RESEARCH

QSAR modeling, molecular docking, dynamic simulation and ADMET study of novel tetrahydronaphthalene derivatives as potent antitubercular agents

Anguraj Moulishankar[1] and T. Sundarrajan[1*]

Abstract

Background Tuberculosis is an air-borne contagious disease caused by slow-growing *Mycobacterium tuberculosis* (Mtb). According to Global Tuberculosis Report 2022, 1.6 million people were infected by tuberculosis in 2021. The continuing spread of drug-resistant tuberculosis (TB) is one of the most difficult challenges to control the tuberculosis. So new drug discovery is essential to the treatment of tuberculosis. This study aims to develop a QSAR model to predict the antitubercular activity of tetrahydronaphthalene derivatives. The QSARINS was used in this study to develop the QSAR predictive model.

Results A number of tetrahydronaphthalene derivatives with MIC_{90} values were obtained from the literature to develop the QSAR predictive model. The compounds were divided into two sets: a training set consisting of 39 compounds and a test set containing 13 compounds. The best predictive Model 4 has R^2: 0.8303, Q^2_{LOO}: 0.7642, LOF: 0.0550, Q^2-F_1: 0.7190, Q^2-F_2: 0.7067, Q^2-F_3: 0.7938 and CCC_{ext}: 0.8720. Based on the developed QSAR equation, the new compounds were designed and subjected to molecular docking, molecular dynamics and ADMET analysis.

Conclusion In the QSAR model, the molecular descriptors of MATS8s, Chi4, bcutv8, Petitjeant and fr_aniline were highly influenced the antitubercular activity. The developed QSAR model helps to predict the antitubercular activity of tetrahydronaphthalene derivatives.

Keywords QSAR, Tetrahydronaphthalene derivatives, *Mycobacterium tuberculosis*, ATP synthase, Molecular docking, MD simulation, ADMET study

1 Background

Mycobacterium tuberculosis is an air-borne bacterial infectious agent that causes tuberculosis (TB), a chronic disease that spreads through the air [1]. It is one of the most common causes of mortality from a single pathogenic agent that mainly affects the lungs. It is an acid-fast, slow-growing gram-positive bacteria with a very impermeable cell wall. According to Global Tuberculosis Report 2022, 1.6 million people were diagnosed with TB in 2021, an increase of 4.5% from 1.1 million people in 2020. In 2021, there were an estimated 14 lakh TB patient death among HIV-negative people and 1.87 lakhs TB patients' deaths among HIV-positive people [2].

The World Health Organization (WHO) has issued guidelines for treating tuberculosis through a program known as Directly Observed Treatment (DOTS short-course). This approach involves a six-month treatment

*Correspondence:
T. Sundarrajan
chemistrysundar@gmail.com
[1] Department of Pharmaceutical Chemistry, SRM College of Pharmacy, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamil Nadu 603203, India

regimen of four first-line TB drugs: rifampicin, isoniazid, pyrazinamide and ethambutol. The treatment begins with a two-month course of all four drugs, followed by a four-month course of rifampicin and isoniazid [3]. One of the major issues with the treatment of TB is the emergence of multidrug-resistant tuberculosis. Genetic mutations mainly cause resistance [4]. The emergence of multidrug-resistant (MDR) and extensively drug-resistant (XDR), and totally drug-resistant-TB (TDR), along with the toxicity of existing antitubercular drugs leading to adverse side effects, leads to reviving the efforts of antitubercular drug discovery [5].

Therefore, it is imperative to develop new therapeutics with improved toxicity profiles in order to overcome these challenges. After more than 40 years, the discovery of delamanid (mycolic acid biosynthesis inhibitor) and bedaquiline (mitochondrial ATP synthase inhibitor) provided a degree of relief in the treatment of MDR-TB. However, the two medications have some notable side effects, such as ADME issues and hERG toxicity [6]. The discovery of new antitubercular agents with low toxicity is the only choice to face this difficult situation. But it is a costly and time-consuming process. Nowadays, various in silico tools are used to speed up the drug discovery process. So in this work, we used the quantitative structure–activity relationship (QSAR) to develop the predictive model. The QSAR is an adorable computational tool; it helps to find the relationship between the physiochemical properties of the compounds and its biological activity. Throughout the past thirty years, lead optimization has been sped up using the QSAR (quantitative structure–activity relationship) model. It helps to predict the pharmacological activity of novel compounds; it aids in the development of mathematical QSAR models. It gives ideas for lead optimization and virtual screening priority [7].

The current study deals with QSARINS, an adorable computational software that helps develop a robust QSAR model, which allows the design of novel compounds with higher mycobacterial tubercular inhibitory activity [8, 9]. In this study, we are investigating the *in silico* theoretical studies to develop novel mycobacterium tuberculosis inhibitors. We present here the results of a QSAR study on derivatives of tetrahydronaphthalene that are effective inhibitors of mycobacterium tuberculosis. Tetrahydronaphthalene is one of the excellent building blocks in medicinal chemistry. It has a wide range of biological properties, including antitubercular activity [10], anticancer [11, 12], anti-inflammatory [13], antiviral [14], anti-Alzheimer [15], antimicrobial [16] and antimalarial [17].

With the significant biological activity of tetrahydronaphthalene and its derivatives, we chosen this scaffold for our QSAR study. The current study aimed to construct a QSAR model of new tetrahydronaphthalene derivatives for their mycobacterium tuberculosis inhibitor and to explain how structural variation affects the antitubercular activity. The tetrahydronaphthalene derivatives showed good mycobacterium tuberculosis inhibition activity with a range of MIC_{90} values of 0.21μg/mL to 13.5μg/mL. The QSAR model development was performed to investigate and explore the structural characteristics of tetrahydronaphthalene analogs required to inhibit Mycobacterium tuberculosis. According to the OECD regulatory guidelines, thorough validation was done, and its parameters were well-fit. The applicability domain was also developed. We performed ligand-based *in silico* studies to interpret the relationship between structural features and the biological activity of the compounds. In this paper, we present an analysis of 2D-QSAR studies conducted on the tetrahydronaphthalene derivatives that have been identified as effective inhibitors against mycobacterium tuberculosis [18–20].

2 Materials and methods

2.1 Softwares

QSARINS software was used to develop the 2D QSAR predictive model. The 2D structure of the compounds was sketched by ACD/Labs ChemSketch (Freeware) 2021.1.1. The energy minimization of the 2D structure was done using the MM2 force field in Chem3D pro version 12.0.2.1076 [21, 22]. PaDEL Descriptor software (version 2.20) was used to construct the molecular descriptors [23, 24]. The physiochemical and pharmacokinetic properties (ADME) of the compounds were evaluated using the SWISSADME online server. The quality of the selected protein was evaluated by the Ramachandran plot using a procheck webserver. The various toxicity studies including hepatotoxicity and carcinogenicity were predicted using the Protox-II webserver [39, 40]. The active site of the selected target protein was predicted by the Supercomputing Facility for Bioinformatics & Computational Biology (SCFBio), IIT- Delhi as DBT'S center of excellence. The molecular docking of the ATP synthase was performed using AutoDock Vina.

2.2 Molecular sketching and 3D optimization

QSARINS software was used to develop the MLR-QSAR predictive models by ordinary least squares. The development of QSAR was pioneered at the University of Insubria, Italy [25]. QSARINS uses a hybrid approach that combines genetic algorithms (GAs) and multiple linear regressions (MLRs) [25]. It helps build QSAR models that are both highly predictive and comprehensible [26]. The compounds structure was drawn by ChemSketch and saved as a mol file format. The energy minimization of

the 2D structure was done by using the MM2 force field in Chem3D pro and saved as the mol file format. The molecular descriptor for 55 compounds was calculated by the PaDEL Descriptor version 2.20 software [27, 28].

2.3 Dataset collection, splitting and model development

For the studies, 55 molecules with antitubercular activity on H37Rv were collected from the literature. Wide variations in activity further enhance the performance of compounds with good activity [10]. The obtained data set was split into two sets, namely the training set or modeling set (70%) and the test set or prediction set (30%). The training set is used to model development and the test set is used to validation of the developed model. This work uses the MLR using the ordinary least-square approach to construct a good QSAR model. Multiple linear regression was used to correlate the relationship between the biological activity (MIC_{90}) and molecular descriptors. Here biological activity is the dependent variable and molecular descriptors are the independent variables. A genetic algorithm was used to selection of good molecular descriptors to build the QSAR model. It shows the direct correlation between independent variable X and dependent variable Y (descriptors). The dependent variable Y (biological activity) in an MLR analysis depends on independent variable X (molecular descriptors) [29].

$$Y = k1 * x1 + k2 * x2 + k3 * x3 + C$$

Y is the dependent variable (biological activity), x is the independent variable (molecular descriptor), 'C' is a regression intercept and 'k's are the regression coefficients [30].

2.4 Applicability domain

The applicability domain of a QSAR model is a theoretical spatial region defined by the molecular properties or structural information of the compounds used in the QSAR model development. In which the William plot was used to define the applicability domain so that reliable prediction of the QSAR model has a leverage value (h) of the compounds below the critical leverage value (h^*) with a standard deviation value of ± 3. The compounds that are outside the spatial region of the applicability domain may be considered an outlier. This method used to find the influential molecules reveals the model's ability to make predictions in the limited region. The term "outlier" refers to any molecule that deviates from the domain space or range. If the cross-validated standardized residual produced by the model is greater than ± 3, then the data are said to be an outlier. The leveraged method goes into more detail on the QSAR model's applicability domain [31, 32]. The formula for the warning leverage (H*) is as follows:

$$H* = 3(p + 1) / n$$

H^* = The warning leverage value to check the effective molecule.

p = The number of molecular descriptors to build the QSAR model.

n = The number of molecules in the training dataset.

2.5 Y-randomization

It serves as an external validation factor to assess the validity and reliability of the developed QSAR model. The Y-randomization test was performed on the training dataset. In the Y-randomization test, the dependent variables (biological activity data) were shuffled, while the independent factors remained unchanged to create the MLR-QSAR model [33]. The developed QSAR model should have low R^2 and Q^2 values, in order to ensure that the model equation is reliable. R^2_{Yscr} needs less to get a reliable model. It confirmed that the developed model was not obtained by the random chance correlation. If the developed QSAR model's internal validation process was good, then the values of R^2 and Q^2 of each iteration, and their averages (R^2_{Yscr} and Q^2_{Yscr}) should be lesser with respect to the values of the model [34, 35].

2.6 Validation of the developed QSAR model

The stability, reliability, predictive strength, fitting criteria and the robustness of the QSAR models were validated by internal validation parameters and external validation parameters using QSARINS software. We used various statistical parameters like correlation coefficient (R^2), adjusted correlation coefficient (R^2_{adj}), leave-one-out squared correlation coefficient (Q^2_{LOO}) and leave-many-out squared correlation coefficient (Q^2_{LMO}) to evaluate the internal validation of the model. We also used some statistical parameters like Q^2-F_1, Q^2-F_2, Q^2-F_3 and concordance correlation coefficient (CCC) to evaluate the internal validation of the model. Higher R^2 for goodness of fit, Higher Q^2_{LOO} for robustness, lowest difference between R^2 and Q^2_{LOO} for stability, low K_{xx} for less correlation among the descriptor and high K_{xy} for high correlation among the descriptor and biological response. The lower $RMSE_{Tr}$, $RMSE_{CV}$, and $RMSE_{ext}$ must be less and as close as possible to get good model. The minimum suggested statistical values tabulated must be met for the QSAR model to be considered acceptable [36–38].

