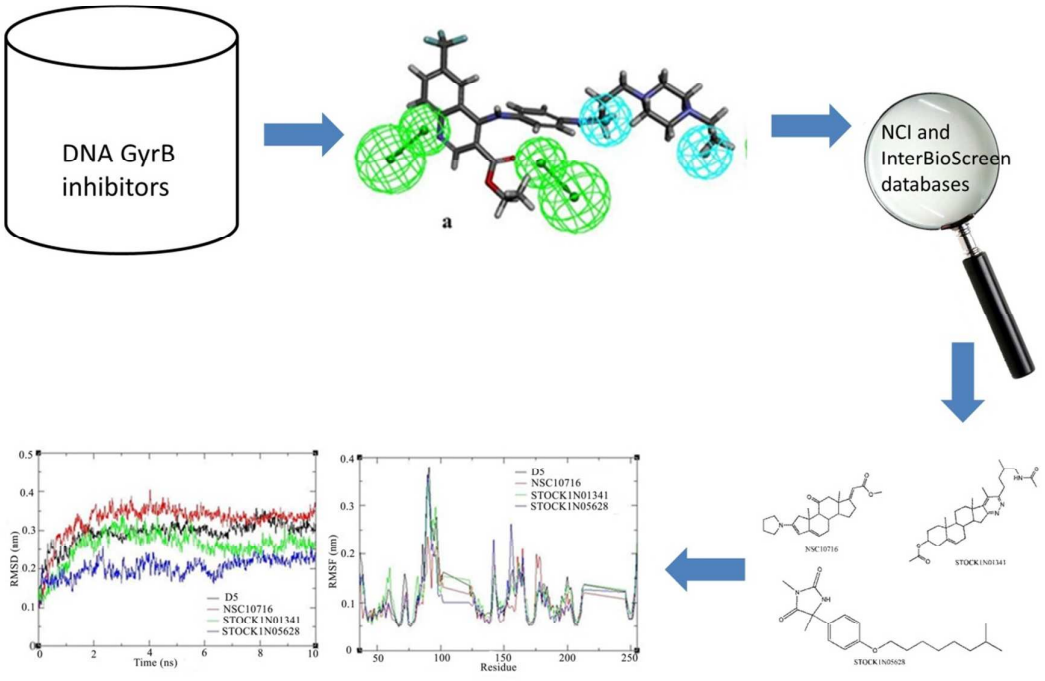
Graphical table of contents

Identification of promising DNA GyrB inhibitors for Tuberculosis using pharmacophore-based virtual screening, molecular docking and molecular dynamics studies

Md Ataul Islam[a], Tahir S. Pillay[*]

Pharmacophore based virtual screening was performed on DNA GyrB inhibitors. The best selected pharmacophore model explained that two each of hydrogen bond acceptor and hydrophobicity were critical for inhibition of DNA GyrB. On virtual screening of molecular databases using the pharmacophore model, three molecules were found to be promising as antibacterial agents which also confirmed by molecular docking and molecular dynamics studies.



Identification of promising DNA GyrB inhibitors for Tuberculosis using pharmacophore-based virtual screening, molecular docking and molecular dynamics studies

Md Ataul Islam[a], Tahir S. Pillay[*a,b]

[a]*Department of Chemical Pathology, Faculty of Health Sciences, University of Pretoria and National Health Laboratory Service Tshwane Academic Division, Pretoria, South Africa.*
[b]*Division of Chemical Pathology, University of Cape Town, , South Africa.*

Correspondence should be addressed to T.S. Pillay, Department of Chemical Pathology, Faculty of Health Sciences, University of Pretoria, Private Bag X323, Arcadia, Pretoria, 0007
Email: tspillay@gmail.com
Phone: +27-12-319-2114
Fax: +27-12-3283600

Keyword: DNA gyrase, pharmacophore; molecular docking; virtual screening; molecular dynamics

Abstract

In the current study, we searched for potential DNA GyrB inhibitors using pharmacophore-based virtual screening followed by molecular docking and molecular dynamics simulation approaches. For this purpose, a set of 248 DNA GyrB inhibitors were collected from the literature and a well-validated pharmacophore model was generated. The best pharmacophore model explained that two each of hydrogen bond acceptor and hydrophobicity were critical for inhibition of DNA GyrB. Good statistical results of the pharmacophore model indicated that the model was robust in nature. Virtual screening of molecular databases revealed three molecules as potential antimycobacterial agents. The final screened promising compounds were evaluated in molecular docking and molecular dynamics simulation studies. In the molecular dynamics studies, RMSD and RMSF values undoubtedly explained that the screened compounds formed stable complexes with DNA GyrB. Therefore it can be concluded that the compounds identified may have potential for the treatment of TB.

Introduction

Tuberculosis (TB) is a prevalent infectious disease with high morbidity and mortality rates globally(1). It is an air-borne disease, caused by *Mycobacterium tuberculosis* (MTB) which mainly affects lungs but can also affect other organs. According to World Health Organization (WHO), in 2014 alone, there were about 1.5 million deaths caused by TB with 9.6 million new cases of TB(2). 95% of TB deaths in 2014 occurred in low income countries. Further challenges arise from the increasing incidence of drug resistant tuberculosis (DR-TB) including multi-DR tuberculosis (MDR-TB), extensive-DR tuberculosis (XDR-TB) and latest total-DR tuberculosis (TDR-TB). MDR-TB was first reported in the 1990s with resistance to rifampicin and isoniazid(3). XDR-TB denotes strains of *MTB* resistant to rifampicin and isoniazid along with at least three of second-line anti-TB drugs included aminoglycosides, polypeptides, fluoroquinolones, thioamides, cycloserine and para-aminosalicylic acid(4, 5). Each of 105 countries has had at least one case of XDR TB in 2014(6) including United Kingdom and the United States of America. An estimated 9.7% of people with MDR TB have XDR-TB(6). TDR-TB was reported in 2007 in Italy but has also been found in Iran, India and South Africa(7) and this refers to MTB strains that show resistant to all the first and second line drugs of TB(8). In South Africa, with a 2.2% burden of TB, a total US $158 million(44% of the total national costs of diagnosing and managing all forms of TB) was consumed by DR-TB in 2011(9). Treatment of DR-TB is very complex and is characterised by long duration of treatments that are expensive and often the recommended medicines are not always available(10). Moreover, most of the first-line drugs used for the treatment of TB are more than forty years old and have serious side effects(11). Although there are a few newer chemical agents in clinical trial, including delamanid and fluoroquinolones, still there are no effective drugs available in the market for therapeutic treatment of DR-TB(12-14). Thus there is an urgent need to design and develop new anti-tuberculosis drug candidates for TB. Currently, the DNA Gyrase enzyme, a type II topoisomerase from the gyrase, HSP 90, histidine kinase, MutL enzyme family is one of the important, most explored and corroborated targets for the development of novel and safer anti-mycobacterial leads(15).

DNA gyrase is a crucial enzyme that interacts with negative supercoils of double-stranded closed-circular DNA and is part of the topoisomerases involved in the control of topological transitions of DNA. DNA gyrase is a good potential target to develop potent and safer lead compounds as this enzyme is absent in eukaryotic cells. It was observed that inhibitors of

DNA gyrase enzyme were effective against non-replicating, persistent mycobacteria and this is crucial in reducing the time for TB therapy(16). A novel and potential MTB gyrase inhibitor should be therapeutically effective against DR-TB. DNA gyrase performs enhanced relaxation, DNA cleavage and de-catenation activities(17). DNA gyrase is generally found as a heterotetramer (A_2B_2) with two subunits each of GyrA and B(18). Generally GyrA plays a role in breaking and reuniting of DNA, whereas GyrB involved in ATPase activity(19). It is reported that GyrA is one of the most understood targets and mostly targeted by the fluoroquinolones. Less is known about GyrB subunit as a receptor molecule for the development of potential inhibitors for TB. A previous FDA approved DNA GyrB entity (novobiocin) was developed but was later withdrawn from the market due to severe safety issues and poor pharmacological properties(20). Additionally, DNA GyrB has been genetically confirmed to be a bactericidal drug target in MTB, but to date there has not been any active drug therapy established against this target(21). With this in mind the current study was initiated to identify potent and safer anti-TB molecules using pharmacoinformatics approaches.

The concept of a pharmacophore is defined as collection of steric and electronic functional groups that provides an intuitive way of depicting and understanding the binding properties of small molecules along with an account of optimum macro-molecular contacts with a particular biological target, to stimulate or inhibit its biological response(22, 23). Pharmacophore features hydrogen bond acceptor (‘a’) and donor (‘d’), hydrophobic (‘p’) and aromatic ring (‘r’) were found to be the crucially related to selectivity and potency in different enzymes including GyrB. Pharmacophore models have wider applicability in the field of drug discovery by providing important facts to study structure-activity relationships (SAR) and exposing the mechanism of ligand-target relationships from inferences of the nature of functional groups and non-covalent bonding patterns(24-26). In the current work, an effort was made to find a well-validated pharmacophore hypothesis using key chemical features of DNA GyrB inhibitors with inhibitory activity. The final selected model was considered for virtual screening to select virtual hits from small molecular structural databases. The potential of the work is demonstrated by the credentials of finding three promising molecules as DNA GyrB inhibitors. The final proposed screened molecules were docked inside the receptor cavity of DNA GyrB to explain the binding interactions and preferred orientation for higher potency. Finally, to investigate stability and detailed binding

interactions of the final screened molecules inside the receptor cavity of DNA GyrB, a molecular dynamics study was performed.

