Knowledge Based Identification of Potent Antitubercular Compounds Using Structure Based Virtual Screening and Structure Interaction Fingerprints

Ashutosh Kumar,[†] Vinita Chaturvedi,[‡] Shalini Bhatnagar,[‡] Sudhir Sinha,[‡] and
Mohammad Imran Siddiqi*,[†]

Molecular and Structural Biology Division and Drug Target Discovery and Development Division, Central
Drug Research Institute, Lucknow 226001, India

In view of the worldwide spread of multidrug resistance of *Mycobacterium tuberculosis*, there is an urgent need to discover antitubercular agents with novel structures. Thymidine monophosphate kinase from *M. tuberculosis* (TMPK*mt*) is an attractive target for antitubercular chemotherapy. We report here the identification of potent antitubercular compounds targeting TMPK*mt* using virtual screening methods. For this purpose we have developed a pharmacophore hypothesis based on the substrate and known TMPK*mt* inhibitors and employed it to screen the Maybridge small molecule database. The molecular docking was then performed in order to select the compounds on the basis of their ability to form favorable interactions with the TMPK*mt* active site. In addition, we applied straightforward weighting using structure interaction fingerprints to include additional knowledge into structure based virtual screening. Eight compounds were acquired and evaluated for antitubercular activity against *M. tuberculosis* H37Rv *in vitro*, and out of these 3 compounds showed MIC of 3.12 μg/mL whereas 2 compounds showed MIC of 12.5 μg/mL. All the active compounds were found to be nontoxic in Vero cell lines and mice bone marrow macrophages. All the identified hits highlighted a key hydrogen bonding interaction with Arg74. The observed π-stacking interaction with Phe70 was also produced by the identified hits. These hits represent promising starting points for structural optimization in hit-to-lead development.

1. INTRODUCTION

Tuberculosis (TB) is the most prevalent infectious disease worldwide and remains the leading cause of mortality, with about 2 million deaths annually.[1,2] Increasing reports of drug resistant *Mycobacterium tuberculosis* strains, namely, the virtually incurable extensive drug resistant TB (XDR-TB) and serious multidrug resistant TB (MDR-TB), are the cause of great concern, as they are resistant to both first-line and second-line TB drugs.[3] Over the decade, it is estimated that as many as 50 million people worldwide have been infected with MDR-TB strains. According to WHO, from 2002 to 2020, there will be about 1 billion more people newly infected with TB and approximately 36 million deaths if the worldwide ravage of tuberculosis is left unchecked.[2] Despite the increasing worldwide incidence of TB and its alarming threat toward the public health, no novel antituberculosis drugs have been introduced into clinical practice over the past 4 decades. The impact of ever increasing drug resistance, the serious side effects of some current anti-TB drugs, and the lack of efficacy of current treatments in immunodepressed patients combine to make the development of new antimycobacterial agents an urgent priority.

M. tuberculosis thymidine monophosphate kinase (TMPK*mt*) recently emerged as an attractive target for the design of new generation antitubercular entities. TMPK*mt* belongs to the nucleoside monophosphate kinase (NMPK) family and catalyzes the reversible phosphorylation of deoxythymidine 5′-monophosphate (dTMP) to deoxythymidine 5′-diphosphate (dTDP) in the presence of adenosine triphosphate (ATP). Situated at the junction of the *de novo* and salvage pathways for the synthesis of deoxythymidine 5′-triphosphate (dTTP), TMPK is the last specific enzyme in these pathways, therefore being essential for cell growth and survival.[4] Also, the sequence of TMPK*mt* when compared with that of its human isozyme shows only 22% sequence identity.[5] These characteristics make TMPK*mt* one of the potential targets for the design of new antitubercular drugs.

Initially it was considered that the potential inhibitors of TMPK*mt* should resemble the natural substrate dTMP and therefore should incorporate a monophosphate moiety. Significant strides have been made in the development of dTMP analogues that exhibit inhibitory potency against TMPK*mt*.[6−10] However, such compounds suffer from several drawbacks, such as degradation by phosphatases. Also, the low permeability of the cell wall for phosphorylated compounds (i.e., nucleotides) prompted medicinal chemists to focus on nucleosides as new drug leads. Subsequent studies lead to the identification of nucleoside thymidine as a moderately potent inhibitor of TMPK*mt* (K_i = 27 μM). Further, to enhance the affinity and potency of thymidine based compounds for the mycobacterial enzyme, both the sugar[6,9,11] and the base[7,12] moieties of thymidine have been the subject of different modifications. The most active TMPK*mt* inhibitors reported so far show K_i values in the low micromolar range, but biological testing performed so far with these inhibitors has not led to potent mycobacterial

* Corresponding author e-mail: imsiddiqi@yahoo.com.
† Molecular and Structural Biology Division.
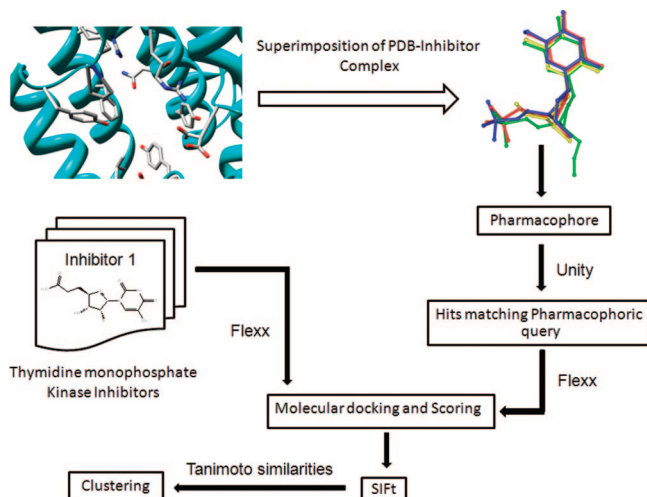‡ Drug Target Discovery and Development Division.

