J Mol Model (2011) 17:1607–1620
DOI 10.1007/s00894-010-0861-y

ORIGINAL PAPER

Anti-tubercular drug designing by structure based screening of combinatorial libraries

Payel Ghosh · Manish C. Bagchi

Received: 17 June 2010 / Accepted: 24 September 2010 / Published online: 16 October 2010

Abstract In the current study, the applicability and scope of descriptor based QSAR models to complement virtual screening using molecular docking approach have been applied to identify potential virtual screening hits targeting DNA gyrase A from *Mycobacterium tuberculosis,* an effective and validated anti-mycobacterial target. Initially QSAR models were developed against *M. fortuitum* and *M. smegmatis* using a series of structurally related fluoroquinolone derivatives as DNA gyrase inhibitors. Both the QSAR models yielded significant cross validated Q(2) values of 0.6715 and 0.6944 and R(2) values of 0.7250 and 0.7420, respectively. The statistically significant models were validated by a test set of 22 compounds with predictive R(2) value of 0.7562 and 0.7087 for *M. fortuitum* and *M. smegmatis* respectively. To aid the creation of novel antituberculosis compounds, combinatorial library was developed on fluoroquinolone template to derive a data set of 5280 compounds whose activity values have been measured by the above models. Highly active compounds predicted from the models were subjected to molecular docking study to investigate the mechanism of drug binding with the DNA gyrase A protein of *M. tuberculosis* and the compounds showing similar type of binding patterns with that of the existing drug molecules, like sparfloxacin, were finally reported. It is seen that hydrophobic characteristics of molecular structure together with few hydrogen bond interactions are playing an essential role in antimicrobial activity for the fluoroquinolone derivatives. A representative set of seven compounds with high predicted MIC values were sorted out in the present study.

Keywords Fluoroquinolone compounds · Genetic algorithm (GA) · Molecular docking · Quantitative structure activity relationship (QSAR) · Virtual screening

P. Ghosh · M. C. Bagchi (✉)
Structural Biology and Bioinformatics Division,
Indian Institute of Chemical Biology (C.S.I.R.),
4 Raja S.C. Mullick Road,
Jadavpur, Kolkata 700032, India
e-mail: mcbagchi@iicb.res.in

Introduction

Tuberculosis is one of the major causes of death worldwide. The number of individuals succumbing with this disease has increased vastly due to the HIV/AIDS pandemic, and as a consequence of the emergence of multidrug resistant (MDR) and extensively drug-resistant (XDR) tuberculosis strain. Thus, there is an urgent need for new drugs that are potent inhibitors of *M. tuberculosis* exhibiting favorable resistance profiles and that are well tolerated by patients. Fluoroquinolones are active against *Mycobacterium tuberculosis*, causative agent of tuberculosis and are the first new antimycobacterial drugs to be available since the discovery of rifampin [1–3]. Fluoroquinolones are part of the drug regimens now recommended for treating rifampin-resistant tuberculosis [4, 5]. Genome studies suggest that DNA gyrase is the sole type II topoisomerase and is likely to be the unique target of quinolones in *Mycobacterium tuberculosis* [6, 7]. The fact that it is essential in all bacteria and absent from eukaryotes, furthermore makes it an ideal drug target. The mechanism of binding for quinolone drugs is extensively studied in *Escherichia coli* system mainly based on the observation like alterations in DNA gyrase that confer quinolone resistance reside in the quinolone-resistant determining region or QRDR (between residues 67 and 106 of GyrA in *E. coli*) by Maxwell and co-workers [8–10] and it

reveals that quinolone inhibition of DNA gyrase in *Escherichia coli* occurs through the formation of a stable ternary complex between DNA gyrase, DNA, and the quinolone molecule that blocks the progression of DNA replication [11, 12]. Though it is known that DNA gyrase is a target of quinolone antibacterial agents, the molecular details of the quinolone–gyrase interaction are not clear in case of *Mycobacterium tuberculosis.*

In the present paper, a series of quinolone derivatives with substitutions at N-1 and C-7, as well as at the 8 positions, is subjected to examine the relationships between structural modifications and activities against *Mycobacterium fortuitum* and *Mycobacterium smegmatis* [13, 14] with the help of quantitative structure-activity relationship (QSAR). The activities of these compounds against *M. fortuitum* and *M. smegmatis* are considered primarily due to the fact that these two mycobacteria are used as standard of *M. tuberculosis* activity. Mainly theoretical descriptors were used for the QSAR model development and biological activity prediction for tuberculostatic drug design similar to our previous investigations [15–18]. Subsequently, a vast number of analogs were generated having common fluoroquinolone template by the application of *in-silico* approach of combinatorial chemistry and an attempt has been made to explore their potency to become a drug like molecule from the standpoint of quantitative structure-activity relationship (QSAR) as well as molecular docking. A virtual library was generated with regard to the specified substituents at different substituting sites. The library constitutes over five thousands analogs and thus, to deal with such a large amount of data and to facilitate the drug discovery process, initially a number of molecules were considered having high activity profiles according to the prediction from QSAR models and then the selected compounds were subjected to molecular docking study to examine their interaction patterns. Finally, seven compounds have been predicted to be potent agents against tuberculosis from the library according to their dock score and interactions with DNA gyrase (subunit A).

Methodology

Biological activity data and descriptor calculation

The actions of the quinolone antibacterials against *Mycobacterium fortuitum* and *Mycobacterium smegmatis* have been studied by Renau et al. by considering the effect of structural modifications at N-1 and C-7 as well as at 8 position of quinolone moiety. The biological activity data in the form of minimum inhibitory concentration (MIC in μg/mL) were determined experimentally [13, 14] against different mycobacterial organisms. The activity of the compounds against *M. fortuitum* was used as a barometer of *M. tuberculosis* activity [13] and these activities have been considered for the construction of QSAR models. Due to the unavailability of anti- *M. tuberculosis* activity values for fluoroquinolone derivatives at this moment, the current study has been performed on close relatives of the concerned species, as evidenced from previous reports in this field of study [15, 16]. The work relies on the evolutionary relatedness of *M. tuberculosis* with *M. smegmatis* and *M. fortuitum*. It may also be noted that the DNA gyrase (A subunit) protein from *M. tuberculosis,* the main target for fluoroquinolone derivatives, exhibits more than 92% sequence similarity with that of *M. smegmatis*, enough to suggest their structural homology. We have considered 110 compounds for model development against *M. fortuitum* while 117 compounds were considered for QSAR analysis against *M. smegmatis*. Of the compounds 75% were considered as the training set for developing the models while the rest of the molecules were used for model validation. The splitting was done by using sphere exclusion method [19].

A large number of theoretical descriptors such as constitutional, physicochemical, electrostatic, topological and semi-empirical type have been computed from the chemical structures of the compounds referred to above with a view to develop structure-activity relationship of fluoroquinolone compounds against *M. fortuitum* and *M. smegmatis*. A total of 1056 descriptors were calculated by using VLife Sciences Molecular Design Suite [20] which was subsequently reduced to 221 in case of *M. fortuitum* and 223 for *M. smegmatis.* The descriptors having the same value or almost same value or highly correlated with other descriptors were removed initially. The reduced set of descriptors was then treated by genetic algorithm for further reduction of non-significant descriptors and finally the optimum models with eight significant descriptors were considered in our QSAR analysis.