Materials and methods

In order to discover novel and potent scaffolds of DNA GyrB inhibitors the pharmacophore space modelling approach was adopted. This is one of the crucial and most widely used versatile pharmacoinformatics techniques. The pharmacophore models can be developed using ligand-based and structure-based approaches and furthermore can be used for screening of novel chemical compounds from molecular databases. Ligand-based pharmacophore approaches use a set of ligands with their biological activity, while receptor-based approaches consider information of amino residues present in the receptor site to identify chemical functionalities important for binding interactions with small molecules. In the current research work, a ligand-based approach was implemented for a set of DNA GyrB inhibitors with known inhibitory activity (IC_{50}).

We used Discovery Studio 4.0 (DS)(27) for the pharmacophore model generation and virtual screening followed by molecular docking study. DS is a commercial software package consisting of several modules and widely used for pharmacoinformatics studies(28-30). In DS, *3D QSAR pharmacophore generation* module was used in which a set of DNA GyrB inhibitors and activity data were taken as input to create potential hypotheses. In the ligand-based approach basically two modules, *HypoGen* and *HipHop* incorporated in DS were used. The *HipHop* module identifies the hypotheses conjoint in the 'active' molecules of training set but not present in the 'inactive' compounds, whereas, *HypoGen* module develops hypotheses present both in 'active' and 'inactive' molecules. In order to develop a pharmacophore model from DNA GyrB inhibitors the *HypoGen* module was used.

Dataset

A set of 248 compounds belonging to the class of DNA GyrB inhibitors(1, 15, 20, 31-34) along with inhibitory activity (IC_{50}) were collected from the literature. In order to develop the pharmacophore model and subsequently validate the model, the molecular dataset was randomly distributed into training and test set compounds respectively. Molecules in the dataset have a wide range of IC_{50} values, from 1.470 to 137.500 μM.

The entire dataset was classified into highly active (IC_{50} < 20 μM, +++), moderately active (20 ≤ IC_{50} < 50.000 μM, ++) and least active (IC_{50} ≥ 50.000 μM, +) based on the IC_{50} values. In order to develop pharmacophore model in DS the training set molecules from the dataset was considered according to the basic guidelines laid down by Li et al.(35). These explain that (a) compounds present in the set should provide strong and brief information with chemical features and range of activity, (b) a minimum of 16 diverse compounds should be considered in the set to confirm statistical significance and elude chance correlation, (c) the most and least active compounds of the dataset must be present in the set and (d) the activity of the compounds should span 4 orders of magnitude. Based on the above criteria 30 molecules were selected in the training set along with most and least active compounds. The rest of the compounds of the dataset (total 218) were used as test set molecules and further used to evaluate the performance of the pharmacophore model. The 2D/3D visualizer module of DS was used to generate three-dimensional (3D) coordinates of the compounds. Each and every compound was considered to correct the coordinates and minimization of energy using the modified CHARMm force field(36, 37). The pharmacophore modelling, virtual screening and molecular docking studies were performed using the different packages of DS. Gromacs5.0.6(38) was used for molecular dynamics simulation.

Pharmacophore model generation

The *3D QSAR Pharmacophore Model Generation* module of DS was used to develop the pharmacophore hypotheses. The training set molecules were converted into 3D structures and conformations were generated by *Cat-Conf* program of the DS software package. In order to generate various acceptable conformations, the BEST conformational generation method was used. This method offers complete and improved coverage of conformational space with help of rigorous energy minimization and optimizing the conformations with the help of poling algorithms(39). In the BEST algorithm, the chemical functionalities are arranged in space instead of simply by the arrangement of atoms(40). In order to search for the favourable features of the highly active compounds of the dataset the *Feature mapping* module was adopted and mapped features of the compounds were given as input features for pharmacophore model generation. Using the conformer along with chemical features the algorithm operates in two modes such as *HipHop* and *HypoGen*. Pharmacophore models develop from the active set compounds using the *HipHop* approach, whereas the *HypoGen* approach generates models from both active and inactive compounds and finds a hypothesis

which is common in the active molecules and absent in the inactive compounds(40). From the training set molecules the *HypoGen* module generates top ten hypotheses by considering conformational models and chemical features through following steps: constructive, subtractive and optimization(41). In the constructive step, hypotheses are created that are common in the most active molecules; in the second phase, inactive compounds are removed from those that fit the hypotheses. In the last phase, using the small perturbations, the rest of the hypotheses improve the score(40, 42). Finally, the best pharmacophore hypothesis was considered based on the best correlation coefficient (*R*), low root mean square deviation (RMSD), cost function analysis and good predictive ability for further analysis.

Validation of pharmacophore model

Validation of any *in-silico* model is a crucial step to verify predictive ability and applicability along with checking the robustness of the model. In this work, the selected pharmacophore model from the training set was validated by five different approaches, (a) internal validation, (b) cost function analysis, (c) Fischer's randomization test, (d) test set prediction and (e) decoy set.

Internal validation

Internal validation is an important validation protocol to explain the predictivity and robustness of the model using the training set molecules. In this regard, the leave-one out (LOO) cross-validation approach was used, in which the model was developed by deleting one compound randomly from the training set and parameters used were the same as the original model. The new developed model was used to calculate estimated inhibitory activity of the deleted compound. Similarly, activity of all molecules of the training set were estimated and recorded for further analysis. Using the estimated inhibitory activity of the training set molecules two important statistical parameters, the LOO cross-validated correlation coefficient (Q^2) and error of estimation (*se*) were calculated. It is reported that high Q^2 (>0.5) and low *se* explain better predictive ability (43). In order to endorse the good predictive capability of the training set molecules the modified r^2 ($r^2_{m(LOO)}$) reported by Roy *et al.*(44, 45) was also calculated which measures the degree of deviation of the estimated activity from the experimental ones. It was reported that model may be considered with $r^2_{m(LOO)}$>0.5.

Cost function analysis

In order to select the best pharmacophore hypothesis, several statistical parameters were noted at the time of hypothesis generation, such as spacing, uncertainty, and weight variation. The spacing signifies the minimum inter-features distance that may be permissible in the final hypothesis. The weight variation is the certain order of magnitude explored by the hypothesis in which every feature indicates some degree of magnitude of the compound's inhibitory activity. In general 3.0 and 0.3 are the default values of spacing and weight variation respectively but it may vary from 4.0 to 1.0 and 1 to 2 respectively depending upon the different cases. Another parameter, uncertainty gives the error of prediction which indicates the standard deviation of the error cost. The default value of this parameter is 3 but in some cases it may vary from 1.5 to 4.0. The cost function was analysed by minimization of three cost factors, *viz*., weight cost, error cost, and configuration cost. The weight cost depends on the weight variation and is directly proportional to the deviation from input value. Error cost is the deviation between the predicted activities of the training set and their experimentally determined activities. A fixed cost penalizes the complexity of the hypothesis space. The configuration cost equals entropy of hypothesis space and it is reported that value should be <17 for a good pharmacophore model. The total cost, which is overall cost of a hypothesis is the summation of all three cost factors. For generation of a pharmacophore hypotheses, the *HypoGen* algorithm also develops the null hypothesis which is the postulation that there is no association in the data, and the experimental activities are distributed about their mean. It is reported that the higher (>60) cost difference ($\Delta cost$ = null cost - total cost) indicated that the hypothesis does not reflect a chance correlation.

Fischer's randomization test

The selected pharmacophore hypothesis was assessed for quality using CatsScramble(42) which is based on Fischer's randomization test. The CatsScramble mainly verifies the strong association between the chemical compound and the biological activity of the training set molecules. In this approach, the new pharmacophore hypotheses were generated after scrambling activity and assigning new values of the training set molecules. Constraint and input features remain the same as original hypothesis. If any randomized run generates pharmacophore hypothesis with better statistical parameters then the original hypothesis may be considered to be developed by chance. Based on statistical significance a number of spreadsheets are generated. The statistical significance is given by following equation.

$$Significance = [1 - (1 + a) / b] \quad (1)$$

Here, *a* describes the number of hypotheses with a total cost less than the original hypothesis, whereas *b* denotes a collection of *HypoGen* and random runs. For example, in case of 95% confidence level total number of random spreadsheets are generated as 19 ($b = 20$) when each generated spreadsheet is submitted to *HypoGen* using the same parameters as the initial run. In the current research, the developed pharmacophore model was tested at 95% confidence level which produced 19 spreadsheets.

Test set prediction

It is important to verify the external predictivity of the model beyond the molecules involved in the model generation. In this purpose, 218 test compounds were estimated by fitting with the selected pharmacophore model using *Ligand Pharmacophore Mapping* protocol in DS, given in Table S1 (Supplementary file). Different statistical parameters including R^2_{pred} (correlation coefficient) and s_p (error of prediction)(46, 47) were calculated to check the quality of prediction of the model. As the value of R^2_{pred} based on the mean observed activity of the training set may be attained for molecules with a wider range of activity value, this may not be guaranteed that the predicted activity values are very close to the observed activity. Therefore there is a chance of a significant numerical difference between the two values instead of a good overall correlation being maintained. To overcome above this, the modified r^2 [$r^2_{m(test)}$](48, 49) value was calculated (threshold value=0.5).

Virtual screening

Virtual screening based on a well validated pharmacophore model can be efficiently used to identify novel potent compounds from molecular databases that can bind to a particular receptor site to block or trigger activity. In the current research, the best pharmacophore hypothesis was used to screen the NCI (National Cancer Institute) (https://cactus.nci.nih.gov/ncidb2.2/) and IBS (InterBioScreen) (http://www.ibscreen.com/) databases to retrieve novel chemical entities for DNA GyrB inhibitors. The NCI and IBS databases contain 265,242 and 523,031 compounds respectively. Best pharmacophore hypothesis was submitted separately to the NCI and IBS databases with set 'Limit Hits' as 'Best N' and 'Maximum Hits' as 600. The initial hit molecules were further screened with a number of criteria to achieve final potential molecules for the DNA GyrB. Furthermore, to analyse binding interactions between the final screened molecules and catalytic amino residues molecular docking was performed.

