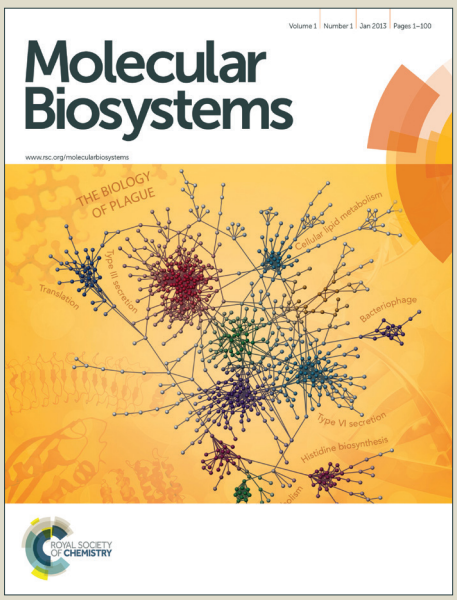
Molecular BioSystems

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Recognizing drug targets using evolutionary information: Implications for repurposing FDA-approved drugs against *Mycobacterium tuberculosis* H37Rv

Gayatri Ramakrishnan[1, 3], *Nagasuma R. Chandra*[2] *and Narayanaswamy Srinivasan*[3]*

[1]Indian Institute of Science Mathematics Initiative, [2]Department of Biochemistry, [3]Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, India.

[‡] *Corresponding author: N. Srinivasan*

Email: ns@mbu.iisc.ernet.in

Phone: +91-80-2293-2837

Fax: +91-80-2360-0535

Abstract

Drug re-purposing to explore target space has been gaining pace over the past decade with the upsurge in the use of systematic approaches for computational drug discovery. Such a cost and time-saving approach gains immense importance for pathogens of special interest, such as *Mycobacterium tuberculosis* H37Rv. We report a comprehensive approach to repurpose drugs, based on exploration of evolutionary relationships inferred from comparative sequence and structural analyses between targets of FDA-approved drugs and the proteins of *M. tuberculosis*. This approach has facilitated identification of several polypharmacological drugs that could potentially target unexploited *M. tuberculosis* proteins. A total of 130 FDA-approved drugs, originally intended against other diseases, could be repurposed against 78 potential targets in *M. tuberculosis*. Additionally, we have also made an attempt to augment the chemical space by recognizing compounds structurally similar to FDA-approved drugs. For three of the attractive cases we have investigated the probable binding modes of the drugs in their corresponding *M. tuberculosis* targets by means of structural modelling. Such prospective targets and small molecules could be prioritized for experimental endeavours, and could significantly influence drug-discovery and drug-development programmes for tuberculosis.

Keywords

Drug repurposing, Drug targets, Mycobacterium tuberculosis, Protein-ligand docking, Sequence analysis.

Introduction

Unlike the conventional *in vitro* "blind" screening programmes, drug discovery efforts over the past decade are increasingly becoming rational and are concentred on target-driven lead discovery. Novel technological initiatives, such as high-throughput screening, structure-based lead design and combinatorial chemistry have resulted in successful identification of compounds with desired therapeutic effect. Albeit promising, the present-day drug discovery efforts yield very few positive results against the enormous cost and time invested, thus forcing many biopharmaceutical companies to leverage investments on a more promising strategy- to re-purpose known drugs[1]. Such a strategy substantially reduces the time, effort and risk involved right from processing and development of the drug till it is commercialized. The significance of drug repurposing, i.e. the identification of new use for existing drugs, with or without any chemical modification, has been realized by various groups. In the recent past, several methodologies to repurpose drugs including *in silico* approaches based on binding site similarity[2, 3], side-effect similarity[4], ligand similarity[5, 6], structure-based virtual-screening[7] exploiting genomic and pharmacological spaces for drug repurposing[8-11], as well as *in vitro* techniques based on experimental screening of FDA-approved drugs against pathogens[12] have been well demonstrated. Indeed, such an effective strategy has been successfully implemented by several groups in an attempt to discover viable therapies for pathogenic infections[13, 14] as well as other diseases[15-17]. One of the several advantages of repurposing an FDA-approved drug for a pathogen is the lower probability of obtaining an "anti-target" in the host, given that the drug was originally developed to treat another pathogenic infection. The approach presented in this article delineates the computational identification and analyses of FDA-approved small molecules which could be repurposed for unexplored potential target proteins in *M. tuberculosis*, a causative agent of tuberculosis.

A persistent leading cause of mortality, tuberculosis continues to be a global burden, with high prevalence in Asia and Africa. Despite the availability of drug regimens, an estimated 9 million new cases and 1.5 million deaths were reported in the year 2013[18]. Current chemotherapeutic regimens include the use of four first-line drugs i.e. isoniazid, rifampicin, pyrazinamide and ethambutol which essentially inhibit the biosynthesis of key constituents of mycobacterial cell wall (mycolic acids, peptidoglycan and arabinogalactan) and inhibit bacterial

RNA synthesis. The first-line treatment is effective against drug-sensitive bacteria alone. Use of combination of second-line drugs i.e. amikacin, kanamycin, capreomycin, viomycin and fluoroquinolones is favoured only when the first-line treatment proves ineffective. Albeit clinically efficacious, the second-line drugs are reported to cause significant side-effects[19]. One of the major shortcomings with the current drug regimen is the emergence of multidrug (MDR) and extensively drug-resistant (XDR) strains that render first-line and second-line drug treatment futile. Such an emergence of drug-resistance strains usually arises either due to acquisition of resistance in the course of a patient's treatment or due to transmission of drug-resistant bacteria. This poses serious complications in the quest of eradicating the epidemic. The success rate of the existing drugs on MDR and XDR tuberculosis is minimal[20], thus entailing a need to explore drug target space in *M. tuberculosis,* explore the potential of existing drugs for repurposing[21, 22] and formulate selective combination of drugs for such a heterogeneous bacterial population; which continues to remain a tremendous challenge.

Established on the grounds of drug repurposing, we present a target identification methodology based on exploration of the evolutionary relationship between targets of FDA-approved drugs (approved for use against diseases other than tuberculosis) and *M. tuberculosis* proteins. A similar integration of pharmacological data and homology information has been demonstrated earlier[23], in order to determine within-target-family selectivity of small-molecules. With the help of sensitive remote-homology detection techniques it is possible to explore evolutionary relationships across species and draw inferences on related proteins and probable small molecule binding-sites. Such an approach has been pursued in the current study using FDA-approved drugs, their corresponding target proteins and their homologues in *M. tuberculosis*. The FDA-approved drugs were initially subjected to a filter to eliminate the ones known to act on human proteins, since the efficacy of a drug known to target proteins in human is questionable in its use as a repurposed drug against a pathogen. Results of enhanced structural and functional characterization of gene products encoded in *M. tuberculosis* proteins[24] served as a guiding tool to explore the functional importance of the potential targets identified. Employing sensitive profile-based homology detection approaches coupled with structural elucidations on binding sites and literature support, a total of 78 proteins in *M. tuberculosis* were recognized which could serve as potential targets for 130 FDA-approved drugs. Additionally, an attempt has been made to increase the chemical expanse by incorporating prospective inhibitors/compounds

which could be taken up further in the drug discovery pipeline against tuberculosis. For a selected drug-target association, protein-ligand docking was performed to assess the molecular details of the predicted interaction, thereby providing cues on binding mode and binding site for compounds structurally similar to the drug. The shortlisted FDA-approved drugs, inhibitors and potential targets identified in *M. tuberculosis* could provide an efficient platform to experimentally investigate the drug-target interactions predicted. The presented methodology is fairly generic and its applicability as a general framework for drug as well as target identification in other pathogens can be explored.

Methodology

Dataset

Sequences of gene products encoded in *M. tuberculosis* H37Rv genome were retrieved from TubercuList (http://tuberculist.epfl.ch/). The updated release of the database (release 27) as on March 2013 contains information on 4018 gene products, which has been used in the current study. DrugBank[25] (version 4.2), a database constituting comprehensive drug and target information, was used in extracting details on FDA-approved drugs and their associated target protein sequences. For a total of 1600 FDA-approved drugs, sequences of 1522 target proteins from diverse source organisms were retrieved. ChEMBL[26] (version 18) database, a collection of bioactive drug-like small molecules, was used to obtain details on inhibitors/compounds along with useful annotations on their clinical progress. This release holds 1,359,508 distinct compound records.

Approach to identify potential drug-target interactions in M. tuberculosis

i) Sequence analysis

Prior to the use of known target protein sequences associated with 1600 FDA-approved drugs, a filtering step was employed to eliminate drugs known to act on human proteins based on details pertaining to source organism of each target protein as well as affected organisms reported for a drug, as detailed in DrugBank 4.2. We also excluded drugs currently under use against tuberculosis. Thence 196 drugs linked with 155 target protein sequences were considered for further analysis. In order to identify evolutionary relationship between these target proteins and proteins encoded in *M. tuberculosis* genome, a sensitive profile-based iterative sequence search, jackhmmer availed through HMMER3.0 suite of programs[27], was employed with conditions on e-value cut-off of 0.0001 and 5 rounds of iteration. An alignment coverage cut-off of 70% or an alignment encompassing at least one functional and/or structural domain was used as an additional criterion to ensure reliability of the relationships recognised. The motive behind selection of such criteria was to capture the most of target sequence information in the context of functional and/or structural domains and eliminate short stretches of alignment. This step facilitated credible inferences on binding sites in the potential targets identified in *M. tuberculosis*.

ii) Structural elucidations: analyses on binding sites and evaluation

Structural information for all the target proteins considered was obtained from PDB[28]. Of 155 target proteins, molecular details on ligand-binding sites could be obtained for 73 targets (Supplementary Table 1). Structural information for the *M. tuberculosis* proteins homologous to known targets was retrieved from PDB. For *M. tuberculosis* proteins with no known structure, reliable structural models were obtained from ModBase[29]. ModBase is a large-scale collection of comparative protein structure models which are generated using an automated modelling pipeline ModPipe. This pipeline primarily relies on MODELLER[30] for fold assignment, sequence-structure alignment, model building and assessment. Models retrieved from ModBase were checked for reliability based on z-DOPE score (<0), ModPipe Quality Score (MPQS) cut-off of 1.1 and a query coverage threshold of 90%. DOPE score[31] or Discrete Optimized Protein Energy score is essentially an atomic distance-dependent statistical potential used to analyse quality of model. Criteria such as model coverage, compactness of the model, sequence identity and gaps in sequence-structure alignment are also considered to investigate the quality of the models[32, 33] which are summed up as MPQS. Quality of the models chosen from ModBase is reliable and therefore such models have the potential to be used with reasonable confidence in the assessment of putative ligand-binding sites. Structural models were also obtained using ModPipe through a web resource ModWeb (http://salilab.org/modweb) for instances where ModBase models are unavailable.

Comparative evaluation of binding sites across known targets and their homologues in *M. tuberculosis* were established at several levels based on availability of molecular details of target proteins:

a) Target-inhibitor complexes

   Cases where crystal structure of target-inhibitor complexes are available, a structural alignment algorithm, TM-align[34] facilitated identification and assessment of conservation of ligand-binding residues in the corresponding homologue in *M. tuberculosis*. TM-align recognizes local structural matches between protein pairs and assigns TM-score which provides a measure of structural similarity. TM-score typically acquires a value between 0 and 1, wherein a TM-score of > 0.50 depicts structural similarity corresponding to same fold and a TM-score < 0.30 corresponds to unconvincing structural similarity[34, 35].

b) Target-substrate complexes

For drugs which are substrate analogues known to exhibit competitive mechanism of inhibition, elucidation on binding-sites in predicted targets were established using TM-align with the help of substrate-bound complexes of known targets.

c) Unknown ligand-binding site

Instances where binding site of an inhibitor could not be determined directly from structural information, SiteMap[36, 37] availed through Schrödinger suite of programs was used to predict binding sites in potential targets identified in *M. tuberculosis*. SiteMap identifies probable ligand-binding sites and ranks them based on the site's propensity for ligand binding with the help of a scoring function, SiteScore. This scoring function, based on weighted sum of properties such as size of site, solvent exposure, hydrophobicity, hydrophilicity, average grid contact strength with the protein, and donor or acceptor characteristics, has the potential to capture pharmaceutically relevant binding sites. A SiteScore threshold of 0.80 was used. Additionally, SiteMap predicts and scores druggability of each potential binding site primarily based on hydrophobicity[37]. Such high confidence binding sites could be conveniently used to generate receptor grids for protein-ligand docking studies using Glide (Grid-based ligand docking with energetics)[38-40], provided by Schrödinger. The docking protocol of Glide includes four major stages. Briefly, the first two stages involve an extensive site search and conformational search for a given ligand, followed by series of hierarchical filters that examine the spatial fit of the ligand to a defined receptor site. The complementarity of receptor-ligand interaction is evaluated using a scoring function, GlideScore[38]. At the third stage, the ligand binding poses are refined and subsequently energy minimized using OPLS-AA (Optimized potentials for liquid simulation-all atom) force field. Finally, the resultant poses are ranked using a model energy score Emodel, which combines GlideScore and non-bonded interaction energies of the ligand binding poses generated.