Model development by GA-PLS method

Feature selection is a key step in quantitative structure activity relationship (QSAR) analysis to eliminate the problems like chance correlations and multicollinearity. Utilizing every available descriptor may produce a predictive model with a good correlation coefficient, but the models are difficult to interpret and do not stand up to external validation. An integral aspect of model development is to build the model with a small but appropriate set of descriptors in a view to interpret the relationships. This process forms the basis of a technique known as feature selection [21] or variable selection. Among several search algorithms, genetic algorithms (GA) based feature selection procedures is the most popular for building QSAR models

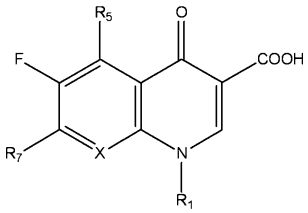


Fig. 1 Fluoroquinolone template indicating four substituting sites

and can explain the situation more effectively [22–24]. Genetic algorithms (GA) described by Holland, is a stochastic optimization technique that mimic natural evolution and selection [25]. The GA begins by generating a set of random solutions (the population), which are analogous to a set of chromosomes in a biological system. The set of variables indicated with a value of 1 in the chromosome is then used as input for model building by partial least square method. PLS was employed as a statistical method for the evaluation of fitness in the GA scheme. PLS has been widely employed to solve multivariate structure-activity relationships in QSAR [26, 27]. The final model obtained is further refined by removing descriptors which do not affect predictive accuracy significantly. Internal validations of the models in all the cases are made in terms of cross-validated Q^2 and external predictability of the developed models are performed by calculating predictive R^2 (R_{pred}^2) using the following equations [28].

$$Q^2 = 1 - \frac{\sum \left(Y_{pred} - Y\right)^2}{\sum \left(Y - \overline{Y}\right)^2} \quad (1)$$

Where Y_{pred} and Y indicate predicted and observed activity values respectively and $\overline{Y}$ indicates mean activity

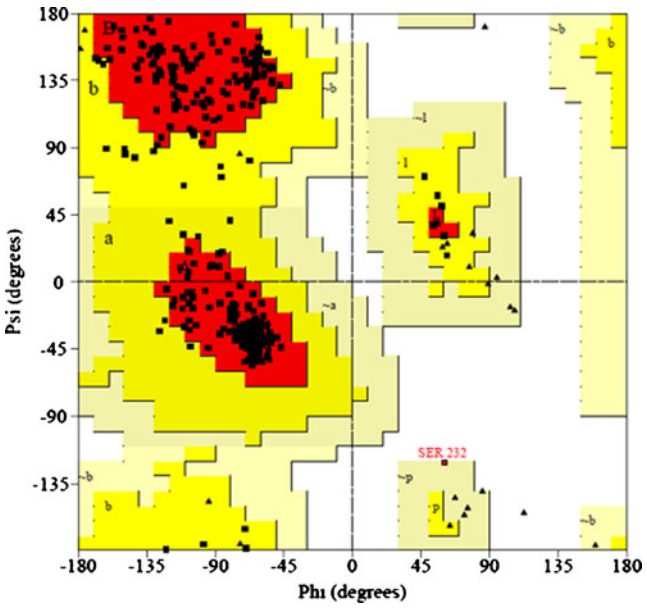


Fig. 2 Phi-Psi plots of the modeled DNA gyrase A for *M. smegmatis* obtained by Procheck. One residue in the generously allowed regions, Ser232, is labeled

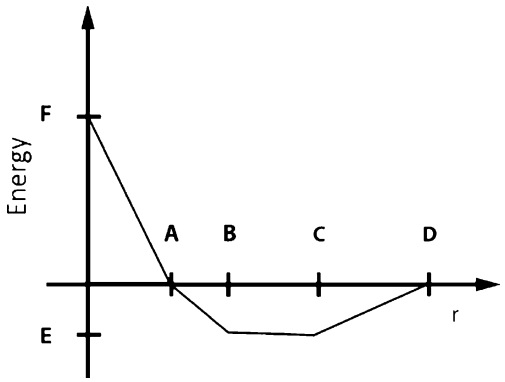


Fig. 3 Graphical representation of the functional form of PLP scoring

value. A model is considered acceptable when the value of Q^2 exceeds 0.5.

$$R^2_{pred} = 1 - \frac{\sum \left(Y_{pred(test)} - Y_{(test)}\right)^2}{\sum \left(Y_{(test)} - \overline{Y}_{training}\right)^2} \quad (2)$$

In Eq. 2, $Y_{pred(test)}$ and $Y_{(test)}$ indicate predicted and observed activity values respectively of the test set compounds and $\overline{Y}_{training}$ indicates mean activity value of the training set. For a predictive QSAR model, the value of R^2_{pred} should be more than 0.5.

Generation of combinatorial library and virtual screening

Computational methods are being increasingly used to assist the combinatorial library design, focusing, and virtual screening by introducing selection criteria such as molecular diversity, drug likeness, receptor binding analysis, and ADME properties of analogs. We have drawn heavily on the LeadGrow Module of MDS software for the generation of such a virtual library. Selection and focusing methods using these descriptors are employed to reduce the size of the combinatorial libraries to be prepared and screened [29–31]. Computational approaches can thus significantly reduce the cost, time, and labor required to synthesize and screen large libraries, as well as enhance the success rate in lead compound generation.

In the area of rational drug design, since QSAR modeling for the series of fluoroquinolone derivatives was developed based on mathematical descriptors which can be calculated rapidly, the synthetic chemists can use these models as a decision support tool in synthesis planning. For example, in the fluoroquinolone template indicating four substituting sites, one can visualize a huge number of groups as substitution at each possible site. One can refer to the work of Hansch and Leo [32] in this regard where they had tabulated a list of 230 substituents for rational drug design. If one wishes to substitute each of R_1, R_5, R_7 and X

Table 1 QSAR models developed by GA-PLS method

Activity (MIC) measured against	Genetic algorithm based PLS models/ equations
Mycobacterium fortuitum	p(MIC)=7.21297+0.0467927 (G_C_N_2)−0.11934 (RadiusOfGyration) − 0.824105 (T_O_F_6)+0.316546 (T_N_F_7)+2.18357 (chiV3chain)+0.785047 (SssssCcount) + 0.13796 (G_2_Br_7) … … … (1)
	Optimum Components=5, $N_{Training}$=88, R^2=0.7250, Q^2 = 0.6715, F test=43.2395, R^2se=0.2885, Q^2se = 0.3153.
	N_{test} =22, Pred_R^2=0.7562, Pred_R^2 se=0.2361.
	Best Rand R^2 (Y-scrambling)=0.28452 Best Rand Q^2 (Y-scrambling) = 0.12879
Mycobacterium smegmatis	p[MIC]=8.12488+0.00844162 (T_2_N_1)+0.245728 (T_N_F_5) -1.36824 (IdAverage)+1.07926 (SssssCcount) -0.675572 (T_O_F_6)+0.0503437 (1PathCount)+3.03102 (chi3chain)+0.288192 (BrominesCount) ... … … (2)
	Optimum Components=5, $N_{Training}$=95, R^2=0.7420, Q^2 = 0.6944, F test=51.1804, R^2(se)=0.3014 , Q^2(se) = 0.3280.
	N_{test} =22, Pred_R^2=0.7087, Pred_R^2 se=0.2903.
	Best Rand R^2 (Y-scrambling)=0.19876, Best Rand Q^2 (Y-scrambling) = 0.00993

positions of fluroquinolone template (shown in Fig. 1) by a small number 50, the possible number of compounds generated will be 50^4=6,250,000. One cannot handle such a large number of chemicals intuitively; but the high quality QSAR of fluoroquinolones formulated in the present investigation can be used to screen such a large virtual library so swiftly that the selected compounds which are predicted to be promising by the QSAR model can be synthesized and tested. Another possible way of tackling the huge virtual library of 6,250,000 derivatives could be to cluster the large set into a small number of clusters by the method developed by Lajiness [33] where one chemical from each cluster may be recommended for synthesis and testing. Such a subset of fluoroquinolone derivatives will be structurally diverse and possess the chance of containing significant bioactivity profiles that might help to discover few lead compounds.