Molecular docking

For the screening of molecules, molecular docking is one of the best techniques and a crucial step in the drug design process. In order to understand how the molecules identified by virtual screening ("drug-like hits") bind to DNA GyrB, molecular docking was performed using the *LigandFit* protocol of DS. First of all this protocol detects the cavity to identify and select the region of the protein as the receptor site followed by docking the molecules in the selected site. The crystal structure of the DNA GyrB was collected from RCSB Protein Data Bank (RCSB-PDB) for the molecular docking study. The receptor size, resolution and date of deposit were considered to select the crystal structure of DNA GyrB and finally PDB ID: 4B6C(16) was downloaded for further study. Both protein and ligands were prepared by using the *Prepare Protein* and *Prepare Ligand* tools of DS respectively. The CHARMm force field(50) was used for minimization of both protein and ligand. In the *Prepare Protein* module of DS the 'Build Loop' and 'Protonate' parameters were fixed to 'True' while, dielectric constant, pH, ionic strengths and energy cut-off were considered as default value. In case of *Prepare Ligand* module, preparation 'Change ionization', 'Generate Tautomers' and 'Generate isomers' were set to 'False', and 'Generate Coordinates' was fixed to '3D'. After preparing the protein, the receptor cavity was identified on the basis of volume occupied by the ligand. In order to overcome the false positive results of a molecular docking study validation is an essential step. In this purpose, the co-crystal small molecule bound in the receptor site was initially redrawn and the same docked into the active site of DNA GyrB (PDB ID: 4B6C). The binding interactions were analysed of the best docked pose of co-crystalized ligand followed by superimposing the docked pose and the co-crystal. The RMSD value was calculated from superimposed ligands to examine docking parameters that were capable of reproducing a similar conformation to that of the co-crystal at the active site of DNA GyrB. The same parameters as were present in the co-crystalized docking were used in molecular docking studies of potential molecules retrieved from databases. For further analysis of binding interactions and dock score values, the top ten conformations for each ligand were considered.

Molecular Dynamics

It is not sufficient to analyse static complexes to completely understand protein-ligand complex as one also needs to understand the dynamic information generated by simulating their internal motions or dynamic processes. In this respect, molecular dynamic (MD)

simulation was performed on the best docked poses of DNA GyrB and the most active compound and final screened compounds from databases. MD simulation was carried out in Gromacs 5.0.6. For all ionisable residues the protonation states were fixed to their normal states at pH 7. For all complexes, simulations were carried out with GROMOS96 43A1 force field of the Gromacs 5.0.6 package installed on an in-house Linux-based desktop. In the MD simulation all proteins were surrounded by a cubic water box of SPC3 water molecules and it was extended 10Å from the protein. The PRODRG was used to generate the topology for small molecules. The periodic boundary conditions (PBC) were applied in all directions. Furthermore, the role of the Na^+ and Cl^- counter ions which replace the water molecules, system was neutralized. The steepest descent algorithm was used to minimize each system for 10,000 steps. Each system was considered for 100 ps position-restrained MD simulations. Further MD simulation was performed for a 10ns production with a time step of 2fs at constant pressure (1 atm) and temperature (300 K). The snapshots were recorded at every 1ps for further analyses of MD simulations. The RMSD, root mean-square fluctuation (RMSF) and radius of gyration were recorded to analyse the behaviour of each system.

Drug-likeness analysis of screened compounds

The drug-likeness attributes of the final screened compounds were compared with existing Food and Drug Administration (FDA) approved DNA GyrB inhibitors. For this purpose different parameters including dockscore, estimated activity, fit value, molecular weight, logP, violation of Lipinski's rule of five (LoF), molecular volume, molecular refractivity, number of H-bonds and number of bump interactions were recorded. In order to calculate the above properties DS(11) and an online program Molinspiration (www.molinspiration.com) were used. Dockscore refers to the pool of internal energy of ligand-receptor complex and ligand only. Both estimated activity and fit score were predicted by fitting ligands onto the best pharmacophore model. logP reflects the hydrophobicity of the molecules. The LoF explain that for a drug-like molecule logP value should not be more than 5, hydrogen bond acceptor and donor should not be more than 10 and 5 respectively, and molecular weight should not be more than 500. Molecular volume defines the transport features of molecules, such as intestinal absorption or blood-brain barrier penetration.

Results and discussion

In order to develop pharmacophore models, the *HypoGen* module of DS was used on the training set molecules selected from whole dataset. The training set molecules are given in Figure 1 (**D1** – **D30**) and the inhibitory activity values pIC_{50} (log[(1/IC_{50})x1000]) are provided within the parentheses.

The pharmacophoric features of the molecules were identified using the *Feature mapping* protocol of DS, and 'a', 'd', 'p' and 'r' were found to be common chemical features in the dataset and were considered as inputs to the 3D QSAR pharmacophore generation. Minimum and maximum feature values were set to '0' and '5' respectively. Based on excellent statistical parameters the top ten hypotheses were considered for further analysis. The statistical parameters along with correlation coefficient were noted and are depicted in Table 1. Debnath's analysis(25, 51) was used to select the best hypothesis which explains that best model should have the low RMSD, high correlation coefficient, low cost value and high cost difference. A well validated hypothesis should have the overall cost of the hypothesis distant from the null cost and close to the fixed cost. It is stated that differences between null cost and total known as *Δcost* in the range of 40–60 bits explains the probability of the predictive correlation of 75–90%, while the cost difference more than 60 bits defines the hypothesis and has a correlation probability of more than 90%(52). In the current work, the cost difference of *Hypo 1* (Table 1) was found to be 100.070 which clearly explained that selected hypothesis has more than 90% chance of being able to select DNA GyrB inhibitors.

Table 1 explains that the best hypothesis (*Hypo 1*) was found to have a good correlation coefficient value (R = 0.883), which explains good predictive ability of the selected hypothesis. The total cost and fixed cost were found to be 199.281 and 87.247 respectively, along with the cost difference (*Δcost*) which was found to be 100.070. The top 10 hypotheses were selected to analyse the results and it was found that only *Hypo 1* possessed high correlation coefficient, less RMSD, highest cost difference and minimum error values in comparison to other hypotheses. Consequently, *Hypo 1* was selected as the best pharmacophore model for advanced analyses.

The best model (*Hypo 1*, Fig. 2a) revealed the importance of two of each hydrogen bond acceptor and hydrophobic region. *Hypo 1* mapped with the most active molecule (**D5** in Figure 1) of the training set and inter-feature distances is depicted in Figure 2.

Observed and predicted activity of individual compounds of training set molecules were analysed and revealed that one highly active (IC_{50} < 20 μm, +++) was overestimated as moderately active molecule, six moderately active molecules (20 ≤ IC_{50} < 50 μm, ++)

underestimated as highly active and two least active compounds underestimated as moderately active. The rest of the compounds in the training set were estimated correctly within their range. From the above discussion it can be explained that the *Hypo 1* predicted inhibitory activity of the training set molecules correctly which is echoed by the high correlation between experimental and predicted biological activities.

The most active compound of the dataset was mapped in *Hypo 1* using *Ligand Pharmacophore Mapping* module of DS and delineated in Figure 2a. Mapping of the most active ligand (**D5** in Figure 1) with pharmacophoric features explained that amine group of double ring system and oxo group attached to the same behave as hydrogen bond acceptors which suggest that these groups are crucial for forming hydrogen bond interactions with the catalytic amino residues of DNA GyrB. The presence of an alkyl group between two single rings and ethyl group attached to the pyperazine ring impart the hydrophobicity of the molecules. Therefore, the above observation and discussion clearly indicate that to design or synthesize new potential DNA GyrB inhibitors, hydrogen bond and hydrophobic features along with inter-feature distances will be crucial factors.

Validation of pharmacophore model

A Selected pharmacophore hypothesis should not only be statistically robust but also well predictive for internal and external compounds. The best pharmacophore hypothesis (*Hypo 1*) was considered for validation using internal validation, cost function analysis, test set prediction, Fischer's randomization test and decoy set. Its ability to consistently predict external data sets and discriminate active inhibitors from the inactive is an important criterion for high-quality models.

Internal validation

Predicted inhibitory activities calculated based on *Hypo 1* along with observed activities of training set molecules are given in Table 2 and Figure 3. The error value is the ratio between observed and predicted activities which reflects the consistency between both activities. Table 2 explains that all compounds of the training set have error values within reasonable range. In order to further validate the model using training set compounds, cross-validated correlation coefficient was calculated and it was observed that the best hypothesis gives Q^2 of 0.786 and *se* of 0.242. Further r^2_m and Δr^2_m values were calculated and were found to be 0.763 and 0.192 respectively. It is reported that for acceptance of the model r^2_m and Δr^2_m

should be more than 0.500 and less than 0.200 respectively. The high Q^2 and r^2_m, and low *se* and Δr^2_m of the *Hypo 1* suggested that model is statistically robust in nature.

Cost value analysis

The *HypoGen* algorithm in DS computes and gives a number of parameters for preliminary assessment of the model. In this regard, the Δ_{cost} which is the difference between null and total costs, the configuration cost, and RMSD between the estimated and the experimental inhibitory activities of the training set molecules were analysed as important parameters to judge statistical significance. As per Table 1, the cost difference of *Hypo 1* was found to be 100.070 which clearly explained that the selected hypothesis was not generated by chance. A consistent and robust pharmacophore model should also have a configuration cost value less than 17. *Hypo 1* generated configuration cost of 15.117. It is also reported that lower differences between total and fixed costs is an indication of a robust model. For *Hypo 1* the difference between total and fixed cost was found to be 11.964 and that is significant for the model.