Figure 1. Scheme of the structure based virtual screening for identification of antitubercular compounds targeting TMPK*mt*. Initially, Maybridge small molecule database was subjected to 3D pharmacophore search based on the common features of known inhibitors. In the next step, resulting hits were subjected to molecular docking in order to select the compounds on the basis of their ability to form favorable interactions with TMPK*mt* active site. In the final step, selection of potent compounds was done based on visual inspection and structure interaction fingerprints based weighting.

inhibition. The utility of targeting TMPK*mt* for the development of antitubercular entities was unclear considering no correlation between TMPK*mt* inhibition and antimycobacterial activity. To probe the consequences of TMPK*mt* inhibition on mycobacterial viability, we sought to identify novel TMPK*mt* inhibitors to discover its potential therapeutic indications against tuberculosis.

In the present study, we have performed a structure-based virtual screen against TMPK*mt* employing a combination of 3D pharmacophore filtering, docking, and structure interactions fingerprints based weighting to identify chemotypes that can be used to develop novel chemical series as antitubercular agents. Starting from 59275 Maybridge structures we selected, acquired, and tested 8 compounds for activity against *M. tuberculosis*. This resulted in the identification of five novel compounds displaying antitubercular properties.

2. MATERIALS AND METHODS

2.1. Data Set. To evaluate the performance of knowledge based weighting using structure interaction fingerprints and to aid in the selection of potent virtual screening hits, a total of 110 TMPK*mt* inhibitors ranging from low micromolar to low nanomolar inhibition were collected from the literature,[6−12] and duplicates were removed. The 110 compounds are all nucleotides and nucleosides analogues and encompass acyclic thymidine analogues, bicyclic thymidine analogues, 2′ and 3′ modified thymidine analogues, and thymidine 5′-*O*-monophosphate analogues. The 3D structures of all ligands were built using the Builder module of INSIGHTII 2000.1,[13] and the geometry was optimized using the CVFF force field.[13]

2.2. 3D Pharmacophore Search. A common feature based 3D pharmacophore query with distance constraints was defined using the crystal structure bound conformation of the substrate deoxythymidine monophosphate (dTMP, PDB entry code 1N5L) and three potent inhibitors viz. deoxythymidine, azidothymidine monophosphate, and deoxyuridine
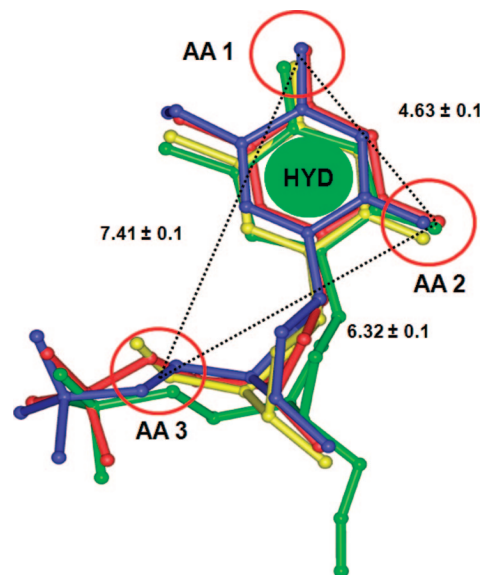

Figure 2. 3D pharmacophore query generated based on the common structural features of substrate deoxythymidine monophosphate (dTMP) and PDB complexed inhibitors deoxythymidine, azidothymidine monophosphate, and deoxyuridine monophosphate (shown in red, yellow, green, and blue, correspondently). The hydrogen bond acceptor feature is denoted by AA and the hydrophobic feature by HYD. The size of the sphere denotes the steric tolerance, and the hydrogen bond acceptor features are constrained by using distance constraints.
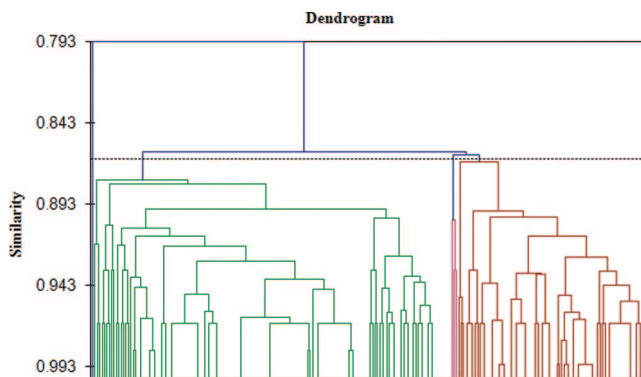

Figure 3. Dendrogram derived by agglomerative hierarchical clustering of structure interaction fingerprints (SIFt) of TMPK*mt* inhibitors and virtual screening hits. Tanimoto similarity coefficient was used to calculate similarity between the SIFts.

monophosphate (PDB entry codes 1W2G, 1W2H, and 1MRS, respectively). The 3D pharmacophore search was performed by using the Unity flexible search protocol as implemented in SYBYL 7.1, with all options set as default.[14] In the Unity search, the conformations of the screening database were generated on the fly by means of the Directed Tweak method.[15]

2.3. Molecular Docking and Scoring. Hits obtained from 3D pharmacophore search were docked into the binding pocket of TMPK*mt* using the FlexX program interfaced with SYBYL7.1.[16] Standard parameters of the FlexX program as implemented in SYBYL7.1 were used during docking. To further evaluate the docking experiment, the G_Score,[17] PMF_Score,[18] D_Score,[19] and ChemScore[20] values were estimated using the CScore module of SYBYL7.1.[14]