2.7 ADMET study

The physicochemical and pharmacokinetic properties (ADME) of the compounds were evaluated using SWISSADME online server. The druglikeness property was evaluated by Lipinski's rule of five. According to the Lipinski rule of five, molecular weight should be less than

500 dalton, hydrogen bond donor should be less than 5, hydrogen bond acceptor should be less than 10 and partition coefficient (LogP) should be less than 5. The various toxicity studies including AMES toxicity, hepatotoxicity, hERG I inhibitor and hepatotoxicity were predicted using an online server called pkCSM server [39, 40].

2.8 Protein and ligand preparation
The drug target proteins were retrieved from the RCSB-PDB database with higher resolution. The crystal structure of the ATP synthase (PDB ID 7NK7) was predicted by the electron microscopy method, resolution of 2.11 Å [41, 42]. This protein is mainly involved in the catalysis for the synthesis of ATP from ADP and phosphate in mycobacterium tuberculosis. The retrieved proteins were prepared by removal of cocrystallized ligand and water molecules using Molegro Molecular Viewer and saved as the PDB file format. The 3D structure of ligand molecules was sketched by ChemSketch and saved in mol file format. These 2D structures of the ligand molecules were energy-minimized by Chem3D pro 12.0 software and saved as the PDB file format. The physiochemical properties of the molecules were calculated using SWISSADME online server [43–45].

2.9 Active site prediction and molecular docking
The active site of the selected target protein was predicted by the Supercomputing Facility for Bioinformatics & Computational Biology (SCFBio), IIT- Delhi as DBT'S center of excellence. The newly designed molecules were docked to the drug target protein. The molecular docking was to find the potential binding affinity of the designed compounds with the target protein. The process of molecular docking the ligands to the target protein encompasses several stages: preparation of the protein, preparation of the ligand, generation of the receptor grid, and actual molecular docking. The molecular docking study of the ATP synthase and tetrahydronaphthalene derivatives was performed using AutoDock Vina with grid coordinates of Center x (246.074), center y (194.992) and center z (251.111), respectively. The grid spacing was set to 70 Å with default docking parameters. The binding interaction of the protein and ligand was visualized and analyzed using Biovia Discovery Studio Visualizer 2021 [46, 47].

2.9.1 Molecular dynamic simulation
Molecular dynamics (MD) simulation was carried out using GROMACS 2022.2. The following steps were utilized.

2.9.2 Preparation of enzyme
The three-dimensional (3D) model of the ligand–protein complex was exported to.pdb format using Pymol. The dynamic behavior of the complex was evaluated using molecular dynamic (MD) simulation in the GROMACS package program (version 2022.2) [48–50]. Protein topology was constructed by pdb2gmx with the CHARMM27 force field [51], and ligand topology was generated using the SwissParam server [52].

2.9.3 Setting up a system for simulation
After applying the force field, the complex was inserted into the system. It was solvated with the TIP3P water model [53] in a cubic box greater than 1 nm from the edge of the protein with periodic boundary conditions. The system was neutralized by adding Na+ions, and energy minimization was done for 50,000 steps using the steepest descent algorithm. It was then followed by 100 ps of NVT simulation at 300 K and 100 ps of NPT simulation to equilibrate the whole system. Leapfrog algorithm was employed in the constant-temperature, constant-pressure (NPT) ensemble to separately couple each component like protein, ligand, water molecules, and ions [54]. The Berendsen temperature and pressure coupling constants were set to 1 and 2, respectively, to keep the system in a stable environment (300 K temperature and 1 bar pressure) [55]. The NPT ensemble step in molecular dynamics (MD) simulations plays a vital role by controlling pressure and temperature. This equilibrium-seeking step enhances the accuracy and stability of simulations, ensuring that they reflect the true thermodynamic behavior of molecules. It also facilitates efficient sampling of phase space, which is critical for obtaining statistically relevant results. Additionally, the NPT ensemble maintains realistic density and compressibility, which is crucial for systems with different phases or phase transitions. Finally, MD simulation for 100 ns was performed in isothermal and isobaric condition ensemble at 300 K. The pressure coupling with time constant was set at 1 ps to maintain pressure constant at 1 bar, and the LINCS algorithm [56] was used to constrain the bond lengths. The van der Waals and Coulomb interactions were truncated at 1.2 nm, and the PME algorithm [57] inbuilt in GROMACS was used to minimize the error from truncation.

2.9.4 Visualization and analysis of simulation
The trajectory file was visualized through VMD (Visual Molecular Dynamics) 1.9.2. [58] and analyzed by indigenously developed tool HeroMDAnalysis and Xmgrace 5.1.25 [59, 60].

2.9.5 Molecular mechanics generalized born and surface area (MM/GBSA) calculations

The binding free energies of the protein–ligand complex were calculated at the time of 0 ns, 50 ns and 100 ns by using molecular mechanics/generalized born surface area (MM/GBSA) methods. All the water molecules and ions were removed before MM/GBSA calculations. MM/GBSA free energy decomposition analysis was also used to calculate the binding free energy between the protein–ligand complex.

3 Results

3.1 QSAR model development and validation

The tetrahydronaphthalene and its antitubercular activity ($pMIC_{90}$) were used for this QSAR modeling study. The QSARINS software was utilized to generate multiple MLR-QSAR models. In this study, we focus exclusively on the most optimal QSAR models. For the studies, 55 molecules with antitubercular activity on H37Rv were collected from the literature. Wide variations in activity further enhance the performance of compounds with good

activity. The molecular descriptors were calculated using PaDEL descriptor software and ChemDes online software. It calculated around 2813 molecular descriptors for all the compounds. During the prereduction of the molecular descriptor step, it removed around 2413 molecular descriptors which have more than 80% of constant value and more than 95% of intercorrelated among the molecular descriptors. The compounds THN-13, THN-14 and THN-56 were entirely different from the rest of the compounds in the applicability domain. So they were excluded as outliers. The obtained data set was split into two sets, namely the training set or modeling set (70%) and the test set or prediction set (30%). The training dataset (39 compounds) was used to QSAR model development, and the test dataset (13 compounds) was used for validation of the developed QSAR model. Table 1 lists the various statistical parameters for developed QSAR models. Table 2 shows the correlation matrixes among the molecular descriptors in developed QSAR model 4. Table 3 enlists the calculated and predicted antitubercular activity by model equation and LOO method. The Williams plot helped to find the

Table 1 Statistical parameters of QSAR model 1, 2, 3 and 4

Statistical parameters	Threshold value	QSAR Model 1	QSAR Model 2	QSAR Model 3	QSAR Model 4
Fitting criteria					
R^2	> 0.60	0.9103	0.8747	0.8680	0.8303
S	< 0.30	0.1422	0.1654	0.1698	0.1895
R^2_{adj}	> 0.60	0.8900	0.8512	0.8432	0.8046
$R^2 - R^2_{adj}$	< 0.30	0.0203	0.0235	0.0248	0.0257
LOF	< 0.30	0.0391	0.0468	0.0494	0.0550
$RMSE_{tr}$	Better < 0.3	0.1268	0.1498	0.1538	0.1743
CCC_{tr}	> 0.85	0.9530	0.9332	0.9293	0.9073
F	Higher than the theoretical value	44.9295	37.2277	35.0570	32.3028
Internal validation criteria					
Q^2_{loo} (r^2_{cv})	> 0.50	0.8629	0.8205	0.8274	0.7642
$R^2 - Q^2_{LOO}$	< 0.30	0.0473	0.0542	0.0406	0.0661
Q^2_{LMO}	> 0.50	0.8333	0.7836	0.7526	0.7473
$RMSE_{cv}$	< 0.30	0.1567	0.1793	0.1759	0.2055
R^2_{Yscr}	< R^2 (smallest is better)	0.1837	0.1578	0.1540	0.1300
$Q2_{Yscr}$	< Q^2 (smallest is better)	− 0.4265	− 0.3807	− 0.4455	− 0.2260
External validation criteria					
R^2_{ext}	> 0.60	0.7229	0.6696	0.6056	0.7751
$RMSE_{ext}$	< 0.30	0.2357	0.2349	0.2349	0.1922
Q2-F1	> 0.70	0.5776	0.5803	0.5804	0.7190
Q2-F2	> 0.70	0.5590	0.5618	0.5619	0.7067
Q2-F3	> 0.70	0.6900	0.6920	0.6921	0.7938
CCC_{ext}	> 0.85	0.8282	0.8106	0.7107	0.8720
r^2m aver	> 0.50	0.5578	0.5419	0.3699	0.6769
Δr^2m	< 0.20	0.2531	0.1841	0.3567	0.1894
K′(pred by model equation)	0.85 < k or k′ < 1.15	0.9961	0.9988	1.0102	1.0040
K(pred by model equation)	0.85 < k or k′ < 1.15	1.0021	0.9994	0.9883	0.9949

Table 2 Correlation matrixes between the descriptors

	MATS8s	bcutv8	Chi4	Petitjeant	fr_aniline
MATS8s	1.0000				
bcutv8	−0.0142	1.0000			
Chi4	0.02630	0.7653	1.0000		
Petitjeant	−0.1946	0.0436	−0.1388	1.0000	
fr_aniline	0.2573	−0.03447	−0.0095	0.2257	1.0000

applicability domain of the developed QSAR model. The applicability domain was defined within a squared area of leverage threshold of $h^* = 0.423$ (x-axis) and ± 3 standard residuals (y-axis) for QSAR model 4. $h^* = 3p'/n$, where n is the number of molecules used in the developed QSAR model, p' is the number of the molecular descriptors in the model plus one, and h^* is the warning leverage value.

Figure 1 shows a scatter plot of the experimental value versus the predicted value by the model equation. In this diagram, the pale circles depict the compounds belongs to the training set, while the deep circles symbolize the compounds belonging to prediction set. Figure 2 depicts the scatter plot of experimental value versus predicted value by the leave-one-out (LOO) method. Figure 3 shows the scatter plot of leave-many-out (LMO) validation. This plot explains the intercorrelation among the molecular descriptors. Figure 4 illustrates the scatter plot of residual predicted by leave one out (LOO). Figure 5 shows the Williams plot of the Hat diagonal leverage value versus standardized residuals. It helps to find the outliers of the developed QSAR model. In this developed model, only one compound shown as an outlier. Figure 6 illustrates the scatter plot of the Y-scrambled model. This Y-scramble plot explains the external validation parameters.)