The diversity in building blocks eventually determines the chemical space coverage of a library. In the present study, the diverse substituents were attached at four different positions of fluoroquinolone ring using the following criteria: they must (1) have structural diversity as determined by calculated physicochemical properties of the virtual product, (2) form products that obey Lipinski's "rule of five" [34], and (3) generate products with synthetic feasibility.

Homology modeling

Homology model can be developed for DNA gyrase A for *M. smegmatis* based on the structural template of the *M.*

Table 2 Significance of different descriptors used in the models

Descriptors name	Descriptor meaning & significance
RadiusOfGyration	Distance based topological descriptors signifies size descriptor for the distribution of atomic masses in a molecule.
chiV3chain/ chi3chain/ 1PathCount	Molecular connectivity descriptor
SssssCcount	Defines the total number of carbon connected with four single bonds
BrominesCount	Defines the total number of bromine atoms in the molecule
IdAverage	Information theory based descriptors
G_C_N_2/ G_2_Br_7	Geometrical descriptor (requires 3D conformations of molecule) G_C_N_2 which is summation of $d_i/2$ over total number of C_N_2 fragments (carbon and nitrogen atoms separated by 2 bond distance) in the molecule, where two things are computed: total number of C_N_2 fragments in molecule and for each such C_N_2 fragment d_i is the corresponding actual distance between C and N atom. Similarly, G_2_Br_7 is summation of $d_i/2$ over total number of 2_Br_7 fragment (i.e. any double bond separated by 7 bond distance with any other Br atom).
T_O_F_6/ T_N_F_7/ T_2_N_1/ T_N_F_5	Topological descriptor (T) says T_O_F_6 is that number of O and F atoms connected by 6 different bonds (in shortest path and 6 different bonds could be of any type ie, single, double etc), similarly T_N_F_7 or T_N_F_5 signifies the count of number of N_F_7 & N_F_5 fragments respectively (means any N atoms (single double or triple bonded) separated from any other F atom by 7 & 5 bond distance respectively in a molecule). Descriptor T_2_N_1 means total number of double bonded atom connected to N atom by one bond distance.

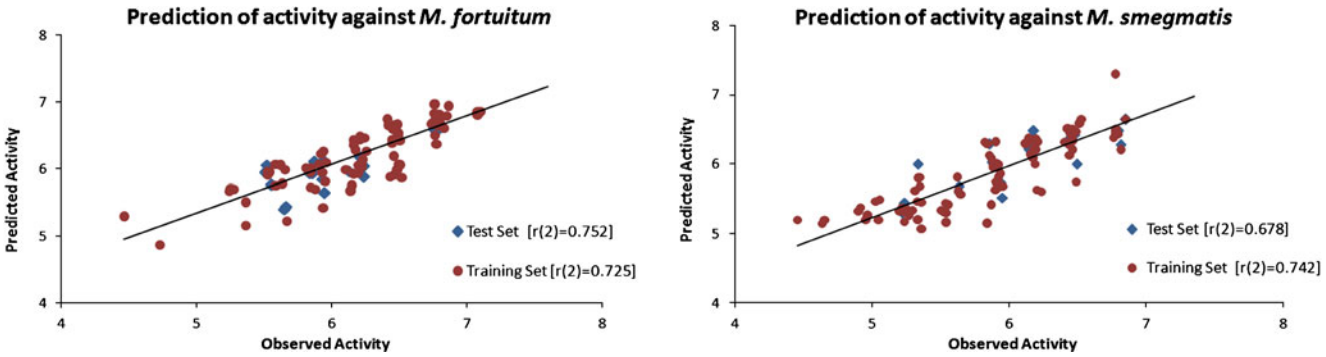


Fig. 4 Actual and predicted activity obtained from QSAR models are represented in a plot for both *M. fortuitum* and *M. smegmatis*

tuberculosis DNA gyrase to further dock the derivatives. *M. tuberculosis* DNA gyrase protein shows maximum sequence identity in PDB repository. For, *M. fortuitum*, genomic information is not yet available in GenBank/other public databases, thus preventing the same experiment on DNA gyrase of this species.

DNA gyrase subunit A has good homology with *M. tuberculosis* protein, showing above 92% sequence identity; therefore, it was sensible to construct a homology model based on these two structures. The gyrase subunit sequence was subjected to automated modeling of SWISSMODEL program [35, 36]. The best model determined by the program, was used for the study. The resulting model had 92.9% of the residues within the most favored region of the Ramachandran plot [Fig. 2] obtained by PROCHECK [37]. No residues fall into the disallowed region, suggesting the acceptability of the model for docking study.

Molecular docking study

Piecewise Linear Pairwise potential (PLP)-based molecular docking function [38, 39] has been applied to fluoroquinolone derivatives using the docking module of Molecular Design software, which involves the use of the PLP function summed over energy interactions between all pairs of protein and ligand atoms. GRIP docking utilizes the PLP scoring function in a novel way for fast and accurate capturing of ligand receptor interactions in the active site of

Table 3 Structures of three fluoroquinolone compounds with available experimental data and dock score

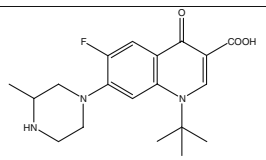
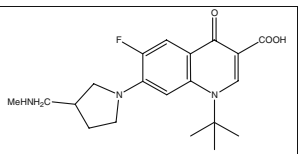
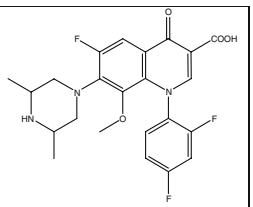
Structure			
Activity	Highly active agent (4c in ref 13). Activity against *M. fort* and *M. smeg* are 0.03 and 0.06 μg/ml respectively.	Highly active agent (4c in ref 13). Activity against *M. fort* and *M. smeg* are 0.13 and 0.25 μg/ml respectively.	Moderate or lowly active agent (11c in ref 14). Activity against both *M. fort* and *M. smeg* are 16 μg/ml.
Dock Score	**-30.01 in GRIP docking**	**-33.47 in GRIP docking**	**0 in GRIP docking**
Remarks on Substituent variations	Presence of tertiary butyl group at N1 may act as favouring substituent for anti-mycobacterterial activity	Presence of pyrrolidine group at C7 gives better dock score as piperazine-substituted derivatives were slightly less active than pyrrolidine-substituted derivatives, suggested by Renau et al. [14]	Presence of 2,4-difluorophenyl at N1 probably does not favour good activity [14] and also the presence of carboxy methyl substituent at C8 position disfavor interactions giving rise to poor dock score.