2.4. Construction of Structure Interaction Fingerprints (SIFt). SIFt stands for Structural Interaction Fingerprint, which is a 1D binary fingerprint representation of the
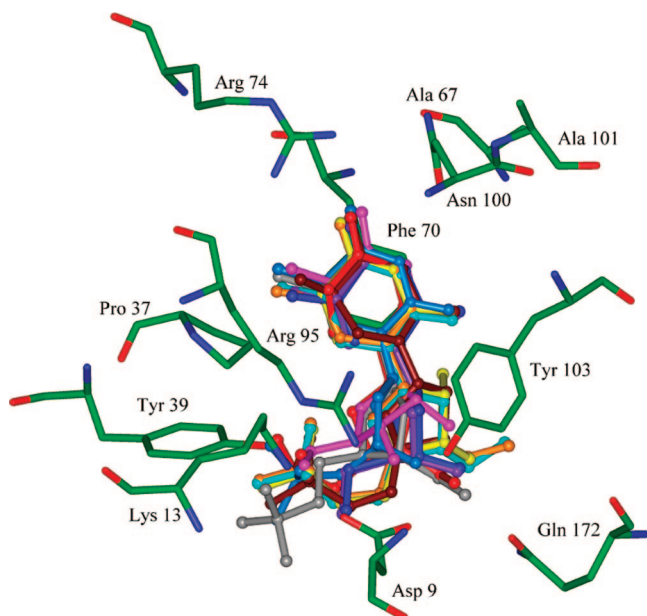
IDENTIFICATION OF POTENT ANTITUBERCULAR COMPOUNDS

J. Chem. Inf. Model., Vol. 49, No. 1, 2009 **37**

$$Tc(A, B) = |A \cap B| / |A \cup B|$$

where $A \cap B$ is the number of ON bits common in both A and B, and $A \cup B$ is the number of ON bits present in either A or B. The SIFt represents the binding mode of a ligand to a target protein; similar fingerprints imply that the corresponding ligands make similar interactions with the protein. We applied a hierarchical clustering methodology to analyze the fingerprints for each test case. Interaction fingerprints were clustered by using an agglomerative hierarchical clustering approach,[27] applying the Tanimoto coefficient as similarity measurements. SYSTAT12[28] was used for hierarchical clustering of SIFt. Clusters of protein−ligand complex structures were manually selected based on the dendrogram of their interaction fingerprints.

2.6. Assay for Antituberculosis Activity and in Vitro Cytotoxicity. The 'proportion method', described by McClachy,[29] was followed. In brief, 0.1 mL of vehicle or serial dilution (in vehicle) of antituberculosis drugs (isoniazid and
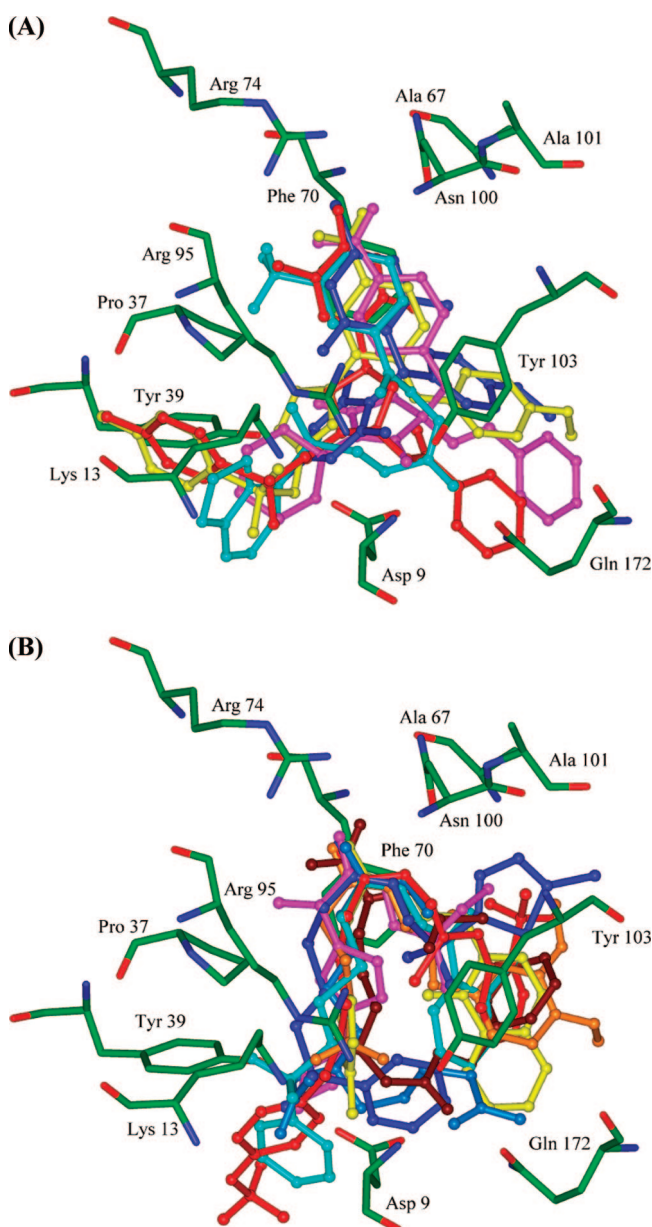

Figure 4. Active site of TMPK*mt*, displaying the binding mode of some TMPK*mt* inhibitors.

