recognized as a major challenge for

global TB control measures [2]. Consequently, there is a need to improve the existing standard treatment for MDR-TB to improve patient survival rates.

The enzymes that manage DNA reorganization (DNA topoisomerases), act by creating transient local DNA breaks in one or both strands of the DNA double helix (type I and type II topoisomerases respectively) and passing second DNA segments through the breaks. Whereas type I DNA topoisomerases are ATP-independent monomeric enzymes, type II topoisomerases are large macromolecular machines (homodimers or heterotetramers) that use ATP to drive conformational changes associated with the double-stranded DNA transfer events. Bacterial type IIA topoisomerases are made up of two subunits, GyrA or ParC and GyrB or ParE, for DNA gyrase and Topo IV respectively, which form the catalytically active heterotetrameric enzymes [3].

DNA gyrase and Topo IV subunits share similar functional and structural arrangements and are thought to be paralogues that resulted from gene duplication of an ancestral type IIA topoisomerase. Subunits GyrB and ParE consist of two domains, the N-terminal ATPase and the C-terminal Toprim domain, whereas GyrA and ParC consist of the N-terminal BRD (breakage–reunion domain) followed by the CTD (C-terminal domain). Dynamic interactions between these two subunits couple ATP binding. Hydrolysis with DNA binding, cleavage and strand transport physically move one DNA duplex through another [4]. Although Topo IV and DNA gyrase are homologous, their substrate selection and activity are strikingly different. DNA gyrase introduces a right-handed wrap in the DNA which, after strand passage, results in the introduction of negative supercoils thereby assisting in the underwinding of bacterial genomes. In contrast, Topo IV preferentially relaxes DNA supercoils, being more active on positively rather than on negatively supercoiled DNA, and is a robust DNA decatenase [5].

In addition to playing a crucial role in cellular events such as replication, recombination and transcription, bacterial DNA gyrase and Topo IV are the targets of a major family of antibiotics, the fluoroquinolones. Fluoroquinolones targeting gyrase A subunit have been facing a major hurdle of their resistance developed by Mtb which makes gyrase B subunit a druggable target for the discovery of potent anti-tubercular agents [6]. DNA gyrase B subunit is involved in the process of ATP hydrolysis which in turn provides energy to gyrase A subunit for maintaining the DNA topological state [7]. These antibiotics form the cornerstone of several treatments for infections with major pathogens, such as multidrug-resistant tuberculosis.

For this reason, inhibitors targeting different steps of this pathway are capturing the attention of TB drug research. TMC207 (diarylquinoline), an antitubercular drug in Phase II trials acts by inhibiting the enzyme mycobacterial adenosine triphosphate (ATP) synthase. Many of quinoline and quinolone series were prescribed in combination with first-line and second-line antimycobacterial drugs [8]. One of the rational and successful methods in drug design is quantitative structure-activity relationship, playing a key role in optimizing leads and thereby improving their biological activity [9]. We explore the possibilities of finding new quinoline-based antimycobacterial leads. In this context, we have reported thoroughly validated *in-silico* studies on triazole benzene sulphonamide as carbonic anhydrase IX inhibitors [10]. It is important to note, developed and validated Qsar model also helps in the prediction of the biological activity of new chemical entities even before its synthesis.

In continuation to the work in identification and development of novel leads using QSAR, molecular docking as well as dynamic simulation tools, we herein report a series of novel 4-aminoquinoline derivatives as inhibitors of Mtb DNA gyrase B. The dataset was subjected to QSAR study for predicting the role of substituent on biological activity by generating efficient models, which are capable of estimating bioactivities with extensive parameters available within the software. We were also interested in investigating the structural features in optimizing the lead for the inhibition of Mtb DNA gyrase B. Molecular dynamics simulations were also performed at a time scale of 30 ns for selected compounds to identify stability of the protein-ligand complex, fluctuations and H-bond network.

## MATERIAL AND METHODS

'QSARINS' allows developing multiple linear regression models by ordinary least squares, attentively verified and in detail validated according to the chemometric approach. A series of 46 compounds dataset (Table 1) of 4-aminoquinoline derivatives with Mtb gyrase B inhibitory values were taken from reported literature [11]. MIC values were converted to corresponding pMIC and used as the dependent variable.

### *Molecule structure preparation and 3D geometry optimization*

Mentioned molecular structures were constructed and geometry optimized by Avogadro V1.2.0 [12] on adding hydrogens. MMFF94, the molecular mechanics force field was employed along with the steepest descent algorithm. For each compound, best conformer with global minimum energy is obtained from Avogadro tool by genetic algorithm with the scoring function 'energy' and the same conformer was used throughout the study.

***Data setup***

Fore mentioned compounds were calculated for molecular descriptor values from online padel tool of chemdes - chemopy servers [13]. The variables were organized and pre-filtered by excluding all-zero value, missing value and constant value (>50%) descriptors. The pairwise correlation is used to filter out descriptors with more than 0.85 values. From the correlation matrix developed using all filtered descriptors, 3D WHIM, MoRSE and RDF descriptors were shown better correlation towards activity and thus, they were preferred for this study. From the screening of descriptors, a total of 32 variables with cut-off correlation value of greater than 0.36, were selected for the study. Forty-six molecules included in the study were divided into training and prediction set in the ratio of 5:1 based on order by the response. Out of many trials and models developed, we present here only the best models.

***Variable selection and model calculation***

The QSARINS software explores all combinations of selected descriptors as defined by user options [14]. The genetic algorithm is used for descriptor selection relational to biological activities of molecules along with Friedman's 'lack-of-fit' (LOF) function, to evaluate the fitness of models. LOF smoothness level is kept at default level of 1.0. Along with genetic algorithm, user-defined parameters like mutation probability (0.1), population size (200) and maximum generations (5000) explore more combinations [15].

***Model validation***

The models developed in QSARINS undergo rigorous validation both internal and external and also applicability domain of the model is checked. Internal validation by cross-validation leave-one-out ($Q2_{LOO}$), cross-validation leave-many-out ($Q2_{LMO}$), root mean squared error (RMSE), Y-Scrambling and external validation by $Q^2_{F1}$, $Q^2_{F2}$, $Q^2_{F3}$ and CCC were applied on selected models. $Q2_{LMO}$ is repeated 5000 times with 30% objects left randomly from the set of training each time. Y-scrambling by 5000 iterations method involves shuffling of response data, to exclude chance correlation in the original model. It is worthy to note, $R^2$ and $Q^2$loo of the model must be reasonably higher than scrambled ones and RMSE of the model under prediction must be reasonably smaller than scrambled ones. To evaluate reproducibility of models, concordance correlation coefficient ($CCC_{ext}$) is analyzed. Applicability domain is a theoretical region in modeling defined by descriptors employed and is evaluated by leverage analysis. The leverage (hat) is calculated by $h_i = x_i (X^T X)^{-1} x^T_i$ (i=1, 2, 3, upto m), where $x_i$ is the descriptor row-value of the query compound i and m is the number of query compounds. X is n x p matrix of training set, where n is number of training set samples and p is number of model descriptors. The limit of model domain, leverage cut-off value, h* is 3(p+1)/n. Leverage greater than h* for the training set means the sample is highly influential in determining the model, while the test set (X outlier), the prediction is an extrapolation of the model. Any compound with standardized residual more than 3σ (3 standard deviation units) is identified as Y outlier.

***Molecular docking studies***

In combination with the developed qsar model, affinities and interactions of Mtb gyrase B inhibitors were estimated using Autodock V4.2.6 [16]. The initial structure of the 3D structure of Mtb GyrB ATPase domain, pdb id: 3ZKD is obtained from protein data bank (w3.rcsb.org). Using Modeller V9.23 program [17] missing residues were fixed along with addition of hydrogen atoms and removal of existing ligands. Redocking co-crystallized ligand to 3ZKD was performed to identify docking parameters which will be helpful for docking designed compounds. All the ligand structures were constructed and geometry optimized using Avogadro V1.2.0 by the MMFF94 force field, steepest descent algorithm with convergence parameter set to$10e^{-7}$.

Protein was prepared by adding polar hydrogens geometrically and Kollman's united atom charges were assigned to create pdbqt file. Protonation states of histidine residues were addressed by assigning ND1 state to zinc bound histidines and NE2 to remaining histidines. Ligand preparation by adding gasteiger charges is done along with the addition of polar hydrogens. Torsions were identified in ligands and pdbqt file is generated.

Autogrid option renders a selection of active site and grid size was set to 60*60*60 points with spacing of 0.375 Å and a distance-dependent function of dielectric constant was used for the calculation of energetic map. The grid box includes the active binding site of the enzyme with enough space for the ligand rotational and translational walk.