Table 4 Results of GRIP docking for 24 training set compounds. The compounds are listed according to the ascending order of dock-scores

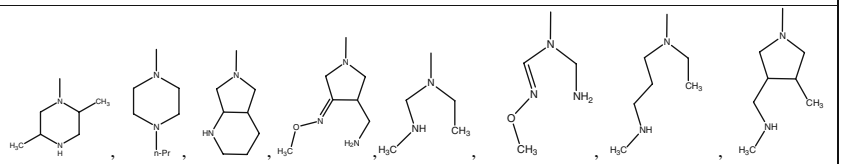
Compound_ID	Reference	Dock_score	Activity (*M. fort.*)	Activity (*M. smeg.*)
54	AAC-1f [14]	−52.02	6.79864	6.462848
1	JMC-1a [13]	−50.62	6.742131	6.122343
56	AAC-2b [14]	−38.82	6.801425	6.465633
93	AAC-9f [14]	−37.35	6.797534	6.461742
90	AAC-9b [14]	−35.13	7.082072	6.781042
36	JMC-4k [13]	−33.47	6.460601	6.176604
32	JMC-4c [13]	−30.01	7.080887	6.779857
68	AAC-5b [14]	−28.78	6.796343	6.796343
91	AAC-9c [14]	−28.73	7.098564	6.461742
53	AAC-1e [14]	−28.22	6.765094	6.765094
60	AAC-3d [14]	−26.84	6.849491	6.513699
51	AAC-1c [14]	−26.76	6.79864	6.462848
59	AAC-3c [14]	−26.7	6.863617	6.527825
55	AAC-2a [14]	−26.07	6.785082	6.44929
101	AAC-11b [14]	−19.3	5.650709	5.048649
33	JMC-4d [13]	−17.93	6.796393	6.796393
69	AAC-5c [14]	−17.15	6.812275	6.476482
72	AAC-6b [14]	−5.72	6.812275	6.476482
75	AAC-6e [14]	−1.47	6.796343	6.460551
66	AAC-4f [14]	−0.09	4.727743	5.028773
64	AAC-4d [14]	0	5.630833	5.329803
70	AAC-5e [14]	0	6.779804	6.779804
73	AAC-6c [14]	0	6.827642	6.49185
102	AAC-11c [14]	0	4.459996	4.459996

proteins. PLP scoring function in GRIP docking method includes interactions like hydrogen bonding, repulsions and dispersion. The PLP score is designed to enable flexible docking of ligands to perform a full conformational and positional search within a rigid binding site.

The functional form of the ligand–protein interaction energy in PLP scoring function is shown in Fig. 3. Parameters of the atomic pairwise ligand–protein potential are: for steric interactions, A=3.4, B=3.6, C=4.5, D=5.5, E=-0.4, F=20.0; for hydrogen bond interactions, A=2.3, B=2.6, C=3.1, D=3.4, E=-2.0, F=20.0. The units of A, B, C, and D are Angstrom; for E and F the units are arbitrary energy units [40].

All optimized ligands were docked into active binding sites of DNA gyrase target protein that can be obtained in a co-crystallized state with glycerol (protein data bank, PDB

Table 5 Possible substituents to develop the virtual library

Substituting Sites	Substituents (Structures)
R_1	Ethyl, propyl, iso-propyl, butyl, iso-butyl, t-butyl, cyclo-propane, cyclo-butane, CH_2-cyclo-propane, CH_2-cyclo-butane, CH_2CH_2F, CH_2CH_2Cl.
R_5	H, Methyl, Ethyl, Amine, $NHCH_3$
R_7	 CS1, CS2, CS3, CS4, CS5, CS6, CS7, CS8
X	CH, N, CF, CCL, CBr, CCH_2F, CCH_2CL, CCH_2Br, COMe, COEt, CNHMe

entry 3ILW) [41], which have been considered as the reference to define the active binding sites in the present investigation. From the previous literature [42, 43], it is evident that the principal target site for fluoroquinolone derivatives is N-terminal of DNA gyrase protein, and thus the present work based on the docking of selected compounds on DNA gyrase A subunit. The docking study was also performed on the homology modeled structure of *M. smegmatis* DNA gyrase A protein and the results were compared. Water molecules and HET-ATOM (representing glycerol molecules) were removed from the co-crystallized PDB file of DNA gyrase protein for the purpose of docking. A rotation angle of 30◦ was set so that the ligand would be rotated inside the receptor cavity to generate different ligand poses. After completion of the docking process, the minimum interaction energy between each ligand and DNA gyrase A protein for the best ligand pose inside the receptor cavity was obtained as the PLP score, which is discussed in the next section.

Results and discussion

GA-PLS modeling

A total number of 117 fluoroquinolone derivatives have been considered for the QSAR study against *M. smegmatis* while 110 compounds were considered for *M. fortuitum*. For both the organisms, the dataset was divided into training and test set by sphere exclusion method [19] and the models are validated by both internal and external validation procedures. Table 1 lists the two QSAR equations with fluoroquinolone derivatives against *M. fortuitum* and *M. smegmatis* respectively generated by PLS analysis in conjunction with the feature selection criteria based on genetic algorithm. In each of the generated models, care was taken to exclude the appearance of correlated descriptors within the same equation. Several statistical parameters to ensure the quality of models such as (i) regression coefficient (R^2), (ii) standard error of estimate (R_{se}^2), and (iii) variance ratio (F test) have been considered seriously in our study. The consistency and robustness of the model are being reflected by the estimation of cross-validated R^2 (Q^2) and standard error of prediction (Q^2_{se}). The widely accepted parameter to reflect the true correlation is data randomization also known as y-scrambling [44]. In this technique, the existing values of dependent variable are shuffled within the data set, and the models are regenerated with this scrambled data set. The result, if worse than the original unscrambled data set, truly reflects the absence of chance correlation, else it strongly suggests a significant correlation by chance. The predictive confidence of the models has been taken care of estimating predictive R^2 (R^2pred) for test set molecules. The significance of molecular descriptors considered in our models is listed in Table 2. All the above statistical parameters (given in Table 1) associated with our QSAR models ensure about an excellent fit and their high level of predictability for such fluoroquinolone derivatives.

The actual and predicted activities derived from the above models are plotted in Fig. 4 for the training and test sets of compounds in both cases. The correlation coefficients (R^2) for the training set compounds against *M.*

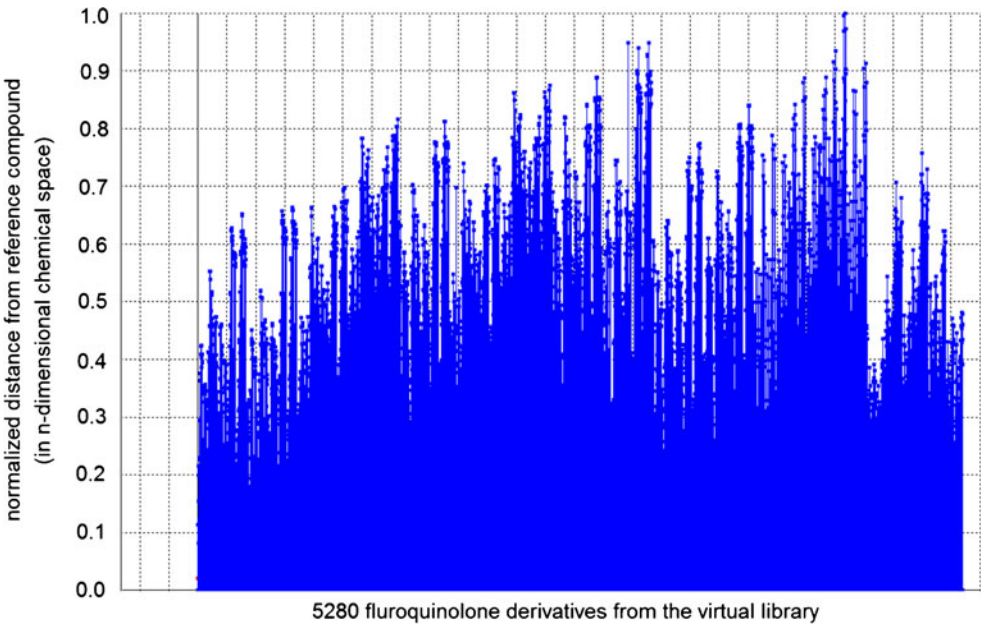


Fig. 5 Diversity plot for the combinatorial library generated for the study

Table 6 Comparative results of PLP Dock scores for selected compounds against *M. tuberculosis* DNA gyrase A (crystal structure) and *M. smegmatis* DNA gyrase A (homology model), indicating the residues participating in important interactions