intermolecular interactions in a 3D protein-inhibitor complex. The SIFt protocol and implementation reported previously[21−24] was used to generate the interaction fingerprints. The SIFt is generated by first defining the union of those residues that are in contact between the protein and the small molecule complex. A uniform list of amino acid residues involved in ligand binding was used for calculating the SIFt for all the ligands. In addition, we identified the protein atoms that were involved in hydrogen bonding interactions with the ligands, using the program HBPLUS[25] with default settings. The program calculated and listed all possible hydrogen bond donor and acceptor pairs in the structure that satisfy predefined geometric criteria. The hydrogen bonding pairs between protein and ligand were extracted for subsequent analysis. After all the ligand binding site residues are identified and all the protein−ligand intermolecular interactions are calculated, the next step was to classify these interactions. Current implementation of SIFt uses seven bits for each binding site residue, representing seven different types of interactions. The seven bits are switched on or off if the following interactions are observed: 1) if a contact is involved at this position; 2) whether it involves the main-chain atoms; 3) whether it involves the side chain; 4) whether it is a polar interaction; 5) whether it is a nonpolar interaction; 6) whether it is a hydrogen bond acceptor; and 7) whether it is a hydrogen bond donor. By doing so, each residue is represented by a seven-bit-long bit string. The whole interaction fingerprint of the complex is finally constructed by sequentially concatenating the bit string of each binding site residue together, according to the ascendant residue number order. This results in each interaction fingerprint being the same length, enabling easy comparison of interactions at a particular binding site position across a series of complexes.

2.5. Similarity Analysis and Hierarchical Clustering of Structure Interaction Fingerprints. We have used the Tanimoto coefficient (Tc), commonly used for binary data,[26] as the quantitative measure of bit string similarity. The Tc between two bit strings A and B is defined as

(A)

(B)

Figure 5. Binding mode of virtual screening hits belonging to major cluster or cluster 1: (a) compounds displaying π-π interactions and (b) compounds displaying C−H···π and N−H···π interactions. Some representatives of structures are shown.
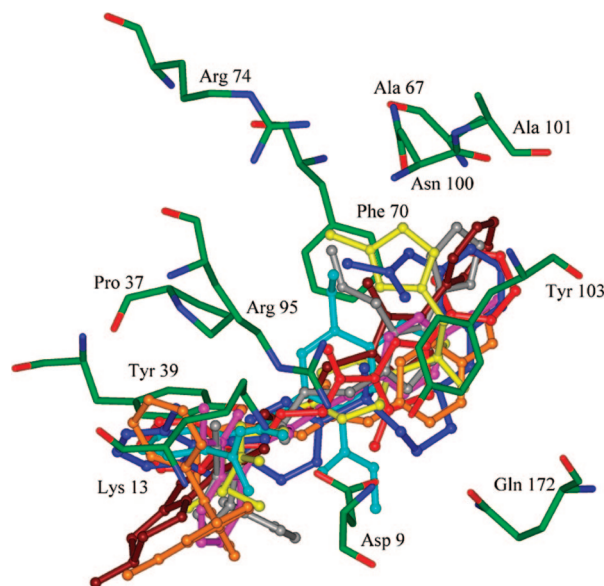
Figure 6. Binding mode of virtual screening hits belonging to cluster 2 or inactive cluster.

rifampicin) or test compounds were mixed with 1.9 mL of culture medium (Middlebrook 7H10 agar) and allowed to solidify in tubes as 'slants'. A culture of *M. tuberculosis* H37Rv was inoculated in the tubes (10^5 bacilli/ tube) and incubated for 4 weeks at 37 °C. The lowest concentration of a drug/compound which caused no visible growth of the inoculated bacilli was its minimal inhibitory concentration (MIC). *in Vitro* cytotoxicity was determined by the method of Mosmann[30] using Vero cells and mouse macrophages. In brief, cells were dispensed in 96-well culture plates (20,000 cells/well) and allowed to adhere for 48 h in a CO_2 incubator. Test compounds (in different concentrations) were added to the wells and incubated for a further 24 h. Finally, MTS (tetrazolium compound) solution was added to the wells and incubated for 2 h. Optical densities were read at 490 nm using a plate reader. A compound was considered as potentially toxic if its IC_{50} (concentration causing 50% reduction in Vero or macrophage cell viability) was ≤10 times its MIC for *M. tuberculosis* H37Rv.

3. RESULT AND DISCUSSION

The virtual screening procedure employed in this study is similar to previous studies where pharmacophore filtering has been used in conjunction with docking and scoring[31−33] with addition of the utility of structure interaction fingerprints to prioritize compounds for biological testing. The general strategy for the pharmacophore based virtual screening using the structure interaction fingerprints based weighting pursued in the present study is presented in Figure 1. In summary, a 3D pharmacophore query was initially used to screen the Maybridge small molecule database consisting of 59275 compounds. Subsequently, compound matching desired pharmacophoric features were docked and prioritized using receptor interaction information.

3.1. 3D Pharmacophore Search. We first adopted a 3D model of the TMPK*mt* binding pocket by superimposing the four PDBs (PDB entry codes 1N5L, 1W2G, 1W2H, and 1MRS) of TMPK*mt* in complex with substrate deoxythymidine monophosphate and inhibitors deoxythymidine, azidothymidine monophosphate, and deoxyuridine monophos-

phate. Common features were deduced from four ligands, and a pharmacophore hypothesis was constructed comprising a hydrophobic feature of the nucleobase and three hydrogen bond acceptors representing two carbonyl groups and a phosphate oxygen (Figure 2). One out of three hydrogen bond acceptors in the hypothesis makes hydrogen bonding interaction with Arg74, and a hydrophobic feature makes significant hydrophobic interactions with Pro37 and Phe70. In addition to hydrophobic interactions, the aromatic group representing the hydrophobic feature here is involved in π-π stacking interactions with Phe70. The pharmacophore hypothesis was generated using the UNITY program available with Sybyl7.1, and three acceptor features were constrained using distance constraints. The pharmacophore hypothesis was then used as a 3D structural query for retrieving potent molecules from the Maybridge small molecule database consisting of 59275 molecules. Virtual screening was carried out using flex utility of the UNITY module available with Sybyl7.1. Pharmacophore based virtual screening yielded 1156 hits that met the specified requirements.