QSAR Model 1

$$\begin{aligned} pMIC_{90} =\ & 94.2403 + 1.0546 * MDEC - 11 \\ & + 2.9961 * bcutv8 - 0.8756 \\ & * bcutm6 - 30.7053 * bcute1 \\ & + 9.7121 * bcutp1 - 17.4800 \\ & * petitjeant - 0.6645 * fr\_aniline \end{aligned}$$

$$\begin{aligned} R^2 =\ & 0.9103,\ Q^2_{LOO} = 0.8629,\ Q^2 - F_1 \\ & : 0.5776,\ Q^2 - F_2 : 0.5590, \\ & Q^2 - F_3 : 0.6900, \\ & CCC_{ext} : 0.8282 \end{aligned}$$

QSAR Model 2

$$\begin{aligned} pMIC_{90} =\ & 94.6383 + 3.5063 * bcutv8 \\ & - 0.0115 * LogP2 - 30.0538 \\ & * bcute1 + 7.6727 * bcutp1 \\ & - 15.2670 * petitjeant \\ & - 0.6416 * fr\_aniline \end{aligned}$$

$$\begin{aligned} R^2 =\ & 0.8747,\ Q^2_{LOO} = 0.8205,\ Q^2 - F_1 \\ & : 0.5803,\ Q^2 - F_2 : 0.5618, \\ & Q^2 - F_3 : 0.6920, \\ & CCC_{ext} : 0.8106 \end{aligned}$$

QSAR Model 3

$$\begin{aligned} pMIC_{90} =\ & 97.3640 + 0.1292 * SHBint8 \\ & + 2.3249 * bcutv8 \\ & - 30.4088 * bcute1 \\ & + 8.1841 * bcutp1 \\ & - 15.8477 * petitjeant \\ & - 0.6230 * fr\_aniline \end{aligned}$$

$$\begin{aligned} R^2 =\ & 0.8680,\ Q^2_{LOO} = 0.8274,\ Q^2 - F_1 \\ & : 0.5804,\ Q^2 - F_2 : 0.5619,\ Q^2 - F_3 : 0.6921, \\ & CCC_{ext} : 0.7107 \end{aligned}$$

QSAR Model 4

$$\begin{aligned} pMIC_{90} =\ & 7.1879 + 1.1936 * MATS8s \\ & 3.5538 * bcutv8 - 0.2939 \\ & * Chi4 - 16.1015 * petitjeant \\ & - 0.4963 * fr\_aniline \end{aligned}$$

$$\begin{aligned} R^2 :\ & 0.8303,\ Q^2_{LOO} : 0.7642,\ Q^2 - F_1 \\ & : 0.7190,\ Q^2 - F_2 : 0.7067,\ Q^2 - F_3 \\ & : 0.7938,\ CCC_{ext} : 0.8720 \end{aligned}$$

3.2 Design, molecular docking and ADMET results

Based on the QSAR model, we designed some structural analogs of the tetrahydronaphthalene compounds, and its antitubercular activity was predicted. The *Mycobacterium smegmatis* ATP synthase F1 state 1 (PDB ID:7NK7) was chosen as the drug target, which is determined by the electron microscopy method

Table 3 Calculated and predicted antitubercular activity by model equation and LOO

S.No.	Code	Exp. endpoint	Pred. by model eq	Pred.Mod. Eq Res	Pred. LOO	Pred. LOO Res
1.	THN-1	6.1549	6.1091	− 0.0458	6.1043	− 0.0506
2.	THN-2	6.0458	6.1333	0.0875	6.1447	0.099
3.	THN-3	5.7212	5.6667	− 0.0545	–	–
4.	THN-4	5.4089	5.8036	0.3947	–	–
5.	THN-5	5.8125	5.9075	0.095	5.9224	0.1099
6.	THN-6	6.0269	5.7034	− 0.3234	5.6733	− 0.3536
7.	THN-7	5.7212	5.7354	0.0142	5.7368	0.0156
8.	THN-8	6.0044	5.8261	− 0.1783	–	–
9.	THN-9	5.8416	5.7658	− 0.0758	5.7525	− 0.0892
10.	THN-10	5.4437	5.4471	0.0034	5.4476	0.0039
11.	THN-11	5.71	5.9024	0.1925	5.9369	0.2269
12.	THN-12	5.1367	5.4535	0.3168	5.4915	0.3549
13.	THN-13	Excluded				
14.	THN-14	Excluded				
15.	THN-15	5.5467	5.2989	− 0.2478	5.2456	− 0.3011
16.	THN-16	5.719	5.6229	− 0.0961	5.6048	− 0.1141
17.	THN-17	5.6615	5.6545	− 0.007	–	–
18.	THN-18	6.0706	6.1091	0.0385	–	–
19.	THN-19	5.9066	5.9071	0.0005	5.9071	0.0005
20.	THN-20	5.7645	5.8604	0.0959	5.8697	0.1052
21.	THN-21	5.7077	5.8602	0.1525	5.8936	0.1858
22.	THN-22	5.7595	5.8026	0.0431	5.8074	0.048
23.	THN-23	6.3188	6.2812	− 0.0376	6.2735	−0.0453
24.	THN-24	6.1487	6.261	0.1122	6.2772	0.1285
25.	THN-25	5.4237	5.6746	0.2509	–	–
26.	THN-26	6.3372	6.2	− 0.1373	6.184	− 0.1532
27.	THN-27	6.6778	6.2596	− 0.4181	6.1984	− 0.4793
28.	THN-28	6.0223	5.6841	− 0.3381	5.6523	− 0.37
29.	THN-29	5.1427	5.5862	0.4435	5.6363	0.4937
30.	THN-30	5.4191	5.4753	0.0562	5.4827	0.0637
31.	THN-31	5.4078	5.4436	0.0358	–	–
32.	THN-32	5.1308	5.2398	0.1091	5.2512	0.1205
33.	THN-33	5.1367	5.2107	0.0741	–	–
34.	THN-34	5.4214	5.307	− 0.1143	5.2851	− 0.1363
35.	THN-35	5.1487	5.2162	0.0675	5.2229	0.0741
36.	THN-36	5.1549	5.3746	0.2197	5.4158	0.2609
37.	THN-37	5.1367	5.1372	0.0005	5.1373	0.0006
38.	THN-38	5.0809	4.9796	− 0.1013	4.9621	− 0.1188
39.	THN-39	5.4547	5.4032	− 0.0515	5.3887	− 0.066
40.	THN-40	4.8539	4.7545	− 0.0994	4.7286	− 0.1253
41.	THN-41	5.4584	5.5108	0.0524	–	–
42.	THN-42	4.8697	4.6874	− 0.1822	–	–
43.	THN-43	5.7167	5.7594	0.0427	5.7695	0.0528
44.	THN-44	5.7352	5.4611	− 0.2741	5.3474	− 0.3877
45.	THN-46	5.433	5.4085	− 0.0245	5.4036	− 0.0294
46.	THN-47	5.4283	5.1707	− 0.2576	–	–
47.	THN-48	6.1024	6.3042	0.2018	6.3377	0.2354
48.	THN-49	6.0177	5.8456	− 0.1721	–	–
49.	THN-50	6	5.8997	− 0.1003	5.8275	− 0.1725

Table 3 (continued)

S.No.	Code	Exp. endpoint	Pred. by model eq	Pred.Mod. Eq Res	Pred. LOO	Pred. LOO Res
50.	THN-51	6	6.2941	0.2941	–	–
51.	THN-52	6.2147	6.2198	0.0051	6.2203	0.0057
52.	THN-53	5.1805	5.313	0.1325	5.3649	0.1844
53.	THN-54	5.4202	5.3236	− 0.0967	5.3057	− 0.1145
54.	THN-55	5.1805	5.3702	0.1898	5.4581	0.2776
55.	THN-56	Excluded				
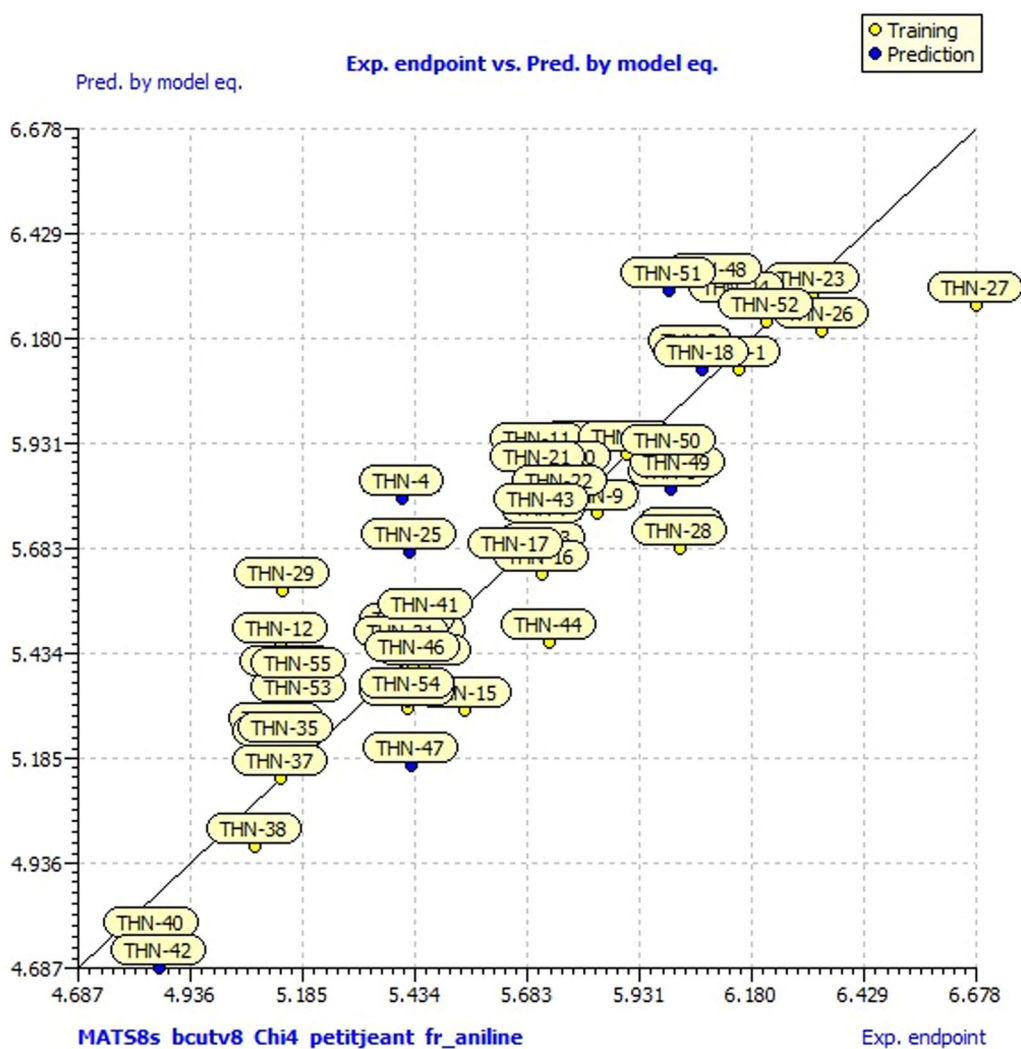
Fig. 1 Scatter plot of experimental pMIC$_{90}$ versus predicted value by model equation
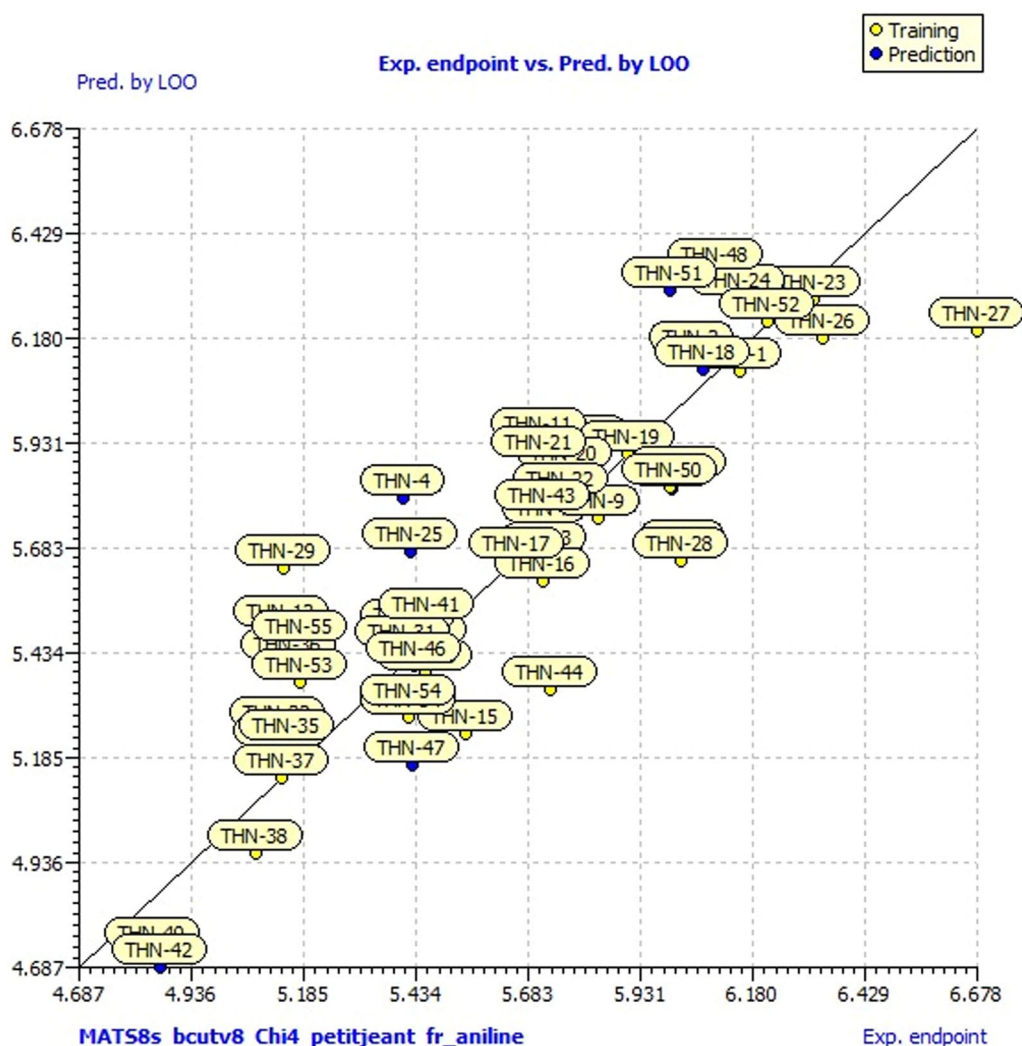
Fig. 2 Scatter plot of experimental pMIC$_{90}$ versus predicted by leave one out (LOO)