Comd_ ID	MIC (μg/ml) / *M. fort* (pred. activity)	MIC (μg/ml) / *M. smeg* (pred. activity)	Dock Score (GRIP) against *M. tub* DNA gyrase	Dock score (GRIP) against *M. smeg (modeled)* DNA gyrase	Hydrogen bond interactions			Hydrophobic interactions		Pi-stacking inter-actions With the quinolone ring
					with F6 atom	with COOH gr.	with C7 substs.	with N1 subts.	with C7 substs.	
Sparfloxacin	0.06	0.13	−16.7	−48.9	Lys49 (2.093), His52 (1.978)	Asn172	Lys49 (2.484)	Arg98, Ser104, Leu105, Asn176, Ser178,	Lys49, Arg98, Gly179, Gly180, Met185	His52 (5.978)
1365	0.035	0.016	−15.2	−53.3	Arg98 (1.584)	Asn172 (2.161), Asn176 (2.317)	Lys49 (2.289)	Leu48, Lys49, His52 , Arg98, Met99, Leu105, Leu109	Lys49, Arg98, Gly179, Met185	–
455	0.08	0.03	−18.4	−52.7	His52 (2.288)	Asn172	Lys49 (2.407), Gly179 (1.896)	Arg98, Ser104, Leu105, Asn176, Gly177, Ser178,	Lys49, Arg98, Gly179, Gly180, Met185	His52 (5.108)
1395	0.022	0.046	−20.2	−40.5	Arg98 (1.620)	Asn172 (1.945)	Lys49 (2.526)	Leu48, Lys49, His52 , Arg98, Met99, Leu105, Leu109	Lys49, Arg98, Gly179, Gly180, Met185	–
2355	0.024	0.067	−16.6	−41.4	Arg98 (1.740)	Asn172 (2.312), Asn176 (2.331)	Lys49 (2.428), Gly179 (2.166)	Leu48, Lys49, His52 , Arg98, Met99, Leu105, Leu109	Lys49, Arg98, Gly179, Gly180, Met185	–
402	0.82	2.83	−57.3	−53.5	Lys49 (1.761), His52 (2.392)	Asn172		Arg98, Ser104, Leu105, Gly177, Ser178,	Arg98, Gly179, Gly180, Met185	–
3665	0.18	0.21	−32.4	−46.4	Arg98 (2.108)	Asn176 (2.504)	Lys49 (2.058),	Pro42, Gly47,	Arg98, Ser178,	–

							Gly179 (2.309)	Leu48, Lys49, His52 ,	Gly179, Met185	
1152	0.35	0.15	−28.1	−46.8	Arg98 (1.931)	Asn176 (2.322)	Lys49 (2.058), Ser178 (1.730), Gly179 (2.248)	Pro42, Leu48, Lys49, His52 , Ser178, Gly179	Arg98, Ser178, Gly179, Met185	–

fortuitum and *M. smegmatis* are 0.725 and 0.742 and that of test set compounds are 0.752 and 0.678 respectively.

To understand the influence of variations of functional groups in the four substituting sites of fluoroquinolone derivatives on their anti-bacterial activity, a comparative study is performed by observing experimental activity values and dock score on few selected compounds in the series. The Table 3 tabulated three compounds from the training set among which two compounds (compound 32 and 36) give high activity against both the organisms and the last one, compound 102, gives comparatively poor activity against *M. fortuitum* and *M. smegmatis.* Explanation of their high or low activity in accordance to the different substituents has been remarked in the table, and this elucidation also matches with the obtained dock score by the GRIP docking methodology. Table 4 lists the dock scores of 24 fluoroquinolone compounds obtained from Renau et al. [13, 14] with their p[MIC] values. Among the 24 compounds, 20 fluoroquinolone derivatives could be classified as high active compounds (showing good activity for both *M. fortuitum* and *M. smegmatis*) and the rest (64, 66, 101 and 102) can be classified as low active compounds considering their poor activity values. Although, the results do not show any clear correlation between dock-scores and activity values, all but four compounds exhibit reasonably good dock scores (less than -15) for high active compounds. The four actives showing poor docking scores probably because of the presence of carboxy-methyl or carboxy-ethyl group in C8 position, which hinder the formation of H-bonds with the neighboring residues of target. The low-active compounds do not give good dock scores, as expected according to our supposition.

Combinatorial library generation

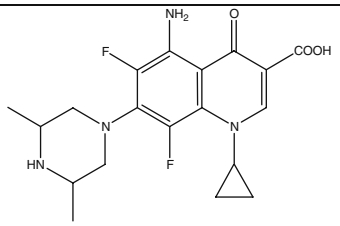
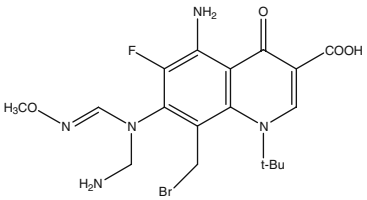
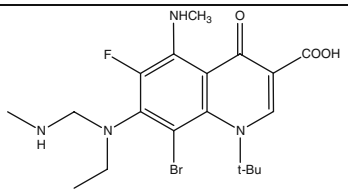
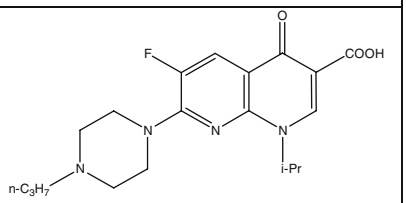
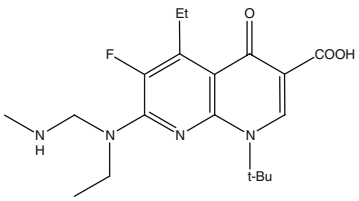
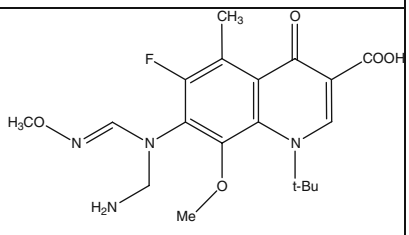
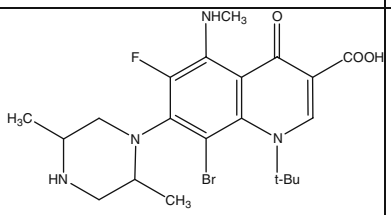
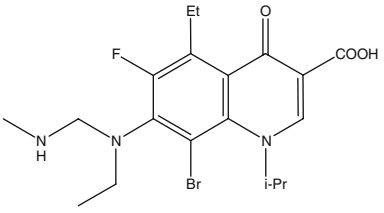
In this article, we have employed computer-assisted combinatorial chemistry methods to design and screen a virtual library of R_1, R_5, R_7 and X substituted fluoroquinolone derivatives. The structure of fluoroquinolones template and the predefined connection sites R_1, R_5, R_7 and X are shown in Fig. 1. The fragments that we used for construction of virtual fluoroquinolone library are included in the Table 5. A total of 12, 5, 8, and 11 number of chemical fragments or substituents were used at R_1, R_5, R_7 and X sites respectively to generate the virtual library. Subsequently the LeadGrow module generated 5280 compounds to form a virtual library having a basic fluoroquinolone template. Using Lipinski's rule of five selection criteria, 144 compounds were eliminated from the library. The diversity of the important molecular properties considered in Lipinski's rule for the virtual analogs is represented in Fig. 5 that indicates theoretically "drug-like" albeit on the basis of physicochemical properties. The diversity analysis helps to analyze chemical diversity of