Test set prediction

In order to check the predictive ability of compounds outside the training set molecules *Pharmacophore Mapping* module of DS was used to predict inhibitory activities of test set compounds. A total 218 compounds from the dataset were considered in the test set and the inhibitory activity values converted into logarithm values [pIC_{50} = log((1/IC_{50})x1000)]. Chemical compounds in SMILES format along with pIC_{50} values are given in the supplementary file (Table S1). The test compounds were categorised according to their pIC_{50} values such that highly active (pIC_{50} >1.700), moderately active (1.300 < pIC_{50} ≤ 1.700) and least active/inactive (1.30 ≤ pIC_{50}). On analysis of predicted and experimental inhibitory activities of the test set molecules it was observed that one highly active and six moderately active compounds were underestimated as moderately active and least active respectively. Fourteen moderately active and two least active compounds were overestimated as highly active compounds, whereas five least active compounds were overestimated as moderately active compounds. Therefore, among 218 test compounds twenty eight molecules were either underestimated or overestimated from their original activity, whilst the remaining 190 were estimated correctly within their range. Furthermore the correlation coefficients (R^2_{pred}) between observed and predicted inhibitory activities and error of prediction (*sp*) were calculated and values found to be 0.701 and 0.232 respectively. The above observations

clearly indicate that *Hypo 1* is competent enough to estimate the inhibitory biological activity of the compounds beyond the training set.

In order to verify better determination of predictive ability of the selected model another important statistical parameter $r^2_{m(test)}$ was calculated which explains how the predicted inhibitory activities are adjacent to the equivalent experimental values as a high correlation coefficient value (R^2_{pred}) cannot always put forward a low residual between the experimental and predicted activity data. Two parameters $r^2_{m(test)}$ and $\Delta r^2_{m(test)}$ were recorded and values found to be 0.616 and 0.030 correspondingly which explains that selected hypothesis (*Hypo 1*) has adequate predictive potential. Therefore, the above observations indicated that the selected hypothesis can reasonably predict the biological activities of new molecules.

Fischer randomization test

The best model was considered to adjudge the quality of the hypothesis through Fischer's randomization in which statistical quality of the model was checked by assigning a particular confidence level. In current situation *Hypo 1* was under taken for 95% confidence level. At 95% confidence level, experimental activities of training set compounds are reorganised and generated into 19 random spreadsheets by creating a hypothesis on each spreadsheet. The consequence of the hypothesis was calculated as per equation (1). Total costs and the correlation values of all 19 spreadsheets were recorded and given in Table 3. From the Table 3 it can be noted that not a single randomized run perceived predictive powers similar to or better than that of *Hypo 1*.

The mean value of all 19 trials was found to be 0.700 and *Hypo 1* achieved much lower total cost value compare to other 19 runs. The total costs of *Hypo 1* of all 19 trials are given in Figure 3 and Table 4. The above observations of Fischer's randomization approach undoubtedly indicated superiority of the hypothesis and *Hypo 1* was not generated by chance.

Decoy set

In order to check the screening capability of the selected pharmacophore model, decoy set validation was performed. In this purpose, a set of 630 DNA GyrB decoys were collected through DecoyFinder1.1. The obtained decoy molecules were amalgamated with 30 active DNA GyrB inhibitors and screened using the *Hypo 1*. The model was further used to find out and discriminate between actives and decoys with good accuracy of 0.860. The true positive (TP), false positive (FP), true negative (TN) and false negative (FN) values were found to be 26, 90, 530 and 4 respectively. The ROC plot of pharmacophore model was derived by

plotting true positive rate of actives vs. false positive rate of inactive compounds and delineated in Figure 5. The ROC curve clearly explained that actives and decoys are well-classified. Further, the area under curve (AUC) was calculated and the value was found to be 0.83 that undoubtedly explained that more true positives have been verified. The enrichment factor (EF 1%) and Boltzmann-enhanced discrimination of ROC were also calculated and the value of average EF (1%) value for pharmacophore model was found to be 15.48 which showed that hypothesis has acknowledged active compounds very well and the top 1% hit is enriched with active molecules. The mean BEDROC was found to be 0.650 which implies that the top hits is not only enriched with active compounds but also ranked higher than the decoys. The aforementioned observations of decoy validation strongly explain that the developed pharmacophore features in the selected model are impeccably acceptable for the mapping of DNA GyrB inhibitors.

Virtual screening

Screening of small molecular databases using pharmacophore model is a powerful technique to obtain novel and potential inhibitors. This is also an effective alternative to the approach of high-throughput screening methodologies. In order to obtain potent DNA GyrB inhibitors, *Hypo 1* was used to explore the NCI and IBS databases. In the DS package the '*Search Database*' protocol under '*Pharmacophore*' module was used for screening of both databases. In the parameter list the protocol 'Search Method' and 'Limit Hits' were set to 'Best' and 'Best N' respectively. 'Maximum Hits' was set to 600 for each screening method. The best pharmacophore model retrieved 596 hits from the NCI database while 595 retrieved from the IBS database. Molecules obtained from both databases were merged and there was 1 redundant molecule found in the merged file. The remaining 1190 compounds of the dataset were considered under "*Ligand Pharmacophore Mapping*" protocol of DS with "Maximum Omitted Feature" set to '0' to calculate the predicted inhibitory activity. After successful prediction, we compared the predicted activity of screened compounds with the experimental inhibitory activity (IC_{50} = 1.47 μM) of the most active compound (**D5** in Figure 1) of the dataset. Compounds with estimated inhibitory activity less than 1.47 were considered for further analysis. It was observed that 657 compounds fulfilled the above criteria and were considered to pass the Lipinski's Rule of Five and Viber's rule. 430 compounds failed to pass both rules. Therefore the remaining 227 molecules along with **D5** were analysed in a molecular docking study. On completion of molecular docking study it was observed that 32

molecules failed to fit into the active site of DNA GyrB. Dock scores of the remaining 195 and **D5** were compared, and we considered molecules that gave a higher dock score than **D5** (25.844). It was observed that 190 compounds had dock score higher than 25.844 and these then considered for further study. In addition, ADMET descriptors were calculated for more screening. The human intestinal absorption (HIA), aqueous solubility and blood brain barrier (BBB) were analysed and we found that one compound from NCI (**NSC10716**) and two from IBS (**STOCK1N01341** and **STOCK1N05628**) databases show good absorption, aqueous solubility and penetration values (Figure 6). Therefore the above three molecules were considered to be promising DNA GyrB inhibitors and were analysed further to assess the critical interactions with the catalytic amino residues of DNA GyrB.

Molecular docking

In order to observe preferred orientation and binding interactions of final screened molecules along with most active compound of the dataset, the dock complexes of the same were analysed. Best docked poses of **D5**, **NSC10716**, **STOCK1N01341** and **STOCK1N05628** are given in Figure 7.

In order to perform molecular docking the crystal structure of DNA GyrB (PDB ID: 4B6C) was collected from RCSB-Protein Data Bank. Validation of molecular docking protocol is an important phase and in this regards self-docking(53) is one of the crucial approaches in which already bound small molecule is re-docked at the catalytic site of macro molecule and the conformer of the original bound small molecule is superimposed to the re-docked pose (Figure S1 in supplementary file) to calculate RMSD value. As per report RMSD < 2 Å value of original bound ligand validates the docking procedure(53). In the current study the RMSD value between co-crystal and docked conformer was found to be 1.305 Å, which clearly indicated that the protocol selected in the docking method was validated.

It was observed from the docked pose between most active compound (**D5**) of the dataset and DNA GyrB that three catalytic amino residues (Arg82, Ile84 and Pro85) in the receptor cavity interacted with **D5**. In detail, two hydrogen bonds were formed between Arg82 and **D5**, while Ile84 and Pro85 formed one bump interaction each with **D5**. All three screened compounds were found to interact with Glu56 and Val99. Each of **NSC10716** and **STOCK1N01341** interacted with Glu56 and Val99 via one hydrogen bond and two bump interactions, whereas **STOCK1N05628** was found to clash with Glu56 and Val99 by one bump interaction separately. In addition Glu56 formed one hydrogen bond with

STOCK1N05628. Asn52, Asp79, Arg82 and Thr169 amino residues were crucial for interaction with **NSC10716** via two, one, four and one bump interactions respectively. Moreover, Arg82 was formed from two potential hydrogen bonds with **NSC10716**. Both **STOCK1N01341** and **STOCK1N05628** interacted with Asp97 via one bump interaction separately. **STOCK1N01341** was able to form hydrogen bond with Arg41, His89 and Gln102 via one, one and two connections respectively. Two amino residues, Arg82 and Pro85 were found to be important at the receptor site of DNA GyrB to form two and three bump interactions respectively with **STOCK1N05628**.

Molecular dynamics

For an in depth analysis of complex stability and to explain the inhibition mechanisms of the screened compounds and the most active compounds of the dataset, molecular dynamics study was performed. In this regards, Gromacs5.0.6 was used for a time span of 10ns. Backbone RMSD, RMSF and radius of gyration (Rg) were explored to analyse the complex constancy during simulation time and portrayed in Figure 7.