with a resolution of 2.11 Å. The quality of the protein was determined by Ramachandran plot using the Prockeck webserver. Figure 7 depicts the Ramachandran plot of the ATP synthase protein. Based on the developed QSAR model, we designed 10 compounds enlisted in Table 4. All the newly designed compounds showed good docking scores ranging from −7.5 kcal/mol to −9.1 kcal/mol. The standard drug Bedaquiline had a docking score of −7.5 kcal/mol toward the protein. The THN_RD9 had docking score of −9.1 kcal/mole also had various Hydrogen bond interactions with ARG A:174, LYS A:175, LYS A:178 and GLN A:435, Pi-Cation interaction with LYS A:276 and Pi-Alkyl interaction with ALA A:180, VAL A:183, ALA

A:222 and VAL A:219 amino acids of ATP synthase protein. Figure 8 illustrates the 3D and 2D interaction of THN_RD9 with ATP Synthase protein. All the designed compounds don't have any Lipinski violation and toxicity (hepatotoxicity and carcinogenicity) (Tables 5, 6, 7).

3.3 Molecular dynamic simulations results

In order to understand the conformational changes and evaluate the binding of THN_RD9 against Mtb ATP synthase (PDB ID: 7NK7), we have carried out MD simulation for a period of 100 ns for THN_RD9-Mtb ATP synthase complex. The simulation was evaluated using various statistical parameters including
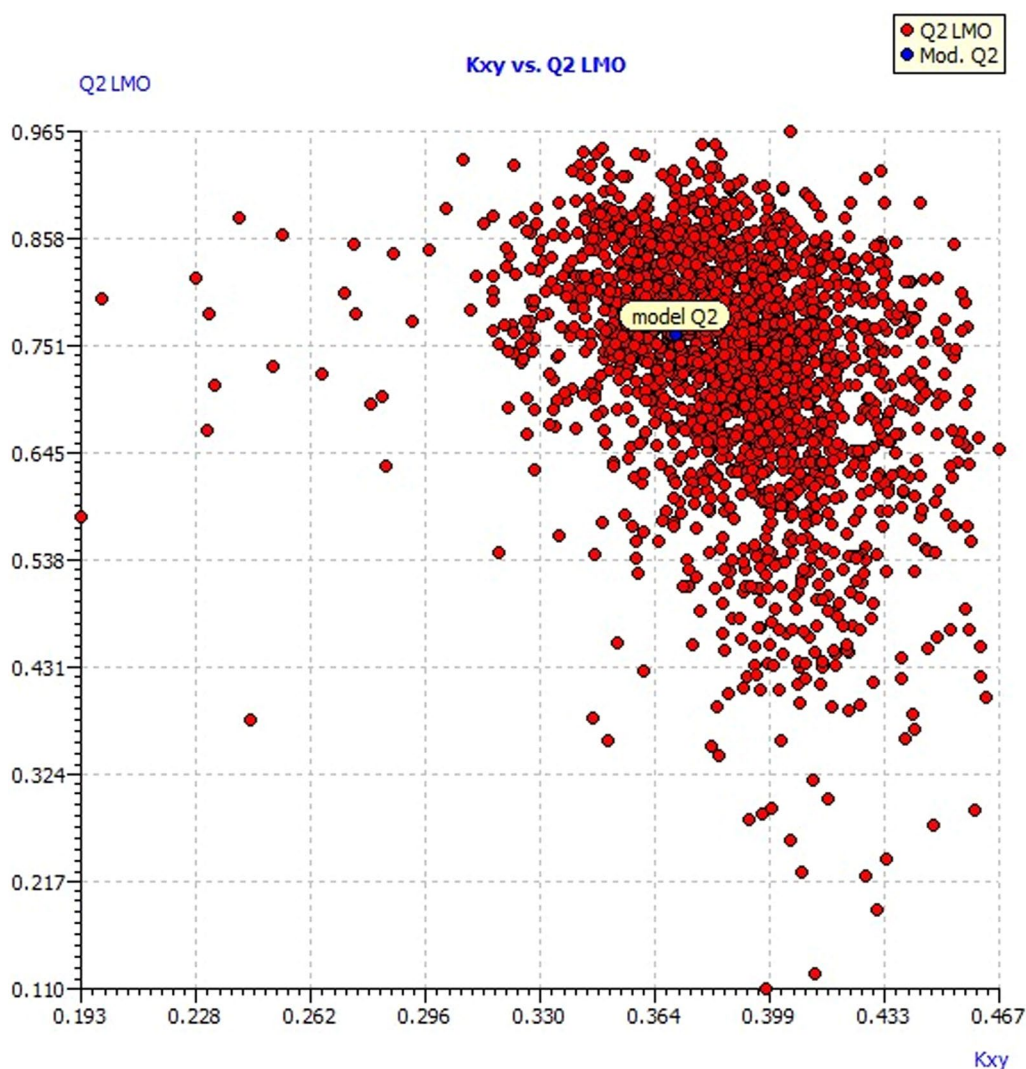
Fig. 3 Scatter plot of Leave many out (LMO) validation compared with the original model

root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), h-bond interactions, and its % occupancies over the time. Figure 9 illustrates the protein–ligand complex of Mtb ATP synthase protein with THN_RD9 for molecular dynamics simulation.

3.3.1 RMSD analysis
Analyzing the RMSD can give insights into any structural conformation that protein and ligand undergo during the simulation. The multiplot for protein Cα versus time for the simulation is shown in Fig. 10. Both the ligand and protein in the complex form have

attained a plateau in RMSD values (of less than 0.3 nm), which indicated that the ligand–protein complex was stable during the simulation.

3.3.2 RMSF analysis
The protein-RMSF is useful for characterizing local changes along the protein chain. The plot for protein-RMSF (nm) versus residue number index is shown in Fig. 11. Notably, the plot describes a fluctuation of less than 0.3 nm for the protein. Particularly the residues, which were involved in binding with the ligand, have
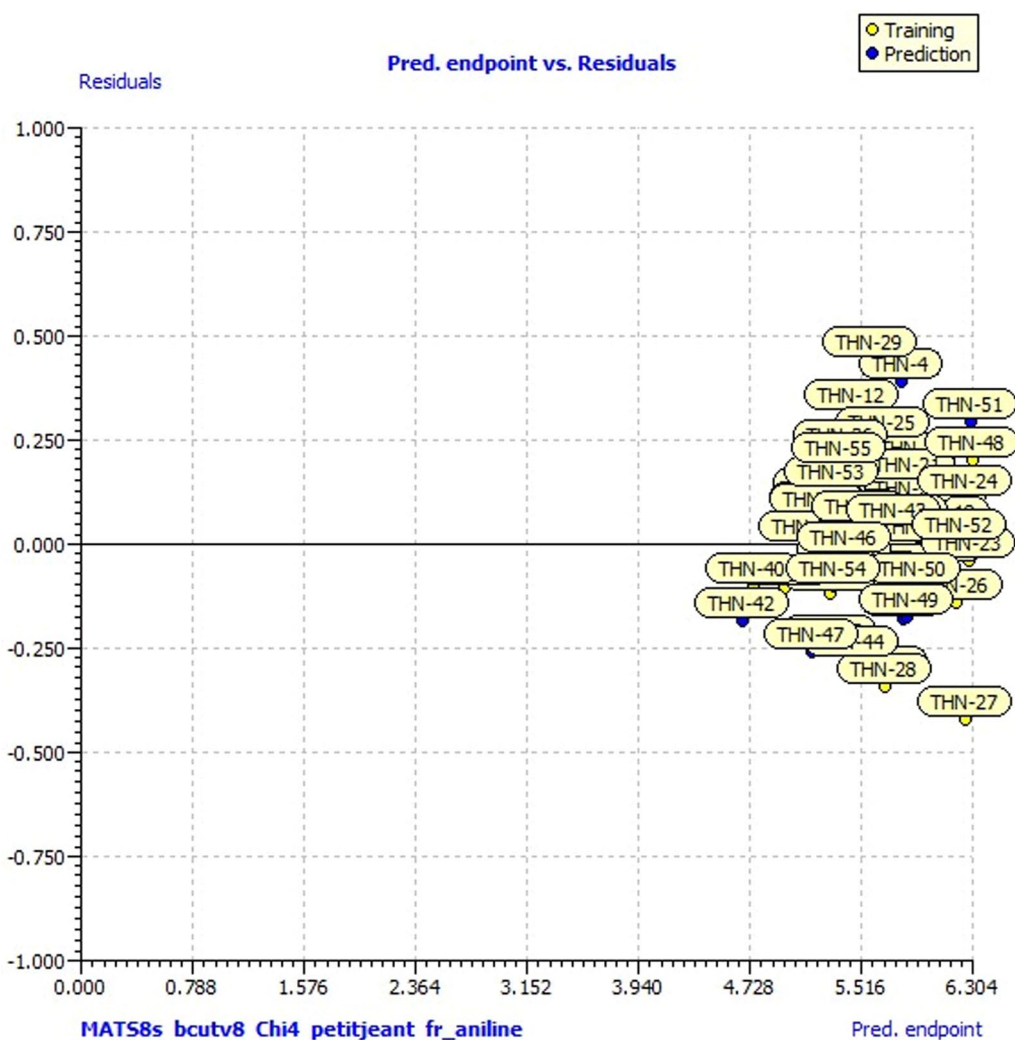
Fig. 4 Scatter plot of residual predicted by leave one out (LOO)

shown very small fluctuations. It also indicated the stability of ligand–protein in complex form.