Lamarckian genetic algorithm was used for exploring ligand conformation poses, orientations inside the active site of Mtb gyrase B. The optimized parameters were as follows; maximum number of energy evaluations was increased to 25,000,000 per run, the number of individuals in the population was 150, maximum number of generations was 2700 along with rate of gene mutations to 0.02. All other parameters were set to default. Results differing by <2 Å in a positional RMSD were clustered together. In each group, the lowest binding energy configuration with the highest percentage frequency was selected

as the group representative. Representations of ligand poses and interactions were generated and presented in the figure by Biovia Discovery studio visualizer version 2021 [18] program.

***Molecular Dynamics Simulations***

Based on Autodock binding energy of protein-ligand complexes, the best ligands from original dataset as well as designer series were identified and taken over for molecular dynamics simulations to study the physical changes in the atoms and molecules upon interacting with solvent environment using GROMACS 2022 package [19]. The topology parameters for 3ZKD (DNA Gyrase B) protein were generated by GROMACS using Charmm36-Jul-2020.ff force field [20], where as for ligands, external tools were applied to create stream file from CGenFF web server followed by python module to generate itp files. Protein-ligand complex file is generated using obtained topology and coordinates information following solvation by TIP3P water molecules in a defined unit cell. Counter ions were added to balance the solvated system. The complex was placed in a cubic periodic box and the minimum distance between the complex and the box walls was set to larger than 10 Å. The complex was comprised of 5740 atoms in the 3ZKD protein.

Prior to simulation, energy minimization was performed to full system without constraints using steepest descent integrator for 50,000 steps. Finally, the system upon converging resulted required negative potential energy mark and Fmax not greater than 1000 KJ / mol / $nm^2$. The system was then equilibrated for 200 ps of NVT & NPT ensemble each, applying the position restraints on protein, inhibitors and counter ions at 300 K with periodic boundary conditions.

The temperature was kept constant by a Berendsen thermostat, while the pressure was maintained at 1 bar using a Parrinello-Rahman scheme [21]. The electrostatic interactions were calculated using the Particle Mesh Ewald method & cut-off distances for the calculation of coulomb and van der Waals interactions were 1.2 nm during the equilibration. Finally, the system was subjected to 30 ns molecular dynamics simulations at a temperature of 300 K (V-rescale thermostat) and a pressure of 1 bar (Parrinello-Rahman barostat). A periodic boundary condition was imposed on the system and the motion equations were integrated by applying the leap-frog algorithm with a time-step of 2 fs. The dynamic trajectories were recorded every pico-second during the production stage for further analysis.

## RESULTS AND DISCUSSION

***Model information***

A dataset of 46 molecules from quinolone series with inhibitory activity were considered for the study. The molecules were optimized for geometry and by using MMFF94. ~ 3000 descriptors were calculated using Padel, Chemopy-chemdesc, RDKit and Bluedesc server. The dataset was divided into training and test based on their chemical and biological diversity. Several qsar model equations were generated using QSARINS. Some of the models showed higher $R^2$ and $Q^2_{LOO}$ values but their external validation was not good and they also showed a good number of outliers. For practical use of QSAR model, cross validated performance (robustness) and external predictive capacity (predictability) were good parameters. The model 1 developed with fore mentioned options gave statistical parameters as,

Model 1:

pMIC = 3.2790 + 0.0338 (L1i) + 3.0485 (MoRSEN26) + 0.0477 (RDFM5) + 0.0349 (RDFE25)

$n_{tr}$ = 39, $n_{pred}$ = 07, $R^2$ = 0.4936, $R^2_{adj}$ = 0.4340, $R^2$-$R^2_{adj}$ = 0.0596, LOF = 0.1312, $RMSE_{tr}$ = 0.2879, $MAE_{tr}$ = 0.2199, $RSS_{tr}$ = 3.2319, $CCC_{tr}$ = 0.6609, s=0.3083, F=8.2844, $Q^2_{LOO}$ = 0.3423, $Q2_{LMO}$ = 0.3022, $R^2_{Yscr}$ = 0.1071, $Q^2_{Yscr}$ = -0.1839, $RMSE_{cv}$ = 0.3281, $MAE_{cv}$ = 0.2528, $PRESS_{cv}$ = 4.1977, $CCC_{cv}$ = 0.5635, $R^2_{ext}$ = 0.3987, $MAE_{ext}$ = 0.1909, $PRESS_{ext}$ = 0.4735, $RMSE_{ext}$ = 0.2601, $CCC_{ext}$ = 0.6169, $Q^2_{F1}$ = 0.3673, $Q^2_{F2}$ = 0.3618, $Q^3_{F3}$ = 0.5866.

This model showed up two outliers (compound 17 and 37) in William's plot, with high RMSE values and low $Q^2_{LOO}$, $Q^2_{LMO}$ values. On removal of two structures models were generated and validated.

Model 2:

pMIC = 2.9281 + 0.0434 (L1i) + 3.1282 (MoRSEN26) + 0.0514 (RDFM5) + 0.0344 (RDFE25)

$n_{tr}$ = 37, $n_{pred}$ = 07, $R^2$ = 0.6404, $R^2_{adj}$ = 0.5954, $R^2$-$R^2_{adj}$ = 0.0450, LOF = 0.0891, $RMSE_{tr}$ = 0.2340, $MAE_{tr}$ = 0.1798, $RSS_{tr}$ = 2.0252, $CCC_{tr}$ = 0.7808, s=0.2516, F=14.2442, $Q^2_{LOO}$ = 0.5271, $Q2_{LMO}$ = 0.4851, $R^2_{Yscr}$ = 0.1122, $Q^2_{Yscr}$ = -0.1976, $RMSE_{cv}$ = 0.2683, $MAE_{cv}$ = 0.2082, $PRESS_{cv}$ = 2.6628, $CCC_{cv}$ = 0.7139, $R^2_{ext}$ = 0.4105, $MAE_{ext}$ = 0.2244, $PRESS_{ext}$ = 0.6026, $RMSE_{ext}$ = 0.2934, $CCC_{ext}$ = 0.5451, $Q^2_{F1}$ = 0.2780, $Q^2_{F2}$ = 0.2494, $Q^3_{F3}$ = 0.4344.

Even though after removing outliers from model 1, model 2 showed up compound 10 and 36 as outliers in William's plot, which provoked to study the effect of removing these outliers. Additionally, the external validation parameters $Q_2F_1$, $Q_2F_2$, $Q_2F_3$, and $CCC_{ext}$ gave lessened statistical values compared to model 1.

Model 3:

pMIC = 2.8467 + 0.0431 (L1i) + 3.1683 (MoRSEN26) + 0.0539 (RDFM5) + 0.0412 (RDFE25)

$n_{tr}$ = 35, $n_{pred}$ = 07, $R^2$ = 0.6790, $R^2_{adj}$ = 0.6376, $R^2$-$R^2_{adj}$ = 0.0414, LOF = 0.0797, $RMSE_{tr}$ = 0.2196, $MAE_{tr}$ = 0.1741, $RSS_{tr}$ = 1.7364, $CCC_{tr}$ = 0.8088, s=0.2367, F=16.3933, $Q^2_{LOO}$ = 0.5777, $Q2_{LMO}$ = 0.5489, $R^2_{Yscr}$ = 0.1159, $Q^2_{Yscr}$ = -0.2047, $RMSE_{cv}$ = 0.2519, $MAE_{cv}$ = 0.2023, $PRESS_{cv}$ = 2.2846, $CCC_{cv}$ = 0.7486, $R^2_{ext}$ = 0.7890, $MAE_{ext}$ = 0.1472, $PRESS_{ext}$ = 0.1639, $RMSE_{ext}$ = 0.1653, $CCC_{ext}$ = 0.8789, $Q^2_{F1}$ = 0.7321, $Q^2_{F2}$ = 0.7314, $Q^3_{F3}$ = 0.8789.

Two outlier compounds 12, 40 in williams plot were found and statistical values were improved when compared to previous models and hence further studies were carried by removing outliers.

Model 4:

pMIC = 2.8063 + 0.0475 (L1i) + 3.3770 (MoRSEN26) + 0.0456 (RDFM5) + 0.0394 (RDFE25)

$n_{tr}$ = 34, $n_{pred}$ = 06, $R^2$ = 0.7430, $R^2_{adj}$ = 0.7075, $R^2$-$R^2_{adj}$ = 0.0355, LOF = 0.0608, $RMSE_{tr}$ = 0.1886, $MAE_{tr}$ = 0.1559, $RSS_{tr}$ = 1.2096, $CCC_{tr}$ = 0.8525, s=0.2042, F=20.9550, $Q^2_{LOO}$ = 0.6461, $Q2_{LMO}$ = 0.6189, $R^2_{Yscr}$ = 0.1226, $Q^2_{Yscr}$ = -0.2213, $RMSE_{cv}$ = 0.2213, $MAE_{cv}$ = 0.1842, $PRESS_{cv}$ = 1.6655, $CCC_{cv}$ = 0.7972, $R^2_{ext}$ = 0.8294, $MAE_{ext}$ = 0.1417, $PRESS_{ext}$ = 0.1590, $RMSE_{ext}$ = 0.1628, $CCC_{ext}$ = 0.8898, $Q^2_{F1}$ = 0.7550, $Q^2_{F2}$ = 0.7488, $Q^3_{F3}$ = 0.8086.