The computational modelling of protein-ligand complexes has become an inevitable element in drug discovery and development[41, 42]. The significant advances made over the past few decades has resulted in the availability of more than 60 docking algorithms, that are broadly based on shape complementarity, fragment-based, evolutionary-based and stochastic Monte Carlo methods[43]. Most of the algorithms avoid systematic search of receptor flexibility and exploration of all degrees of freedom in the

ligand to reduce computational burden, and use a simplified scoring function (empirical-based) for computationally efficient evaluation of protein-ligand affinity at the cost of accuracy[44]. Devising a docking protocol which combines best possible scoring function, protein flexibility, protein-ligand conformational sampling, characteristics of protein-ligand binding and an efficient search algorithm, remains a challenge in the field of computational drug discovery[41, 45, 46].

The choice of the molecular docking program, Glide, has been made on the basis of numerous published studies which report the performance evaluation of widely used docking programs including Glide, GOLD[47-49], FlexX[50], AutoDOCK[51], DOCK[52, 53], ICM[54], Surflex[55] and LigandFit[56], through molecular docking and virtual screening trials[43, 57-66]. In most of the studies, Glide was reported to consistently outperform other docking programs, making it the preferred program for molecular docking studies. For a selected case, protein-ligand docking study was performed using Glide XP (extra precision) scoring function[40] to evaluate the predicted binding sites of a potential *M. tuberculosis* target and the putative effects of the small molecule on protein function were investigated. A schematic overview of the protocol followed is illustrated in Figure-1.

Augmenting the chemical space

The resultant shortlisted FDA-approved drugs were subjected to *in silico* screening against 1,359,508 compounds from ChEMBL (version 18) database by employing two-dimensional and three-dimensional molecular similarity calculations. This step enabled identification of compounds structurally similar to FDA-approved drugs, which can be investigated for their usefulness as anti-tubercular agents.

Fingerprint similarity searching was primarily used to determine similarity between the molecules by means of 2D molecular fingerprints. The simplicity of such fingerprints facilitates similarity calculations on a large-scale. A molecular fingerprint is a bit string or set of features representing substructures and properties of a molecule. In order to quantify the molecular similarities, Tanimoto coefficient[67, 68] was used, which quantifies the fraction of features common to pair of molecules in question to the total number of features in both the molecules. Open Babel (version 2.3.1)[69], an open-source toolbox to handle chemical data, was used to perform 2D molecular similarity searches for the shortlisted FDA-approved drugs against

ChEMBL compounds at a Tanimoto score cut-off of 0.9. Self-hits and salt formulations of FDA-approved drugs were discarded from ChEMBL dataset. The high scoring ChEMBL compounds were further re-screened against their corresponding FDA-approved drugs using a 3D molecular similarity program, SHAFTS (SHApe-FeaTure Similarity)[70, 71] which takes shape and pharmacophore features into account. This program provides a hybrid similarity measure which is a weighted sum of pharmacophore feature fit values (FeatureScore) and shape-densities overlap (ShapeScore), scaled to [0, 2]. The hybrid similarity score cut-off used was 1.0.

Results and Discussion

Potential drug candidates for repurposing

A total of 130 FDA-approved drugs were identified which are attractive candidates to repurpose against 78 potential *M. tuberculosis* targets. While majority of the approved drugs constituted antibacterial agents we could also recognize several antifungals, anti-malarial agents, leprostatic agents and anti-infectives which can be explored for their anti-tubercular activity. Based on the information on hierarchical rule-based structural classification of drugs obtained from DrugBank, the set of 130 FDA-approved drugs could be grouped into 19 classes. Figure-2 delineates the distribution of 130 drugs across drug classes and the number of potential *M. tuberculosis* targets predicted to be associated with each class. As depicted in the bar chart, the most populated drug class in terms of number of drugs pertains to lactams comprising of 51 drugs associated with 12 targets. This finding clearly implies the possibility of combination therapy using lactams, one of the well-studied strategies[22] that could stand tolerance against development of drug resistance in the pathogen. In addition, the effectiveness of other drug classes such as quinolones and derivatives, macrolides and analogues and carbohydrate conjugates in combination therapy can also be explored. On the other hand, we also noted several instances where a single drug is capable of acting on multiple targets. This observation is evident in the Figure-2 where 5 single entity drug classes in particular, steroids and steroid derivatives (DB02703), furans (DB00336), benzothiophenes (DB01153), benzimidazoles (DB00730) and azolidines (DB00698) are predicted to target multiple proteins in *M. tuberculosis*. Indeed, modulating activity of multiple targets with the help of single drug (polypharmacology) has been increasingly gaining interest in rationalizing drug design, discovery and development[7]. An initial filter on elimination of FDA-approved drugs known to act on human proteins excludes "promiscuous" drugs with undesirable polypharmacological potential. Hence, the multi-target drugs recognized in the analysis are capable of mainly targeting multiple *M. tuberculosis* proteins. Such multi-target drugs have an immense potential as a complement to the current drug regimen and as an alternative to combination therapy. Table-1 summarizes details on few of the interesting examples. The complete list is provided in Supplementary Table 2.

Potential drug targets identified

The potential *M. tuberculosis* targets could be classified under four functional categories: a) intermediary metabolism and respiration, b) information pathways, c) cell wall and cell processes and d) lipid metabolism, based on the annotations detailed in TubercuList. Notably, of the 78 *M. tuberculosis* proteins identified as targets, 40 are recognized as genes active during log-phase growth of the pathogen[72] and 42 are reportedly essential for its growth and survival[73, 74]. A chord layout of these observations sorted based on functional categories is illustrated in Figure-3. The rectangular blocks tagged for respective proteins depict essential genes, while the dots indicate their involvement in log-phase growth of the pathogen. Apparent from the figure, more than half (41) of the potential targets identified pertain to the category- intermediary metabolism and respiration. This finding gains importance as metabolic proteins of *M. tuberculosis* are one of the highly represented functional classes in the host lung that play a significant role in successful establishment of bacterial infection and persistence[75]. Various factors contribute to its success, several of which can be attributed to the proteins involved in biosynthesis of cell wall associated proteins (category: cell wall and cell processes) and the proteins involved in fatty acid biosynthesis (category: lipid metabolism) that promote complex formation of cell wall components[76]. Evidence on implications of such components in evading host immune response and pathogenesis[77, 78] makes the proteins associated with such components suitable for targeting. We could recognize a total of 17 proteins pertaining to the aforementioned functional categories as probable targets (Figure-2). Also important in the development of infection in the host, are the proteins categorized under information pathways. These proteins participate in protein biosynthesis during growth of the pathogen primarily under aerobic conditions[79]. Thus, it is of no surprise that majority of the potential targets recognized under this category are essential to the pathogen and are actively involved in its growth. Such proteins indispensable for the growth of the pathogen qualify as attractive targets for chemotherapeutic intervention and thus, their identification as prospective targets forms a crucial step in anti-tubercular drug discovery. Details on some of the predicted drug-target associations are provided in Table-2. The complete list of potential targets and their corresponding drugs are detailed in the Supplementary Table 3.

In order to obtain a distinctive viewpoint of the drug-target associations predicted for *M. tuberculosis*, a drug-target network was generated using Cytoscape v.3.0[80] which is a tool for

data integration, visualization and analyses of data in the form of networks. Figure-4 illustrates the connectivity between 78 potential targets (denoted as coloured nodes), and their corresponding 130 FDA-approved drugs (denoted as white nodes). The drug-target network delineates 21 connected components. Evidently, three kinds of associations can be interpreted from the figure:

i) Single drug, single target: 9 instances of such drug-target associations can be recognized from the figure. Some notable examples of this type include thymidylate synthase *thyA* (Rv2764c) with its potential drug candidate sulfadoxine (DB01299), urease alpha subunit *ureC* (Rv1850) with acetohydroxamic acid (DB00551), isoleucyl-tRNA synthetase *ileS* (Rv1536) with mupirocin (DB00410) and RNA polymerase sigma factor *sigA* (Rv2703) with fidaxomicin (DB08874). Few of these proteins have been regarded as desirable targets due to their recognition as critical determinants of infection[81-83]. Furthermore, the experimental evidence on potent anti-mycobacterial activities exhibited by a couple of associated drugs provides credible inferences on the predictions made[84, 85].

ii) Single drug, multiple targets: 92 drugs were identified which can be explored for their polypharmacological potential. Several noteworthy cases include nitrofural (DB00336), a compound proven to be effective against mycobacteria[86], thiabendazole (DB00730), an antihelminthic drug and sulfamethoxazole (DB01015), a bacteriostatic agent effective against *M. tuberculosis*[87]. One of the interesting examples in this category is the drug triclosan (DB08604), an antimicrobial agent widely used in personal care products, which has been demonstrated to inhibit enoyl reductase *inhA* of *M. tuberculosis*[88]. However, based on our analysis we could identify multi-target potential of triclosan. Figure-4 highlights multiple associations, including proteins belonging to the category "intermediary metabolism and respiration" and "lipid metabolism", predicted for triclosan apart from *inhA*.

iii) Multiple drugs, single/multiple target(s): Drugs which can be investigated for combination therapy are represented by several drug classes in the figure. As described in the previous section and also apparent in the figure are drugs of various drug classes, theorized to target proteins categorized under "intermediary metabolism and respiration", "information pathways" and "cell wall and cell processes".

In depth investigations on two of the instances are provided in the following sections.

Case study 1: ureC as a target for acetohydroxamic acid

Urease, a nickel-containing enzyme, is responsible for assimilation of urea through ureolytic activity, thus generating ammonia, a nitrogen source, utilized by *M. tuberculosis* during its growth[89]. Although, the pathogen has the ability to adapt to varied host environmental conditions, the urease activity could be indispensable for its survival under conditions where nitrogen sources are limited[89]. Thus, inhibiting this enzyme has the potential to impair the growth of *M. tuberculosis*. Based on our analysis we could identify *M. tuberculosis* urease as a feasible target for acetohydroxamic acid.

A prokaryotic urease enzyme is composed of three subunits: α-subunit, β-subunit and γ-subunit, which assemble as trimer of trimers arranged in a triangular fashion. The active site of the enzyme is constituted by the α-subunit which contains two nickel ions[90]. Acetohydroxamic acid, a structural analogue of urea, is a competitive inhibitor of urease. The inhibitor occupies the active site of the enzyme and coordinates with both nickel ions and active site residues[91]. With the help of crystal structure of acetohydroxamate-bound urease of *Enterobacter aerogenes* we were able to probe the molecular details of probable inhibitor-binding site in *ureC* (α-subunit) of *M. tuberculosis*. The α-subunit of urease enzyme of both *M. tuberculosis* and *E. aerogenes* were recognized to be closely related (see Table-2). Figure-5(a) shows the superposition of the modelled structure of *M. tuberculosis ureC* on the crystal structure of inhibitor-bound *Enterobacter aerogenes* urease, the inhibitor-binding residues and their conservation is exemplified in Figure-5(b). Interestingly, the inhibitor-binding residues and the metal coordinating residues were recognized to be completely conserved in *ureC* of *M. tuberculosis,* thus implying the possibility of enzyme inhibition in *M. tuberculosis* by acetohydroxamic acid.

Case study 2: Fusidic acid as a drug candidate for fusA1 and fusA2

Protein factors, along with ribosomal machineries, play important roles in the efficient synthesis of a protein. Two such critical roles are brought about by elongation factor G (EF-G); one during translation, by promoting guanosine 5′-triphosphate (GTP)-dependent translocation of nascent peptide chain from A-site to the P-site of the ribosome; and second during ribosome disassembly

by interacting with ribosome recycling factor[92]. *M. tuberculosis* encodes two EF-Gs *fusA1* (Rv0684) and *fusA2* (Rv0120c), which have been identified as potential targets for fusidic acid. The *in vitro* susceptibility of *M. tuberculosis* to fusidic acid has been demonstrated earlier, and its effectiveness against susceptible and resistant strains of the pathogen is noteworthy[93]. Fusidic acid, an antibacterial agent, has been shown to exhibit its bacteriostatic activity by trapping EF-G in a post-translocational state, thus, preventing its release from the ribosome during translation[94]. Crystal structure of ribosome with EF-G in this arrested state in the presence of fusidic acid[94] was utilized to determine the putative residues in *M. tuberculosis* targets that may be essential for fusidic acid activity (see Table-2). Figure-6(a) shows the multiple sequence alignment of *fusA1* and *fusA2* with respect to *fusA* of *Thermus thermophilus* pertaining to fusidic acid-binding region. The key residues in *fusA* of *T. thermophilus,* which allow fusidic acid to exert its inhibitory effect, are conserved in *fusA1* and *fusA2* of *M. tuberculosis* as illustrated in Figure-6(b).

For majority of the predicted targets (53 of 78) information on inhibitor-binding sites could be directly obtained from the crystal structures of their corresponding homologous known targets. Instances where binding sites could not be readily inferred, SiteMap (see methodology section) was employed to obtain pharmaceutically relevant binding sites in the predicted targets. These sites provide a resource of putative pockets in the target proteins which could be probed using their respective potential drug candidates. A list of high scoring binding sites estimated for 20 targets is provided in Supplementary Table 4.