3.3.3 H-bond interaction

Molecular interactions, particularly the h-bond interactions, are distance and angle depend and liable to disrupt under dynamic conditions. Herein, we have analyzed the ligand–protein interactions. The plot for the number of hydrogen vs time is shown in Fig. 12. From the plot, it was observed that THN_RD9 displayed 1–2 h-bond contact in binding Mtb ATP synthase. To access the residues involved in such interactions and their stabilities, the %occupancies vs the residues were also calculated.

Figure 13 represents the histogram of % occupancies of the h-bond contacts formed by THN_RD9 in binding Mtb ATP synthase. This graph has displayed the ability of THN_RD9 to form stable h-bond contacts with residues VAL139, LYS212, LYS175 and THR215, with occupancies of 16.16, 6.86, 1.59 and 1.04%, respectively. So overall, it can be concluded that the THN_RD9 can be efficient in binding with Mtb ATP synthase.
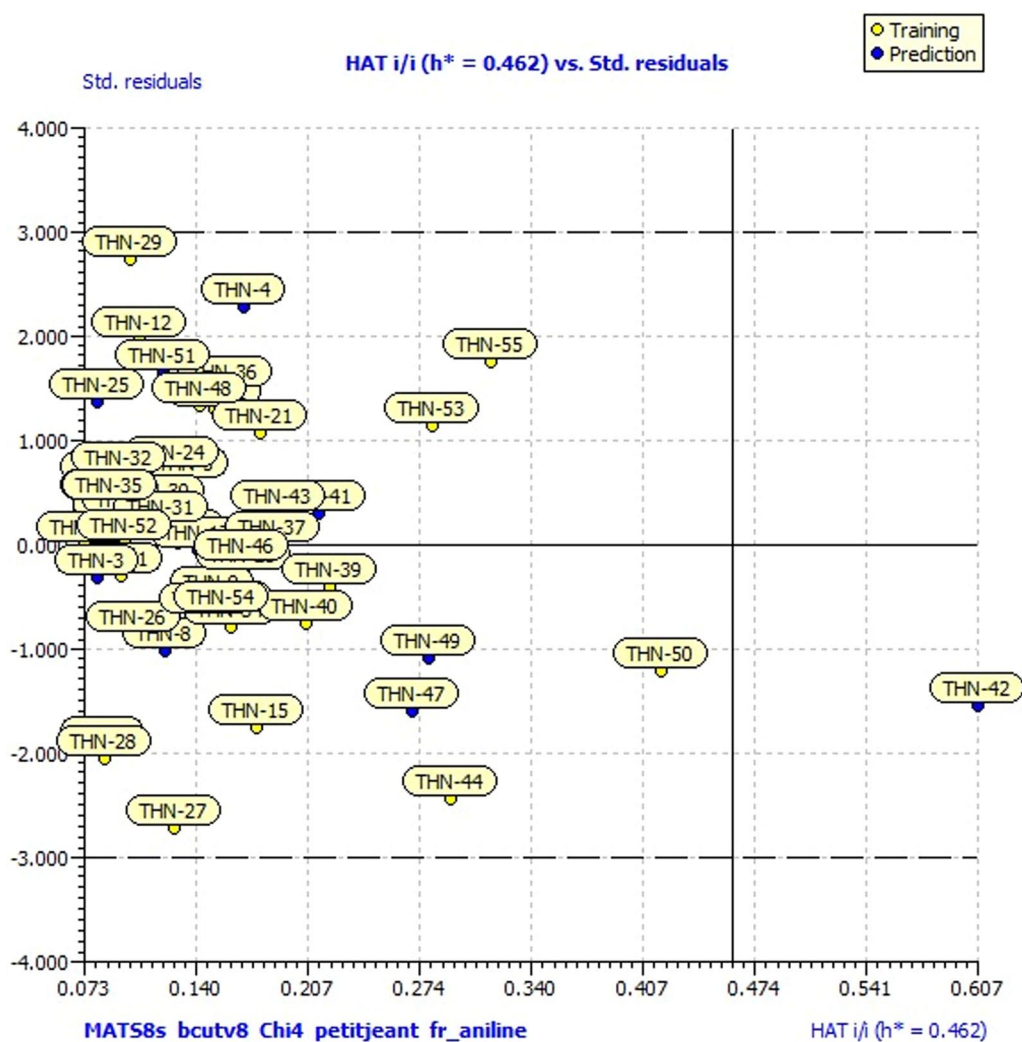
Fig. 5 Williams plot of Hat diagonal leverage value versus standardized residuals

3.4 MM/GBSA calculation

Further, to understand the energy contributions of individual residues inside the active site, the per-residue interaction energies were calculated for the Mtb ATP synthase in complex with THN_RD9. The major contributing residues include ASP241, ASP272 and GLU331 (Fig. 14) The binding free energy of the protein–ligand complex was calculated using MM/GBSA calculation. In this result, the binding free energy of the initial time (0 ns) of the protein-THN RD9 complex was −31.57 kcal/mol, at 50 ns the binding free energy was -39.85 kcal/mol; at the final stage(100 ns), the binding free energy was -34.29 kcal/mol.

4 Discussion

We created four QSAR models that exhibited strong model fitting and adhered to the OECD guidelines for producing reliable predictive QSAR models. However, we selected model 4 as the best model based on the external validation parameters. Hence, we are focusing on the validation, significance and predictive ability of the developed QSAR model 4. This model exhibits strong internal predictive capability, as indicated by a Q^2 value of 0.7642. The goodness of fit of the developed model was the correlation coefficient ($R^2 = 0.8303$). The developed QSAR model's predictive ability was evaluated by an external correlation coefficient ($R^2_{adj} = 0.7751$). The lower R^2_{Yscr}
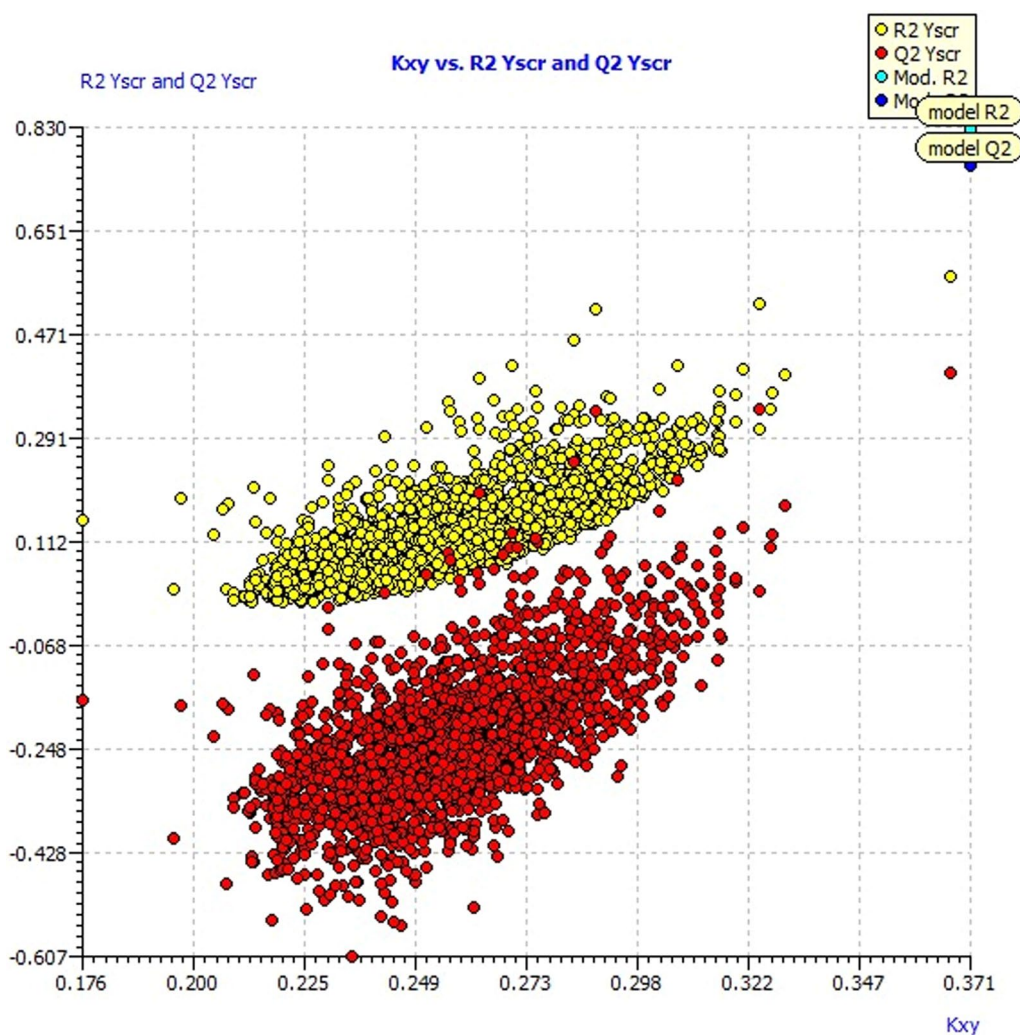
Fig. 6 Scatter plot of Y-scrambled model compared with the original model

values indicate that the proposed QSAR models are not obtained by the chance correlation. The importance of the quantity of molecular descriptors in the developed QSAR model was validated by the minimal difference of 0.0257 (less than 0.3) observed between R^2 and R^2_{adj}. The stability of the developed QSAR model was confirmed by the lowest difference of 0.0661(less than 0.3) between the R^2 and Q^2_{LOO}. In this method, we excluded one compound and computed the model with the remaining compounds and then we predicted the activity of the excluded one. The next stronger technique included in the QSARINS is the leave-many-out (LMO) method, in

which 30% of the compounds were left out to develop the predictive model and to study the behavior of the model.

All the statistical parameters of the four models are given in Table 1, in which the fourth model's coefficient of determination, denoted as R^2, is 0.8303, which helps to evaluate the goodness of fit. A good model should have an R^2 value greater than 0.6. In the fourth model, the coefficient of determination denoted as R^2 is 0.8303, aiding in the assessment of goodness of fit. A robust model should exhibit an R^2 value greater than 0.6. R^2_{adj} represents the adjusted coefficient of determination; it gives insight into the suitability by addition of new molecular descriptors
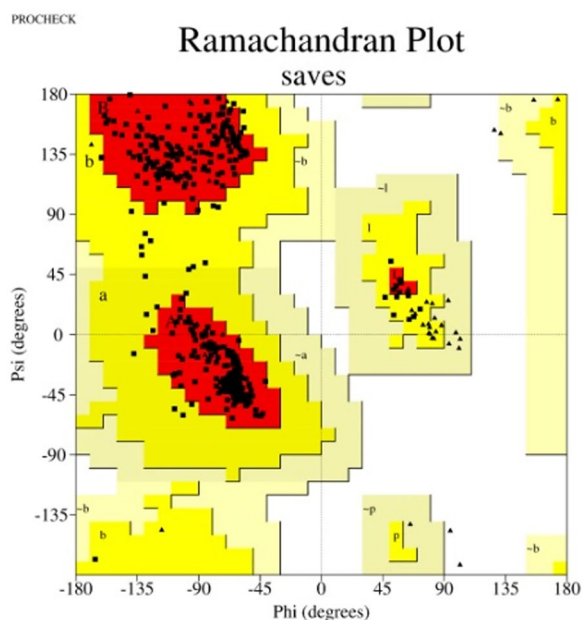
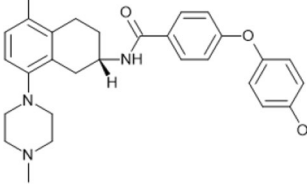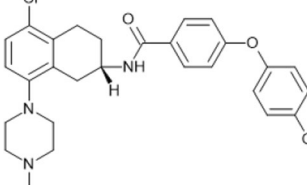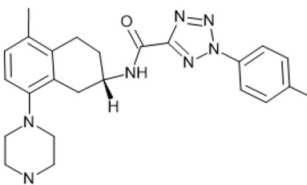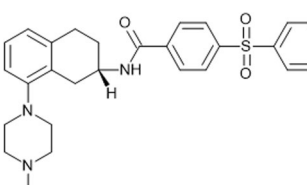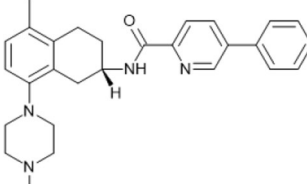**Fig. 7** Ramachandran plot of the ATP synthase
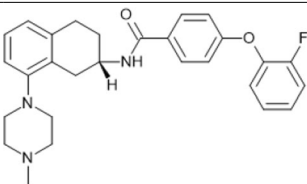
to the QSAR model. For a model to be considered good, R^2_{adj} should exceed 0.6. LOF refers to Friedman's lack of fit criteria, which assesses overfitting in the QSAR model. For a model to be deemed good, LOF should be below 0.3. Kxx represents the overall correlation among the descriptors. Delta K indicates the difference in correlation between the descriptor (kx) and the correlation difference between the descriptor and the response (kxy).