Model 4 obtained shows good fitting criteria, both internal and external validation values. Compared with previous models, it shows betterment in external validation parameter values with no outliers in William's plot. The descriptor correlation matrix of model 4 is presented in Supplementary Table 1. The scatter plot of the experimental *vs.* the calculated Mtb gyrase B inhibitory activities of 4-aminoquinolines is shown in figure 1(left), displaying predicted values similar to corresponding experimental values. The resulted $K_{xy}$ (the inter-correlation among descriptors and response) *vs.* $Q^2_{LMO}$ of final model were plotted in figure 1(right) which shows the LMO parameter values were around the model parameters, meaning the model is robust and stable. Figure 2(left) displays the Y-scramble plot of $K_{xy}$ vs. $R^2_{Yscr}$ and $Q^2_{Yscr}$, which shows correlation coefficients of the final model are much higher than those after endpoint scrambling and a broken relationship can be noticed between structure and responses.

The Williams plot, standardized residuals *vs.* leverage values shown in Figure 2(right), was used to illustrate the prediction and express the applicability domain of the model. From William's plot, it can be seen that all the molecules are located in the applicability domain of the model with leverage values lower than the warning $h^*$ of 0.441. The values of Q2F1, Q2F2 and Q2F3 are greater than threshold of 0.70 and of better value along with with elevated CCC (concordance correlation coefficient) parameter values greater than 0.80. All these results state the best model obtained is not by chance and truly there is a relationship between structures of 4-aminoquinolines analogs with corresponding Mtb gyrase B inhibitory activity.

***Molecular descriptors information***

WHIM (*W*eighted *H*olistic *I*nvariant *M*olecular) descriptors [22] are 3D molecular indices that represent different sources of chemical information. WHIM descriptors like L1i, 1st component size directional WHIM index / weighted by relative first ionization potential, contain information about the whole 3D molecular structure in terms of size, shape, symmetry and atom distribution [23]. These indices are calculated from x,y,z-coordinates of a 3D structure of the molecule, usually from a spatial conformation of minimum energy, within different weighting schemes in a straightforward manner and represent a very general approach to describe molecules in a unitary conceptual framework. The WHIM descriptor approach has also been extended to treat interaction scalar fields: G-WHIM (*G*rid-*W*eighted *H*olistic *I*nvariant *M*olecular) descriptors are defined and calculated from the coordinates of the grid-points where an interaction energy field between the molecule and a probe has been evaluated [24].

WHIM descriptors are built in such a way that they capture relevant molecular 3D information regarding molecular size, shape, symmetry and atom distribution with respect to invariant reference frames. The algorithm consists of performing a principal component analysis on the centered molecular coordinates by using a weighted covariance matrix **S** obtained from different weighting schemes for the atoms. The elements of the covariance matrix are:

$$S_{jk} = \frac{\sum_{i=1}^{n} w_i (q_{ij} - \bar{q_j})(q_{ik} - \bar{q_k})}{\sum_{i=1}^{n} w_i}$$

Where $n$ is the number of atoms, $w_i$ the weight of the $i$-th atom, $q_{ij}$ represents the $j$-th coordinate ($j$ = 1, 2, 3) of the $i$-th atom and $q_j^-$ is the average of the $j$-th coordinates.

Six different weighting schemes have been proposed: (1) the unweighted case **U** ($w_i$ = 1, $i$ = 1, $n$, where $n$ is the number of atoms for each compound), (2) atomic masses **M** ($w_i = m_i$), (3) the van der Waals volumes **V** ($w_i = vdw_i$), (4) the Mulliken atomic electronegativities **E** ($w_i = eln_i$), (5) the atomic polarizabilities **P** ($w_i = pol_i$) and (6) the electrotopological indices of Kier and Hall S ($w_i = S_i$). All the weights (1) to (5) are also scaled with respect to the carbon atom.

3D-MoRSE [25] meaning 3D molecular representations of structure based on electron diffraction descriptors were predominant in a number of QSAR/QSPR studies. These descriptors provide molecule structure information derived from euclidean interatomic distances, scattering parameter (0-31 integer values) and weighting by atomic properties like atomic charges (MoRSEC9, where 9 is scattering parameter), atomic mass (MoRSEM15), atomic number (MoRSEN26), atomic van der Waals volume (MoRSEV23), Sanderson electronegativity (MoRSEE12) and atomic polarizability (MoRSEP4) as well as unweighted (MoRSEU13).

Simplified equation from electron diffraction studies to determine MoRSE function:

$$I(s) = \sum_{i=2}^{N} \sum_{j=1}^{i-1} A_i \times A_j \frac{\sin sr_{ij}}{sr_{ij}}$$

where *s* is the scattering parameter, $r_{ij}$ is the euclidean distance between *i* th and *j* th atoms, *N* is the total number of atoms and $A_i$ and $A_j$ are different atomic properties used as weights. Each term of this function depends on distance and thus may be viewed as a radial basis function itself.

Weighting descriptors make the compound sensitive to the presence of specific molecule fragments. For instance, weighting atomic partial charge reflects distance between atoms with excessive or deficient electron density. The 3D-MoRSE descriptors weighted with schemes where the role of hydrogen is diminished should exhibit lower variation with increasing scattering parameter. The lowest relative variation is observed for atomic mass, atomic number, van der Waals volume and polarizability weightings as reported in literature. The impact of different interatomic distances can be identified by dynamics of cumulative sum of 3D-MoRSE terms ordered by interatomic distance. If increasing values of some 3D-MoRSE descriptors lead to increase in biological activity, then atomic groups that contribute to the 3D-MoRSE descriptor mostly are preferential for activity. 3D-MoRSE descriptors interpretation can be done by estimating favourable optimal range MoRSE values needed for best activity and finding favourable interatomic distance inactive compounds that are compared to non-actives.

The 3D coordinates of the atoms of molecules can be transformed into a structure code that has a fixed number of descriptors irrespective of the size of a molecule. This task is performed by a structure coding technique referred to as a radial distribution function code (RDF code) [26]. In general, there are some prerequisites for a structure code:

- independent from the number of atoms, i.e., the size of a molecule,
- unambiguity regarding the three-dimensional arrangement of the atoms, and
- Invariance against translation and rotation of the entire molecule.

Formally, the radial distribution function of an ensemble of N atoms can be interpreted as the probability distribution to find an atom in a spherical volume of radius r. The equation represents the radial distribution function code as it is used in this investigation:

$$g(r) = f \times \sum_{i}^{N-1} \sum_{j}^{N} A_i \times A_j \times e^{-B(r-r_{ij})^2}$$

Where f is a scaling factor and N is the number of atoms. By including characteristic atomic properties, a of the atoms i and j, the RDF codes can be used in different tasks to fit the requirements of the information to be represented. The exponential term contains the distance $r_{ij}$ between the atoms i and j and the smoothing parameter B, which defines the probability distribution of the individual distances. g(r) was calculated at a number of discrete points with defined intervals.

The atomic properties $A_i$ and $A_j$ used in this equation enable the discrimination of the atoms of a molecule for almost any property that can be attributed to an atom [27]. Such distribution function provides, besides information about interatomic distances in the whole molecule, the opportunity to gain access to other valuable information, e.g., bond distance, ring types, planar and non-planar systems and atom types. This fact is the most valuable consideration for a computer-assisted code elucidation. The radial distribution function in this form meets the entire requirement mentioned above, especially invariance against linear translations. In total, the best model molecular descriptors contributed to structural information relating predicted bioactivity [28].

***SAR Studies of the original dataset***

Utilizing molecular descriptors information obtained from QSAR studies, a self-explorative study on the dataset was performed to identify the model efficiency and accuracy. Residuals from the difference calculated between experimental and predicted bioactivity values were almost near to zero, meaning

validated model generation and effective prediction (Table 2). Further using structure function relationship values from dataset, meaningful information was depicted.

Substitutions on R1 position of quinoline ring with ethoxy, R position with trifluoromethyl, X with amino and Y with ethylamino groups lead to best-predicted activity as depicted in compound 16. Upon changing amino group of X position to oxygen group, loss of activity is seen except in case of methoxy substituent at R position of compound 10. Replacing alkyl amino group at Y position to oxygen group leads to diminished activity as seen in compound 4 and 8.