3.2. Molecular Docking and Scoring. All the hit compounds retrieved from the pharmacophore based screening were then subjected to molecular docking to the TMPK*mt* binding site along with 110 known actives. The hydrophobic binding pocket of TMPK*mt* kinase is made up of key residues Asp9, Lys13, Pro37, Tyr39, Ala67, Phe70, Arg74, Arg95, Asn 100, Ala101, Tyr103, and Gln172. The TMPK*mt* crystal structure in complex with azidothymidine monophosphate (PDB entry code 1W2H) was used for docking studies. The molecular docking was carried out using the FlexX program, and 30 distinct poses of each ligand in the active site were generated. In the case of the known TMPK*mt* inhibitors only the best scoring pose for each ligand was further selected. The best pose was determined using the CScore module of Sybyl7.1 which generates the consensus of FlexX_Score and four other scoring functions (G_Score, PMF_Score, D_Score, and ChemScore). The docked pose with the highest CScore and FlexX_Score value was finally selected as the best pose for known TMPK*mt* inhibitors. In the case of virtual screening hits retrieved using common feature based pharmacophore query, only those hits were selected for which the top FlexX scoring pose had a good FlexX score (−36.475 to −20 KJ/mol) with a CScore value of 5 i.e. holds true for all five scoring functions (FlexX_score, G_Score, PMF_Score, D_Score, and ChemScore). Finally 112 virtual hits were taken as promising candidates for further investigation.

3.3. Structure Interaction Fingerprint Based Prioritization. A common problem in structure based virtual screening is that some compounds are ranked well by scoring functions postdocking, although their respective pose is barely in the binding site leading to false enrichments.[34] Second, and importantly, Warren et al. have recently observed that, from an assessment of 35 scoring functions, none were able to reliably identify the best-docked pose against a set of different targets.[35] Obviously, this leads to difficulty in prioritizing leads in a structure based virtual screening campaign against a particular target. To overcome this pitfall, we incorporated a receptor knowledge based weighting approach using structure interaction fingerprints (SIFt) in our virtual screening study to ensure that only realistic binders are prioritized. The knowledge based scoring scheme presented in this study is based on the incorporation of
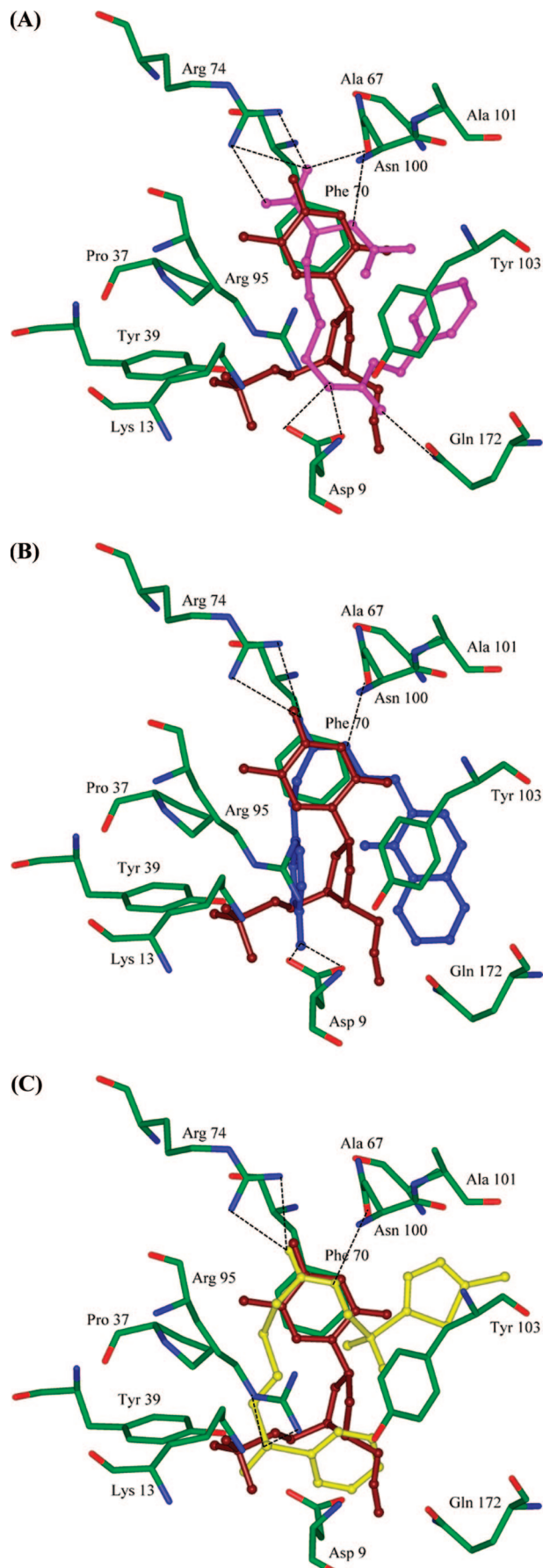
IDENTIFICATION OF POTENT ANTITUBERCULAR COMPOUNDS

J. Chem. Inf. Model., Vol. 49, No. 1, 2009 **39**

receptor−ligand interaction information from known actives to the receptor. Patterns of interactions were modeled using binary ligand−receptor fingerprints or SIFt by the method described by Singh and co-workers, explained in brief in Materials and Methods section. The SIFts were generated for 110 known actives and 112 virtual screening hits selected after molecular docking and conventional scoring, and SIFts were clustered using agglomerative hierarchical clustering algorithm. The hierarchical clustering groups similar SIFts together and is able to group structures into meaningful clusters where variable ligands have similar interactions with receptor. After clustering the SIFts, one could trace the active clusters (cluster where the known actives fall). This will enable the straightforward short listing of compounds and the binding modes as it would allow the users to focus attention on just the lead clusters.