Inferences from augmented chemical space

In order to provide a larger scope of exploration in the chemical space, we acquired a two-step screening of the ChEMBL compound library against the FDA-approved drugs. 1367 ChEMBL compounds were recognized to show high molecular similarity to 101 FDA-approved drugs. By virtue of molecular similarity to FDA-approved drugs, it is possible to recognize prospective unexplored inhibitors that might exhibit potent mycobacterial activity. Table-4 details few notable examples. The entire list of compounds similar to approved drugs is provided as a supplementary material (Supplementary Table 5). The feasibility of couple of such compounds as potential anti-tuberculars is investigated further.

Sulfamethoxazole is a bacteriostatic agent belonging to a class of organic compounds which constitute benzenesulfonamide moiety with an amine group attached to the benzene ring. Our approach could recognize three ChEMBL compounds structurally similar to sulfamethoxazole (see Table-4 and Figure-7), which differ in the group attached to the benzene ring. Of the three, two compounds CHEMBL610753 and CHEMBL608998 have been evaluated experimentally for their activity against *Leishmania infantum*[95]. Figure-7 illustrates the 2D representation of these small molecules and their 3D alignment with respect to sulfamethoxazole. Owing to their high structural similarity to sulfamethoxazole, the probable mode and the site of binding of these ChEMBL compounds in the associated target(s) can be inferred from the structural aspects of a sulfamethoxazole-bound protein. We pursued a protein-ligand docking study in order to determine the characteristics of the binding site and the binding mode of sulfamethoxazole in one of its associated targets *folC* (Rv2447c).

The activity of folylpolyglutamate synthetase *folC* of *M. tuberculosis* is essential for the growth and survival of *M. tuberculosis,* as it aids the conversion of folates to polyglutamate derivatives which are in turn vital in diverse biosynthetic pathways. In addition, this enzyme is also thought to exhibit dihydrofolate synthetase activity. This bifunctional enzyme comprises of two distinct domains: an ATPase domain at the N-terminus and a C-terminal Rossmann-fold domain, which flank the active site region[96]. This region hosts functionally important loops including dihydropteroate-binding loop (α1-α2 loop), P-loop and a β5-α6 loop (res: 197-210) which is proposed to provide a mechanistic switch to regulate enzyme activity[96]. Such a functionally important region was recognized as a pharmaceutically relevant pocket by SiteMap with a SiteScore of 1.058, which was used as a receptor grid to dock sulfamethoxazole using Glide with XP scoring function. Investigations on the binding poses of the drug revealed a docked pose with a best possible score of -6.75 kcal/mol. Figure-8 exemplifies the binding pocket, the docked pose of sulfamethoxazole and the proximity of the drug to the functionally important loops. The probable residues influenced are highlighted in ball and stick representation. Notably, these residues include Lys218, Lys77, Asp202, His356, Glu176 and Ser100 which are involved in stabilizing activated conformation of the *M. tuberculosis* enzyme and are also engaged in metal ion coordination. Thus, the proposed site and mode of binding of sulfamethoxazole could be theorized to impede proper functioning of the enzyme. While the details on binding site of sulfamethoxazole can be extrapolated to the ChEMBL compounds

structurally similar to the drug, thorough investigations on binding affinity of these compounds and their bioactivities is required to validate their potential as anti-tubercular compounds.

Comparison with previously published computational studies

Several computational efforts have been pursued in the past decade, including our previously published studies, in order to outline strategies for identification of attractive *M. tuberculosis* targets and potential drug candidates that could serve as anti-tubercular agents[2, 97, 98]. A comparison with previously reported targets and drug candidates is discussed further.

A list of 451 *M. tuberculosis* proteins of high targetability was proposed previously by our group[97] with the use of an extensive target identification pipeline, targetTB, which employs systems, sequence and structural level analysis of *M. tuberculosis* proteins coupled with data-intensive filters. Our analysis could identify nine such proteins of high targetability as feasible targets for 14 FDA-approved drugs (Table-3). A recently published study by our group on structural characterization of binding sites in *M. tuberculosis* on genome-wide basis has provided cues on proteins that could serve as polypharmacological targets[98]. Four such targets with polypharmacological potential overlapped with our analysis and indeed, were recognized as proteins belonging to a set of targets which could be potentially targeted by single FDA-approved drug (Table-3).

We also compared our predicted drug-target associations with TB-drugome[2], a resource which provides a list of approved drugs and druggable *M. tuberculosis* targets that have been computationally determined on the basis of binding-site similarity between targets of approved drugs and *M. tuberculosis* proteins. No overlapping drug-target associations were identified; however we identified 11 *M. tuberculosis* targets in our analysis which were identified as druggable targets in TB-drugome, and 6 approved drugs (chloramphenicol, ketoconazole, mupirocin, novobiocin, triclosan and tetracycline) in TB-drugome which have also been recognized in our analysis as drug candidates for repurposing against tuberculosis. Table-3 summarizes the list of *M. tuberculosis* proteins recognized as high-confidence targets in previously published studies.

Proof of the principle

Apart from several computational studies, experimental investigations on exploiting the use of available drugs as possible repurpose-able candidates for tuberculosis, has been demonstrated by numerous groups. However, we approached the question of identifying FDA approved drugs that may be effective against tuberculosis without considering previously known information. Interestingly, 74 out of 130 approved drugs proposed in our analysis have been experimentally supported earlier, as anti-tubercular agents capable of inhibiting the pathogen either alone or in combination with other drugs. Supplementary Table-2 summarizes literature reports on 74 of 130 approved drugs whose anti-tubercular activity has been tested either *in vitro, in vivo* or in clinical studies.

In addition, several studies have also exploited the pharmaceutical relevance of proteins encoded in *M. tuberculosis* genome. Mutational studies[73, 74], use of transposon site hybridization techniques[99, 100], analyses on response of the pathogen to nutrient starvation[101], biochemical studies under different conditions[102, 103], drug-induced alterations in gene expression[104, 105], genes involved in non-replicating persistence[106], conditional gene inactivation[107, 108] and numerous gene-knockout experiments[109-118] have aided in comprehending genes essential for growth, survival, virulence and antibiotic resistance in *M. tuberculosis,* and thus qualify as attractive targets. Interestingly, we could identify 59 of 78 potential targets (Supplementary Table-3) from our analysis which concurred with such targets of interest.

Identification of such pharmaceutically relevant targets and drugs in our analyses justifies the strength of the computational approach used; therefore the rest of the drug-target associations have the potential to serve as a guiding tool for experimental endeavours.

Conclusion

Drug repurposing strategies have been increasingly gaining importance over the past decade and several pharmaceutical companies and research groups have successfully established new uses for old drugs. Over and above the high-throughput screening techniques, used to determine new indications for approved drugs, there has been considerable interest in the usage of computational approaches to extract, integrate and assess relevant biological and chemical information, in order to design protocols for an accelerated drug repurposing pipeline. Such time and cost-saving strategies are of immense use particularly for pathogenic microorganisms such as *M. tuberculosis* which have acquired antibiotic resistance over the course of co-evolution with their human host.

Besides the recognition of approved drugs which could be repurposed against tuberculosis, comprehension of the target space of the pathogen also acquires equal stance. Towards the exploration of druggable target space in *M. tuberculosis* our previously published study[97] demonstrated the use of sequence, structural and systems level analysis in concert with integration of diverse datasets to recognize high priority targets. Another noteworthy study on computational identification of *M. tuberculosis* proteins as targets for drug-like compounds experimentally active against *M. tuberculosis,* has provided interesting insights on compound-target associations that can be readily pursued for biochemical studies[119]. Furthermore, recognition of *M. tuberculosis* proteins involved in interaction with the host cellular components could aid in prioritization of targets that could be determinants of pathogenesis. Recent published works on recognition of protein-protein interactions across *M. tuberculosis* and human provide useful resource of attractive targets[120, 121].

Through exploration of evolutionarily relationship between known targets of FDA-approved drugs and proteins in *M. tuberculosis,* we were able to demonstrate the recognition of 78 *M. tuberculosis* proteins which could serve as potential targets for 130 approved drugs. By elimination of approved drugs capable of acting on human proteins, we have minimized the chances of obtaining an anti-target in the host. Indeed, repurposing drugs against tuberculosis has considerable advantages, especially as a time and cost saving strategy and the less likelihood of an approved drug exhibiting adverse and/or toxic side effects in human. However, there can be a possibility where a drug predicted as an anti-tubercular drug candidate is unsuccessful in

exhibiting anti-tubercular activity. This can be due to distinct features of the binding site of *M. tuberculosis* protein and its homologous known target protein, or due to the inadequate penetration of the drug to the site of infection in the host. The latter case is largely reliant on *in vivo* experimental investigations or clinical studies on viability of the drug, whilst the former case can be tackled by identifying similarities and dissimilarities in the binding pocket-lining residues between established ligand-bound targets and their homologous *M. tuberculosis* proteins, which is attempted in the current study. Sequence and structure-based assessment of *M. tuberculosis* proteins homologous to known targets has yielded credible inferences on putative binding sites in potential targets. We also recognized potentially druggable sites for instances where information on binding sites could not be readily inferred from known targets. Identification of target-specific drugs and drugs with polypharmacological potential, capable of acting on multiple *M. tuberculosis* proteins, could be clearly comprehended from drug-target network. Such drug-target associations can be investigated for their chemotherapeutic relevance. In addition, we have made an attempt to recognize prospective anti-tubercular compounds by virtue of their structural similarity to FDA-approved drugs. The approved drugs predicted as promising candidates for repurposing against *M. tuberculosis,* in concert with the structurally similar compounds could serve as an effective resource for an experimental follow-up towards anti-tubercular drug development and drug discovery.

Competing Interests

None declared.

Author Contributions

N.S. and N.R.C. conceived and designed the project. G.R. performed the experiments, analysed the data and wrote the manuscript. The manuscript was finalised through contributions of all the authors.

Acknowledgements

Authors would like to thank Prof. R. Sowdhamini for her generous support in providing access to Schrödinger suite of programs. This research was supported by Mathematical Biology programme, Department of Science and Technology and by the Department of Biotechnology, Government of India. N.S. is a J.C. Bose National Fellow.

References

1. T. T. Ashburn and K. B. Thor, *Nature reviews. Drug discovery*, 2004, **3**, 673-683.
2. S. L. Kinnings, L. Xie, K. H. Fung, R. M. Jackson, L. Xie and P. E. Bourne, *PLoS computational biology*, 2010, **6**, e1000976.
3. S. L. Kinnings, N. Liu, N. Buchmeier, P. J. Tonge, L. Xie and P. E. Bourne, *PLoS computational biology*, 2009, **5**, e1000423.
4. M. Campillos, M. Kuhn, A. C. Gavin, L. J. Jensen and P. Bork, *Science*, 2008, **321**, 263-266.
5. M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin and B. K. Shoichet, *Nature biotechnology*, 2007, **25**, 197-206.
6. M. J. Keiser, V. Setola, J. J. Irwin, C. Laggner, A. I. Abbas, S. J. Hufeisen, N. H. Jensen, M. B. Kuijer, R. C. Matos, T. B. Tran, R. Whaley, R. A. Glennon, J. Hert, K. L. Thomas, D. D. Edwards, B. K. Shoichet and B. L. Roth, *Nature*, 2009, **462**, 175-181.
7. D. L. Ma, D. S. Chan and C. H. Leung, *Chemical Society reviews*, 2013, **42**, 2130-2141.
8. J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander and T. R. Golub, *Science*, 2006, **313**, 1929-1935.
9. Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda and M. Kanehisa, *Bioinformatics*, 2008, **24**, i232-240.
10. S. Zhao and S. Li, *PloS one*, 2010, **5**, e11764.
11. G. V. Paolini, R. H. Shapland, W. P. van Hoorn, J. S. Mason and A. L. Hopkins, *Nature biotechnology*, 2006, **24**, 805-815.
12. S. Chopra, K. Matsuyama, C. Hutson and P. Madrid, *The Journal of antimicrobial chemotherapy*, 2011, **66**, 1533-1536.
13. J. Kouznetsova, W. Sun, C. Martinez-Romero, G. Tawa, P. Shinn, C. Z. Chen, A. Schimmer, P. Sanderson, J. C. McKew, W. Zheng and A. Garcia-Sastre, *Emerg Microbes Infec*, 2014, **3**., e84.
14. K. Kim, L. Zilbermintz and M. Martchenko, *Annals of clinical microbiology and antimicrobials*, 2015, **14**, 32.
15. C. H. Leung, D. S. Chan, M. H. Kwan, Z. Cheng, C. Y. Wong, G. Y. Zhu, W. F. Fong and D. L. Ma, *ChemMedChem*, 2011, **6**, 765-768.
16. D. S. H. Chan, H. Yang, M. H. T. Kwan, Z. Cheng, P. Lee, L. P. Bai, Z. H. Jiang, C. Y. Wong, W. F. Fong, C. H. Leung and D. L. Ma, *Biochimie*, 2011, **93**, 1055-1064.
17. H. J. Zhong, L. J. Liu, D. S. Chan, H. M. Wang, P. W. Chan, D. L. Ma and C. H. Leung, *Biochimie*, 2014, **102**, 211-215.
18. W. H. Organization, 2014.
19. Y. L. Janin, *Bioorganic & medicinal chemistry*, 2007, **15**, 2479-2513.
20. C. E. Maus, B. B. Plikaytis and T. M. Shinnick, *Antimicrobial agents and chemotherapy*, 2005, **49**, 3192-3197.
21. K. E. Lougheed, D. L. Taylor, S. A. Osborne, J. S. Bryans and R. S. Buxton, *Tuberculosis*, 2009, **89**, 364-370.
22. J. C. Palomino and A. Martin, *The Journal of antimicrobial chemotherapy*, 2013, **68**, 275-283.
23. F. A. Kruger and J. P. Overington, *PLoS computational biology*, 2012, **8**, e1002333.
24. G. Ramakrishnan, B. Ochoa-Montano, U. S. Raghavender, R. Mudgal, A. G. Joshi, N. R. Chandra, R. Sowdhamini, T. L. Blundell and N. Srinivasan, *Tuberculosis*, 2015, **95**, 14-25.
25. V. Law, C. Knox, Y. Djoumbou, T. Jewison, A. C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu, A. Tang, G. Gabriel, C. Ly, S. Adamjee, Z. T. Dame, B. Han, Y. Zhou and D. S. Wishart, *Nucleic acids research*, 2014, **42**, D1091-1097.
26. A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani and J. P. Overington, *Nucleic acids research*, 2012, **40**, D1100-1107.