The average RMSD values of protein backbone were found to be 0.294, 0.327, 0.260 and 0.202 nm for the DNA GyrB complex with **D5**, **NSC10716**, **STOCK1N01341** and **STOCK1N05628** respectively. On analysis of RMSD trajectories (Figure 8) it was observed that average RMSD value of **NSC10716** was found to be higher compared to the others. It was observed that compounds screened from IBS database achieved lower RMSD values compared to the most active compound (**D5**) of the dataset. Further it can be noticed that RMSD of **STOCK1N01341** and **STOCK1N05628** fluctuated initially but both achieved stability with new confirmations after about 7ns at around 0.25nm. The complex with **D5** also attained stability after approximate 9ns of time span. In the case of the complex of protein with **NSC10716**, it was found that after up to about 3ns the RMSD value increases after it has attained stability, but at around 9.5ns the value increases which explains the instability of the complex. Average RMSF of complexes with **D5**, **NSC10716**, **STOCK1N01341** and **STOCK1N05628** were found to be 0.380, 0.235, 0.295 and 0.360 nm respectively, while differences between maximum and minimum RMSF were perceived as 0.331, 0.189, 0.300 and 0.313 nm correspondingly. The RMSF trajectories (Figure 8) revealed that backbone of the all complexes fluctuated more around amino acid Pro85 to Thr95, Leu140 to Trp146 and Gln160 to Lys165. For complexes with **STOCK1N05628** and **NSC10716** respectively RMSF fluctuation was found more around Pro155 and Ala175. In order to correlate

conformational variations of the protein with respect to the protein's initial structure Rg was recorded from all the complexes.

The trajectories of the Rg is depicted in Figure 9. The trajectories explained that complexes with both **STOCK1N01341** and **STOCK1N05628** achieved almost similar Rg, while in case of complex with **D5,** Rg is reduced up to 6ns and increases afterwards up to the mark of the complex with **STOCK1N01341**. It was also noticed that in the complex with **NSC10716** the Rg value decreases and after 10ns of time it is almost unchanged. The above findings of MD simulation clearly explained that the complex with both **STOCK1N01341** and **STOCK1N05628** achieved stable conformation at low RMSD while complex with **D5** reached stability at higher RMSD value. Moreover the complex with **NSC10716** failed to show stability up to 10ns of time span. It was observed that both RMSD and RMSF analysis successfully correlate with findings of the molecular docking study. High Rg value of both complexes with **STOCK1N01341** and **STOCK1N05628** indicates the accessiblity of the ligand to receptor cavity by opening of the binding pocket of the protein molecule. The lower Rg value of the complex with **NSC10716** indicates that there was no change of compactness of the protein molecule.

Comparison of drug-likeness with FDA approved DNA GyrB inhibitors

A few standard DNA Gyrase inhibitors approved by the FDA including **Bedaquiline**, **Moxifloxacin** and **Novobiocin** were considered for a comparison of the drug-likeness of screened compounds from databases. Different drug-like properties of FDA-approved and final screened compounds were calculated and given in Table 4. Parameters are included as dock score, estimated activity, fit value, hydrophobicity, molecular weight, violation of Lipinski's rule of five, molecular volume, molecular refractivity, number of H-bonds and number of bump interactions.

Analysis of different drug-like properties (Table 4) revealed that with the exception of **Bedaquiline** all molecules docked inside the receptor cavity of the DNA GyrB. The dock score of **Novobiocin** was found to be highest (100.694) whereas dock score of screened compounds were recorded higher than **Moxifloxacin** and **D5**. The inhibitory activity and fit value were calculated by fitting compounds on the pharmacophore model. Both estimated activity and fit value explained that all three screened compounds may have more potential than **Bedaquiline**, **Moxifloxacin** and **D5** as promising DNA GyrB inhibitors. **Novobiocine**,

Bedaquiline and **D5** also violated ROF 3, 2 and 2 rules respectively whilst final screened compounds satisfied all the rules. The molecular docking study explained that **NSC10716**, **STOCK1N01341** and **STOCK1N05628** interacted with a greater number of hydrogen bonds and bump interactions. The molecular weight, molecular volume and molecular refractivity were also recorded and listed in Table 4. The above comparison results revealed that **NSC10716**, **STOCK1N01341** and **STOCK1N05628** may be promising DNA GyrB inhibitors for therapeutic application in TB.

Conclusions

A chemical feature-based pharmacophore model was developed from a set of DNA GyrB inhibitors to explore the structural and orientational factors important for potential inhibition of DNA GyrB. Among the several generated pharmacophore models from a set of 30 training set molecules, the best ten hypotheses were selected for further evaluation. Hypotheses were internally and externally validated using R, Q^2, se, r_m^2, R_{pred}^2, sp, $r_{m(test)}^2$, $\Delta r_{m(test)}^2$, Fischer's randomization and decoy set. The selected model explained that two of each of hydrogen bond acceptors and hydrophobic regions were crucial for inhibitory activity. The best hypothesis was used for virtual screening of NCI and IBS databases to retrieve promising DNA GyrB inhibitors. The initial 1191 hits from both databases were passed through a number of criteria and finally one from NCI (**NSC10716**) and two from IBS database (**STOCK1N01341** and **STOCK1N05628**) were found to be promising for inhibition of DNA GyrB. The final selected three molecules with the most active molecule of the dataset were subjected to molecular docking study to explore binding interactions in the receptor cavity of the protein. Molecular docking study revealed that screened compounds were able to form a number of binding interactions with the catalytic amino residues of DNA GyrB. Furthermore, the complex between DNA GyrB and screened compounds along with most active compound of the dataset were subjected to molecular dynamics simulation study. RMSD, RMSF and Rg values from MD simulation study explained that both screened compounds from the IBS database might be promising inhibitors but **NSC10716** requires some modification to achieve the goal. Finally it can be concluded that the final screened compounds might be promising lead candidates for the treatment of TB but further confirmation will require experimental validation, *in vitro*.

Acknowledgment

MA Islam and TS Pillay were funded by the National Research Foundation (NRF), South Africa Innovation post-doctoral fellowship scheme.

Conflict of Interest

The authors declare that they have no conflicts of interest with the contents of this article.

Figure 1: 2D chemical structures of the training set compounds and the inhibitory activity values (pIC_{50}) are given in the parentheses. (pIC_{50} = log[(1/IC_{50})x1000]).

Figure 2: a) Mapped pharmacophore features (*Hypo 1*) with most active compound; b) Inter-feature distances of *Hypo 1*

Figure 3: Observed and predicted inhibitory activity of DNA GyrB inhibitors as per *Hypo 1*.

Figure 4: Comparison of total costs of best model (*Hypo 1*) and 19 random hypotheses generated in the Fischer's randomization test.

Figure 5: ROC curve for pharmacophore model derived from true positive rate of actives vs. false positive rate of inactive compounds.

Figure 6: Screened promising DNA GyrB inhibitors from NCI (**NSC10716**) and IBS (**STOCK1N01341** and **STOCK1N05628**) databases.

Figure 7: Binding modes of the most active molecules of the dataset and final screened compounds from databases.

Figure 8: Plot of RMSD vs. simulation time (left) and RMSF vs. residue number (right).

Figure 9: Radius of gyration of Cα atoms of DNA GyrB over the simulation time

References

1. Reddy KI, Srihari K, Renuka J, Sree KS, Chuppala A, Jeankumar VU, et al. (2014) An efficient synthesis and biological screening of benzofuran and benzo[d]isothiazole derivatives for Mycobacterium tuberculosis DNA GyrB inhibition. *Bioorganic & medicinal chemistry;***22**: 6552-63.
2. http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf.
3. Frieden TR, Sterling T, Pablos-Mendez A, Kilburn JO, Cauthen GM, Dooley SW (1993) The emergence of drug-resistant tuberculosis in New York City. *The New England journal of medicine;***328**: 521-6.
4. Centers for Disease C, Prevention (2006) Emergence of Mycobacterium tuberculosis with extensive resistance to second-line drugs--worldwide, 2000-2004. *MMWR Morbidity and mortality weekly report;***55**: 301-5.
5. Gandhi NR, Moll A, Sturm AW, Pawinski R, Govender T, Lalloo U, et al. (2006) Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. *Lancet;***368**: 1575-80.
6. (2015).
7. Parida SK, Axelsson-Robertson R, Rao MV, Singh N, Master I, Lutckii A, et al. (2015) Totally drug-resistant tuberculosis and adjunct therapies. *Journal of internal medicine;***277**: 388-405.
8. Migliori GB, Loddenkemper R, Blasi F, Raviglione MC (2007) 125 years after Robert Koch's discovery of the tubercle bacillus: the new XDR-TB threat. Is "science" enough to tackle the epidemic? *The European respiratory journal;***29**: 423-7.
9. Pooran A, Pieterson E, Davids M, Theron G, Dheda K (2013) What is the cost of diagnosis and management of drug resistant tuberculosis in South Africa? *PloS one;***8**: e54587.
10. Orenstein EW, Basu S, Shah NS, Andrews JR, Friedland GH, Moll AP, et al. (2009) Treatment outcomes among patients with multidrug-resistant tuberculosis: systematic review and meta-analysis. *The Lancet Infectious diseases;***9**: 153-61.
11. Kale MG, Raichurkar A, P SH, Waterson D, McKinney D, Manjunatha MR, et al. (2013) Thiazolopyridine ureas as novel antitubercular agents acting through inhibition of DNA Gyrase B. *Journal of medicinal chemistry;***56**: 8834-48.
12. Wong EB, Cohen KA, Bishai WR (2013) Rising to the challenge: new therapies for tuberculosis. *Trends in microbiology;***21**: 493-501.
13. Skripconoka V, Danilovits M, Pehme L, Tomson T, Skenders G, Kummik T, et al. (2013) Delamanid improves outcomes and reduces mortality in multidrug-resistant tuberculosis. *The European respiratory journal;***41**: 1393-400.
14. Migliori GB, Langendam MW, D'Ambrosio L, Centis R, Blasi F, Huitric E, et al. (2012) Protecting the tuberculosis drug pipeline: stating the case for the rational use of fluoroquinolones. *The European respiratory journal;***40**: 814-22.
15. Medapi B, Renuka J, Saxena S, Sridevi JP, Medishetti R, Kulkarni P, et al. (2015) Design and synthesis of novel quinoline-aminopiperidine hybrid analogues as Mycobacterium tuberculosis DNA gyraseB inhibitors. *Bioorganic & medicinal chemistry;***23**: 2062-78.
16. Shirude PS, Madhavapeddi P, Tucker JA, Murugan K, Patil V, Basavarajappa H, et al. (2013) Aminopyrazinamides: novel and specific GyrB inhibitors that kill replicating and nonreplicating Mycobacterium tuberculosis. *ACS chemical biology;***8**: 519-23.
17. Aubry A, Fisher LM, Jarlier V, Cambau E (2006) First functional characterization of a singly expressed bacterial type II topoisomerase: the enzyme from Mycobacterium tuberculosis. *Biochemical and biophysical research communications;***348**: 158-65.