Substitution on R1 position of quinoline ring with hydrazine group, R position with trifluoromethyl, X with amino and Y with ethylamino groups lead to diminished activity as depicted in compound 32. Replacing amino group of X position to oxygen group and alkyl-amino group at Y position to oxygen group leads to loss of activity as seen in compound 20. Substitution on R1 position of quinoline ring with hydroxyl group, R position with methoxy, X with oxygen and Y with ethylamino groups lead to increased activity as depicted in compound 42 compared to compound 46.

New compounds were designed with varying substituents on the core structure and molecular descriptor values were calculated from online padel tool of chemdes - chemopy servers. Table 3 display predicted activities of best designed compounds from the QSAR model 4 (Supplementary table 2 display full list of designed compounds). The best activity compounds from the original dataset (11, 12, 16, 17 and 31) and the designer series (8e, 9c, 9d, 10c, 11a, 11b, 12a, and 12c) were selected for molecular docking studies to predict favourable H-bonds interactions and binding scores.

***Molecular docking studies***

H-bond interactions were seen in compound 11 with the residues Asn309, Trp47, Thr371, His311 and Glu312 of 3zkd protein. Compound 11 with ethoxy group at R1 position interacted with Asn309 and Thr371 by hydrogen bonding followed by oxygen of ligand ethoxy group with Trp47. Alkyl group near X position interacted with Glu312 using H-bond. The best active site conformation from Autodock result the binding affinity of compound 11 with 3ZKD protein as -8.4 K Cal / mol. Similarly, H-bond interactions were predicted in compound 12 with the residues Asn309, His44, Arg40, Thr371, His311, Glu312 and Gln370 of 3zkd protein. Compound 12 display H-bonds near to R1 position with Arg40 and Thr371 followed by oxygen at X position interaction with His44. Asn309 interacted with piperazine N atom forming a favorable H-bond. Trifluoro group of compound 12 interacted well with Gly369. The best active site conformation from Autodock result the binding affinity of compound 12 with 3ZKD protein as -7.6 K Cal / mol.

Again, H-bond interactions were seen in compound 16 with the residues Asn309 and Trp47 of 3zkd protein. Compound 16 with ethoxy group at R1 position interacted with Asn309 and Thr371 by hydrogen bonding followed by H-bond of alkyl group near X position interacted with Glu312. Trifluoro group of compound 16 interacted well with Glu39 and Arg40. The best active site conformation from Autodock result the binding affinity of compound 16 with 3ZKD protein as -8.7 K Cal / mol. Similarly, H-bond interactions were predicted in compound 17 with the residues Asn309, Arg40, Thr371, His311 and Trp47 of 3zkd protein. Compound 17 display H-bonds near to R1 position with Trp47, Asn309 and Thr371. The best active site conformation from Autodock result the binding affinity of compound 17 with 3ZKD protein as -8.1 K Cal / mol. Further, H-bond interactions were predicted in compound 31 with the residues Asn309, Trp47, Glu312 and His44 of 3zkd protein. Compound 31 display Pi interactions with Arg40 and His311. The best active site conformation from Autodock result the binding affinity of compound 31 with 3ZKD protein as -8.3 K Cal / mol.

With various substituents at R, R1, X positions and keeping Y position fixed to N-ethyl group, designer series resulted few best leads based on PMIC values using QSAR model 4. These leads were evaluated for H-bond interactions and binding scores using molecular docking studies. Residue interactions with His44, Trp47, Asn309 and His311 were common in almost all the designed compounds giving preference to H-bonding. The results were presented in Table 4 with interacting residues, binding affinity scores from Autodock as well as PMIC values from QSAR model 4. The results were favorable giving importance to model equation descriptors supporting QSAR studies. In general, the molecular descriptors (weighted or unweighted) of obtained qsar model and type of molecular docking interactions (charged, hydrogen bond, dipole-dipole, van der Waals, pi-cation, etc.), were correlated and found to be in association with each other. RDF, WHIM and MoRSE descriptors weighted with Ionization potential, electonegativity indices were in relation to fluorine group charged interactions with arginine, histidine residues. Figure 3 shows the docking interactions of best compounds with 3ZKD from Discovery studio visualiser (Remaining docking results were displayed in Supplementary figures 1 to 5).

To confirm the interactions and stability in the active site of Mtb DNA Gyrase B, compound 16 of dataset and designed compound 8e were taken over for molecular dynamics simulations using GROMACS package.