27. S. R. Eddy, *PLoS computational biology*, 2011, **7**, e1002195.
28. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic acids research*, 2000, **28**, 235-242.
29. U. Pieper, B. M. Webb, D. T. Barkan, D. Schneidman-Duhovny, A. Schlessinger, H. Braberg, Z. Yang, E. C. Meng, E. F. Pettersen, C. C. Huang, R. S. Datta, P. Sampathkumar, M. S. Madhusudhan, K. Sjolander, T. E. Ferrin, S. K. Burley and A. Sali, *Nucleic acids research*, 2011, **39**, D465-474.
30. A. Sali and T. L. Blundell, *Journal of molecular biology*, 1993, **234**, 779-815.
31. M. Y. Shen and A. Sali, *Protein science : a publication of the Protein Society*, 2006, **15**, 2507-2524.
32. D. Eramian, M. Y. Shen, D. Devos, F. Melo, A. Sali and M. A. Marti-Renom, *Protein science : a publication of the Protein Society*, 2006, **15**, 1653-1666.
33. F. Melo, R. Sanchez and A. Sali, *Protein science : a publication of the Protein Society*, 2002, **11**, 430-448.
34. Y. Zhang and J. Skolnick, *Nucleic acids research*, 2005, **33**, 2302-2309.
35. J. Xu and Y. Zhang, *Bioinformatics*, 2010, **26**, 889-895.
36. T. Halgren, *Chemical biology & drug design*, 2007, **69**, 146-148.
37. T. A. Halgren, *Journal of chemical information and modeling*, 2009, **49**, 377-389.
38. R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *Journal of medicinal chemistry*, 2004, **47**, 1739-1749.
39. T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks, *Journal of medicinal chemistry*, 2004, **47**, 1750-1759.
40. R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin and D. T. Mainz, *Journal of medicinal chemistry*, 2006, **49**, 6177-6196.
41. D. B. Kitchen, H. Decornez, J. R. Furr and J. Bajorath, *Nature reviews. Drug discovery*, 2004, **3**, 935-949.
42. D. S. Chan, H. M. Lee, F. Yang, C. M. Che, C. C. Wong, R. Abagyan, C. H. Leung and D. L. Ma, *Angewandte Chemie*, 2010, **49**, 2860-2864.
43. D. Plewczynski, M. Lazniewski, R. Augustyniak and K. Ginalski, *J Comput Chem*, 2011, **32**, 742-755.
44. N. Moitessier, P. Englebienne, D. Lee, J. Lawandi and C. R. Corbeil, *British journal of pharmacology*, 2008, **153 Suppl 1**, S7-26.
45. S. F. Sousa, P. A. Fernandes and M. J. Ramos, *Proteins*, 2006, **65**, 15-26.
46. A. R. Leach, B. K. Shoichet and C. E. Peishoff, *Journal of medicinal chemistry*, 2006, **49**, 5851-5855.
47. G. Jones, P. Willett, R. C. Glen, A. R. Leach and R. Taylor, *Journal of molecular biology*, 1997, **267**, 727-748.
48. M. L. Verdonk, G. Chessari, J. C. Cole, M. J. Hartshorn, C. W. Murray, J. W. M. Nissink, R. D. Taylor and R. Taylor-, *J Med Chem*, 2005, **48**, 6504-6515.
49. G. Jones, P. Willett and R. C. Glen, *Journal of molecular biology*, 1995, **245**, 43-53.
50. B. Kramer, M. Rarey and T. Lengauer, *Proteins*, 1999, **37**, 228-241.
51. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J Comput Chem*, 2009, **30**, 2785-2791.
52. T. J. A. Ewing and I. D. Kuntz, *J Comput Chem*, 1997, **18**, 1175-1189.
53. T. J. Ewing, S. Makino, A. G. Skillman and I. D. Kuntz, *Journal of computer-aided molecular design*, 2001, **15**, 411-428.
54. R. Abagyan, M. Totrov and D. Kuznetsov, *J Comput Chem*, 1994, **15**, 488-506.
55. A. N. Jain, *J Med Chem*, 2003, **46**, 499-511.
56. C. M. Venkatachalam, X. Jiang, T. Oldfield and M. Waldman, *J Mol Graph Model*, 2003, **21**, 289-307.

57. Z. Y. Zhou, A. K. Felts, R. A. Friesner and R. M. Levy, *J Chem Inf Model*, 2007, **47**, 1599-1608.
58. R. Tiwari, K. Mahasenan, R. Pavlovicz, C. Li and W. Tjarks, *J Chem Inf Model*, 2009, **49**, 1581-1589.
59. E. Kellenberger, J. Rodrigo, P. Muller and D. Rognan, *Proteins-Structure Function and Bioinformatics*, 2004, **57**, 225-242.
60. E. Perola, W. P. Walters and P. S. Charifson, *Proteins-Structure Function and Bioinformatics*, 2004, **56**, 235-249.
61. M. Kontoyianni, L. M. McClellan and G. S. Sokol, *J Med Chem*, 2004, **47**, 558-565.
62. H. Chen, P. D. Lyne, F. Giordanetto, T. Lovell and J. Li, *J Chem Inf Model*, 2006, **46**, 401-415.
63. M. D. Cummings, R. L. DesJarlais, A. C. Gibbs, V. Mohan and E. P. Jaeger, *J Med Chem*, 2005, **48**, 962-976.
64. M. Kontoyianni, G. S. Sokol and L. M. McClellan, *J Comput Chem*, 2005, **26**, 11-22.
65. J. B. Cross, D. C. Thompson, B. K. Rai, J. C. Baber, K. Y. Fan, Y. B. Hu and C. Humblet, *J Chem Inf Model*, 2009, **49**, 1455-1474.
66. X. Li, Y. Li, T. J. Cheng, Z. H. Liu and R. X. Wang, *J Comput Chem*, 2010, **31**, 2109-2125.
67. P. Willett, J. M. Barnard and G. M. Downs, *Journal of chemical information and computer sciences*, 1998, **38**, 983-996.
68. T. T. Tanimoto, *IBM Internal Report*, NY, 1957.
69. N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *Journal of cheminformatics*, 2011, **3**, 33.
70. X. Liu, H. Jiang and H. Li, *Journal of chemical information and modeling*, 2011, **51**, 2372-2385.
71. W. Lu, X. Liu, X. Cao, M. Xue, K. Liu, Z. Zhao, X. Shen, H. Jiang, Y. Xu, J. Huang and H. Li, *Journal of medicinal chemistry*, 2011, **54**, 3564-3574.
72. L. M. Fu and C. S. Fu-Liu, *BMC microbiology*, 2007, **7**, 37.
73. J. Rengarajan, B. R. Bloom and E. J. Rubin, *Proceedings of the National Academy of Sciences of the United States of America*, 2005, **102**, 8327-8332.
74. J. E. Griffin, J. D. Gawronski, M. A. Dejesus, T. R. Ioerger, B. J. Akerley and C. M. Sassetti, *PLoS pathogens*, 2011, **7**, e1002251.
75. N. A. Kruh, J. Troudt, A. Izzo, J. Prenni and K. M. Dobos, *PloS one*, 2010, **5**, e13938.
76. R. J. Kinsella, D. A. Fitzpatrick, C. J. Creevey and J. O. McInerney, *Proceedings of the National Academy of Sciences of the United States of America*, 2003, **100**, 10320-10325.
77. M. Jackson, G. Stadthagen and B. Gicquel, *Tuberculosis*, 2007, **87**, 78-86.
78. T. Fukuda, T. Matsumura, M. Ato, M. Hamasaki, Y. Nishiuchi, Y. Murakami, Y. Maeda, T. Yoshimori, S. Matsumoto, K. Kobayashi, T. Kinoshita and Y. S. Morita, *mBio*, 2013, **4**, e00472-00412.
79. Y. M. Hu, P. D. Butcher, K. Sole, D. A. Mitchison and A. R. Coates, *FEMS microbiology letters*, 1998, **158**, 139-145.
80. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome research*, 2003, **13**, 2498-2504.
81. D. L. Clemens, B. Y. Lee and M. A. Horwitz, *Journal of bacteriology*, 1995, **177**, 5644-5652.
82. M. Sassanfar, J. E. Kranz, P. Gallant, P. Schimmel and K. Shiba, *Biochemistry*, 1996, **35**, 9995-10003.
83. D. M. Collins, R. P. Kawakami, G. W. de Lisle, L. Pascopella, B. R. Bloom and W. R. Jacobs, Jr., *Proceedings of the National Academy of Sciences of the United States of America*, 1995, **92**, 8036-8040.
84. L. Haller, R. Sossouhounto, I. M. Coulibaly, M. Dosso, M. Kone, H. Adom, U. A. Meyer, B. Betschart, M. Wenk, W. E. Haefeli, L. R. Lobognon, M. Porquet, G. Kabore, F. Sorenson, R. Reber-Liske and D. Sturchler, *Chemotherapy*, 1999, **45**, 452-465.
85. M. Kurabachew, S. H. Lu, P. Krastel, E. K. Schmitt, B. L. Suresh, A. Goh, J. E. Knox, N. L. Ma, J. Jiricek, D. Beer, M. Cynamon, F. Petersen, V. Dartois, T. Keller, T. Dick and V. K. Sambandamurthy, *The Journal of antimicrobial chemotherapy*, 2008, **62**, 713-719.

86. B. Murugasu-Oei and T. Dick, *The Journal of antimicrobial chemotherapy*, 2000, **46**, 917-919.
87. W. Ong, A. Sievers and D. E. Leslie, *Antimicrobial agents and chemotherapy*, 2010, **54**, 2748; author reply 2748-2749.
88. S. L. Parikh, G. Xiao and P. J. Tonge, *Biochemistry*, 2000, **39**, 7645-7650.
89. W. Lin, V. Mathys, E. L. Ang, V. H. Koh, J. M. Martinez Gomez, M. L. Ang, S. Z. Zainul Rahim, M. P. Tan, K. Pethe and S. Alonso, *Infection and immunity*, 2012, **80**, 2771-2779.
90. E. Jabri, M. B. Carr, R. P. Hausinger and P. A. Karplus, *Science*, 1995, **268**, 998-1004.
91. M. A. Pearson, L. O. Michel, R. P. Hausinger and P. A. Karplus, *Biochemistry*, 1997, **36**, 8164-8172.
92. K. Ito, T. Fujiwara, T. Toyoda and Y. Nakamura, *Molecular cell*, 2002, **9**, 1263-1272.
93. C. Cicek-Saydam, C. Cavusoglu, D. Burhanoglu, S. Hilmioglu, N. Ozkalay and A. Bilgic, *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 2001, **7**, 700-702.
94. Y. G. Gao, M. Selmer, C. M. Dunham, A. Weixlbaumer, A. C. Kelley and V. Ramakrishnan, *Science*, 2009, **326**, 694-699.
95. M. A. Dea-Ayuela, E. Castillo, M. Gonzalez-Alvarez, C. Vega, M. Rolon, F. Bolas-Fernandez, J. Borras and M. E. Gonzalez-Rosende, *Bioorganic & medicinal chemistry*, 2009, **17**, 7449-7456.
96. P. G. Young, C. A. Smith, P. Metcalf and E. N. Baker, *Acta crystallographica. Section D, Biological crystallography*, 2008, **D64**, 745-753.
97. K. Raman, K. Yeturu and N. Chandra, *BMC systems biology*, 2008, **2**, 109.
98. P. Anand and N. Chandra, *Scientific reports*, 2014, **4**, 6356.
99. C. M. Sassetti, D. H. Boyd and E. J. Rubin, *Proceedings of the National Academy of Sciences of the United States of America*, 2001, **98**, 12712-12717.
100. C. M. Sassetti, D. H. Boyd and E. J. Rubin, *Molecular microbiology*, 2003, **48**, 77-84.
101. J. C. Betts, P. T. Lukey, L. C. Robb, R. A. McAdam and K. Duncan, *Molecular microbiology*, 2002, **43**, 717-731.
102. V. Singh, D. Chandra, B. S. Srivastava and R. Srivastava, *Microbiol-Sgm*, 2011, **157**, 29-37.
103. S. Watanabe, M. Zimmermann, M. B. Goodwin, U. Sauer, C. E. Barry and H. I. Boshoff, *PLoS pathogens*, 2011, **7**.
104. M. Wilson, J. DeRisi, H. H. Kristensen, P. Imboden, S. Rane, P. O. Brown and G. K. Schoolnik, *Proceedings of the National Academy of Sciences of the United States of America*, 1999, **96**, 12833-12838.
105. H. I. Boshoff, T. G. Myers, B. R. Copp, M. R. McNeil, M. A. Wilson and C. E. Barry, 3rd, *The Journal of biological chemistry*, 2004, **279**, 40174-40184.
106. M. I. Voskuil, K. C. Visconti and G. K. Schoolnik, *Tuberculosis*, 2004, **84**, 218-227.
107. P. Kaur, S. Agarwal and S. Datta, *PloS one*, 2009, **4**.
108. S. Gandotra, D. Schnappinger, M. Monteleone, W. Hillen and S. Ehrt, *Nature medicine*, 2007, **13**, 1515-1520.
109. A. K. Azad, T. D. Sirakova, N. D. Fernandes and P. E. Kolattukudy, *The Journal of biological chemistry*, 1997, **272**, 16741-16745.
110. T. P. Primm, S. J. Andersen, V. Mizrahi, D. Avarbock, H. Rubin and C. E. Barry, *Journal of bacteriology*, 2000, **182**, 4889-4898.
111. R. Colangeli, D. Helb, S. Sridharan, J. C. Sun, M. Varma-Basil, M. H. Hazbon, R. Harbacheuski, N. J. Megjugorac, W. R. Jacobs, A. Holzenburg, J. C. Sacchettini and D. Alland, *Molecular microbiology*, 2005, **55**, 1829-1840.
112. P. Peirs, P. Lefevre, S. Boarbi, X. M. Wang, O. Denis, M. Braibant, K. Pethe, C. Locht, K. Huygen and J. Content, *Infection and immunity*, 2005, **73**, 1898-1902.
113. S. Cowley, M. Ko, N. Pick, R. Chow, K. J. Downing, B. G. Gordhan, J. C. Betts, V. Mizrahi, D. A. Smith, R. W. Stokes and Y. Av-Gay, *Molecular microbiology*, 2004, **52**, 1691-1702.
114. T. D. Sirakova, A. K. Thirumala, V. S. Dubey, H. Sprecher and P. E. Kolattukudy, *Journal of Biological Chemistry*, 2001, **276**, 16833-16839.