The dendrogram derived by clustering SIFts of known actives and virtual screening hits is shown in Figure 3. The dendrogram revealed two major clusters each of which represents a distinct binding pattern in the ligand protein complexes. Cluster 1 composed of 75.45% inhibitors of TMPK*mt* (mostly highly active and medium active) and 50% virtual screening hits interacting with the protein. Similarly cluster 2 is composed of 21.82% TMPK*mt* inhibitors mostly low active and 50% virtual screening hits. The remaining 2.73% of the TMPK*mt* inhibitors do not belong to these two major clusters, and they are either singleton or form a tiny third cluster. Interestingly, each of these clusters is comprised of poses having similar binding modes with the receptor; cluster 1 contains molecules more or less similar to the known X-ray crystal structures. Cluster 2 compounds are similar in position but represent distinct binding mode that result in dissimilar interactions with the active site pocket formed by residues Phe70, Arg74, and Asp 100 and the phosphate binding loop known as P-Loop. Finally, third tiny cluster compounds are outside the inhibitor binding site.
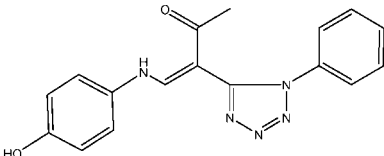
TMPK*mt* inhibitors belonging to major cluster or cluster 1 prefer one single binding position in the binding pocket. In the case of most of the TMPK*mt* inhibitors nucleobase occupies that same space buried deep in the TMPK*mt* active site and makes π-π stacking interactions with Phe70 aromatic side chain (Figure 4). This stacking interaction is very important and has been conserved in all the TMPKs for which crystal structures have been solved in the NAPK family. The tail of the inhibitors points to the outside of the enzyme through an exit channel surrounded by the residues Lys13, Pro37, Tyr39, and Asp9 (Figure 4). Some other interactions are also found to occur in most of the TMPK*mt* inhibitors viz. a hydrogen bond between O4 of the thymine base and the Arg74 side chain which results in a preference for thymine over cytosine, a hydrogen bond between Asn100 and N3 of the thymine ring, and hydrogen bonds with a terminal carboxyl of Asp9 and a side chain nitrogen of Arg95, respectively. A CH-π interaction is also found to occur in most of the inhibitors between the π system of Tyr103 and the C2 of the ribose ring.

Figure 5a,b shows the binding mode of some virtual screening hits belonging to cluster 1 as suggested by molecular docking. The predicted binding mode suggests that the bound conformation of virtual screening hits strongly mimics the observed conformation of TMPK*mt* inhibitors in the TMPK*mt* active site. The moieties like hydrazide,

Figure 7. Binding mode of active hits (a) BTB13566, (b) RF00102, and (c) KM03098 shown with TMPKmt inhibitor azidothymidine monophosphate (brown). Dashed lines indicate a hydrogen bond between hits and active site residues.

Table 1. Biological Activity, Docking Scores, Molecular Weight, and LogP of Identified Virtual Screening Hits[a]

S. No.	Maybridge Id	Structure	FlexX Score	Mol. Wt	LogP	MIC (µg/ml)
1.	JFD 01835*		-34.204	500.528	2.41	-
2.	BTB 13566		-28.907	322.361	1.822	3.12
3.	KM 09131		-27.395	526.482	4.525	12.5
4.	PD 00361		-27.277	342.372	0.575	NA
5.	RF 00102		-26.667	336.347	2.462	3.12
6.	KM 03101		-26.531	412.918	2.688	NA
7.	KM 03098		-25.529	455.000	2.561	3.12
8.	KM 06452		-25.516	459.429	2.693	NA
9.	RJC 02526*		-24.246	392.420	0.451	-

IDENTIFICATION OF POTENT ANTITUBERCULAR COMPOUNDS

J. Chem. Inf. Model., Vol. 49, No. 1, 2009 **41**

Table 1. Continued

S. No.	Maybridge Id	Structure	FlexX Score	Mol. Wt	LogP	MIC (µg/ml)
10.	NRB 05016		-23.405	321.340	1.378	12.5

a The compounds designated with an asterisk were not tested as they were out of stock.

pyrimidine, and phenyl present in the hits belonging to cluster 1 are predicted to occupy the TMPK*mt* catalytic cleft, playing the role of nucleobase in the natural substrate dTMP, burying deep within the binding groove, and reaching the catalytic residue Phe70 and are involved in either π-π interactions (some representative compounds shown in Figure 5a) or other π-hydrogen bonds like C−H•••π and N−H•••π (some representative compounds shown in Figure 5b). These moieties are also predicted to be involved in strong hydrogen bonding interactions with conserved residues Arg74 and Asn100. In most cases the remaining portion of the virtual screening hits protrude toward the outside of the catalytic pocket and occupy the neighborhood of Lys13, Tyr39, Tyr103, and Gln172. Here, the corresponding portions of virtual screening hits are involved in hydrophobic interactions with Pro37, Tyr39, and Tyr103 and hydrogen bonding interaction with Asp9 and Arg95.

Hierarchical clustering of SIFts is able to classify some virtual screening hits and TMPK*mt* inhibitors (low active or inactive) into a different cluster where they are displayed in a different binding mode that separates them from rest of the hits (Figure 6). Interestingly, most of the hits are located in the active site and occupy the same spatial position as the known actives and virtual screening hits of cluster 1; however, these compounds are found to rotate by 45 degrees toward Tyr103. The compounds in this cluster do not have all their hydrogen bonding interactions with the protein in common and are clustered separately. It seems clear that the presence of various groups increases the steric hindrance inside the binding cavity and changes the optimal binding mode of the ligand, thus decreasing the relative stability of the complex.