generated molecules on appropriately chosen chemical space. Fig. 5 represents a bar graph that facilitates to visualize the chemical diversity of the entire data set generated by combinatorial method with reference to a particular compound from entire data set (here, the first compound generated in the combinatorial method, chosen as a reference compound). Here, the bar graph represented in Fig. 5, shows a reasonable amount of diversity of the total set of compounds with respect to reference compound. The Y axis in the Bar Graph is the distance metric in the chosen n dimensional descriptor space.

Prediction of biological activity & molecular docking study

The main objective of the present study is to develop a model for virtual screening for fluoroquinolone derivatives. The activities in terms of p[MIC] were calculated by utilizing the QSAR models against *M. fortuitum* and *M. smegmatis* given by Eqs. 1 and 2 respectively. In our approach we have identified few analogues, which might have high inhibitory activity against tuberculosis from the series of virtual compounds generated with the fluoroquinolone template. The library is initially screened according to their predicted activity obtained from the QSAR models against both the organisms, and those compounds having high activity profiles were considered for the docking study. DNA gyrase is one of the few thoroughly characterized and well-validated targets in anti-tubercular inhibitors like fluoroquinolones. Recently, an X-ray structure of DNA gyrase protein of *M. tuberculosis* has been determined as complexes with glycerol molecule (pdb id 3ILW). For the molecules selected according to their predicted activity from QSAR models, we have made an attempt to verify and locate their affinity toward DNA gyrase protein with the help of molecular docking study. The binding scores of the respective analogs with target protein were obtained by applying specific piecewise linear pairwise potential (PLP)

Table 7 Structures of Sparfloxacin and seven compounds selected from the virtual library

Comd _ID	Structures	Comd_ ID	Structures
Sparfloxacin		**2355**	
1365		**402**	
455		**3665**	
1395		**1152**	

scoring function using a specifically designed grip docking function. The grip docking is a fast scoring algorithm, which can be operated in two formats, fast scoring and exhaustive scoring functions. Grip docking analysis was performed with DNA gyrase A protein and the selected virtual analogs in exhaustive scoring method and the docked complexes were examined based on the scoring function to pick out the best inhibitor based on GRIP energy score.

The basic docking analysis using GRIP functions (Table 6) indicated a promising prospect for these molecules to be used as a second line drug. Table 6 demonstrates the results of sparfloxacin and few selected analogs in terms of lowest dock score and interacting residues of DNA gyrase A protein when exposed to grip docking. The dock scores obtained for these molecules with respect to *M. smegmatis* DNA gyrase A protein (modeled structure) are also tabulated in the table, and the results again suggest the screening of these compounds from the huge list of virtual molecules. Table 7 demonstrates the structures of sparfloxacin and the above analogs identified from the virtual screening analysis. For sparfloxacin and rest of the screened analogs, it is seen that fluorine atom at the six position of the quinolone ring, plays an important role in protein-ligand interaction by forming the hydrogen bond interaction with positively charged amino acids like Lys, His and Arg. The carboxyl group in the third position in quinolone ring in the case of most of the screened molecules interacts with asparagine (Asn) residue while mainly Lys49 residue (sometime, Gly179 is appended with Lys) participates in H-bond interactions with C7 moiety. Besides hydrogen bonding, hydrophobic interactions are one of the major interactive forces in ligand recognition and binding. Here, hydrophobic interactions with different substituting sites take a major role toward the binding affinity with the DNA gyrase protein and the pattern of interactions in the case of sparfloxacin and the other seven virtual analogs have a noticeable resemblance, which implies proper selection of derivatives from the huge library. In grip docking, Sparfloxacin has exhibited a dock score of -16.68 kcalmol^{-1} at optimized pose 16. Figure 6 shows the non-bonded interactions like H-bonds, hydrophobic and pi-stacking interactions formed by sparfloxacin and DNA gyrase A protein in the docked complex. When the docking results of a few selected