115. F. Bigi, A. Gioffre, L. Klepp, M. D. Santangelo, A. Alito, K. Caimi, V. Meikle, M. Zumarraga, O. Taboga, M. I. Romano and A. Cataldi, *Microbes and Infection*, 2004, **6**, 182-187.
116. K. N. Lewis, R. Liao, K. M. Guinn, M. J. Hickey, S. Smith, M. A. Behr and D. R. Sherman, *The Journal of infectious diseases*, 2003, **187**, 117-123.
117. K. G. Papavinasasundaram, B. Chan, J. H. Chung, M. J. Colston, E. O. Davis and Y. Av-Gay, *Journal of bacteriology*, 2005, **187**, 5751-5760.
118. H. I. Boshoff, M. B. Reed, C. E. Barry, 3rd and V. Mizrahi, *Cell*, 2003, **113**, 183-193.
119. F. Martinez-Jimenez, G. Papadatos, L. Yang, I. M. Wallace, V. Kumar, U. Pieper, A. Sali, J. R. Brown, J. P. Overington and M. A. Marti-Renom, *PLoS computational biology*, 2013, **9**, e1003253.
120. G. Ramakrishnan, N. R. Chandra and N. Srinivasan, *IUBMB life*, 2014, **66**, 759-774.
121. H. Zhou, S. Gao, N. N. Nguyen, M. Fan, J. Jin, B. Liu, L. Zhao, G. Xiong, M. Tan, S. Li and L. Wong, *Biology direct*, 2014, **9**, 5.
122. The PyMOL Molecular Graphics System, Version 1.4.1 Schrodinger, LLC.
123. Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J.; Honig, B.; Shaw, D. E.; Friesner, R. A., *Proteins*, 2004, **55**, 351-67.
124. Maestro, version 9.9, Schrodinger, LLC, 2014.
125. S. H. Cho, S. Warit, B. Wan, C. H. Hwang, G. F. Pauli and S. G. Franzblau, *Antimicrobial agents and chemotherapy*, 2007, **51**, 1380-1385.
126. H. E. Seward, A. Roujeinikova, K. J. McLean, A. W. Munro and D. Leys, *The Journal of biological chemistry*, 2006, **281**, 39437-39443.
127. M. Opravil, M. Pechere, A. Lazzarin, A. Heald, S. Ruttimann, A. Iten, H. Furrer, D. Oertle, G. Praz, D. A. Vuitton, B. Hirschel and R. Luthy, *Clinical Infectious Diseases*, 1995, **20**, 244-249.
128. V. Nopponpunth, W. Sirawaraporn, P. J. Greene and D. V. Santi, *Journal of bacteriology*, 1999, **181**, 6814-6821.
129. R. L. Mayer, *Journal of bacteriology*, 1944, **48**, 93-96.
130. M. Naito, R. Watanabe, S. Tsukuma and K. Iba, *Acta tuberculosea Japonica*, 1960, **10**, 47-56.
131. P. Forgacs, N. L. Wengenack, L. Hall, S. K. Zimmerman, M. L. Silverman and G. D. Roberts, *Antimicrobial agents and chemotherapy*, 2009, **53**, 4789-4793.
132. M. D. Driscoll, K. J. McLean, M. R. Cheesman, T. A. Jowitt, M. Howard, P. Carroll, T. Parish and A. W. Munro, *Biochimica et biophysica acta*, 2011, **1814**, 76-87.
133. A. Lutz and M. A. Witz, *Comptes rendus des seances de la Societe de biologie et de ses filiales*, 1955, **149**, 1467-1470.
134. K. Stoffels, H. Traore, F. Vanderbist and M. Fauville-Dufaux, *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease*, 2009, **13**, 1041-1044.
135. W. E. Sanders, R. Cacciatore, H. Valdez, I. Pejovic and F. P. Dunbar, *The Journal of infectious diseases*, 1971, **124 Suppl**, S33-36.
136. C. C. Sanders and W. E. Sanders, Jr., *Antimicrobial agents and chemotherapy*, 1981, **20**, 247-251.
137. B. A. Waisbren and W. W. Spink, *Annals of internal medicine*, 1950, **33**, 1099-1119.
138. Y. I. Ho, C. Y. Chan and A. F. Cheng, *The Journal of antimicrobial chemotherapy*, 1997, **40**, 27-32.
139. T. P. Kanyok, M. V. Reddy, J. Chinnaswamy, L. H. Danziger and P. R. J. Gangadharam, *Antimicrobial agents and chemotherapy*, 1994, **38**, 170-173.
140. T. Udou, *Journal of chemotherapy*, 2006, **18**, 610-616.
141. M. Malik and K. Drlica, *Antimicrobial agents and chemotherapy*, 2006, **50**, 2842-2844.
142. S. Bosne-David, V. Barros, S. C. Verde, C. Portugal and H. L. David, *The Journal of antimicrobial chemotherapy*, 2000, **46**, 391-395.
143. R. A. Slayden, D. L. Knudson and J. T. Belisle, *Microbiology*, 2006, **152**, 1789-1797.

144. J. E. Hugonnet, L. W. Tremblay, H. I. Boshoff, C. E. Barry and J. S. Blanchard, *Science*, 2009, **323**, 1215-1218.
145. K. England, H. I. M. Boshoff, K. Arora, D. Weiner, E. Dayao, D. Schimel, L. E. Via and C. E. Barry, *Antimicrobial agents and chemotherapy*, 2012, **56**, 3384-3387.
146. C. Jagannath, H. S. Allaudeen and R. L. Hunter, *Antimicrobial agents and chemotherapy*, 1995, **39**, 1349-1354.
147. H. Thadepalli, A. H. Niden and J. T. Huang, *Chest*, 1976, **69**, 743-746.
148. J. E. Hugonnet and J. S. Blanchard, *Biochemistry*, 2007, **46**, 11998-12004.
149. S. Chopra, K. Matsuyama, T. Tran, J. P. Malerich, B. J. Wan, S. G. Franzblau, S. Lun, H. D. Guo, M. C. Maiga, W. R. Bishai and P. B. Madrid, *J Antimicrob Chemoth*, 2012, **67**, 415-421.
150. S. Ramon-Garcia, C. Ng, H. Anderson, J. D. Chao, X. Zheng, T. Pfeifer, Y. Av-Gay, M. Roberge and C. J. Thompson, *Antimicrobial agents and chemotherapy*, 2011, **55**, 3861-3869.
151. S. T. Byrne, S. M. Denkin, P. Gu, E. Nuermberger and Y. Zhang, *Journal of medical microbiology*, 2007, **56**, 1047-1051.
152. B. Gold, M. Pingle, S. J. Brickner, N. Shah, J. Roberts, M. Rundell, W. C. Bracken, T. Warrier, S. Somersan, A. Venugopal, C. Darby, X. J. Jiang, J. D. Warren, J. Fernandez, O. Ouerfelli, E. L. Nuermberger, A. Cunningham-Bussel, P. Rath, T. Chidawanyika, H. T. Deng, R. Realubit, J. F. Glickman and C. F. Nathan, *Proceedings of the National Academy of Sciences of the United States of America*, 2012, **109**, 16004-16011.
153. L. W. Tremblay, F. Fan and J. S. Blanchard, *Biochemistry*, 2010, **49**, 3766-3773.
154. K. Prabhakaran, E. B. Harris, B. Randhawa, L. B. Adams, D. L. Williams and R. C. Hastings, *Microbios*, 1993, **76**, 251-261.
155. A. R. Flores, L. M. Parsons and M. S. Pavelka, *Microbiol-Sgm*, 2005, **151**, 521-532.
156. T. B. Sorg and M. H. Cynamon, *The Journal of antimicrobial chemotherapy*, 1987, **19**, 59-64.
157. F. Wang, C. Cassidy and J. C. Sacchettini, *Antimicrobial agents and chemotherapy*, 2006, **50**, 2762-2771.
158. V. Dubee, S. Triboulet, J. L. Mainardi, M. Etheve-Quelquejeu, L. Gutmann, A. Marie, L. Dubost, J. E. Hugonnet and M. Arthur, *Antimicrobial agents and chemotherapy*, 2012, **56**, 4189-4195.
159. J. E. Kasik, M. Weber, E. Winberg and W. R. Barclay, *The American review of respiratory disease*, 1966, **94**, 260-261.
160. C. H. Chen, M. H. Yang, J. S. Lin, Y. C. Lee and R. P. Perng, *Proceedings of the National Science Council, Republic of China. Part B, Life sciences*, 1995, **19**, 80-84.
161. A. R. Flores, L. M. Parsons and M. S. Pavelka, Jr., *Journal of bacteriology*, 2005, **187**, 1892-1900.
162. A. J. Crowle, J. A. Sbarbaro and M. H. May, *Tubercle*, 1988, **69**, 15-25.
163. H. F. Chambers, T. Kocagoz, T. Sipit, J. Turner and P. C. Hopewell, *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 1998, **26**, 874-877.
164. L. B. Heifets, M. D. Iseman, J. L. Cook, P. J. Lindholm-Levy and I. Drupa, *Antimicrobial agents and chemotherapy*, 1985, **27**, 11-15.
165. N. Selvakumar, V. Kumar and C. N. Paramasivan, *The Indian journal of medical research*, 1997, **105**, 58-60.
166. I. Dincer, A. Ergin and T. Kocagoz, *International journal of antimicrobial agents*, 2004, **23**, 408-411.
167. H. F. Chambers, D. Moreau, D. Yajko, C. Miick, C. Wagner, C. Hackbarth, S. Kocagoz, E. Rosenberg, W. K. Hadley and H. Nikaido, *Antimicrobial agents and chemotherapy*, 1995, **39**, 2620-2624.
168. T. Ishibashi, M. Takamoto, A. Shinoda, M. Yoshida, Y. Ichikawa, M. Kido, K. Ninomiya, Y. Kitahara, N. Hara, H. Tsurutani and et al., *The Japanese journal of antibiotics*, 1990, **43**, 239-256.
169. H. Haas, J. Michel and T. Sacks, *Chemotherapy*, 1982, **28**, 1-5.
170. G. Abate and S. E. Hoffner, *Diagn Micr Infec Dis*, 1997, **28**, 119-122.