In order to select potential hits for biological testing through visual inspection of the clusters and compound binding mode, the following criteria was considered: (1) the compound should be from cluster 1, (2) the degree of occupancy of the enzyme in particular related to the achieved protein ligand surface complementarities, (3) the distance and the orientation of the aromatic, aliphatic, or hydrophobic group in relation to Phe70 for π-stacking interaction, (4) the formation of a hydrogen bond with Asp 9, Arg74, Arg95, and Asn100, and (5) the quality of the overall binding conformation, from which 10 Maybridge compounds were selected for biological assay and are presented in Table 1. Two of the selected compounds were not available, and, therefore, finally 8 compounds were tested for antitubercular activity *in vitro*.

3.4. Biological Assay. The selected compounds were evaluated for their *in vitro* antitubercular activity against *M.*

tuberculosis H37Rv in Agar dilution assay. The minimum concentrations of the compounds required for the complete inhibition of the bacterial growth per spot (minimum inhibitory concentration, MIC) are summarized in Table 1. Isoniazid and rifampicin was used as the standard drug. Among the selected hits, Maybridge compounds BTB 13566, RF 00102, and KM 03098 showed activity at MICs 3.12 µg/mL, whereas KM 09131 and NRB 05016 showed activity at 12.5 µg/mL. The compounds were also investigated for cytotoxicity in VERO cell line and in mouse macrophages. None of the test compounds showed cytotoxicity against the two cell types.

3.5. Binding Mode of Active Hits. To obtain structural insight into the inhibitory mechanism by newly identified active compounds, their binding modes in the TMPK*mt* active were investigated. The predicted binding modes of active compounds BTB 13566, RF 00102, and KM03098 are described in Figure 7 along with the crystal structure bound conformation of inhibitor azidothymidine monophosphate. Like already documented TMPK*mt* inhibitors, for example azidothymidine monophosphate shown in brown color in Figure 7, active compounds are all anchored to the cavity by the combination of π-stacking interactions between the ligand and conserved amino acid Phe70, hydrogen bonding interactions with Asp9, Arg74, Arg95, and Asn100, and strong hydrophobic interactions with Pro37, Tyr39, Phe70, and Tyr103. The bound conformation of BTB13566, RF00102, and KM03098 in the TMPK*mt* active site is stabilized by a combination of several weak π-hydrogen bonds between the ligand and Phe70, rather than aromatic face to face π-π stacking interaction with Phe70, as seen in known TMPK*mt* inhibitors. The hydrazide core in KM03098 and RF00102 is involved in N−H•••π interactions with the π-system of Phe70, whereas acetyl amino acetic acid moiety in another Maybridge hit BTB13566 is making C−H•••π and N−H•••π hydrogen bonds with Phe70s π cloud. The presence of these weak π hydrogen bonds may be responsible for the optimal orientation of ligand functional groups to make strong hydrogen bonding interactions with Asp9, Arg74, Arg95, and Asn100. As seen from Figure 7, the hydrazide oxygen in RF00102 and KM03098 makes hydrogen bonding interaction with the side chain nitrogen of Arg74, whereas the hydrazide nitrogen in these compounds hydrogen bonds with the backbone oxygen of Asn100. The sulfate oxygen in KM03098 and phenyl hydroxy in RF00102 interacts with the side chain nitrogen of Arg95 and the side chain oxygen of Asp9, respectively, and contributes to the stability of the complex. BTB13566 interacts to TMPK*mt* via four strong hydrogen bonding interactions. In particular,

the acetylamino acetic acid group was predicted to be in close proximity with Arg74 and Asn100 where it forms hydrogen bonds with side chain nitrogen and backbone oxygen. The terminal phenyl in BTB13566 is also involved in stabilizing hydrophobic interactions with Tyr103.

4. CONCLUSION

Here, we have applied virtual screening protocol including 3D pharmacophore search, molecular docking, and receptor based weighting using Structure Interaction Fingerprints (SIFt) to identify potent virtual screening hits against TMPK*mt*, a potential therapeutic target against tuberculosis. Use of Structural Interaction Fingerprints makes it possible to quickly shortlist promising compounds and automated visualization of putative active binding modes using hierarchical clustering. In addition, we believe that knowledge based weighting can enable researchers to apply virtual screening and docking tools for more upstream drug discovery research projects, where the aim would be to suggest new compounds for old targets. This study should provide further insights to support structure-based design of antitubercular agents with improved activity profiles.

REFERENCES AND NOTES

(1) Kaufmann, S. H.; McMichael, A. J. Annulling a dangerous liaison: vaccination strategies against AIDS and tuberculosis. *Nat. Med.* **2005**, *11*, S33–44.

(2) WHO World Health Organization. Factsheet on tuberculosis; 2005. http://www.who.int/mediacentre/factsheets/fs104/en/index.html (accessed Nov 3, 2008).

(3) CDC. Emergence of *Mycobacterium tuberculosis* with extensive resistance to second-line drugs -worldwide, 2000-2004. *MMWR* **2006**, *55*, 301–305.

(4) Anderson, E. P. In *The Enzymes*; Boyer, P. D.; Ed.; Academic Press: New York, 1973; Vol. 8, pp 49−96.

(5) Sierra, L. D. L.; Munier-Lehmann, H.; Gilles, A. M.; Bârzu, O.; Delarue, M. X-ray structure of TMP kinase from Mycobacterium tuberculosis complexed with TMP at 1.95 Å resolution. *J. Mol. Biol.* **2001**, *311*, 87–100.