$RMSE_{tr}$ stands for root-mean-square error in the training set calculations. For a model to be considered good, its RMSE should be under 0.3. MAE represents the mean absolute error in the adjustment calculated within the training series; for a model to be considered good, it should be below 0.3. S stands for the standard estimation error. A good model should exhibit an S value lower than 0.3. Additionally, both R^2_{Yscr} and Q^2_{Yscr} should be minimized to indicate that the developed model is not obtained by chance correlation. The model 4 fulfills all the required statistical parameters. Q^2-F_1, Q^2-F_2, and Q^2-F_3 values are helpful to quantify the predictive power of the developed model; it should also exceed 0.7 for a

Table 4 The structures and SMILES of newly designed compounds

Code	Compound structure	SMILES
THN_RD1		CN1CCN(CC1)c1ccc(C)c2CC[C@H](NC(=O)c3ccc(cc3)c3ccc(F)cc3C)Cc12
THN_RD2		CN1CCN(CC1)c1ccc(C)c2CC[C@H](NC(=O)c3ccc(cc3)c3ccc(OC)cc3C)Cc12
THN_RD3		CN1CCN(CC1)c1cccc2CC[C@H](NC(=O)c3ccc(Oc4ccc(F)cc4)cc3)Cc12
THN_RD4		CN1CCN(CC1)c1cccc2CC[C@H](NC(=O)c3ccc(Oc4cccc(F)c4)cc3)Cc12

Table 4 (continued)

Code	Compound structure	SMILES
THN_RD5		CN1CCN(CC1)c1cccc2CC[C@H](NC(=O)c3ccc(Oc4ccccc4F)cc3)Cc12
THN_RD6		CN1CCN(CC1)c1ccc(C)c2CC[C@H](NC(=O)c3ccc(Oc4ccc(O)cc4)cc3)Cc12
THN_RD7		CN1CCN(CC1)c1ccc(Cl)c2CC[C@H](NC(=O)c3ccc(Oc4ccc(O)cc4)cc3)Cc12
THN_RD8		CN1CCN(CC1)c1ccc(C)c2CC[C@@H](Cc12)NC(=O)c1nn(nn1)c1ccc(Cl)cc1
THN_RD9		CN1CCN(CC1)c1cccc2CC[C@H](NC(=O)c3ccc(cc3)S(=O)(=O)c3ccccc3)Cc12
THN_RD10		CN1CCN(CC1)c1ccc(C)c2CC[C@H](NC(=O)c3ccc(cn3)c3ccc(F)cc3)Cc12

good predictive model. The CCC represents the concordance correlation coefficient; for a model to be considered good, it must surpass 0.85. This CCC value quantifies the similarity of the predicted and experimental values. While Q^2F_3 and CCC are the most stable and reliable parameters. The value of F denotes Fisher's F value, which should exceed the theoretical threshold. The domain of applicability of the selected model was illustrated in William's plot. It showed that most of the compounds were placed within the theoretical space of the model's applicability domain. One compound (THN-42) with leverage values ($h=0.6130$) is greater than the
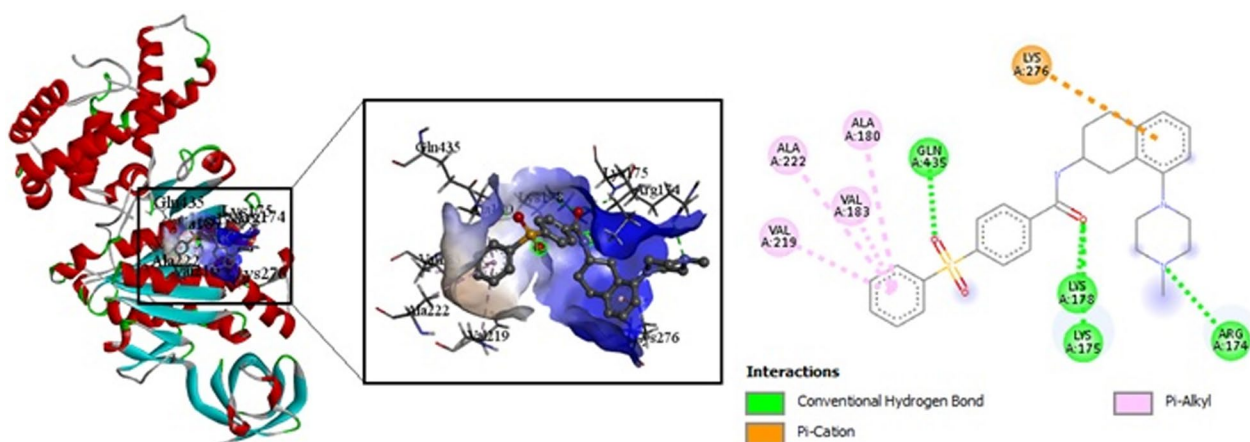
Fig. 8 3D interaction of THN_RD9 with ATP synthase protein

Table 5 ADMET properties of newly designed compounds

S.No.	Compound code	Mol weight	Log P	HBD	HBA	Lipinski rule	Hepatotoxicity	Carcinogenicity
1	THN_RD1	471.61	4.63	1	3	Nil	inactive	inactive
2	THN_RD2	483.64	4.68	1	3	Nil	inactive	inactive
3	THN_RD3	459.56	4.57	1	4	Nil	active	inactive
4	THN_RD4	459.56	4.41	1	4	Nil	inactive	inactive
5	THN_RD5	459.56	4.31	1	4	Nil	inactive	inactive
6	THN_RD6	471.59	4.26	2	4	Nil	active	inactive
7	THN_RD7	492.01	4.24	2	4	Nil	active	inactive
8	THN_RD8	465.98	4.41	1	5	Nil	active	inactive
9	THN_RD9	489.63	4.16	1	4	Nil	inactive	inactive
10	THN_RD10	458.57	3.93	1	4	Nil	inactive	inactive

Table 6 Molecular docking result of the designed compounds

S.No.	Compound code	Docking score (kcal/mol)
1.	THN_RD1	−7.5
2.	THN_RD2	−9.1
3.	THN_RD3	−9.1
4.	THN_RD4	−8.7
5.	THN_RD5	−8.2
6.	THN_RD6	−8.9
7.	THN_RD7	−8.5
8.	THN_RD8	−8.7
9.	THN_RD9	−9.1
10.	THN_RD10	−8.9
11.	Bedaquiline	−7.5

Table 7 Binding free energy of protein–ligand complex using MM/GBSA calculation

Code	0 ns (Kcal/mol)	50 ns (Kcal/mol)	100 ns (Kcal/mol)	Averege (Kcal/mol)
Protein-THN RD9 complex	−31.57	−39.85	−34.29	−35.24

warning leverage value ($h* = 0.462$), However, the plot shows that standardized residual value was within the limits. This compound must be considered as possible outliers; however, its behavior is relatively similar to the remaining compounds, so we did not exclude them from this study.
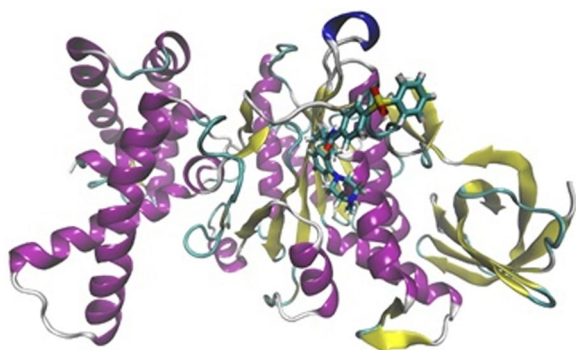
Fig. 9 Protein–ligand complex of THN_RD9-Mtb ATP synthase, where protein is shown in cartoon representation and the ligand is shown in CPK representation
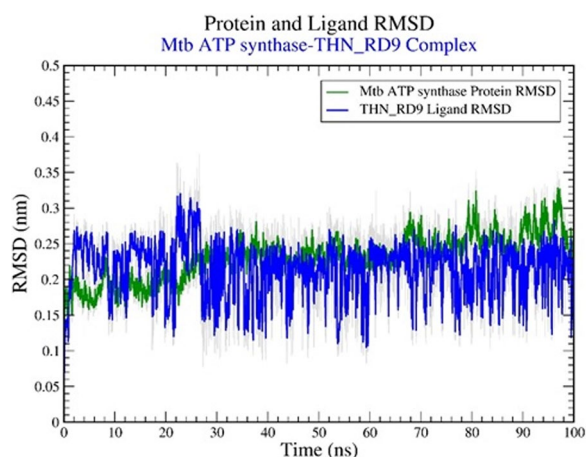

Fig. 10 Graphical representation of the plot showing protein Cα and Ligand RMSD (nm) versus time (100 ns) for (A) Mtb ATP synthase protein (green in color), and (B) THN_RD9 ligand (blue in color)
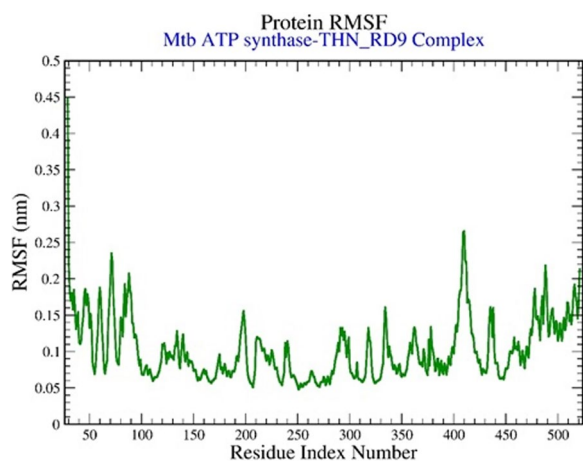