171. S. Solapure, N. Dinesh, R. Shandil, V. Ramachandran, S. Sharma, D. Bhattacharjee, S. Ganguly, J. Reddy, V. Ahuja, V. Panduga, M. Parab, K. G. Vishwas, N. Kumar, M. Balganesh and V. Balasubramanian, *Antimicrobial agents and chemotherapy*, 2013, **57**, 2506-2510.
172. H. F. Chambers, J. Turner, G. F. Schectcr, M. Kawamura and P. C. Hopewell, *Antimicrobial agents and chemotherapy*, 2005, **49**, 2816-2821.
173. N. Veziris, C. Truffot, J. L. Mainardi and V. Jarlier, *Antimicrobial agents and chemotherapy*, 2011, **55**, 2597-2600.
174. C. S. Wong, G. S. Palmer and M. H. Cynamon, *The Journal of antimicrobial chemotherapy*, 1988, **22**, 863-866.
175. S. K. Agarwal, *Chest*, 2004, **126**, 752s-752s
176. N. Rastogi, K. S. Goh, P. Ruiz and M. Casal, *Antimicrobial agents and chemotherapy*, 1995, **39**, 1162-1165.
177. S. J. Cavalieri, J. R. Biehle and W. E. Sanders, Jr., *Antimicrobial agents and chemotherapy*, 1995, **39**, 1542-1545.
178. M. S. Bolhuis, T. van der Laan, J. G. W. Kosterink, T. S. van der Werf, D. van Soolingen and J.-W. C. Alffenaar, *European Respiratory Journal*, 2014, **44**, 808-811.
179. S. Ananthan, E. R. Faaleolea, R. C. Goldman, J. V. Hobrath, C. D. Kwong, B. E. Laughon, J. A. Maddry, A. Mehta, L. Rasmussen, R. C. Reynolds, J. A. Secrist, 3rd, N. Shindo, D. N. Showe, M. I. Sosa, W. J. Suling and E. L. White, *Tuberculosis*, 2009, **89**, 334-353.
180. G. Xu, H. Li and S. Liu, *Zhonghua jie he he hu xi za zhi = Zhonghua jiehe he huxi zazhi = Chinese journal of tuberculosis and respiratory diseases*, 1995, **18**, 175-177, 192.
181. A. Aubry, X. S. Pan, L. M. Fisher, V. Jarlier and E. Cambau, *Antimicrobial agents and chemotherapy*, 2004, **48**, 1281-1288.
182. H. B. Xu, R. H. Jiang and H. P. Xiao, *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 2012, **18**, 1104-1110.
183. M. J. Ruiz-Serrano, L. Alcala, L. Martinez, M. Diaz, M. Marin, M. J. Gonzalez-Abad and E. Bouza, *Antimicrobial agents and chemotherapy*, 2000, **44**, 2567-2568.
184. S. H. Gillespie and O. Billington, *J Antimicrob Chemoth*, 1999, **44**, 393-395.
185. J. C. Rodriguez, M. Ruiz, M. Lopez and G. Royo, *International journal of antimicrobial agents*, 2002, **20**, 464-467.
186. S. H. Gillespie, A. M. Crook, T. D. McHugh, C. M. Mendel, S. K. Meredith, S. R. Murray, F. Pappas, P. P. J. Phillips, A. J. Nunn and R. Consortium, *New Engl J Med*, 2014, **371**, 1577-1587.
187. C. Truffot-Pernot, B. Ji and J. Grosset, *Tubercle*, 1991, **72**, 57-64.
188. M. Satti, F. Faqir, A. Sattar, S. Abbasi, T. Butt, K. A. Karamat and M. Abidi, *Journal of Ayub Medical College, Abbottabad : JAMC*, 2010, **22**, 101-103.
189. C. Piersimoni, V. Morbiducci, S. Bornigia, G. De Sio and G. Scalise, *The American review of respiratory disease*, 1992, **146**, 1445-1447.
190. S. Sulochana, F. Rahman and C. N. Paramasivan, *Journal of chemotherapy*, 2005, **17**, 169-173.
191. E. J. Alvirez-Freites, J. L. Carter and M. H. Cynamon, *Antimicrobial agents and chemotherapy*, 2002, **46**, 1022-1025.
192. C. S. Merle, K. Fielding, O. B. Sow, M. Gninafon, M. B. Lo, T. Mthiyane, J. Odhiambo, E. Amukoye, B. Bah, F. Kassa, A. N'Diaye, R. Rustomjee, B. C. de Jong, J. Horton, C. Perronne, C. Sismanidis, O. Lapujade, P. L. Olliaro, C. Lienhardt and O. F. G. f. T. Project, *The New England journal of medicine*, 2014, **371**, 1588-1598.
193. J. D. Gay, D. R. DeYoung and G. D. Roberts, *Antimicrobial agents and chemotherapy*, 1984, **26**, 94-96.
194. D. C. Perlman, W. M. El Sadr, L. B. Heifets, E. T. Nelson, J. P. Matts, K. Chirgwin, N. Salomon, E. E. Telzak, O. Klein, B. N. Kreiswirth, J. M. Musser and R. Hafner, *Aids*, 1997, **11**, 1473-1478.
195. W. W. Yew, C. K. Chan, C. H. Chau, C. M. Tam, C. C. Leung, P. C. Wong and J. Lee, *Chest*, 2000, **117**, 744-751.

196. J. L. Johnson, D. J. Hadad, W. H. Boom, C. L. Daley, C. A. Peloquin, K. D. Eisenach, D. D. Jankus, S. M. Debanne, E. D. Charlebois, E. Maciel, M. Palaci and R. Dietze, *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease*, 2006, **10**, 605-612.
197. R. K. Shandil, R. Jayaram, P. Kaur, S. Gaonkar, B. L. Suresh, B. N. Mahesh, R. Jayashree, V. Nandi, S. Bharath and V. Balasubramanian, *Antimicrobial agents and chemotherapy*, 2007, **51**, 576-582.
198. E. Rey-Jurado, G. Tudo, J. A. Martinez and J. Gonzalez-Martin, *Tuberculosis*, 2012, **92**, 260-263.
199. A. Lubasch, R. Erbes, H. Mauch and H. Lode, *The European respiratory journal*, 2001, **17**, 641-646.
200. H. Tomioka, K. Sato and H. Saito, *Tubercle*, 1991, **72**, 176-180.
201. N. F. Walker, S. O. Clark, T. Oni, N. Andreu, L. Tezera, S. Singh, L. Saraiva, B. Pedersen, D. L. Kelly, J. A. Tree, J. M. D'Armiento, G. Meintjes, F. A. Mauri, A. Williams, R. J. Wilkinson, J. S. Friedland and P. T. Elkington, *Am J Resp Crit Care*, 2012, **185**, 989-997.
202. R. Fernandez-Roblas, N. Z. Martin-de-Hijas, A. I. Fernandez-Martinez, D. Garcia-Almeida, I. Gadea and J. Esteban, *Antimicrobial agents and chemotherapy*, 2008, **52**, 4184-4186.
203. B. Singh and D. A. Mitchison, *Journal of general microbiology*, 1955, **13**, 176-184.
204. R. M. Cayre and J. Viallier, *Comptes rendus des seances de la Societe de biologie et de ses filiales*, 1956, **150**, 1202-1204.
205. M. Tsukamura, S. Mizuno and H. Toyama, *Kekkaku : [Tuberculosis]*, 1983, **58**, 665-667.
206. P. C. Lombardo and I. Weitzman, *The New England journal of medicine*, 1990, **323**, 916-917.

Tables

Table 1. List of some of the predicted drug-target associations in the context of drug classes. The complete list is provided as a supplementary material (Supplementary Table 2)

Sr. no.	DrugBank ID	Drug name	Drug class	Predicted targets
1.	DB00336	Nitrofural*	Furans	*pstP* (Rv0118c), *mdh* (Rv1240), *ilvG* (Rv1820), *ilvB1* (Rv3003c), *ilvB2* (Rv3470c)
2.	DB00551	Acetohydroxamic acid	Carboxylic acids and derivatives	*ureC* (Rv1850)
3.	DB00560	Tigecycline*	Tetracyclines	*rpsL* (Rv0682), *rpsS* (Rv0705), *rpsN2* (Rv2056c), *rpsI* (Rv3442c), *rpsM* (Rv3460c)
4.	DB00698	Nitrofurantoin*	Azolidines	Rv0306, *rpsJ* (Rv0700)
5.	DB00730	Thiabendazole	Benzimidazoles	Rv0248c, *frdA* (Rv1552), *nadB* (Rv1595), *sdhA* (Rv3318)
6.	DB01421	Paromomycin*	Carbohydrate conjugates	*rpsJ* (Rv0700)
7.	DB01598	Imipenem*	Lactams	*pbpA* (Rv0016c), *ponA1* (Rv0050), *blaC* (Rv2068c), *pbpB* (Rv2163c), Rv2864c, *ponA2* (Rv3682)
8.	DB02703	Fusidic acid*	Steroids and steroid derivatives	*fusA2* (Rv0120c), *fusA1* (Rv0684)
9.	DB04794	Bifonazole	Benzene and substituted derivatives	*cyp138* (Rv0136), *cyp135A1* (Rv0327c), *cyp135B1* (Rv0568), *cyp51* (Rv0764c), *cyp139* (Rv1666c), *cyp136* (Rv3059)
10.	DB08933	Luliconazole	Azoles	*cyp138* (Rv0136), *cyp135A1* (Rv0327c), *cyp135B1* (Rv0568), *cyp51* (Rv0764c), *cyp139* (Rv1666c), *cyp136* (Rv3059)

*Experimentally investigated earlier for their anti-tubercular activity

Table 2. Details on some of the potential targets identified in *M. tuberculosis*. The complete list is provided as a supplementary material (Supplementary Table 3)

Rv ID	Essential gene?	Protein description	Crystal structure/ model [region]	UniProt ID of known target homologous to Mtb protein	Protein description	Crystal structure/ model [region]	Source organism	Seq. identity	TM-score	Drug predicted to be associated with Mtb protein (DrugBank ID:name)
Rv0120c	No	Elongation factor G FusA2	Model (2DY1:A) [21-714]	Q5SHN5	Elongation factor G	4V5F:AY [1-691]	*Thermus thermophilus*	34%	0.68	DB02703: Fusidic acid
Rv0684	Yes	Elongation factor G FusA1	Model (2XEX:A) [3-701]	Q5SHN5	Elongation factor G	4V5F:AY [1-691]	*Thermus thermophilus*	60%	0.80	DB02703: Fusidic acid
Rv1850	No	Urease alpha subunit UreC	Model (4UBP:C) [3-577]	P18314	Urease subunit alpha	1FWE:C [1-567]	*Enterobacter aerogenes*	60%	0.95	DB00551: Acetohydroxamic acid
Rv2139	Yes	Dihydro-orotate dehydrogenase	Model (1F76:A) [1-332]	Q08210	Dihydro-orotate dehydrogenase (quinone)	1TV5:A [158-569]	*Plasmodium falciparum*	32%	0.92	DB01117: Atovaquone
Rv2447c	Yes	Folylpoly-glutamate synthase FolC	2VOS:A [1-487]	P08192	Bifunctional protein FolC	1W78:A [1-422]	*Escherichia coli* K12	29%	0.85	DB01015: Sulfamethoxazole

Table 3. List of *M. tuberculosis* targets identified in our analysis concurring with previously published works.

Sr. no.	Rv ID	Associated drug(s) based on our analysis	targetTB target (Raman et al. 2008)	*Mtb* pocketome polypharmacological target (Anand and Chandra, 2014)	TB-drugome (Kinnings et al. 2010)
1.	Rv0120c	Fusidic Acid	Yes	No	No
2.	Rv0136	Fluconazole, Terconazole, Voriconazole, Ketoconazole, Miconazole, Sertaconazole, Posaconazole, Bifonazole, Luliconazole	No	No	Yes
3.	Rv0327c	Fluconazole, Terconazole, Voriconazole, Ketoconazole, Miconazole, Sertaconazole, Posaconazole, Itraconazole, Bifonazole, Luliconazole	No	No	Yes
4.	Rv0399c	Cefalotin	Yes	No	No
5.	Rv0547c	Triclosan	No	No	Yes
6.	Rv0568	Fluconazole, Terconazole, Voriconazole, Ketoconazole, Miconazole, Sertaconazole, Posaconazole, Itraconazole, Bifonazole, Luliconazole			
7.	Rv0651	Roxithromycin, Clindamycin, Clarithromycin, Quinupristin, Lincomycin	No	Yes	No
8.	Rv0687	Triclosan	No	Yes	No
9.	Rv0706	Azithromycin, Quinupristin	Yes	No	No
10.	Rv0764c	Fluconazole, Terconazole, Voriconazole, Ketoconazole, Miconazole, Sertaconazole, Posaconazole, Itraconazole, Bifonazole, Luliconazole	No	No	Yes
11.	Rv1393c	Clofazimine	No	No	Yes
12.	Rv1484	Triclosan	Yes	No	No
13.	Rv1595	Thiabendazole	No	No	Yes

14.	Rv1666c	Fluconazole, Terconazole, Voriconazole, Ketoconazole, Miconazole, Sertaconazole, Posaconazole, Itraconazole, Bifonazole, Luliconazole	No	Yes	No
15.	Rv1714	Triclosan	Yes	No	No
16.	Rv1820	Nitrofural	No	No	Yes
17.	Rv1850	Acetohydroxamic acid	Yes	No	No
18.	Rv1941	Triclosan	No	Yes	Yes
19.	Rv2139	Atovaquone	Yes	No	Yes
20.	Rv2703	Fidaxomicin	Yes	No	No
21.	Rv3458c	Doxycycline, Lymecycline, Clomocycline, Oxytetracycline, Demeclocycline, Minocycline	Yes	No	No
22.	Rv3559c	Triclosan	No	No	Yes

Table 4. Few examples of ChEMBL compounds similar to FDA-approved drugs. The entire list is provided as a supplementary material (Supplementary Table 5)

Sr. no.	DrugBank ID	Drug name	Molecular formula	ChEMBL compound	Molecular formula	SHAFTS score
1.	DB00487	Perfloxacin	C17H20FN3O3	CHEMBL732	C16H18FN3O4	1.225
				CHEMBL735	C16H16FN3O4	1.26
				CHEMBL1180570	C20H25FN3O3	2.016
2.	DB00817	Rosoxacin	C17H14N2O3	CHEMBL288437	C13H13NO3	1.715
3.	DB01015	Sulfamethoxazole	C10H11N3O3S	CHEMBL610753	C10H9N3O5S	1.349
				CHEMBL608998	C10H10N2O3S	1.499
				CHEMBL1972802	C10H11N3O4S	1.585
4.	DB01051	Novobiocin	C31H36N2O11	CHEMBL1834263	C36H45N3O14S	1.726
				CHEMBL1834269	C37H48N4O12	1.824
5.	DB02703	Fusidic acid	C31H48O6	CHEMBL1552107	C31H48O6	2.309
				CHEMBL1477084	C31H48O6	2.132

Legends to Figures

Figure 1. Workflow adopted. A schematic diagram of the steps adopted for identification of drug-target associations for *M. tuberculosis* is illustrated.