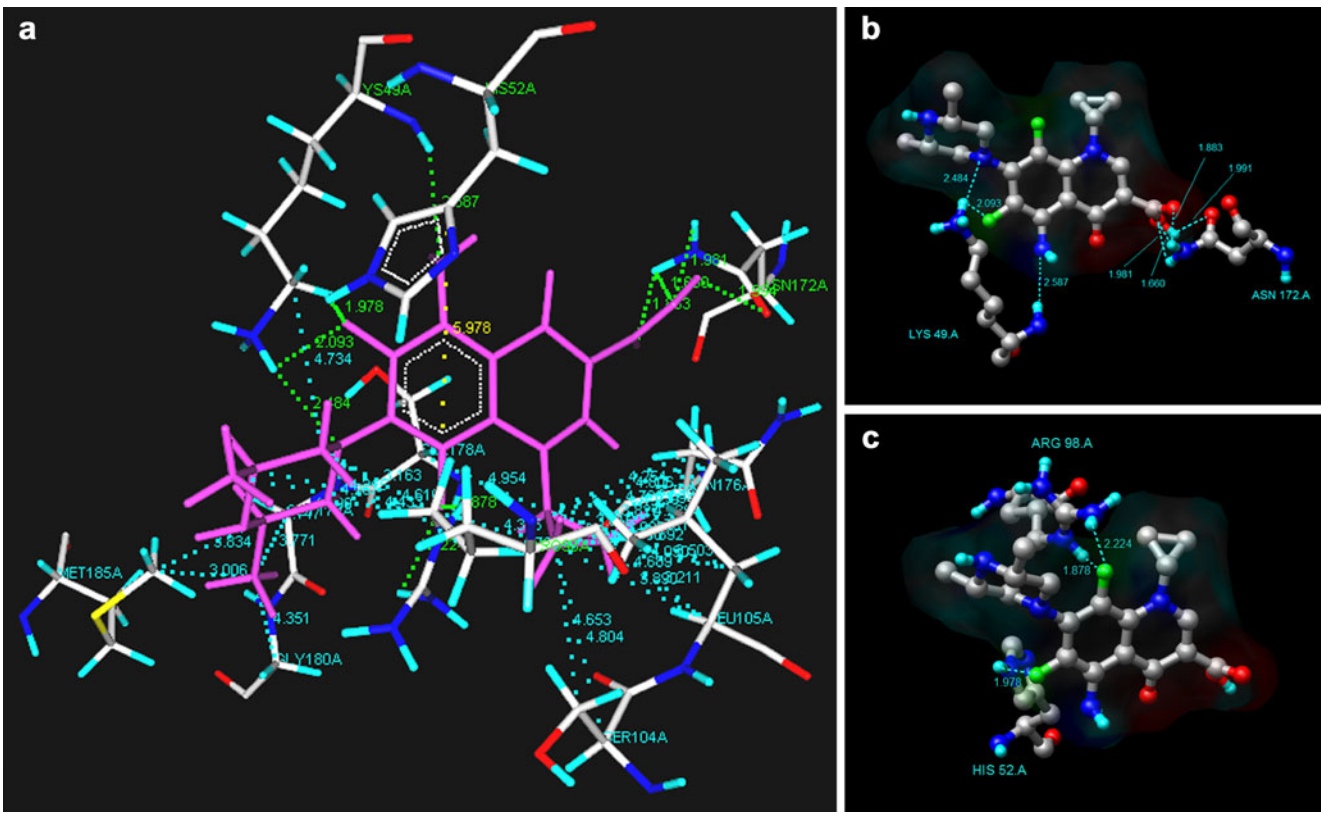


Fig. 6 (**a**) Interactions with various residues of DNA gyrase A protein with Sparfloxacin. The cyan and green colored dotted lines represent the hydrophobic and hydrogen-bond interactions respectively. One pi-stacking interaction is formed between His52 residue and aromatic ring, shown by yellow color dotted line (**b**) H-bonds are shown prominently which are formed with the residues, Lys49(A) and Asn172(A) respectively and the bond lengths are also marked (**c**) H-bonds formations with His52(A) and Arg98(A) are shown from another angle with sparfloxacin

virtual analogs (Table 7) were compared with the docking of sparfloxacin, which is used as anti-TB drug, a comparable performance was noted in terms of dock score. Docking features of compd. no. 3665 and 1152 using Grip gives a score of -32.43 and -28.52 kcalmol^{-1} respectively while compound no. 402 seems to have a higher score of -57.28 kcalmol^{-1}. On the other hand compounds 455, 1365, 2355 and 1395 appear to be quite similar to sparfloxacin as far as binding scores and H-bond interactions are concerned. In fact, the residues responsible in hydrophobic (Lys49, Arg98, Gly179, Gly180, Met185); H-bond (Lys49, His52 and Asn172) and pi-stacking interactions (His52) are almost the same in case of compound no. 455 and sparfloxacin. Additionally, when compared with sparfloxacin, the dock score (-18.67 kcal mol^{-1}) as well as activity predicted by our models in the case of compound 455 lead us to suggest it as a potential agent against tuberculosis. Figure 7 shows the interaction patterns with various residues of DNA gyrase A protein with compound no. 455 in best docked pose. Thus, the virtual screening approach yielded a novel and potent hit class of tuberculosis inhibitors from a limited selection of compounds.

Conclusions

A QSAR study of a set of anti-tuberculosis agents has been performed against *Mycobacterium fortuitum* and *Mycobacterium smegmatis*. Genetic algorithm has been applied for variable selection and the models were developed by partial least square method with limited number of descriptors. Reliability of the models was confirmed by several statistical analyses and thus, the QSAR models were used to predict the anti-mycobacterial activity of the virtual compounds developed by using combinatorial chemistry approach with fluoroquinolone template. The present investigation is an attempt in this direction resulting in the identification of a few compounds having high predicted activity values and subsequently the compounds were further screened according to their dock score and interaction patterns. Among them, one compound (compound no. 455) gives a closer resemblance with sparfloxacin in terms of predicted activity, dock scores and interaction patterns and thus, it is recommended for further synthesis and testing. It is believed that such a rational design of combinatorial chemistry libraries considering binding sites of proteins aided by molecular docking will pave the way for generating new lead compounds that

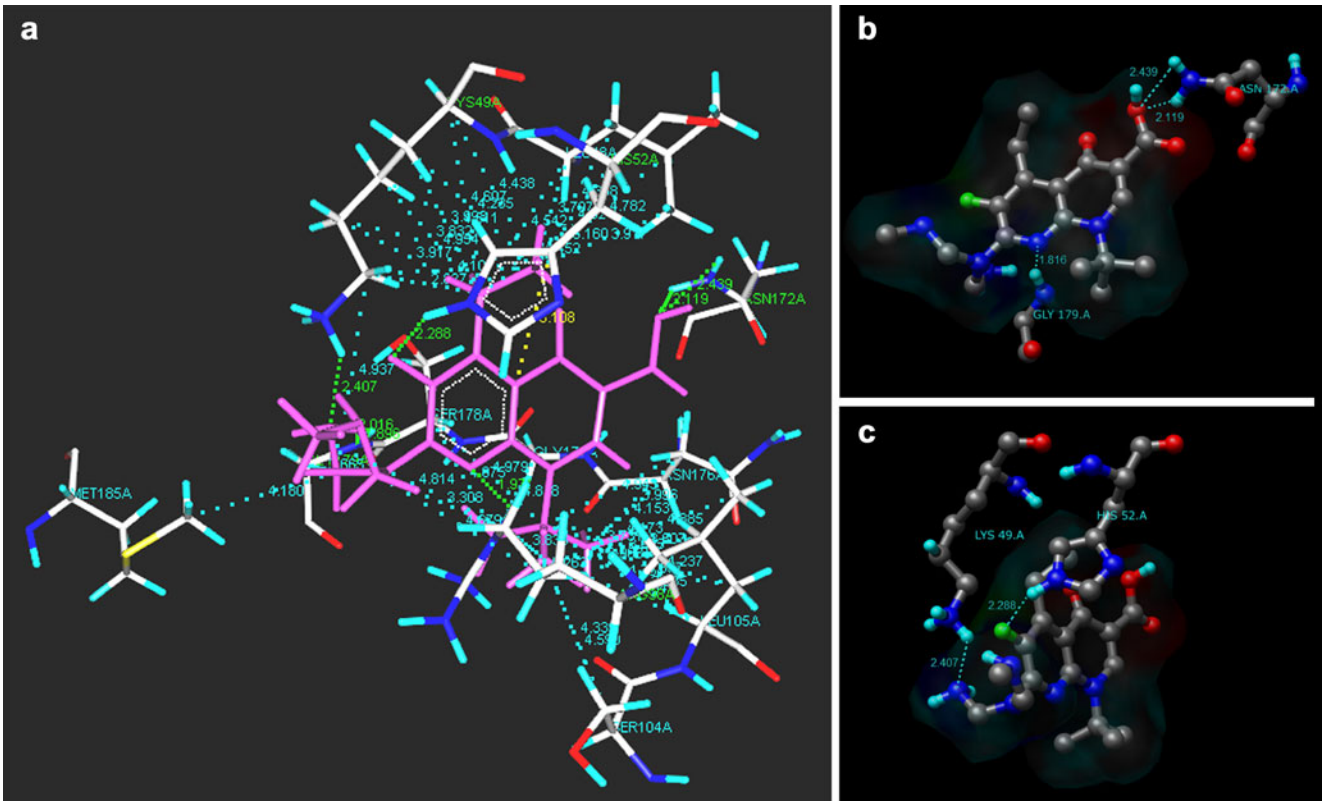


Fig. 7 (**a**) Interactions of DNA gyrase A with compound no. 455, which has a strong resemblance with sparfloxacin (**b**) H-bonds that are formed between compound 455 and the residues, Asn172(A) and Gly179(A) respectively, are shown and the bond lengths are also marked (**c**) H-bonds formations with Lys49(A) and His52(A) are shown from another angle with the same compound

may be of interest for both academia and industry within a rapid period of time which, in turn, may lead to the discovery of a potent anti-tubercular agent.

Acknowledgments Payel Ghosh thanks the Council of Scientific and Industrial Research, New Delhi 110001, India, for the grant of a Senior Research Fellowship to her. MCB acknowledges the Department of Biotechnology, New Delhi, India, for the grant of a project. The infrastructural support received from Bioinformatics Centre, IICB is also acknowledged.