Fig. 11 Graphical representation of the plot showing the protein RMSF (nm) versus residue index number of protein for THN_RD9-Mtb ATP synthase complex
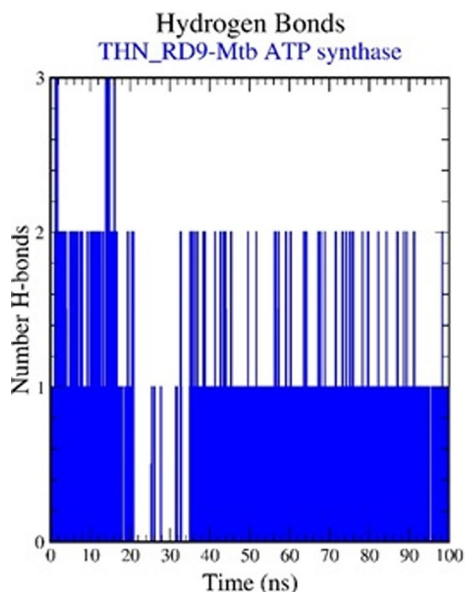
Fig. 12 Pictorial representation of the number of h-bond contacts formed by THN_RD9 in complex with Mtb ATP synthase (PDB ID: 7NK7)
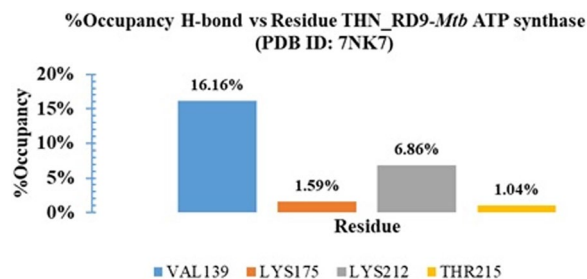

Fig. 13 Histogram representation of %occupancies of the h-bond protein ligand contacts of THN_RD9 in complex with Mtb ATP synthase (PDB ID: 7NK7)
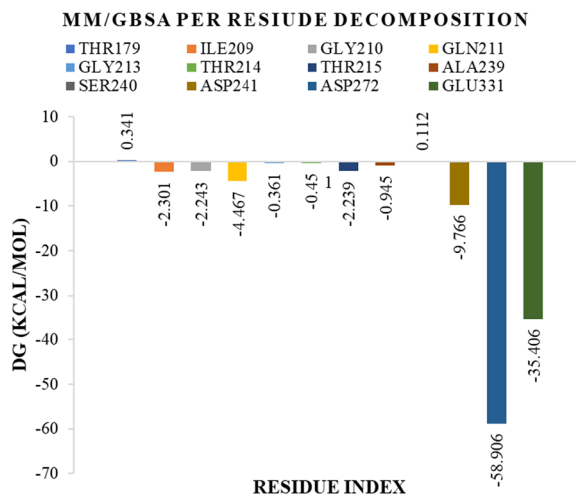

Fig.14 Per-residue energy contribution using molecular mechanics/generalized born and surface area continuum solvation (MM/GBSA)

The Organization for Economic Co-operation and Development (OECD) Community gave five important principles to develop a good QSAR predictive model. OECD principle 1 deals with a defined end point which means it refers to a pharmacological or biological activity that could be measured. OECD principle 2 associated with unambiguous algorithm, it explained about how the molecular descriptor of the compounds there biological activity are related. OECD principle 3 deals with the Applicability Domain. It aids in identifying outliers, which refer to compounds that fall outside the spatial region of the applicability domain. OECD principle 4 deals with robustness, predictive ability and goodness of fit of the developed QSAR model. OECD principle 4 deals with the mechanistic interpretation, which is not mandatory for the model.

In model 4, various molecular descriptors such as MATS8s, bcutv8, Chi4, Petitjeant and fr_aniline influence the antitubercular activity, in which MATS8s and Chi4 were positively correlating and contributing toward the antitubercular activity. Other descriptors like bcutv8, Petitjeant and fr_aniline were negatively contributed to the antitubercular activity, in which bcutv8 was related to the atomic van der Waals volume. The molecular descriptor Chi is the atomic connectivity indices. MATS8s descriptor is the Moran autocorrelation of lag 8 weighted by I-state. Petitjeant is the topological descriptor. Based on the developed QSAR model equation, the newly designed 10 tetrahydronaphthalene compounds were subjected to molecular docking and ADMET analysis. All the designed compounds had good docking scores (greater than -7.1 kcal/mol) and zero Lipinski violation. The THN-RD09 compound has high $pMIC_{90}$, good binding affinity, zero Lipinski violation and no toxicity.

5 Conclusion

In the present study, the QSAR study was carried out on a series of tetrahydronaphthalene derivatives with antitubercular activity. The molecular descriptors were calculated using by the PaDEL Descriptor software and ChemDes webserver. It helps to develop a stable, predictive and robust model with fulfilled OECD principles. The developed QSAR model's robustness was evaluated through various internal (Q^2_{LOO}, Q^2_{LMO} and R^2_{Yscr}) validation parameters, and its predictive ability was assessed using external validation. In the QSAR equation, the molecular descriptors of MATS8s and Chi4 positively correlated with antitubercular activity. While the bcutv8, Petitjeant and fr_aniline were negatively correlated with antitubercular activity. Modification of these molecular descriptor parameters will lead to good antitubercular

activity. The Mtb ATP synthase inhibition was greatly dependent on atomic van der Waals volume, Moran autocorrelation, connectivity indices, topological factor and a fragment of aniline. Based on the developed QSAR model equation, the newly designed 10 tetrahydronaphthalene compounds were subjected to molecular docking and ADMET analysis. This developed QSAR predictive model will help to predict the antitubercular activity of tetrahydronaphthalene derivatives. The molecular dynamics (MD) simulations were performed for 100 ns, using the Gromacs package. It evaluated the conformational stability and alteration of protein–ligand complexes during the simulation. Thus, our findings confirmed that newly designed THN-RD9, a tetrahydronaphthalene derivative and known strong ATP synthase inhibitor, may be used as an important "lead" molecule to be developed as an antitubercular drug in the future. However, we need to do an in vitro and in vivo study to confirm its antitubercular activity.

Abbreviations

QSAR	Quantitative structural activity relationship
ATP	Adenosine triphosphate
ADMET	Absorption, distribution, metabolism, excretion and toxicity
TB	Tuberculosis
MTB	Mycobacterium tuberculosis
MDR	Multidrug-resistant
XDR	Extensively drug-resistant
TDR	Totally drug-resistant
GA	Genetic algorithms
MLR	Multiple linear regressions
PDB	Protein data bank
MD	Molecular dynamics
NPT	Constant number, temperature and pressure

References
1. Badar AD, Sulakhe SM, Muluk MB, Rehman NN, Dixit PP, Choudhari PB, Rekha EM, Sriram D, Haval KP (2020) Synthesis of isoniazid-1, 2, 3-triazole conjugates: antitubercular, antimicrobial evaluation and molecular docking study. J Heterocycl Chem 57(10):3544–3557. https://doi.org/10.1002/jhet.4072
2. Bagcchi S (2023) WHO's global tuberculosis report 2022. The Lancet Microbe 4(1):e20. https://doi.org/10.1016/S2666-5247(22)00359-7
3. Bahuguna A, Rawat DS (2020) An overview of new antitubercular drugs, drug candidates, and their targets. Med Res Rev 40(1):263–292. https://doi.org/10.1002/med.21602
4. Konduri S, Pogaku V, Prashanth J, Siva Krishna V, Sriram D, Basavoju S, Behera JN, Prabhakara RK (2021) Sacubitril-based urea and thiourea derivatives as novel inhibitors for anti-tubercular against dormant tuberculosis. ChemistrySelect 6(16):3869–3874. https://doi.org/10.1002/slct.202004724
5. Sharma K, Tanwar O, Deora GS, Ali S, Alam MM, Zaman MS, Krishna VS, Sriram D, Akhter M (2019) Expansion of a novel lead targeting M. Tuberculosis DHFR as antitubercular agents. Bioorganic Med Chem 27(7):1421–1429. https://doi.org/10.1016/j.bmc.2019.02.053
6. Dogamanti A, Chiranjeevi P, Aamate VK, Vagolu SK, Sriram D, Balasubramanian S, Sarasija M (2021) Indole-fused spirochromenes as potential

anti-tubercular agents: design, synthesis and in vitro evaluation. Mol Diversity 25:2137–2148. https://doi.org/10.1007/s11030-020-10108-z

7. Macalino SJY, Billones JB, Organo VG, Carrillo MCO. In Silico Strategies in Tuberculosis Drug Discovery. Molecules. Multidisciplinary Digital Publishing Institute; 2020; 25:665. https://doi.org/10.3390/molecules25030665

8. Gramatica P, Sangion A (2016) A historical excursus on the statistical validation parameters for QSAR Models: a clarification concerning metrics and terminology. J Chem Inf Model 56:1127–1131. https://doi.org/10.1021/acs.jcim.6b00088

9. Roy K, Narayan DR (2014) A review on principles, theory and practices of 2D-QSAR. Curr Drug Metab 15:346–379. https://doi.org/10.2174/13892 0015666140908102230

10. Sutherland HS, Lu GL, Tong AS, Conole D, Franzblau SG, Upton AM, Denny WA (2022) Synthesis and structure-activity relationships for a new class of tetrahydronaphthalene amide inhibitors of mycobacterium tuberculosis. Eur J Med Chem 229:114059. https://doi.org/10.1016/j.ejmech.2021.114059

11. Wang B, Peng F, Huang W, Zhou J, Zhang N, Sheng J, Han B (2020) Rational drug design, synthesis, and biological evaluation of novel chiral tetrahydronaphthalene-fused spirooxindole as MDM2-CDK4 dual inhibitor against glioblastoma. Acta Pharmaceutica Sinica B 10(8):1492–1510. https://doi.org/10.1016/j.ejmech.2021.114059

12. Hamza EK, Hamdy NA, Zarie ES, Fakhr IM, Elwahy AH, Awada HM (2019) Synthesis and in vitro anticancer evaluation of novel pyridine derivatives bearing tetrahydronaphthalene scaffold. Org Chem. https://doi.org/10.24820/ark.5550190.p011.056

13. Makar S, Saha T, Singh SK (2019) Naphthalene, a versatile platform in medicinal chemistry: sky-high perspective. Eur J Med Chem 161:252–276. https://doi.org/10.1016/j.ejmech.2018.10.018

14. Elrayess AA, Elshihawy H (2023) Naphthalene: an overview. Rec Pharm Biomed Sci 7(1):145–153. https://doi.org/10.21608/rpbs.2023.212630.1228

15. Gümüş M, Yakan M, Koca İ (2019) Recent advances of thiazole hybrids in biological applications. Future Med Chem 11(16):1979–1998. https://doi.org/10.4155/fmc-2018-0196

16. Taniya OS, Kopchuk DS, Khasanov AF, Kovalev IS, Santra S, Rahman M, Chupakhin ON (2019) 2-Azaanthracenes: a chronology of synthetic approaches and bright prospects for practical applications. New J Chem 43(28):11382–11390. https://doi.org/10.1039/C9NJ01813A

17. Cheng HG, Jia S, Zhou Q (2023) Benzo-fused-ring toolbox based on palladium/norbornene cooperative catalysis: methodology development and applications in natural product synthesis. Acc Chem Res 56(5):573–591. https://doi.org/10.1021/acs.accounts.2c00781

18. Hajalsiddig TTH, Osman ABM, Saeed AEM (2020) 2D-QSAR modeling and molecular docking studies on 1HPyrazole-1-carbothioamide derivatives as EGFR kinase inhibitors. ACS Omega Am Chem Soc 5:18662–18674. https://doi.org/10.1021/acsomega.0c01323

19. Er-rajy M, El Fadili M, Hadni H, Mrabti NN, Zarougui S, Elhallaoui M (2022) 2D-QSAR modeling, druglikeness studies, ADMET prediction, and molecular docking for anti-lung cancer activity of 3-substituted5-(phenylamino) indolone derivatives. Struct Chem 33:973–986. https://doi.org/10.1007/s11224-022-01913-3

20. Ravichandran V, Shalini S, Sokkalingam AD, Harish R, Suresh K (2014) QSAR study of 7-chloroquinoline derivatives as antitubercular agents. World J Pharm Pharm Sci 3:1072–1082