18. Chopra S, Matsuyama K, Tran T, Malerich JP, Wan B, Franzblau SG, et al. (2012) Evaluation of gyrase B as a drug target in Mycobacterium tuberculosis. *The Journal of antimicrobial chemotherapy;***67**: 415-21.
19. Champoux JJ (2001) DNA topoisomerases: structure, function, and mechanism. *Annual review of biochemistry;***70**: 369-413.
20. Jeankumar VU, Renuka J, Santosh P, Soni V, Sridevi JP, Suryadevara P, et al. (2013) Thiazole-aminopiperidine hybrid analogues: design and synthesis of novel Mycobacterium tuberculosis GyrB inhibitors. *European journal of medicinal chemistry;***70**: 143-53.
21. Kaur P, Agarwal S, Datta S (2009) Delineating bacteriostatic and bactericidal targets in mycobacteria using IPTG inducible antisense expression. *PloS one;***4**: e5923.
22. Laggner C, Schieferer C, Fiechtner B, Poles G, Hoffmann RD, Glossmann H, et al. (2005) Discovery of high-affinity ligands of sigma1 receptor, ERG2, and emopamil binding protein by pharmacophore modeling and virtual screening. *Journal of medicinal chemistry;***48**: 4754-64.
23. Steindl T, Laggner C, Langer T (2005) Human rhinovirus 3C protease: generation of pharmacophore models for peptidic and nonpeptidic inhibitors and their application in virtual screening. *Journal of chemical information and modeling;***45**: 716-24.
24. van Drie JH (2003) Pharmacophore discovery--lessons learned. *Current pharmaceutical design;***9**: 1649-64.
25. Debnath AK (2003) Generation of predictive pharmacophore models for CCR5 antagonists: study with piperidine- and piperazine-based compounds as a new class of HIV-1 entry inhibitors. *Journal of medicinal chemistry;***46**: 4501-15.
26. Wei J, Wang S, Gao S, Dai X, Gao Q (2007) 3D-pharmacophore models for selective A2A and A2B adenosine receptor antagonists. *Journal of chemical information and modeling;***47**: 613-25.
27. Studio D. (2014). San Diego, USA: Accelrys Inc.
28. Al-Balas QA, Amawi HA, Hassan MA, Qandil AM, Almaaytah AM, Mhaidat NM (2013) Virtual lead identification of farnesyltransferase inhibitors based on ligand and structure-based pharmacophore techniques. *Pharmaceuticals;***6**: 700-15.
29. Huang D, Zhu X, Tang C, Mei Y, Chen W, Yang B, et al. (2012) 3D QSAR pharmacophore modeling for c-Met kinase inhibitors. *Medicinal chemistry;***8**: 1117-25.
30. Chhabria MT, Brahmkshatriya PS, Mahajan BM, Darji UB, Shah GB (2012) Discovery of novel acyl coenzyme a: cholesterol acyltransferase inhibitors: pharmacophore-based virtual screening, synthesis and pharmacology. *Chemical biology & drug design;***80**: 106-13.
31. Medapi B, Meda N, Kulkarni P, Yogeeswari P, Sriram D (2016) Development of acridine derivatives as selective Mycobacterium tuberculosis DNA gyrase inhibitors. *Bioorganic & medicinal chemistry;***24**: 877-85.
32. Medapi B, Suryadevara P, Renuka J, Sridevi JP, Yogeeswari P, Sriram D (2015) 4-Aminoquinoline derivatives as novel Mycobacterium tuberculosis GyrB inhibitors: Structural optimization, synthesis and biological evaluation. *European journal of medicinal chemistry;***103**: 1-16.
33. Chandran M, Renuka J, Sridevi JP, Pedgaonkar GS, Asmitha V, Yogeeswari P, et al. (2015) Benzothiazinone-piperazine derivatives as efficient Mycobacterium tuberculosis DNA gyrase inhibitors. *International journal of mycobacteriology;***4**: 104-15.
34. Jeankumar VU, Renuka J, Pulla VK, Soni V, Sridevi JP, Suryadevara P, et al. (2014) Development of novel N-linked aminopiperidine-based mycobacterial DNA gyrase B

inhibitors: scaffold hopping from known antibacterial leads. *International journal of antimicrobial agents;***43**: 269-78.

35. Li H, Sutter J, Hoffman R. (1999) An automated system for generating 3D predic-tive pharmacophore models. In: Guner OF, editor *An automated system for generating 3D predic-tive pharmacophore models*. La Jolla,CA: International University Line: p. 173–89.
36. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry;***4**: 187–217.
37. Momany FA, Rone R (1992) Validation of the general purpose QUANTA ®3.2/CHARMm® force field. *Journal of Computational Chemistry;***13**: 888–900.
38. Abrahama MJ, Murtolad T, Schulzb R, Pall P, Smithb JC, Hessa B, et al. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX;***1-2**: 19-25.
39. Smellie A, Teig SL, Towbin P (1995) Poling: Promoting conformational variation. *Journal of Computational Chemistry;***16**: 171-87.
40. Kristam R, Gillet VJ, Lewis RA, Thorner D (2005) Comparison of conformational analysis techniques to generate pharmacophore hypotheses using catalyst. *Journal of chemical information and modeling;***45**: 461-76.
41. Sadler BR, Cho SJ, Ishaq KS, Chae K, Korach KS (1998) Three-dimensional quantitative structure-activity relationship study of nonsteroidal estrogen receptor ligands using the comparative molecular field analysis/cross-validated r2-guided region selection approach. *Journal of medicinal chemistry;***41**: 2261-7.
42. Li H, Sutter J, Hoffman R (2000) *Pharmacophore Perception, Development, and Use in Drug Design*. California: International University Line.
43. Kubinyi H, Hamprecht FA, Mietzner T (1998) Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices. *Journal of medicinal chemistry;***41**: 2553-64.
44. Roy K, Mitra I, Kar S, Ojha PK, Das RN, Kabir H (2012) Comparative studies on some metrics for external validation of QSPR models. *Journal of chemical information and modeling;***52**: 396-408.
45. Ojha PK, Mitra I, Das RN, Roy K (2011) Further exploring rm2 metrics for validation of QSPR models. *Chemometrics and Intelligent Laboratory Systems;***107**: 194–205.
46. Golbraikh A, Tropsha A (2002) Beware of q2! *Journal of molecular graphics & modelling;***20**: 269-76.
47. Mitra I, Saha A, Roy K (2010) Pharmacophore mapping of arylamino-substituted benzo[b]thiophenes as free radical scavengers. *Journal of molecular modeling;***16**: 1585-96.
48. Roy PP, Paul S, Mitra I, Roy K (2009) On two novel parameters for validation of predictive QSAR models. *Molecules;***14**: 1660-701.
49. Roy PP, Roy K (2008) On Some Aspects of Variable Selection for Partial Least Squares Regression Models. *QSAR & Combinatorial Science;***27**: 302–13.
50. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of computational chemistry;***31**: 671-90.
51. Debnath AK (2002) Pharmacophore mapping of a series of 2,4-diamino-5-deazapteridine inhibitors of Mycobacterium avium complex dihydrofolate reductase. *Journal of medicinal chemistry;***45**: 41-53.

52. Sakkiah S, Thangapandian S, John S, Kwon YJ, Lee KW (2010) 3D QSAR pharmacophore based virtual screening and molecular docking for identification of potential HSP90 inhibitors. *European journal of medicinal chemistry;***45**: 2132-40.
53. Taha MO, Habash M, Al-Hadidi Z, Al-Bakri A, Younis K, Sisan S (2011) Docking-based comparative intermolecular contacts analysis as new 3-D QSAR concept for validating docking studies and in silico screening: NMT and GP inhibitors as case studies. *Journal of chemical information and modeling;***51**: 647-69.