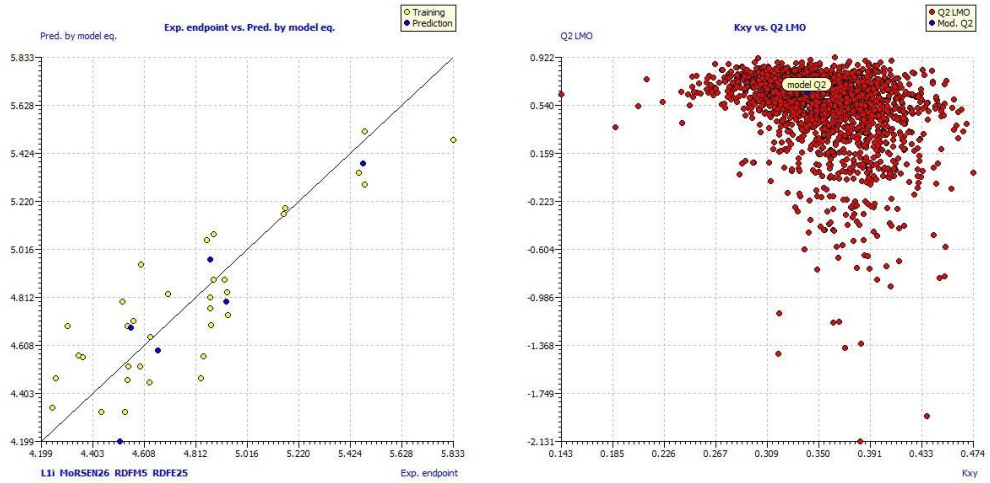



**Figure 1. Scattered plot left and LMO Plot Right**

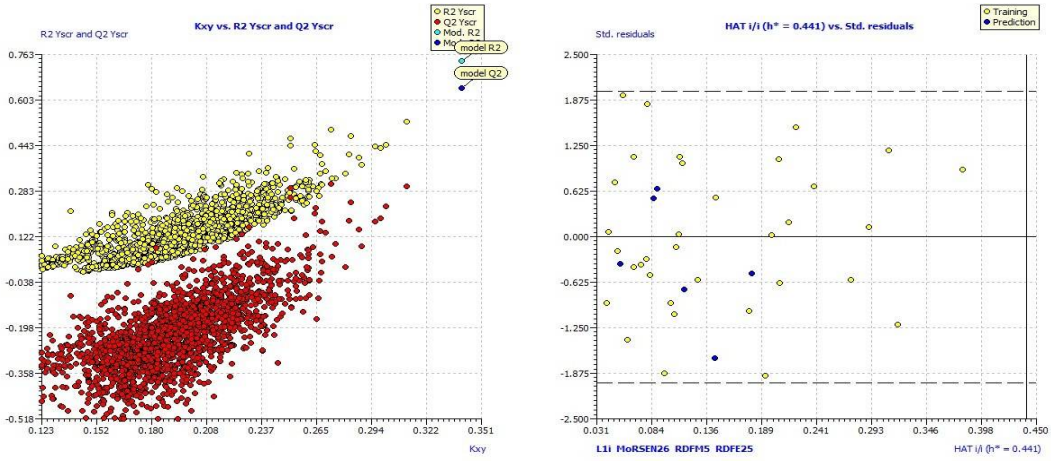



**Figure 2. Y Scrambles plot left and Williams plot right**

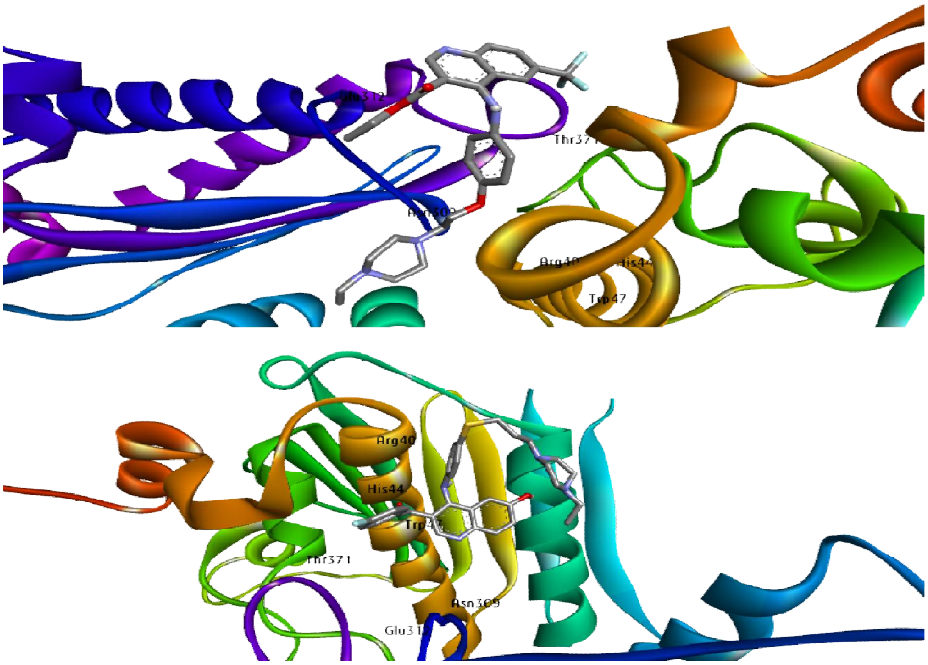



**Figure 3. Molecular docking results a). Compound 6 and b). Compound 8e**

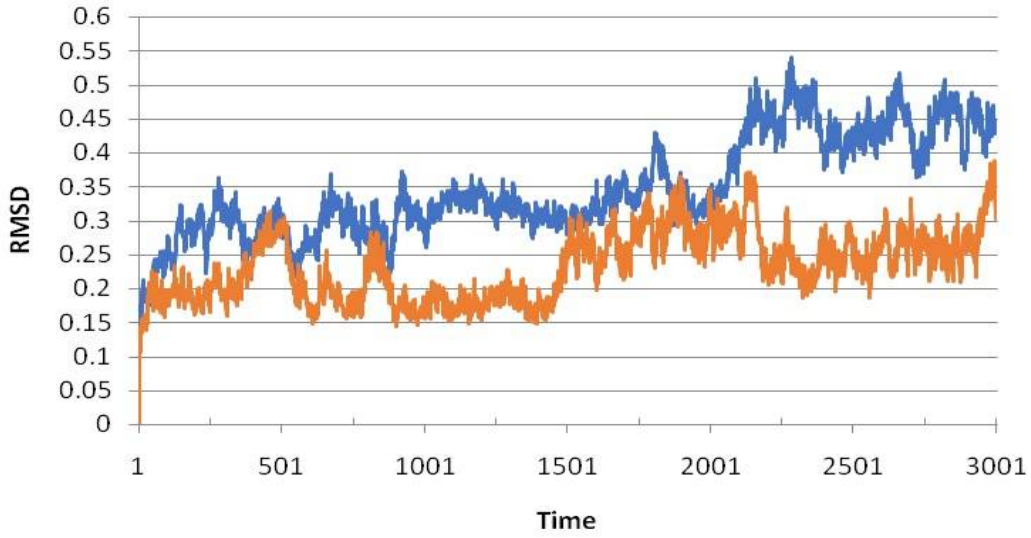



**Figure 4. RMSD**

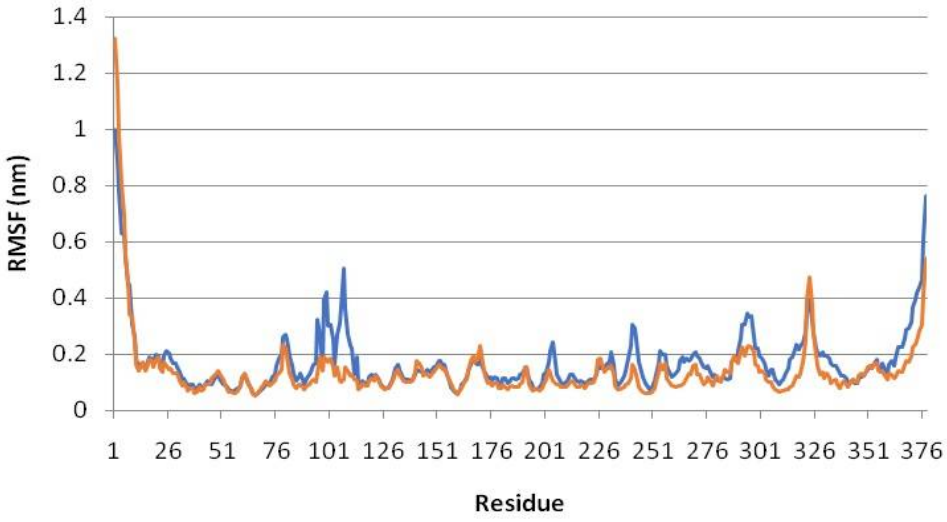



**Figure 5. RMSF**

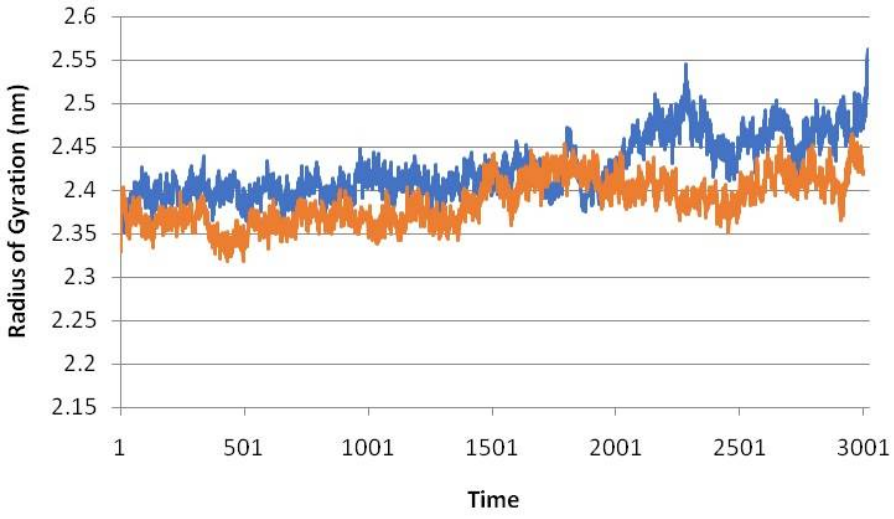
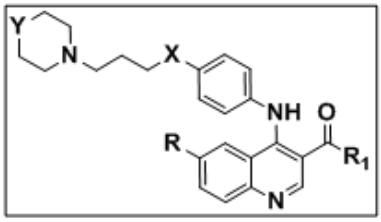