References

1. Cambau E, Sougakoff W, Besson M, Truffot-Pernot C, Grosset J, Jarlier V (1994) Selection of a gyrA mutant of *Mycobacterium tuberculosis* resistant to fluoroquinolones during treatment with ofloxacin. J Infect Dis 170:479–483
2. Grosset JH (1992) Treatment of tuberculosis in HIV infection. Tuberc Lung Dis 73:378–383
3. Tsukamura M, Nakamura E, Yoshii S, Amano H (1985) Therapeutic effect of a new antibacterial substance ofloxacin (DL8280) on pulmonary tuberculosis. Am Rev Respir Dis 131:352–356
4. Blumberg HM, Burman WJ, Chaisson RE, Daley CL, Etkind SC et al. (2003) American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: treatment of tuberculosis. Am J Respir Crit Care Med 167:603–662
5. Crofton J, Chaulet P, Maher D, Grosset J, Harris W et al. (1997) Guidelines for the management of drug-resistant tuberculosis. World Health Organization, Geneva, Switzerland
6. Takiff HE, Salazar L, Guerrero C, Philipp W, Huang WM et al. (1994) Cloning and nucleotide sequence of *Mycobacterium tuberculosis* gyrA and gyrB genes and detection of quinolone resistance mutations. Antimicrob Agents Chemother 38:773–780
7. Aubry A, Pan XS, Fisher LM, Jarlier V, Cambau E (2004) *Mycobacterium tuberculosis* DNA gyrase: interaction with quinolones and correlation with antimycobacterial drug activity. Antimicrob Agents Chemother 48:1281–1288
8. Maxwell A (1997) DNA gyrase as a drug target. Trends Microbiol 5:102–109
9. Willmott CJR, Maxwell A (1993) A single point mutation in the DNA gyrase A protein greatly reduces binding of fluoroquinolones to the gyrase-DNA complex. Antimicrob Agents Chemother 37:126–127
10. Barnard FM, Maxwell A (2001) Interaction between DNA Gyrase and Quinolones: effects of Alanine mutations at GyrA subunit residues Ser83 and Asp87. Antimicrob Agents Chemother 45:1994–2000
11. Hiasa H, Shea ME (2000) DNA gyrase-mediated wrapping of the DNA strand is required for the replication fork arrest by the DNA gyrase-quinolone-DNA ternary complex. J Biol Chem 275:34780–34786
12. Wentzell LM, Maxwell A (2000) The complex of DNA gyrase and quinolone drugs on DNA forms a barrier to the T7 DNA polymerase replication complex. J Mol Biol 304:779–791
13. Renau TE, Sanchiez JP, Gage JW, Dever JA, Shapiro MA et al. (1996) Structure-activity relationships of the Quinolone antibacterials against mycobacteria: effect of structural changes at N-1 and C-7. J Med Chem 39:729–735
14. Renau TE, Gage JW, Dever JA, Roland GE, Joannides ET et al. (1996) Structure-activity relationships of Quinolone agents against mycobacteria: effect of structural modifications at the 8 position. Antimicrob Agents Chemother 40:2363–2368
15. Bagchi MC, Maiti BC, Mills D, Basak SC (2004) Usefulness of graphical invariants in quantitative structure-activity correlations of tuberculostatic drugs of the isonicotinic acid hydrazide type. J Mol Model 10:102–111
16. Ghosh P, Thanadath M, Bagchi MC (2006) On an aspect of calculated molecular descriptors in QSAR studies of quinolone antibacterials. Mol Divers 10:415–27
17. Bagchi MC, Mills D, Basak SC (2007) Quantitative structure-activity relationship(QSAR) studies of antibacterials against *M. fortuitum* and *M. smegmatis* using theoretical molecular descriptors. J Mol Model 13:111–120
18. Ghosh P, Vracko M, Chattopadhyay AK, Bagchi MC (2008) On application of constitutional descriptors for merging of Quinoxaline data sets using linear statistical methods. Chem Biol Drug Design 72:155–162
19. Golbraikh A, Tropsha A (2002) Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. J Comput Aided Mol Des 16:357–369
20. Molecular Design Suite 3.5, VLife Technologies, Pune, India. www.vlifesciences.com
21. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 3:1157–1182
22. Hasegawa K, Kimura T, Funatsu K (1999) GA strategy for variable selection in QSAR studies:enhancement of comparative molecular binding energy analysis by GA-based PLS method. Quant Struct Act Relat 18:262–272
23. Ghosh P, Bagchi MC (2009) QSAR modeling for Quinoxaline derivatives using genetic algorithm and simulated annealing based feature selection. Curr Med Chem 16:4032–4048
24. Ghosh P, Bagchi MC (2009) Comparative QSAR studies of nitrofuranyl amide derivatives using theoretical structural properties. Mol Simulat 35:1185–1200
25. Holland JH (1992) Genetic algorithms. Sci Am 267:66–72
26. Luco JM, Ferretti FH (1997) QSAR based on multiple linear regression and PLS methods for the anti-HIV activity of a large group of HEPT derivatives. J Chem Inf Comput Sci 37:392–401
27. Leonard JT, Roy K (2006) QSAR by LFER model of HIV protease inhibitor mannitol derivatives using FA-MLR, PCRA, and PLS techniques. Bioorg Med Chem 14:1039–1046
28. Leach AR, Gillet VJ (2003) An Introduction to Chemoinformatics. Kluwer, Boston, pp 79–81
29. Martin YC (2001) Diverse viewpoints on computational aspects of molecular diversity. J Comb Chem 3:231–250
30. Böhm HJ, Stahl M (2000) Structure-based library design: molecular modeling merges with combinatorial chemistry. Curr Opin Chem Biol 4:283–286
31. Beavers MP, Chen Z (2002) Structure-based combinatorial library design: methodologies and applications. J Mol Graph Model 20:463–468
32. Hansch C, Leo A (1995) Exploring QSAR:Fundamentals and Applications in Chemistry and Biology. American Chemical Society, Washington, DC
33. Lajiness M (1990) Computational Chemical Graph Theory. Ed Rouvray, DH (Nova, New York), pp 299–316
34. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 23:3–25
35. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL Workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22:195–201
36. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. Nucl Acids Res 31:3381–3385

37. Laskowski RA, MacArthur MW, Moss D, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291
38. Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel D et al. (1995) Molecular recognition of the inhibitor AC-1343 by HIV-l protease: conformationally flexible docking by evolutionary programming. Chem Biol 2:317–324
39. Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Arthurs S et al. (2000) Deciphering common failures in molecular docking of ligand–protein complexes. J Comput Aided Mol Des 14:731–751
40. Ajmani S, Karanam S, Kulkarni SA (2009) Rationalizing protein-ligand interactions for PTP1B inhibitors using computational methods. Chem Biol Drug Des 74:582–595
41. Tretter EM, Schoeffler AJ, Weisfield SR, Berger JM (2010) Crystal structure of the DNA Gyrase GyrA N-terminal domain from *Mycobacterium tuberculosis*. Proteins 78:492–495
42. Madurga S, Sánchez-Céspedes J, Belda I, Vila J, Giralt E (2008) Mechanism of binding of fluoroquinolones to the quinolone resistance-determining region of DNA gyrase: towards an understanding of the molecular basis of quinolone resistance. Chem Bio Chem 9(13):2081–2086
43. Maxwell A (1992) The molecular basis of quinolone action. J Antimicrob Chemother 30:409–416
44. Tropsha A, Gramatica P, Gombar VK (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb Sci 22:69–77