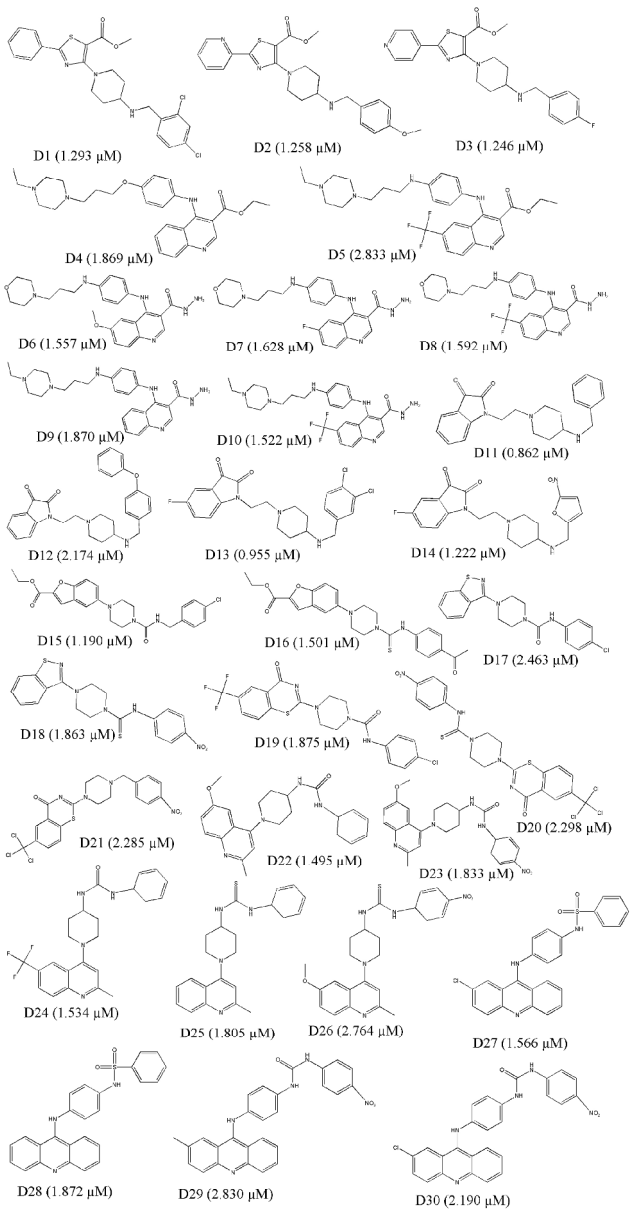


Figure 1: 2D chemical structures of the training set compounds and the inhibitory activity values (pIC50) are given in the parentheses. (pIC50 = log[(1/IC50)x1000]).

Figure 1
170x326mm (300 x 300 DPI)

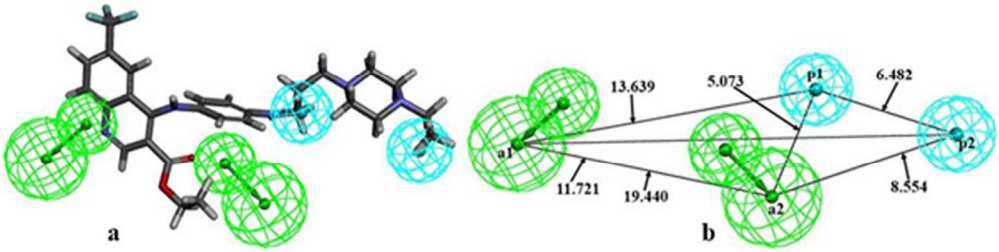


Figure 2: a) Mapped pharmacophore features (Hypo 1) with most active compound; b) Inter-feature distances of Hypo 1
Figure 2
129x34mm (300 x 300 DPI)

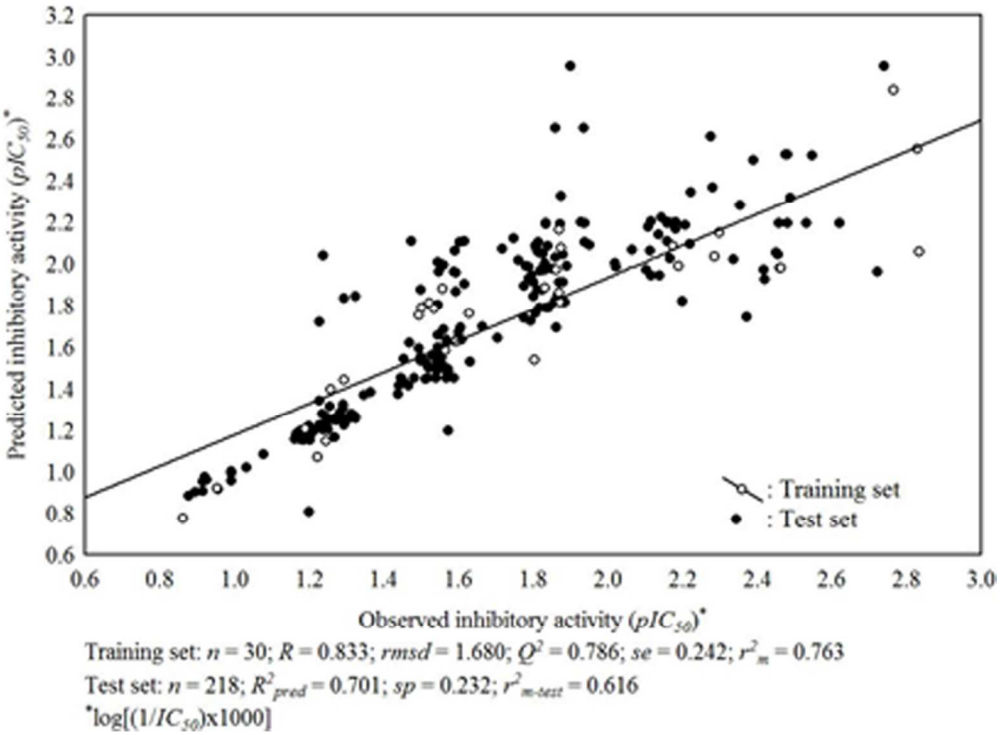


Figure 3: Observed and predicted inhibitory activity of DNA GyrB inhibitors as per Hypo 1.
Figure 3
80x59mm (300 x 300 DPI)

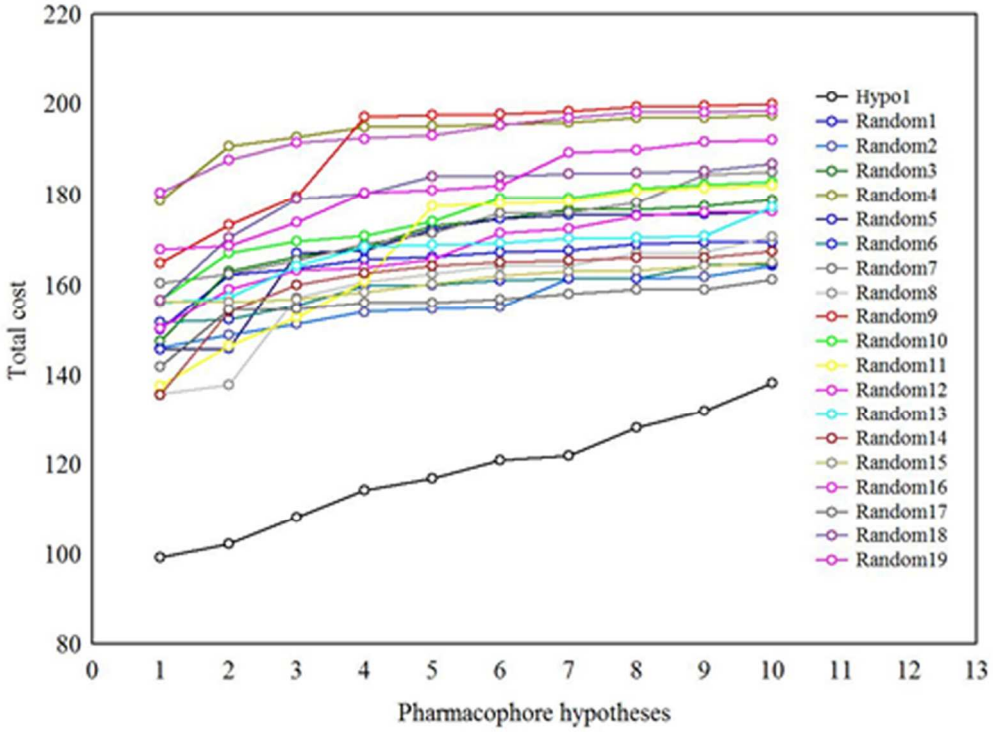


Figure 4: Comparison of total costs of best model (Hypo 1) and 19 random hypotheses generated in the Fischer's randomization test.
Figure 4
80x59mm (300 x 300 DPI)

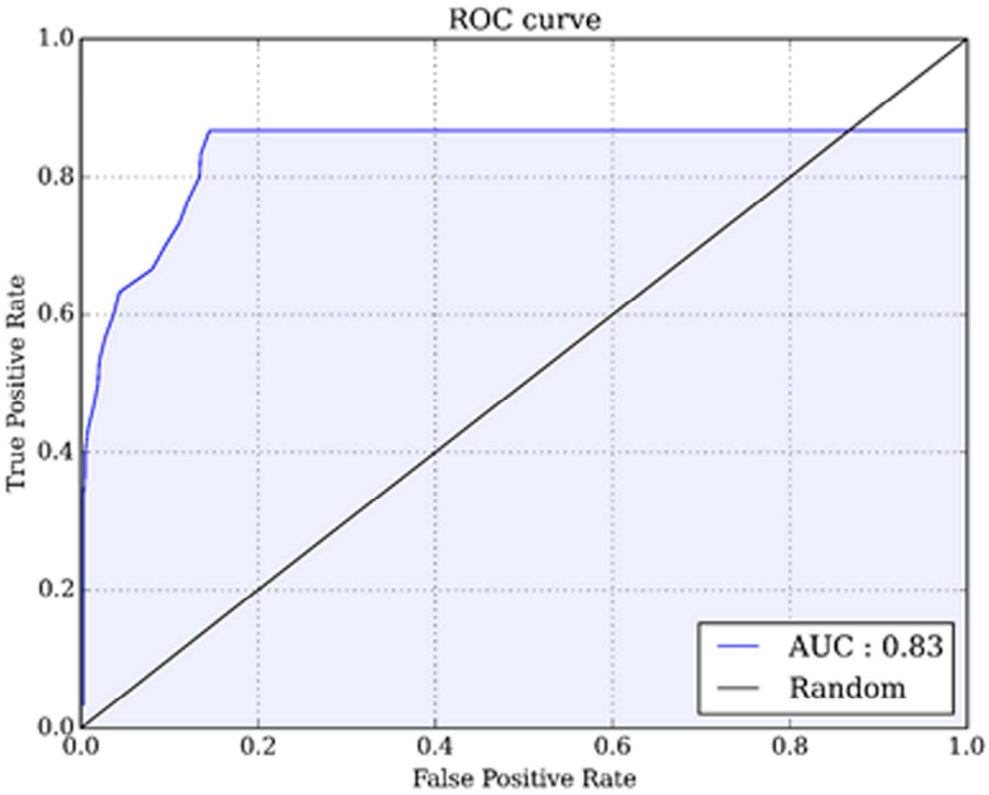


Figure 5: ROC curve for pharmacophore model derived from true positive rate of actives vs. false positive rate of inactive compounds.
Figure 5
80x63mm (300 x 300 DPI)