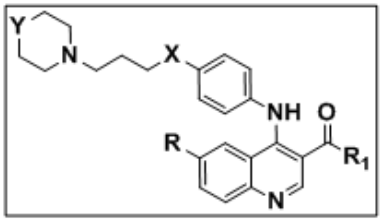

**Figure 6. Radius of Gyration (nm)**

**Table 1. Dataset compounds with Mtb Gyrase B inhibitory activity**




| S.No | R | R1 | X | Y | MIC (μM) | pMIC |
|---|---|---|---|---|---|---|
| 1 | H | $OC_2H_5$ | O | O | 28.7 | 4.542118 |
| 2 | $OCH_3$ | $OC_2H_5$ | O | O | 13.35 | 4.874519 |
| 3 | F | $OC_2H_5$ | O | O | 21.72 | 4.66314 |
| 4 | $CF_3$ | $OC_2H_5$ | O | O | 43.1 | 4.365523 |
| 5 | H | $OC_2H_5$ | NH | O | 14.38 | 4.842241 |
| 6 | $OCH_3$ | $OC_2H_5$ | NH | O | 23.36 | 4.631527 |
| 7 | F | $OC_2H_5$ | NH | O | 13.45 | 4.871278 |
| 8 | $CF_3$ | $OC_2H_5$ | NH | O | 49.75 | 4.303207 |
| 9 | H | $OC_2H_5$ | O | $NC_2H_5$ | 13.51 | 4.869345 |
| 10 | $OCH_3$ | $OC_2H_5$ | O | $NC_2H_5$ | 11.58 | 4.936291 |
| 11 | F | $OC_2H_5$ | O | $NC_2H_5$ | 3.25 | 5.488117 |
| 12 | $CF_3$ | $OC_2H_5$ | O | $NC_2H_5$ | 2.94 | 5.531653 |
| 13 | H | $OC_2H_5$ | NH | $NC_2H_5$ | 6.77 | 5.169411 |
| 14 | $OCH_3$ | $OC_2H_5$ | NH | $NC_2H_5$ | 25.43 | 4.594654 |
| 15 | F | $OC_2H_5$ | NH | $NC_2H_5$ | 13.03 | 4.885056 |
| 16 | $CF_3$ | $OC_2H_5$ | NH | $NC_2H_5$ | 1.47 | 5.832683 |
| 17 | H | $NHNH_2$ | O | O | 2.4 | 5.619789 |
| 18 | $OCH_3$ | $NHNH_2$ | O | O | 55.37 | 4.256725 |
| 19 | F | $NHNH_2$ | O | O | 28.44 | 4.54607 |
| 20 | $CF_3$ | $NHNH_2$ | O | O | 36.38 | 4.439137 |
| 21 | H | $NHNH_2$ | NH | O | 44.86 | 4.348141 |
| 22 | $OCH_3$ | $NHNH_2$ | NH | O | 27.75 | 4.556737 |
| 23 | F | $NHNH_2$ | NH | O | 23.56 | 4.627825 |
| 24 | $CF_3$ | $NHNH_2$ | NH | O | 25.59 | 4.59193 |
| 25 | H | $NHNH_2$ | O | $NC_2H_5$ | 13.93 | 4.856049 |
| 26 | $OCH_3$ | $NHNH_2$ | O | $NC_2H_5$ | 11.63 | 4.93442 |
| 27 | F | $NHNH_2$ | O | $NC_2H_5$ | 3.34 | 5.476254 |
| 28 | $CF_3$ | $NHNH_2$ | O | $NC_2H_5$ | 11.51 | 4.938925 |
| 29 | H | $NHNH_2$ | NH | $NC_2H_5$ | 13.49 | 4.869988 |
| 30 | $OCH_3$ | $NHNH_2$ | NH | $NC_2H_5$ | 13.09 | 4.88306 |
| 31 | F | $NHNH_2$ | NH | $NC_2H_5$ | 3.3 | 5.481486 |
| 32 | $CF_3$ | $NHNH_2$ | NH | $NC_2H_5$ | 30.03 | 4.522445 |
| 33 | H | OH | O | O | 30.68 | 4.513145 |
| 34 | $OCH_3$ | OH | O | O | 57.14 | 4.24306 |
| 35 | F | OH | O | O | 14.69 | 4.832978 |
| 36 | $CF_3$ | OH | O | O | 13.15 | 4.881074 |
| 37 | H | OH | NH | O | 7.69 | 5.114074 |
| 38 | $OCH_3$ | OH | NH | O | 28.64 | 4.543027 |
| 39 | F | OH | NH | O | 29.46 | 4.530767 |
| 40 | $CF_3$ | OH | NH | O | 6.59 | 5.181115 |
| 41 | H | OH | O | $NC_2H_5$ | 27.19 | 4.565591 |
| 42 | $OCH_3$ | OH | O | $NC_2H_5$ | 3.3 | 5.481486 |
| 43 | F | OH | O | $NC_2H_5$ | 6.91 | 5.160522 |
| 44 | $CF_3$ | OH | O | $NC_2H_5$ | 19.75 | 4.704433 |
| 45 | H | OH | NH | $NC_2H_5$ | 11.8 | 4.928118 |
| 46 | $OCH_3$ | OH | NH | $NC_2H_5$ | 3.48 | 5.458421 |

**Table 2: Best designed compounds with QSAR model 4 predicted Mtb Gyrase B inhibitory activity**




| S.No. | R | R1 | X | Y | L1i | MoRSEN26 | RDFM5 | RDFE25 | pMIC |
|---|---|---|---|---|---|---|---|---|---|
| 5a | $-NH_2$ | $-C_6H_5$ | -S | $-NC_2H_5$ | 46.3672 | -0.076 | 13.332 | 19.126 | 6.113594 |
| 5b | $-NO_2$ | $-C_6H_5$ | -S | $-NC_2H_5$ | 49.1495 | -0.149 | 12.913 | 21.996 | 6.093203 |
| 5f | $-OCH_3$ | $-C_6H_5$ | -S | $-NC_2H_5$ | 47.1623 | -0.066 | 12.143 | 22.939 | 6.281145 |
| 6f | $-OCH_3$ | $-C_6H_5$ | -O | $-NC_2H_5$ | 46.0847 | -0.12 | 13.474 | 23.633 | 6.135638 |
| 7b | $-NO_2$ | - p-fluoro phenyl | -O | $-NC_2H_5$ | 46.7933 | -0.124 | 13.889 | 23.224 | 6.158598 |
| 8a | $-NH_2$ | - p-fluoro phenyl | -S | $-NC_2H_5$ | 49.6656 | -0.046 | 12.777 | 12.554 | 6.087333 |
| 8b | $-NO_2$ | - p-fluoro phenyl | -S | $-NC_2H_5$ | 48.3485 | -0.058 | 12.894 | 17.594 | 6.188158 |
| 8e | -OH | - p-fluoro phenyl | -S | $-NC_2H_5$ | 50.0636 | -0.03 | 10.526 | 21.374 | 6.405132 |
| 8f | $-OCH_3$ | - p-fluoro phenyl | -S | $-NC_2H_5$ | 47.9896 | -0.026 | 9.956 | 19.023 | 6.201504 |
| 9b | $-NO_2$ | - p-hydroxy phenyl | -O | $-NC_2H_5$ | 49.4618 | -0.136 | 13.144 | 28.024 | 6.399976 |
| 9c | $-CH_3$ | - p-hydroxy phenyl | -O | $-NC_2H_5$ | 49.9071 | -0.026 | 7.777 | 25.636 | 6.453775 |
| 9d | $-C_2H_5$ | - p-hydroxy phenyl | -O | $-NC_2H_5$ | 47.5355 | -0.011 | 13.044 | 21.974 | 6.487671 |
| 9e | -OH | - p-hydroxy phenyl | -O | $-NC_2H_5$ | 50.8142 | -0.053 | 11.625 | 14.621 | 6.147161 |
| 10a | $-NH_2$ | - p-hydroxy phenyl | -S | $-NC_2H_5$ | 48.1916 | -0.083 | 11.752 | 23.026 | 6.258226 |
| 10c | $-CH_3$ | - p-hydroxy phenyl | -S | $-NC_2H_5$ | 47.9848 | -0.113 | 14.121 | 29.466 | 6.508855 |
| 10d | $-C_2H_5$ | - p-hydroxy phenyl | -S | $-NC_2H_5$ | 46.4965 | -0.095 | 13.937 | 21.868 | 6.191195 |
| 10e | -OH | - p-hydroxy phenyl | -S | $-NC_2H_5$ | 49.5106 | -0.094 | 14.355 | 15.054 | 6.088331 |
| 11a | $-NH_2$ | - p-methoxy phenyl | -O | $-NC_2H_5$ | 52.0165 | -0.1 | 13.868 | 24.528 | 6.538168 |
| 11b | $-NO_2$ | - p-methoxy phenyl | -O | $-NC_2H_5$ | 51.8471 | -0.134 | 12.936 | 26.694 | 6.458144 |
| 11c | $-CH_3$ | - p-methoxy phenyl | -O | $-NC_2H_5$ | 51.9889 | -0.073 | 14.154 | 13.52 | 6.207362 |
| 11d | $-C_2H_5$ | - p-methoxy phenyl | -O | $-NC_2H_5$ | 49.8103 | -0.152 | 13.156 | 26.843 | 6.316513 |
| 11e | -OH | - p-methoxy phenyl | -O | $-NC_2H_5$ | 53.1721 | -0.109 | 11.364 | 20.996 | 6.309323 |
| 11f | $-OCH_3$ | - p-methoxy phenyl | -O | $-NC_2H_5$ | 50.9633 | -0.119 | 14.541 | 19.004 | 6.237021 |
| 12a | $-NH_2$ | - p-methoxy phenyl | -S | $-NC_2H_5$ | 54.5283 | -0.06 | 11.824 | 18.976 | 6.480603 |
| 12b | $-NO_2$ | - p-methoxy phenyl | -S | $-NC_2H_5$ | 52.8259 | -0.104 | 10.987 | 21.555 | 6.314596 |
| 12c | $-CH_3$ | - p-methoxy phenyl | -S | $-NC_2H_5$ | 53.9246 | -0.099 | 11.629 | 22.676 | 6.457112 |
| 12d | $-C_2H_5$ | - p-methoxy phenyl | -S | $-NC_2H_5$ | 51.9382 | -0.12 | 12.321 | 18.933 | 6.175922 |
| 12e | -OH | - p-methoxy phenyl | -S | $-NC_2H_5$ | 55.2092 | -0.128 | 13.254 | 26.201 | 6.633183 |
| 12f | $-OCH_3$ | - p-methoxy phenyl | -S | $-NC_2H_5$ | 53.1851 | -0.053 | 10.625 | 13.506 | 6.170248 |
| 18a | $-NH_2$ | $-C_6H_5$ | -S | $-NCH_3$ | 45.2354 | -0.04 | 12.459 | 23.262 | 6.304555 |
| 19c | $-CH_3$ | - p-fluoro phenyl | -O | $-NCH_3$ | 43.0018 | -0.004 | 15.978 | 11.984 | 6.036144 |
| 19d | $-C_2H_5$ | - p-fluoro phenyl | -O | $-NCH_3$ | 41.2298 | -0.058 | 13.825 | 20.561 | 6.009373 |
| 23b | $-NO_2$ | - p-methoxy phenyl | -O | $-NCH_3$ | 47.7775 | -0.054 | 18.399 | 18.294 | 6.453151 |
| 24c | $-CH_3$ | - p-methoxy phenyl | -O | $-NCH_3$ | 48.6639 | -0.122 | 11.991 | 20.423 | 6.057297 |
| 34c | $-CH_3$ | -p-hyroxy phenyl | -S | -S | 39.2501 | 0.035 | 13.432 | 15.87 | 6.026652 |