21. Mulatsari E, Mumpuni E, Nurhidayati L, Purwanggana A, Pratami DK (2021) Pelatihan Visualisasi Molekul Kimia Dengan Software Chemsketch Untuk Siswa Tingkat Sekolah Menengah Atas. Magistrorum et Scholarium: Jurnal Pengabdian Masyarakat, 2(1), 102–112. https://doi.org/10.24246/jms.v2i12021p102-112

22. Sathishkumar GK, Ibrahim M, Mohamed Akheel M, Rajkumar G, Gopinath B, Karpagam R, Gowri Shankar G (2022) Synthesis and mechanical properties of natural fiber reinforced epoxy/polyester/polypropylene composites: a review. J Nat Fibers 19(10):3718–3741. https://doi.org/10.1080/15440478.2020.1848723

23. Zhao M, Wang L, Zheng L, Zhang M, Qiu C, Zhang Y, Niu B (2017) 2D-QSAR and 3D-QSAR analyses for EGFR inhibitors. BioMed Res Int. https://doi.org/10.1155/2017/4649191

24. Kasmi R, Hadaji E, Chedadi O, El Aissouq A, Bouachrine M, Ouammou A (2020) 2D-QSAR and docking study of a series of coumarin derivatives as inhibitors of CDK (anticancer activity) with an application of the

molecular docking method. Heliyon 6(8):e04514. https://doi.org/10.1016/j.heliyon.2020.e04514

25. Yu X (2021) Prediction of inhibitory constants of compounds against SARS-CoV 3CLpro enzyme with 2D-QSAR model. J Saudi Chem Soc 25(7):101262. https://doi.org/10.1016/j.jscs.2021.101262

26. Gramatica P (2020) Principles of QSAR modeling: comments and suggestions from personal experience. Int J Quant Struct Prop Relationsh (IJQSPR) 5(3):61–97. https://doi.org/10.4018/IJQSPR.20200701.oa1

27. Srimathi R, Kathiravan MK (2021) Lead optimization of 4-(thio)-chromenone 6-O-sulfamate analogs using QSAR, molecular docking and DFT–a combined approach as steroidal sulfatase inhibitors. J Recept Signal Transduct Res 41(2):123–137. https://doi.org/10.1080/10799893.2020.1794004

28. Hassan GS, Georgey HH, Mohammed EZ, George RF, Mahmoud WR, Omar FA (2021) Mechanistic selectivity investigation and 2D-QSAR study of some new antiproliferative pyrazoles and pyrazolopyridines as potential CDK2 inhibitors. Eur J Med Chem 218:113389. https://doi.org/10.1016/j.ejmech.2021.113389

29. Medapi B, Renuka J, Saxena S, Sridevi JP, Medishetti R, Kulkarni P, Sriram D (2015) Design and synthesis of novel quinoline–aminopiperidine hybrid analogues as mycobacterium tuberculosis DNA gyraseB inhibitors. Bioorg Med Chem 23(9):2062–2078. https://doi.org/10.1016/j.ejmech.2021.113389

30. Nour H, Abchir O, Belaidi S, Qais FA, Chtita S, Belaaouad S (2022) 2D-QSAR and molecular docking studies of carbamate derivatives to discover novel potent anti-butyrylcholinesterase agents for Alzheimer's disease treatment. Bull Korean Chem Soc 43(2):277–292. https://doi.org/10.1002/bkcs.12449

31. Chatterjee M, Roy K (2022) Application of cross-validation strategies to avoid overestimation of performance of 2D-QSAR models for the prediction of aquatic toxicity of chemical mixtures. SAR QSAR Environ Res 33(6):463–484. https://doi.org/10.1080/1062936X.2022.2081255

32. Mukherjee RK, Kumar V, Roy K (2022) Chemometric modeling of plant protection products (PPPs) for the prediction of acute contact toxicity against honey bees (A. mellifera): a 2D-QSAR approach. J Hazard Mater 423:127230. https://doi.org/10.1016/j.jhazmat.2021.127230

33. Du M, Zhang D, Hou Y, Zhao X, Li Y (2019) Combined 2D-QSAR, principal component analysis and sensitivity analysis studies on fluoroquinolones' genotoxicity. Int J Environ Res Public Health 16(21):4156. https://doi.org/10.3390/ijerph16214156

34. Younis MH, Mohammed ER, Mohamed AR, Abdel-Aziz MM, Georgey HH, Gawad NMA (2022) Design, synthesis and anti-Mycobacterium tuberculosis evaluation of new thiazolidin-4-one and thiazolo [3, 2-a][1, 3, 5] triazine derivatives. Bioorg Chem 124:105807. https://doi.org/10.1016/j.bioorg.2022.105807

35. Aziz MA, Shehab WS, Al-Karmalawy AA, El-Farargy AF, Abdellattif MH (2021) Design, synthesis, biological evaluation, 2D-QSAR modeling, and molecular docking studies of novel 1 H-3-indolyl derivatives as significant antioxidants. Int J Mol Sci 22(19):10396. https://doi.org/10.3390/ijms221910396

36. Kumar V, Ojha PK, Saha A, Roy K (2020) Exploring 2D-QSAR for prediction of beta-secretase 1 (BACE1) inhibitory activity against Alzheimer's disease. SAR QSAR Environ Res 31(2):87–133. https://doi.org/10.1080/1062936X.2019.1695226

37. Rajasekhar S, Das S, Karuppasamy R, Musuvathi Motilal B, Chanda K (2022) Identification of novel inhibitors for Prp protein of Mycobacterium tuberculosis by structure based drug design, and molecular dynamics simulations. J Comput Chem 43(9):619–630. https://doi.org/10.1002/jcc.26823

38. Shanthakumar B, Kathiravan MK (2020) Insights into structures of imidazo oxazines as potent polyketide synthase XIII inhibitors using molecular modeling techniques. J Recept Signal Trans Res 40(4):313–323. https://doi.org/10.1080/10799893.2020.1742740

39. Garg A, Tadesse A, Eswaramoorthy R (2021) A four-component domino reaction: an eco-compatible and highly efficient construction of 1, 8-naphthyridine derivatives, their in silico molecular docking, drug likeness, ADME, and toxicity studies. J Chem 2021:1–16. https://doi.org/10.1155/2021/5589837

40. Chen X, Li H, Tian L, Li Q, Luo J, Zhang Y (2020) Analysis of the physicochemical properties of acaricides based on Lipinski's rule of five. J Comput Biol 27(9):1397–1406. https://doi.org/10.1089/cmb.2019.0323

41. Montgomery MG, Petri J, Spikes TE, Walker JE (2021) Structure of the ATP synthase from Mycobacterium smegmatis provides targets for treating tuberculosis. Proc Natl Acad Sci 118(47):e2111899118. https://doi.org/10.1073/pnas.2111899118

42. Kelam LM, Wani MA, Dhaked DK (2023) An update on ATP synthase inhibitors: a unique target for drug development in M. tuberculosis. Progr Biophys Mol Biol. https://doi.org/10.1016/j.pbiomolbio.2023.04.009

43. Taghizadeh MS, Niazi A, Moghadam A, Afsharifar A (2022) Experimental, molecular docking and molecular dynamic studies of natural products targeting overexpressed receptors in breast cancer. PLoS ONE 17(5):e0267961. https://doi.org/10.1371/journal.pone.0267961

44. Rivera-Quiroga RE, Cardona N, Padilla L, Rivera W, Rocha-Roa C, Diaz De Rienzo MA, Martinez MC (2020) In silico selection and in vitro evaluation of new molecules that inhibit the adhesion of Streptococcus mutans through Antigen I/II. Int J Mol Sci 22(1):377. https://doi.org/10.3390/ijms22010377

45. Babaeekhou L, Ghane M (2021) Antimicrobial activity of ginger on cariogenic bacteria: molecular networking and molecular docking analyses. J Biomol Struct Dyn 39(6):2164–2175. https://doi.org/10.1080/07391102.2020.1745283

46. Gupta A, Vijayan V, Pant P, Kaur P, Singh TP, Sharma P, Sharma S (2022) Structure prediction and discovery of inhibitors against phosphopantothenoyl cysteine synthetase of Acinetobacter baumannii. J Biomol Struct Dyn 40(21):11405–11417. https://doi.org/10.1080/07391102.2021.1958699

47. El-Hachem N, Haibe-Kains B, Khalil A, Kobeissy FH, Nemer G (2017) AutoDock and AutoDockTools for protein-ligand docking: beta-site amyloid precursor protein cleaving enzyme 1 (BACE1) as a case study. Neuroprot Methods Protocols. https://doi.org/10.1007/978-1-4939-6952-4_20

48. Bekker H, Berendsen HJC, Dijkstra EJ, Achterop S, van Drunen R, van der Spoel D, Sijbers A, Keegstra H (1993) Gromacs: a parallel computer for molecular dynamics simulations. Phys Comp 92:252–256

49. Ganesan A, Coote ML, Barakat K (2017) Molecular dynamics-driven drug discovery: leaping forward with confidence. Drug Discovery Today 22(2):249–269. https://doi.org/10.1016/j.drudis.2016.11.001

50. Thalla M, Kant K, Dalchand, Rawat R, Banerjee S (2020) Merged experimental guided computational strategy toward tuberculosis treatment mediated by alveolar macrophages mannose receptor. J Biomol Struct Dyn 38(17):5195–5203. https://doi.org/10.1080/07391102.2019.1697369

51. Schmid N, Eichenberger AP, Choutko A, Riniker S, Winger M, Mark AE, Van Gunsteren WF (2011) Definition and testing of the GROMOS force-field versions 54A7 and 54B7. Eur Biophys J 40:843–856. https://doi.org/10.1007/s00249-011-0700-9

52. Van Aalten DM, Bywater R, Findlay JB, Hendlich M, Hooft RW, Vriend G (1996) PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. J Comput Aided Mol Des 10:255–262. https://doi.org/10.1007/bf00355047

53. Mark P, Nilsson L (2001) Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. J Phys Chem A 105(43):9954–9960. https://doi.org/10.1021/jp003020w

54. Van Gunsteren WF, Berendsen HJ (1988) A leap-frog algorithm for stochastic dynamics. Mol Simul 1(3):173–185. https://doi.org/10.1080/08927028808080941

55. Berendsen HJ, van der Spoel D, van Drunen R (1995) GROMACS: a message-passing parallel molecular dynamics implementation. Comput Phys Commun 91(1–3):43–56. https://doi.org/10.1016/0010-4655(95)00042-E

56. Hess B, Bekker H, Berendsen HJ, Fraaije JG (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18(12):1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12%3C1463::AID-JCC4%3E3.0.CO;2-H

57. Di Pierro M, Elber R, Leimkuhler B (2015) A stochastic algorithm for the isobaric–isothermal ensemble with Ewald summations for all long range forces. J Chem Theory Comput 11(12):5624–5637. https://doi.org/10.1021/acs.jctc.5b00648

58. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14(1):33–38. https://doi.org/10.1016/0263-7855(96)00018-5

59. Rawat R, Kant K, Kumar A, Bhati K, Verma SM (2021) HeroMDAnalysis: an automagical tool for GROMACS-based molecular dynamics simulation analysis. Future Med Chem 13(05):447–456. https://doi.org/10.4155/fmc-2020-0191

60. Devi S, Rangra NK, Rawat R, Alrobaian MM, Alam A, Singh R, Singh A (2021) Anti-atherogenic effect of nepitrin-7-O-Glucoside: a flavonoid isolated from nepeta hindostana via acting on PPAR – α receptor. Steroids 165:108770. https://doi.org/10.1016/j.steroids.2020.108770

Publisher's Note