(6) Vanheusden, V.; Munier-Lehmann, H.; Pochet, S.; Herdewijn, P.; Van Calenbergh, S. Synthesis and evaluation of thymidine-5′-O-monophosphate analogues as inhibitors of Mycobacterium tuberculosis thymidylate kinase. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2695–2698.

(7) Haouz, A.; Vanheusden, V.; Munier-Lehmann, H.; Froeyen, M.; Herdewijn, P.; Van Calenbergh, S.; Delarue, M. Enzymatic and structural analysis of inhibitors designed against Mycobacterium tuberculosis thymidylate kinase. New insights into the phosphoryl transfer mechanism. *J. Biol. Chem.* **2003**, *278*, 4963–4971.

(8) Munier-Lehmann, H.; Pochet, S.; Dugue, L.; Dutruel, O.; Labesse, G.; Douget, D. Design of Mycobacterium tuberculosis thymidine monophosphate kinase inhibitors. *Nucleosides, Nucleotides Nucleic Acids* **2003**, *22*, 801–804.

(9) Vanheusden, V.; Van Rompaey, P.; Munier-Lehmann, H.; Pochet, S.; Herdewijn, P.; Van Calenbergh, S. Thymidine and thymidine-5′-O-monophosphate analogues as inhibitors of Mycobacterium tuberculosis thymidylate kinase. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3045–3048.

(10) Vanheusden, V.; Munier-Lehmann, H.; Froeyen, M.; Dugue, L.; Heyerick, A.; De Keukeleire, D.; Pochet, S.; Busson, R.; Herdewijn, P.; Van Calenbergh, S. 3′-C-branched-chain-substituted nucleosides and nucleotides as potent inhibitors of Mycobacterium tuberculosis thymidine monophosphate kinase. *J. Med. Chem.* **2003**, *46*, 3811–3821.

(11) Vanheusden, V.; Munier-Lehmann, H.; Froeyen, M.; Busson, R.; Rozenski, J.; Herdewijn, P.; Van Calenbergh, S. Discovery of bicyclic thymidine analogues as selective and high-affinity inhibitors of Mycobacterium tuberculosis thymidine monophosphate kinase. *J. Med. Chem.* **2004**, *47*, 6187–6194.

(12) Pochet, S.; Dugue, L.; Labesse, G.; Delepierre, M.; Munier-Lehmann, H. Comparative study of purine and pyrimidine nucleoside analogues acting on the thymidylate kinases of Mycobacterium tuberculosis and of humans. *ChemBioChem* **2003**, *4*, 742–747.

(13) *INSIGHT II, Version 2000.1*; Accelrys Inc.: San Diego, U.S.A., 2000.

(14) *SYBYL Molecular Modeling System, Version 7.1*; TRIPOS Assoc. Inc.: St. Louis, MO, 2006.

(15) Hurst, T. Flexible 3D searching: The directed tweak technique. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 190–196.

(16) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.

(17) Jones, G.; Willett, P.; Glen, G. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 43–53.

(18) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791–804.

(19) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* **1992**, *13*, 505–524.

(20) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425–445.

(21) Deng, Z.; Chuaqui, C.; Singh, J. Structural Interaction Fingerprint (SIFt): A Novel Method for Analyzing Three dimensional Protein-Ligand Binding Interactions. *J. Med. Chem.* **2004**, *47*, 337–344.

(22) Chuaqui, C.; Deng, Z.; Singh, J. Interaction Profiles of Protein Kinase-Inhibitor Complexes and Their Application to Virtual Screening. *J. Med. Chem.* **2005**, *48*, 121–133.

(23) Deng, Z.; Chuaqui, C.; Singh, J. Knowledge-Based Design of Target-Focused Libraries Using Protein-Ligand Interaction Constraints. *J. Med. Chem.* **2006**, *49*, 490–500.

(24) Singh, J.; Deng, Z.; Narale, G.; Chuaqui, C. Structural Interaction Fingerprints: A New Approach to Organizing, Mining, Analyzing, and Designing Protein-Small Molecule Complexes. *Chem. Biol. Drug. Des.* **2006**, *67*, 5–12.

(25) McDonald, I. K.; Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **1994**, *238*, 777–793.

(26) Rogers, D. J.; Tanimoto, T. T. A Computer Program for Classifying Plants. *Science* **1960**, *132*, 1115–1118.

(27) Dubes, R.; Jain, A. K. Clustering Methodologies in Exploratory Data Analysis. *Adv. Comput.* **1980**, *19*, 113–228.

(28) *SYSTAT for Windows, Version 12*; SYSTAT Software Inc.: Richmound, CA, 2007.

(29) McClachy, J. K. Susceptibility testing of mycobacteria. *Lab. Med.* **1978**, *9*, 47–52.

(30) Mosmann, T. Rapid colorimeteric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *J. Immunol. Methods* **1983**, *65*, 55–63.

(31) Perola, E. Minimizing false positives in kinase virtual screens. *Proteins* **2006**, *64*, 422–435.

(32) Brenk, R.; Naerum, L.; Gradler, U.; Gerber, H. D.; Garcia, G. A.; Reuter, K.; Stubbs, M. T.; Klebe, G. Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. *J. Med. Chem.* **2003**, *46*, 1133–1143.

(33) Muthas, D.; Sabnis, Y. A.; Lundborg, M.; Karlen, A. Is it possible to increase hit rates in structure-based virtual screening by pharmacophore filtering? An investigation of the advantages and pitfalls of post-filtering. *J. Mol. Graphics Modell.* **2007**, *26*, 1237–1251.

(34) Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. Prediction of protein-ligand interactions. Docking and scoring: Successes and gaps. *J. Med. Chem.* **2006**, *49*, 5851–5855.

(35) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; Lalonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A Critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.

CI8003607