Figure 2. Distribution of predicted drug-target associations on the basis of drug classes. A bar chart representation of distribution of approved drugs across 19 drug classes (in grey) and number of targets (in light grey) identified to be associated with each class is depicted.

Figure 3. Functional importance of the potential targets recognized in *M. tuberculosis*. The potential targets could be grouped into four functional categories which are represented in the form of a sorted chord layout. The size of each section of the chord is akin to the number of *M. tuberculosis* proteins under that section, denoted in brackets. Listed towards one end of each section are the Rv IDs pertaining to a functional category. Information on genes which are essential for growth and survival of *M. tuberculosis* are indicated as rectangular blocks coloured according to the functional category, while the blue dots depict the genes active during log-phase growth of the organism.

Figure 4. Drug-target network. The set of 130 approved drugs which have the potential to be repurposed against 78 targets in *M. tuberculosis* are rendered as connected components in the network diagram. The coloured nodes correspond to *M. tuberculosis* targets, coloured according to functional categories, while the white nodes pertain to drugs. The colour key adopted for Rv IDs is identical to that portrayed in Figure-3, which is as follows: yellow, “intermediary metabolism and respiration”; green, “information pathways”; red, “cell wall and cell processes” and violet, “lipid metabolism”. The numbers associated with the connected components denote drug classes ranked based on number of drugs. The descriptions of drug class under each rank are as follows: (1) Lactams; (2) Benzene and substituted derivatives; (3) Quinolines and derivatives; (4) Carboxylic acid and derivatives; (5) Tetracyclines; (6) Macrolides and analogues; (7) Carbohydrate conjugates; (8) Diazines; (9) Naphthyridines; (10) Azoles; (11) Steroids and steroid derivatives; (12) Naphthalenes; (13) Furans; (14) Fatty acyls; (15) Cyclohexylamines; (16) Coumarins and derivatives; (17) Benzothiophenes; (18) Bezimidazoles; (19) Azolidines.

Figure 5. Mycobacterial urease as a viable target for acetohydroxamic acid. A) A cartoon representation of superposition of modelled structure of *M. tuberculosis ureC* (urease subunit alpha), in yellow, on the acetohydroxamate-bound crystal structure of *E. aerogenes* urease subunit alpha, in blue, (PDB code: 1FWE:C) is shown. The bound ligand, acetohydroxamic acid, is shown in stick representation with nickel ions as spheres. The binding site of acetohydroxamic acid in the alpha subunit of urease is indicated with an arrow. B) Inhibitor-binding residues and the metal coordinating residues in *E. aerogenes* urease (in blue) observed to be completely conserved in mycobacterial *ureC* (in yellow) are shown in stick representation. This figure and Figure-6 are generated using PyMOL[122] (http://www.pymol.org).

Figure 6. Inferences on fusidic acid binding residues in *M. tuberculosis* targets. A) Multiple sequence alignment of mycobacterial *fusA1, fusA2* and their corresponding homologue *fusA* (UniProt ID: Q5SHN5|EFG_THET8) of *T. thermophilus* is shown. Only those aligned regions are shown which house residues involved in fusidic acid binding as reported for *fusA,* which are conserved across *fusA1* and *fusA2.* The conserved residues are indicated as triangular symbols. B) Fusidic acid binding residues in *fusA* as indicated in the alignment are represented as sticks (PDB code: 4V5F:AY) with their van der Waals radii represented as dots.

Figure 7. Sulfamethoxazole and its structurally similar ChEMBL compounds. 2D chemical structures of the drug sulfamethoxazole and the three structurally similar ChEMBL compounds (CHEMBL610753, CHEMBL608998, CHEMBL1972802) are depicted in the left panel of the figure, while the right panel depicts their respective pairwise 3D alignment. Visualization and alignment of chemical structures was pursued using Marvin 15.5.18, 2015, ChemAxon (http://www.chemaxon.com)

Figure 8. Characterizing the site and the mode of binding of sulfamethoxazole. A) The bifunctional enzyme *folC* of *M. tuberculosis* (PDB code: 2VOS) is shown in cartoon representation (grey) with a zoomed-in view of the high scoring binding site recognized using SiteMap. This high scoring binding site sandwiched between N-terminal ATPase domain and C-terminal Rossmann fold domain houses functionally important loops as highlighted in B). The proximity of the docked sulfamethoxazole (shown as sticks) to these loops is illustrated in B). Cobalt and magnesium ions are represented as blue and pink spheres, respectively. C) The

probable residues influenced (ball and stick representation) upon binding of sulfamethoxazole are indicated. Note: The missing coordinates of the disordered residues of α1-α2 DHP-binding loop were filled, corrected and refined using Prime[123], availed through Schrodinger suite. Figures were generated using Maestro[124].

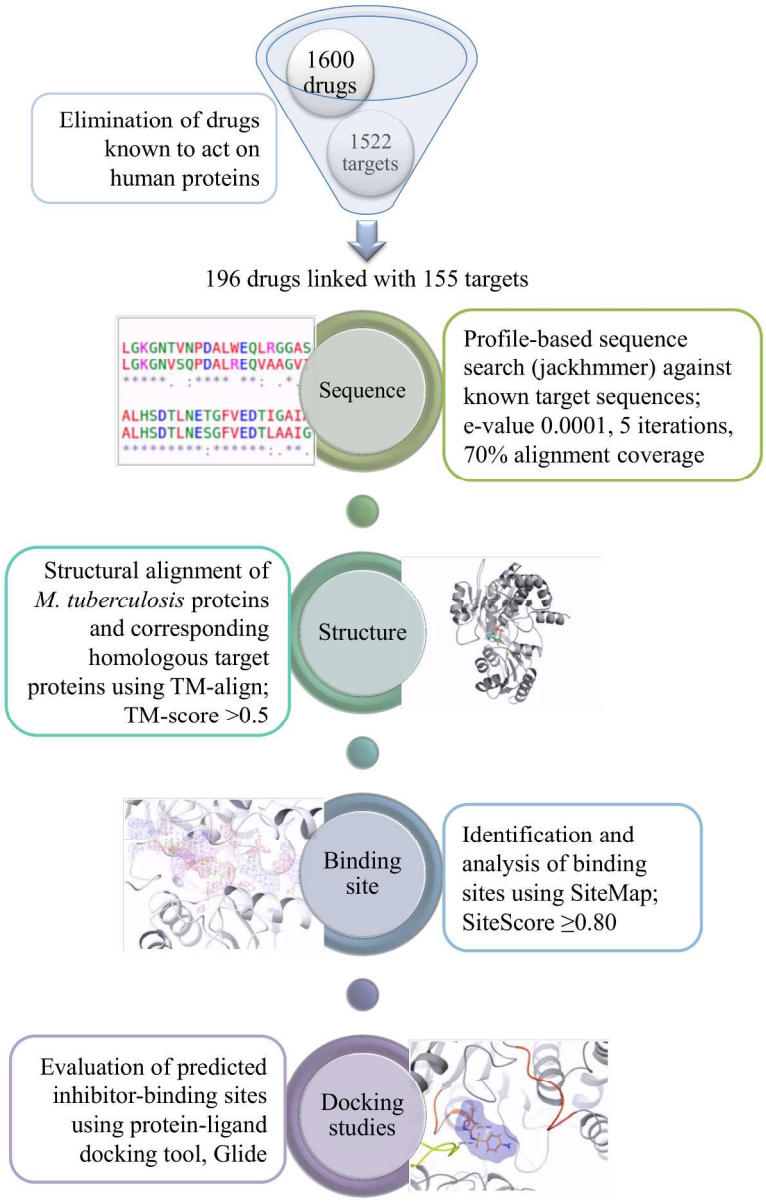


Figure 1. Workflow adopted. A schematic diagram of the steps adopted for identification of drug-target associations for M. tuberculosis is illustrated.
190x279mm (300 x 300 DPI)

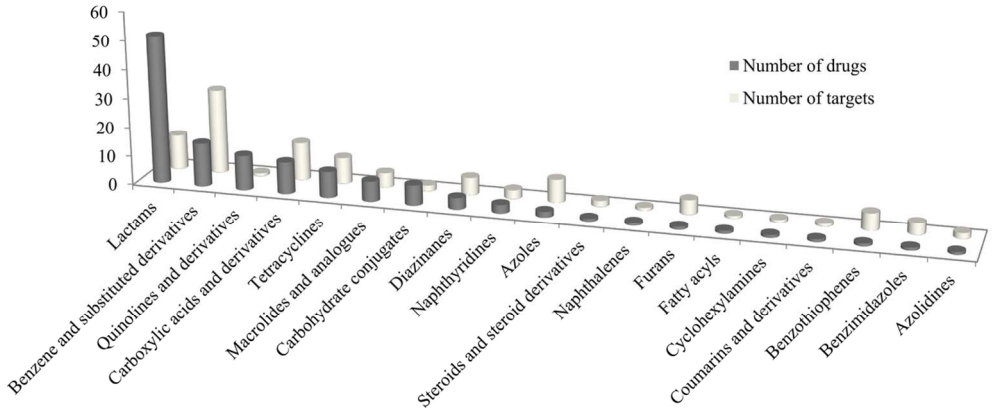


Figure 2. Distribution of predicted drug-target associations on the basis of drug classes. A bar chart representation of distribution of approved drugs across 19 drug classes (in grey) and number of targets (in light grey) identified to be associated with each class is depicted.
114x52mm (300 x 300 DPI)

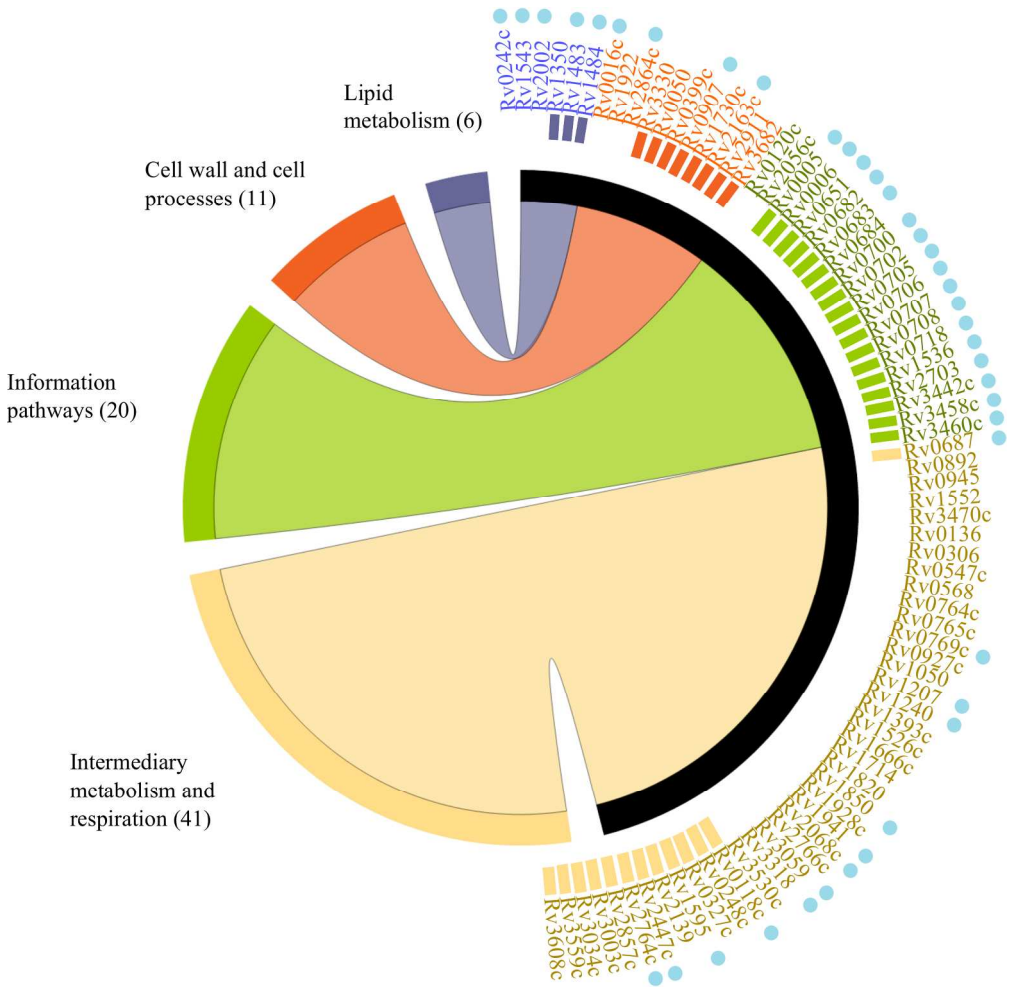


Figure 3. Functional importance of the potential targets recognized in M. tuberculosis. The potential targets could be grouped into four functional categories which are represented in the form of a sorted chord layout. The size of each section of the chord is akin to the number of M. tuberculosis proteins under that section, denoted in brackets. Listed towards one end of each section are the Rv IDs pertaining to a functional category. Information on genes which are essential for growth and survival of M. tuberculosis are indicated as rectangular blocks coloured according to the functional category, while the blue dots depict the genes active during log-phase growth of the organism.
197x193mm (300 x 300 DPI)