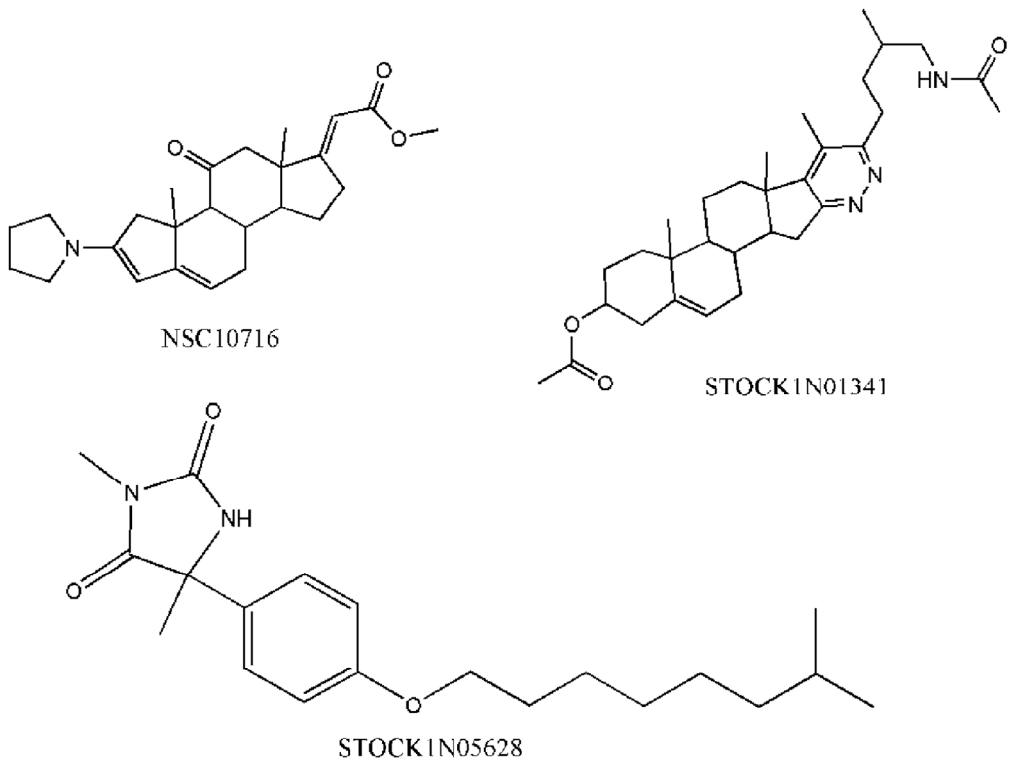


Figure 6: Screened promising DNA GyrB inhibitors from NCI (NSC10716) and IBS (STOCK1N01341 and STOCK1N05628) databases.
Figure 6
129x98mm (300 x 300 DPI)

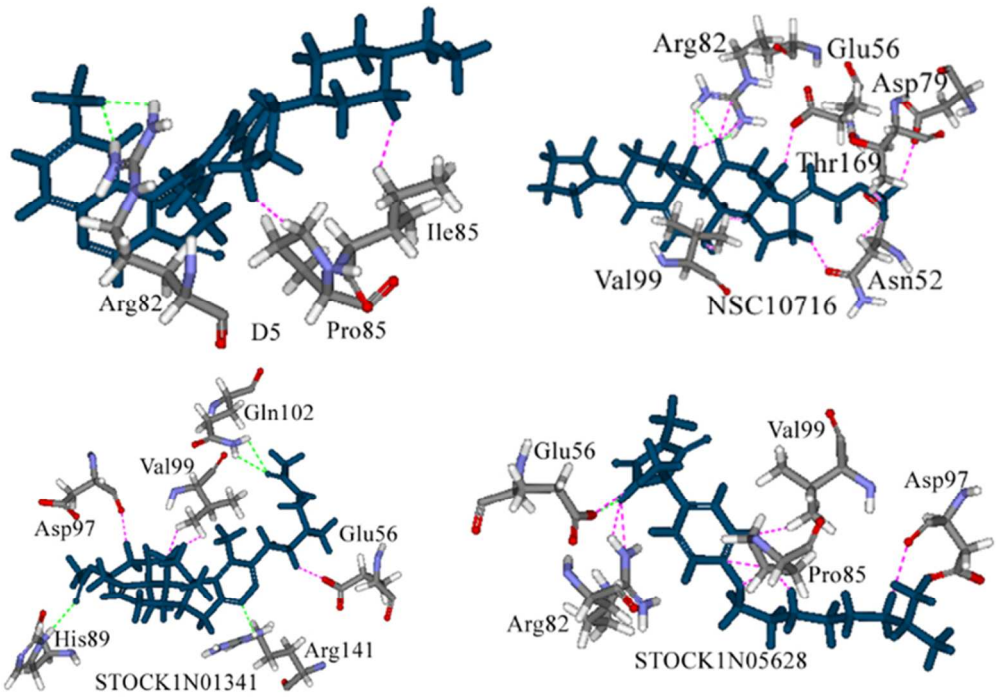


Figure 7: Binding modes of the most active molecules of the dataset and final screened compounds from databases.
Figure 7
80x55mm (300 x 300 DPI)

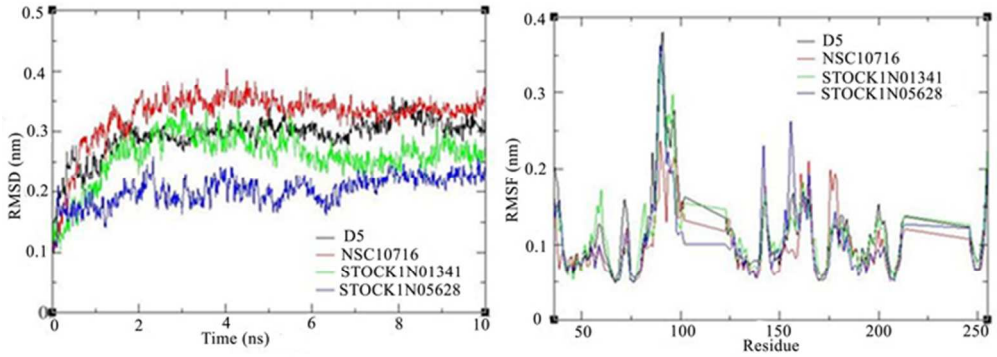


Figure 8: Plot of RMSD vs. simulation time (left) and RMSF vs. residue number (right).
Figure 8
63x23mm (300 x 300 DPI)

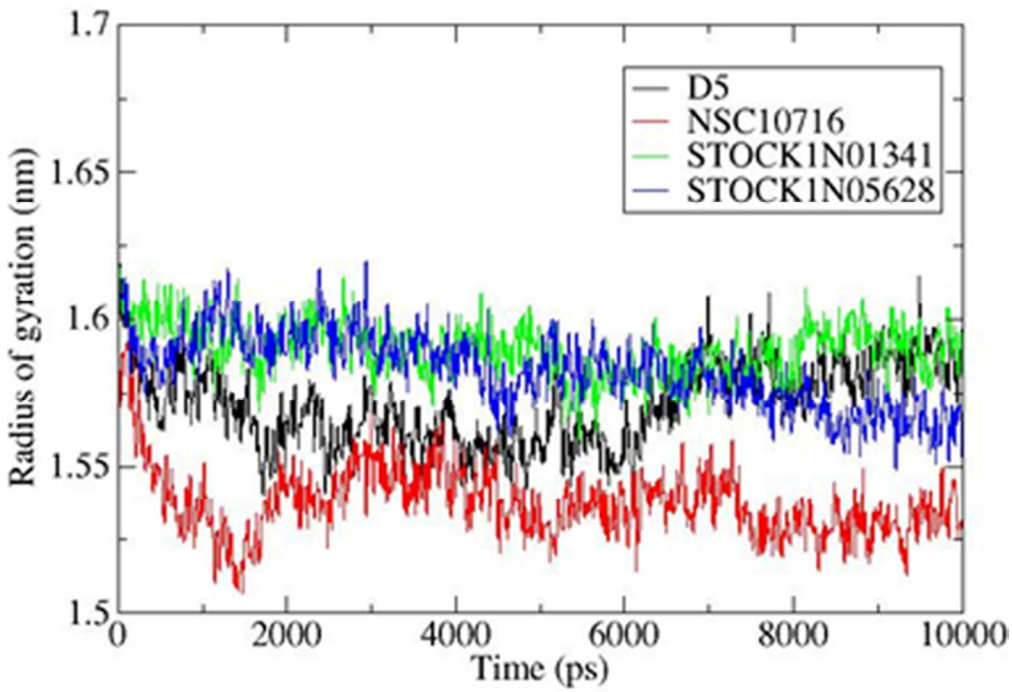


Figure 9: Radius of gyration of Cα atoms of DNA GyrB over the simulation time.
Figure 9
55x38mm (300 x 300 DPI)