**Table 3: Molecular docking results of designed compounds**

| Designed compounds | Binding energy (K Cal / mol) | Binding residue interactions along with distance (Å) |
|---|---|---|
| **8e** | -9.1 | His44 (2.03349), Trp47 (2.041813), Asn309 (2.03184), Glu312 (3.52474) and His311 (3.58504) |
| **9c** | -8.2 | His44 (2.1455), Trp47 (2.29861), Asn309 (1.9318), Glu312 (2.79548) and His311 (3.64649) |
| **9d** | -8.0 | His44 (2.19698), Trp47 (2.30081), Asn309 (1.94101) and His311 (3.61389) |
| **10c** | -8.1 | His44 (2.01258), Trp47 (2.40607), Asn309 (1.95965), Glu39 (2.34003) and His311 (3.58864) |
| **11a** | -8.2 | His44 (1.86867), Trp47 (2.51243), Asn309 (1.95123), Thr189 (3.536) and His311 (3.75156) |
| **11b** | -8.1 | Val301 (3.31381), Trp47 (2.56797), Thr303 (2.98798), Thr310 (3.19984) and Asn309 (2.53557) |
| **12a** | -7.8 | His44 (2.09709), Trp47 (2.41714), Asn309 (1.91631) and His311 (3.60803) |
| **12c** | -7.7 | His44 (2.15838), Trp47 (2.31666), Asn309 (1.9318) and His311 (3.59417) |
| **12e** | -8.6 | His44 (2.03439), Trp47 (2.03571), Asn309 (2.05184), Thr189 (3.328) |
| **23b** | -8.0 | Trp47 (2.04181), Asn309 (2.0318), Val301 (3.30567) |

### *Molecular dynamics simulations*

The potential energy of the protein-ligand complex was analyzed for 30 ns. The potential energy of complex stabilization was found to be approximately -1.2304035 × $10^6$ kcal/mol. The temperature of the system rapidly reached 300 K and then remained stable over the rest of the simulation time. The average temperature of the system was 299.638 K. The pressure fluctuated widely over the 200 ps equilibration phase. The average pressure was 10.1745 bar, and the average density was 981.465 kg $m^{-3}$. Density values were very stable over time, indicating that the system was well-equilibrated with respect to pressure and density.

RMSD and Rg values are widely used for the assessment of macromolecules stability and rigidity, respectively [29]. An Unconstrained simulation of 30 ns was performed on the docked structure of Mtb gyrase B (3ZKD) bound to compound 16 and 8e. Despite the initial structural arrangements of the docked complex, the average RMSD of the trajectories for bound protein backbone atoms showed relative stability. The stability of the protein–ligand complexes were compared within 30 ns of MD simulations. RMSD values of the protein–ligand complexes remained less than 0.4 nm throughout the simulation periods. Interestingly, the lead (compound 8e) - DNA Gyrase B complex showed even lower fluctuation and higher stability than the reference complex. The highest stability in RMSD value was observed in DNA Gyrase B–lead complex which had also obtained the best binding affinity. Figure 4 display stable RMSD values of the atoms for the compounds 16 and 8e with DNA Gyrase B (3zkd) respectively. The RMSD analysis for 3zkd and compound 16 indicates that they reach equilibration and oscillate around an average value after 1ns. The average RMSDs from 1ns to30ns for compounds 16 and 8e bound to 3zkd protein were 0.3nm and 0.25nm respectively. These results show the relative stability of compound 8e during the simulation while compound 16 had higher RMSD.

Analysis of the docking results showed that His44, Trp47, Asn309, Glu312, His311, Glu39, Thr189 and Val301 residues of DNA Gyrase B could be considered as important residues for catalytic activity. Structural fluctuations of the protein residues in the complexes were also evaluated by computing RMSF values during MD simulations. It is documented that residues with higher RMSF values in the protein sequence are more flexible and probably have an unfolded state in the protein structure. Local protein mobility was analysed by measuring the time-averaged RMSF values of selected compounds bound to 3zkd protein against residue numbers based on 30ns trajectory data. Like 'DNA Gyrase B–compound 16' complex from the dataset, lead compound 8e based complex displayed a noticeable diminution in the RMS fluctuation of these residues in comparison with the other residues during MD simulations. The results proposed that upon binding of the compounds to DNA Gyrase B, critical DNA-binding residues of DNA gyrase would be occupied by the inhibitory compounds which consequently results blockage in the activity of DNA gyrase. The average RMSFs measured for compounds bound to 3zkd were 0.012nm and 0.005nm, which reveals the relative stability of the complex upon binding (Figure 5). Upon comparing the RMSFs for both complexes near the binding site region suggest that the designed compound 8e showed better binding pattern and stabilize in favorable conformation for inhibition.

Moreover, Rg values of the protein–ligand complexes were decreased throughout MD simulations. This elucidates that DNA Gyrase B–compound complexes are highly compact under the simulated conditions. Rg value, which provides insight into the overall dimension and shape of the protein, is calculated for both complexes. The average Rg values were 2.45nm and 2.32nm for 3zkd protein (Figure 6). The results of the average Rg values indicate that overall shape of the protein was stable upon binding of the ligand.

## CONCLUSION

Current work aims towards identification and development of novel leads for a series of novel 4-aminoquinoline derivatives as inhibitors of Mtb DNA gyrase B using QSAR, molecular docking as well as dynamic simulation tools. QSAR model developed using WHIM, RDF and MoRSE based descriptors with both internal and external validated values established a meaningful relationship between structure and bioactivity of 4-aminoquinolines with Mtb Gyrase subunit B inhibitory activity. Positively correlated descriptors benefit the model along with leverage analysis. Utilizing molecular descriptors information from QSAR studies, a self-explorative study on the dataset was performed to identify the model efficiency and accuracy. Further, structure bioactivity screening was performed to identify the role of various functional moieties on the core structure, which result a meaningful relationship between dataset compounds and DNA Gyrase B inhibitory activity.

A series of compounds were designed from structure activity relationship information and tested for predicted bioactivity using QSAR model by obtaining molecular descriptor values from fore mentioned web servers. Best designed compounds having predicted bioactivity values above the dataset bioactivity values were considered for docking analysis to rule out hydrogen bond interactions. Docking results

analyzed nonbonding interactions in the active site of enzyme favouring hydrogen bond interactions with Asp186, Glu188, His44, Trp47, Asn309, Glu312 and His 311. Compound 8e with binding energy of -9.1 K Cal/mol have shown better predictions in both QSAR and molecular docking studies compared to compound 16 with binding energy of -8.7 K Cal/mol. Thus, the key designed compound 8e was taken over to simulation studies along with compound 16, taking as reference.

The molecular dynamic studies revealed that compounds 8e bound to 3zkd were stable and exhibited minimal conformational changes, compared to compound 16 of dataset. Overall, molecular dynamics simulation studies of 3zkd with ligand 16, 8e analyzed using RMSD; RMSF and Rg analysis gave better information on complex stability. The average RMSDs for compounds 16 and 8e bound to 3zkd protein were 0.3nm and 0.25nm respectively. These results show the relative stability of compound 8e during the simulation while compound 16 had higher RMSD. The average RMSFs measured for compounds 16 and 8e bound to 3zkd were 0.012nm and 0.005nm respectively, which reveals the relative stability of the complex upon binding. Upon comparing the RMSFs for both complexes near the binding site region suggest that the designed compound 8e showed better binding pattern and stabilize in favorable conformation for inhibition. The average Rg values for compounds 16 and 8e were 2.45nm and 2.32nm for 3zkd protein. The results of the average Rg values indicate that overall shape of the protein was stable upon binding of the ligand. Thus the present computational work reports the compound 8e as scaffold for the development of analogs to target 3zkd, DNA Gyrase B. The overall research work determines compound 8e having a great potential to inhibit DNA gyrase B and thus can be applied for further studies which aim to discover novel therapeutic drugs.

## Abbreviations

TB: Tuberculosis; WHO: World health organisation; MDR: Multi drug-Resistant; XDR: Extremely drug-resistant; BRD: Breakage-reunion domain; CTD: C-terminal domain; ATP: Adenosine tri phosphate; QSAR: Quantitative structural activity relationship; LDF: lack-of-fit; RMSE: Root mean square error; CCC: Concordance correlation coefficient; WHIM: Weighed holistic invariant molecular; G-WHIM: Grid-weighed holistic invariant molecular; RDF: Radical distribution function.

## Acknowledgment

We thank the Research Council of SRMIST and the Dean, SRM College of Pharmacy for support.

## Funding

The present research work was not funded by any funding agency.

## Ethics approval and consent to participate

Not applicable.

## Competing interests

The Authors declare no competing interests.

**Declarations of interest:** None

## REFERENCES

1. https://www.who.int/publications/i/item/9789240037021.
2. Seung K. J., Keshavjee S., Rich M. L. (2015). Multidrug-Resistant Tuberculosis and Extensively Drug-Resistant Tuberculosis. Cold Spring Harbour Perspectives of Medicine. 5(9): a017863. doi: 10.1101/cshperspect. a017863.
3. Hirsch J., Klostermeier D. What makes a type IIA topoisomerase a gyrase or a Topo IV? Nucleic Acids Research. 2021; 49(11): 6027–6042. https://doi.org/10.1093/nar/gkab270.
4. Basu A., Parente A. C., Bryant Z. (2016). Structural Dynamics and Mechanochemical Coupling in DNA Gyrase. Journal of Molecular Biology. 428(9 Pt B): 1833-45. doi: 10.1016/j.jmb.2016.03.016.
5. McKie S. J., Desai P., Seol Y., Maxwell A., Neuman K. (2021). Topoisomerase VI is a chirally-selective, preferential DNA decatenase. BioRxiv. 15: 431225. https://doi.org/10.1101/2021.02.15.431225.
6. Chopra S., Matsuyama K., Tran T., Malerich J. P., Wan B., Franzblau S. G., Lun S., Guo H., Maiga M. C., Bishai W. R., Madrid P. B. (2012). Evaluation of gyrase B as a drug target in Mycobacterium tuberculosis. Journal of Antimicrobial Chemotherapy. 67(2): 415-21. doi:10.1093/jac/dkr449.
7. Weidlich D., Klostermeier D. (2020). Functional interactions between gyrase subunits are optimized in a species-specific manner. Journal of Biological Chemistry. 295(8): 2299-2312. doi: 10.1074/jbc.RA119.010245.
8. Chaudhari K., Surana S., Jain P., Patel H. M. (2016). Mycobacterium Tuberculosis (MTB) GyrB inhibitors: An attractive approach for developing novel drugs against TB. European Journal of Medicinal Chemistry, 124: 160-185.

*9.* Sandeep S., Gurpuneet K., Veenu M., Manish K. G. (2015). Quinoline and quinolones: promising scaffolds for future antimycobacterial agents. Journal of Enzyme Inhibition and Medicinal Chemistry. 30(3): 492-504.
10. Gopinath P., Kathiravan M. K. (2019). QSAR and docking studies on Triazole Benzene Sulfonamides with human Carbonic anhydrase IX inhibitory activity. Journal of Chemometrics. 33(12): e3189. https://doi.org /10.1002/cem.3189.
11. Medapi B., Suryadevara P., Renuka J., Sridevi J. P., Yogeeswari P., Sriram D. (2015). 4-Aminoquinoline Derivatives as Novel Mycobacterium tuberculosis GyrB Inhibitors: Structural Optimization, Synthesis, and Biological Evaluation. European Journal of Medicinal Chemistry. DOI: 10.1016/j.ejmech.2015.06.032.
12. Hanwell M. D., Curtis D. E., Lonie D. C., Vandermeersch T., Zurek E., (2012). Geoffrey R Hutchison. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. Journal of Cheminformatics. 4:17.
13. Dong J., Dong-Sheng C., et. al. (2015). ChemDes: An integrated web-based platform for molecular descriptor and fingerprint computation. Journal of Cheminformatics. 7. DOI: 10.1186/s13321-015-0109-z.
14. Gramatica P., Chirico N., Papa E., Cassani S., Kovarich S. (2013). QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. Journal of Computational Chemistry. 3: 2121–2132. DOI: 10.1002/jcc.23361.
15. Gramatica P., Chirico N., Cassani S. (2014). QSARINS-chem: Insubria datasets and new QSAR / QSPR models for environmental pollutants in QSARINS. Journal of Computational Chemistry 35: 1036–1044. DOI: 10.1002/ jcc.23576.
16. Morris G. M., Huey R., Lindstrom W., Sanner M. F., Belew R. K., Goodsell D. S., Olson A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of computational chemistry. ; 30(16): 2785–2791. DOI:10.1002/jcc.21256.
17. Webb B., Sali A. (2016). Comparative Protein Structure Modeling Using Modeller. Current Protocols in Bioinformatics. John Wiley & Sons, Inc., 54: 5.6.1-5.6.37. https://doi.org/10.1002/cpbi.3.
18. BIOVIA, Dassault Systèmes, Discovery studio visualizer, v20.1.0.19295. San Diego: Dassault Systèmes, San diego.
19. Abraham M.K., Murtola T., Schulz R., Páll S., Jeremy C. Smith, Hess B., Lindahl E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015; 1(2): 19-25. https://doi.org/10.1016/j.softx.2015.06.001.
20. Huang J., Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell AD Jr. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Natural Methods. 2017; 14(1): 71-73. doi:10.1038/nmeth.4067.
21. Bernetti M. and Bussi G. Pressure control using stochastic cell rescaling. Journal of Chemical Physics. 2020; 153: 114107.
22. Todeschini R., Gramatica P. (1998). New 3D Molecular Descriptors: The WHIM theory and QSAR Applications. Perspectives in Drug Discovery and Design. 9 (10/11): 355–380. https://doi.org/10.1007/0-306-46857-3_19.
23. Todeschini R., Gramatica P. (1997). The Whim Theory: New 3D Molecular Descriptors for QSAR in Environmental Modelling. SAR and QSAR in Environmental Research. 7(1-4): 89-115. DOI: 10.1080 /10629369708039126.
24. González M. P., Suárez P. L., Fall Y., Gómez G. (2005). Quantitative structure–activity relationship studies of vitamin D receptor affinity for analogues of 1α,25-dihydroxyvitamin D3.1: WHIM descriptors. Bioorganic & Medicinal Chemistry Letters. 15 (23): 5165-5169, https://doi.org/10.1016/j.bmcl.2005.08.065.
25. Devinyak O., Havrylyuk D., Lesyk R. (2014). 3D-MoRSE Descriptors Explained. Journal of Molecular Graphics and modeling. 54: 194-203. DOI: 10.1016/j.jmgm.2014.10.006.
26. Maykel P. G., Carmen T., Marta T., Aliuska M. H. (2007).QSAR Studies Using Radial Distribution Function for Predicting A1 Adenosine Receptors Agonists. Bulletin of Mathematical Biology.69: 347–359. DOI: 10.1007/s11538-006-9127
27. Maykel P. G., Zoila G., Yagamare F., Generosa G. (2008). Radial Distribution Function descriptors for predicting affinity for vitamin D receptor. European Journal of Medicinal Chemistry.43: 1360-1365. DOI: 10.1016/j.ejmech.2007.10.020
28. González M. P., Terán C., Teijeira M., et al. (2007). QSAR Studies Using Radial Distribution Function for Predicting A1 Adenosine Receptors Agonists. Bulletin of Mathematical Biology. ; 69: 347. https://doi.org/10.1007/s11538-006-9127-3.
29. Gopinath P., Kathiravan M. K. (2021). Docking studies and Molecular Dynamics Simulation of Triazole Benzene Sulfonamide Derivatives with human Carbonic anhydrase IX Inhibition Activity. RSC Advances. 11: 38079-38093. Doi: 10.1039/d1ra07377.


## CITATION OF THIS ARTICLE

G Papichettypalle, P. Aravanan, RSamineni, B B Karkara, S Thumula, G P Muppala, S. Subramanyam. Molecular leads generation for 4-aminoquinolines having Mtb Gyrase B inhibitor activity: QSAR and Molecular dynamics simulation studies. Bull. Env.Pharmacol. Life Sci., Vol 12 [4] March 2023: 58-71