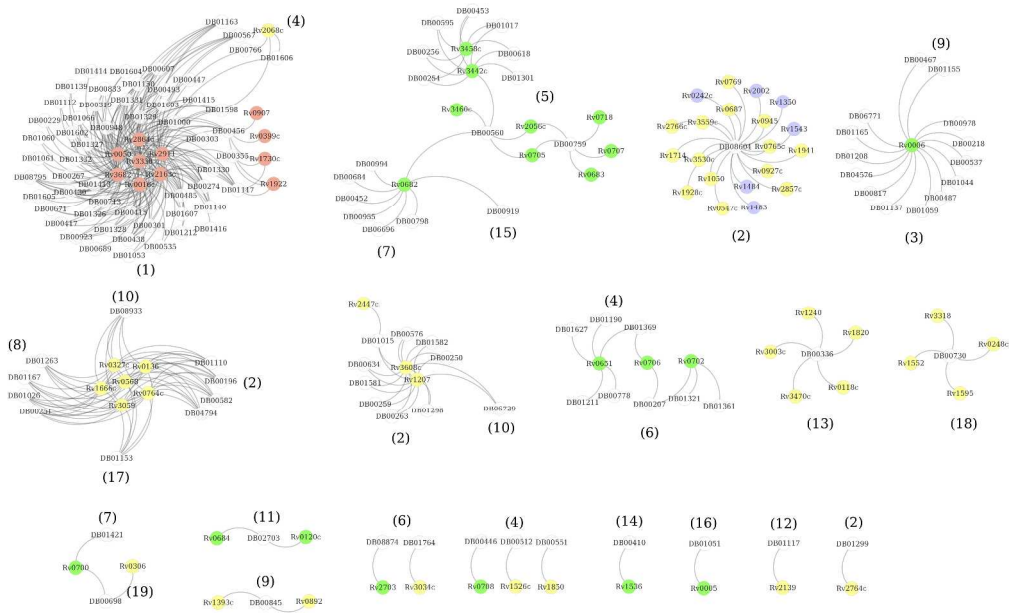


Figure 4. Drug-target network. The set of 130 approved drugs which have the potential to be repurposed against 78 targets in M. tuberculosis are rendered as connected components in the network diagram. The coloured nodes correspond to M. tuberculosis targets, coloured according to functional categories, while the white nodes pertain to drugs. The colour key adopted for Rv IDs is identical to that portrayed in Figure-3, which is as follows: yellow, "intermediary metabolism and respiration"; green, "information pathways"; red, "cell wall and cell processes" and violet, "lipid metabolism". The numbers associated with the connected components denote drug classes ranked based on number of drugs. The descriptions of drug class under each rank are as follows: (1) Lactams; (2) Benzene and substituted derivatives; (3) Quinolines and derivatives; (4) Carboxylic acid and derivatives; (5) Tetracyclines; (6) Macrolides and analogues; (7) Carbohydrate conjugates; (8) Diazines; (9) Naphthyridines; (10) Azoles; (11) Steroids and steroid derivatives; (12) Naphthalenes; (13) Furans; (14) Fatty acyls; (15) Cyclohexylamines; (16) Coumarins and derivatives; (17) Benzothiophenes; (18) Bezimidazoles; (19) Azolidines.

2362x1436mm (72 x 72 DPI)

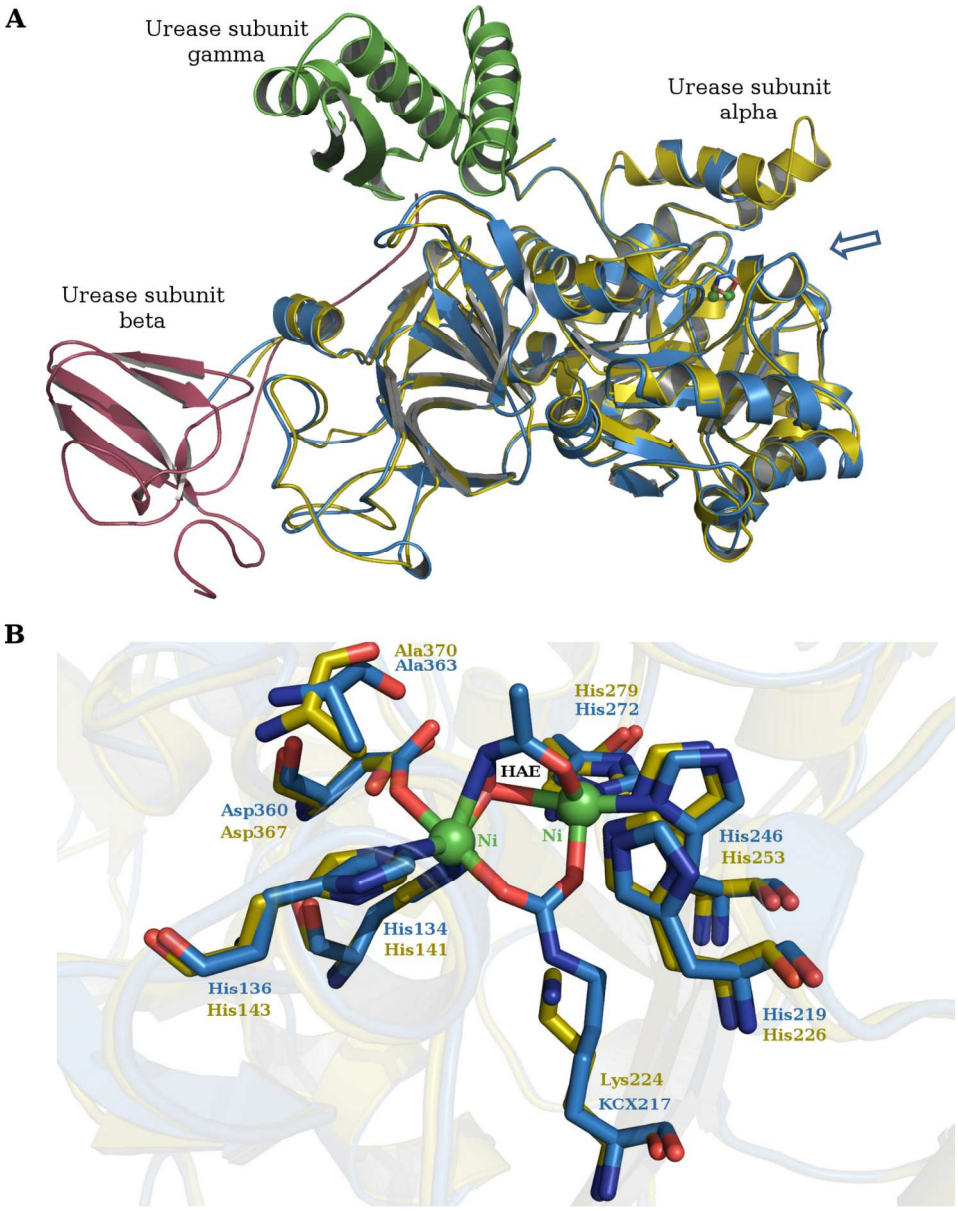


Figure 5. Mycobacterial urease as a viable target for acetohydroxamic acid. A) A cartoon representation of superposition of modelled structure of M. tuberculosis ureC (urease subunit alpha), in yellow, on the acetohydroxamate-bound crystal structure of E. aerogenes urease subunit alpha, in blue, (PDB code: 1FWE:C) is shown. The bound ligand, acetohydroxamic acid, is shown in stick representation with nickel ions as spheres. The binding site of acetohydroxamic acid in the alpha subunit of urease is indicated with an arrow. B) Inhibitor-binding residues and the metal coordinating residues in E. aerogenes urease (in blue) observed to be completely conserved in mycobacterial ureC (in yellow) are shown in stick representation. This figure and Figure-6 are generated using PyMOL122 (http://www.pymol.org).
203x256mm (300 x 300 DPI)

A

```
Rv0120c|fusA2      TVCDFDEAEIRQQRSVGLAVASLAYDGIKVNLVDTPGYADFVGELRAGLRAADCALFVIA   120
Q5SHN5|EFG_THET8   ATMDFMEQERERGITITAAVTTCFWKDHRINIIDTPGHVDFTIEVERSMRVLDGAIVVFD   109
Rv0684|fusA1       ATMDWMEQEQERGITITSAATTTFWKDNQLNIIDTPGHVDFTVEVERNLRVLDGAVAVFD   110
                   :. *: * * .:  ::  *.::  :.  ::*::****:.**. *:.  :*. * *: *:

Rv0120c|fusA2      DTLSDK-AEPLVLKPWTMPEPLLPIAIAAHAKTDEDKLSVGLGRLAAEDPTLRIEQNQET   467
Q5SHN5|EFG_THET8   DTLVGEDAPRVILESIEVPEPVIDVAIEPKTKADQEKLSQALARLAEEDPTFRVSTHPET   446
Rv0684|fusA1       DTLSDP-NQQIVLESMTFPDPVIEVAIEPKTKSDQEKLSLSIQKLAEEDPTFKVHLDSET   450
                   ***       ::*:    .*:*:: :**  ::*:*::*** .: :** ****:::  . **

Rv0120c|fusA2      HQVVLWCMGEAHAGVVLDTLANRYGVSVDTIELRVPLRETFAGNAK-GHGRHIKQSGGHG   526
Q5SHN5|EFG_THET8   GQTIISGMGELHLEIIVDRLKREFKVDANVGKPQVAYRETITKPVD-VEGKFIRQTGGRG   505
Rv0684|fusA1       GQTVIGGMGELHLDILVDRMRREFKVEANVGKPQVAYKETIKRLVQNVEYTHKKQTGGSG   510
                    *.::   *** *   :::* : ..: *..:. : :*  :**:    ..  .   . :*:** *
```

B

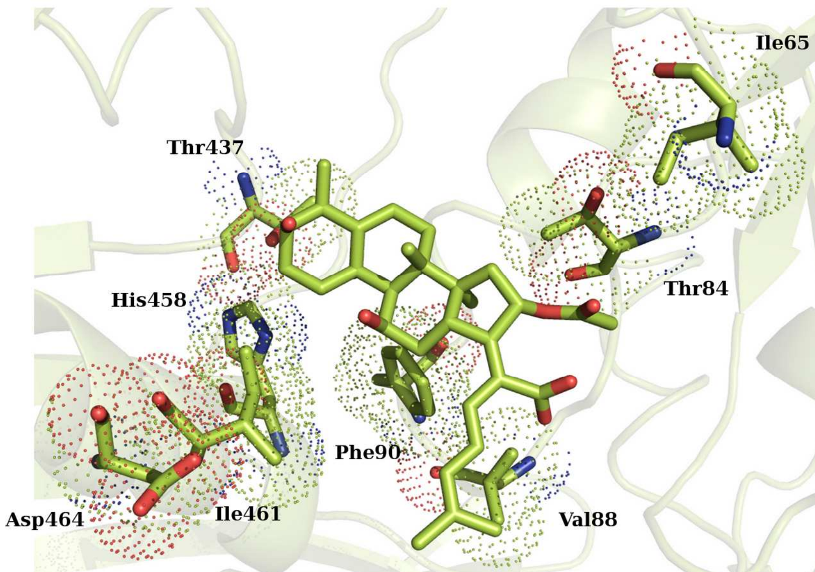


Figure 6. Inferences on fusidic acid binding residues in M. tuberculosis targets. A) Multiple sequence alignment of mycobacterial fusA1, fusA2 and their corresponding homologue fusA (UniProt ID: Q5SHN5|EFG_THET8) of T. thermophilus is shown. Only those aligned regions are shown which house residues involved in fusidic acid binding as reported for fusA, which are conserved across fusA1 and fusA2. The conserved residues are indicated as triangular symbols. B) Fusidic acid binding residues in fusA as indicated in the alignment are represented as sticks (PDB code: 4V5F:AY) with their van der Waals radii represented as dots.

177x162mm (300 x 300 DPI)

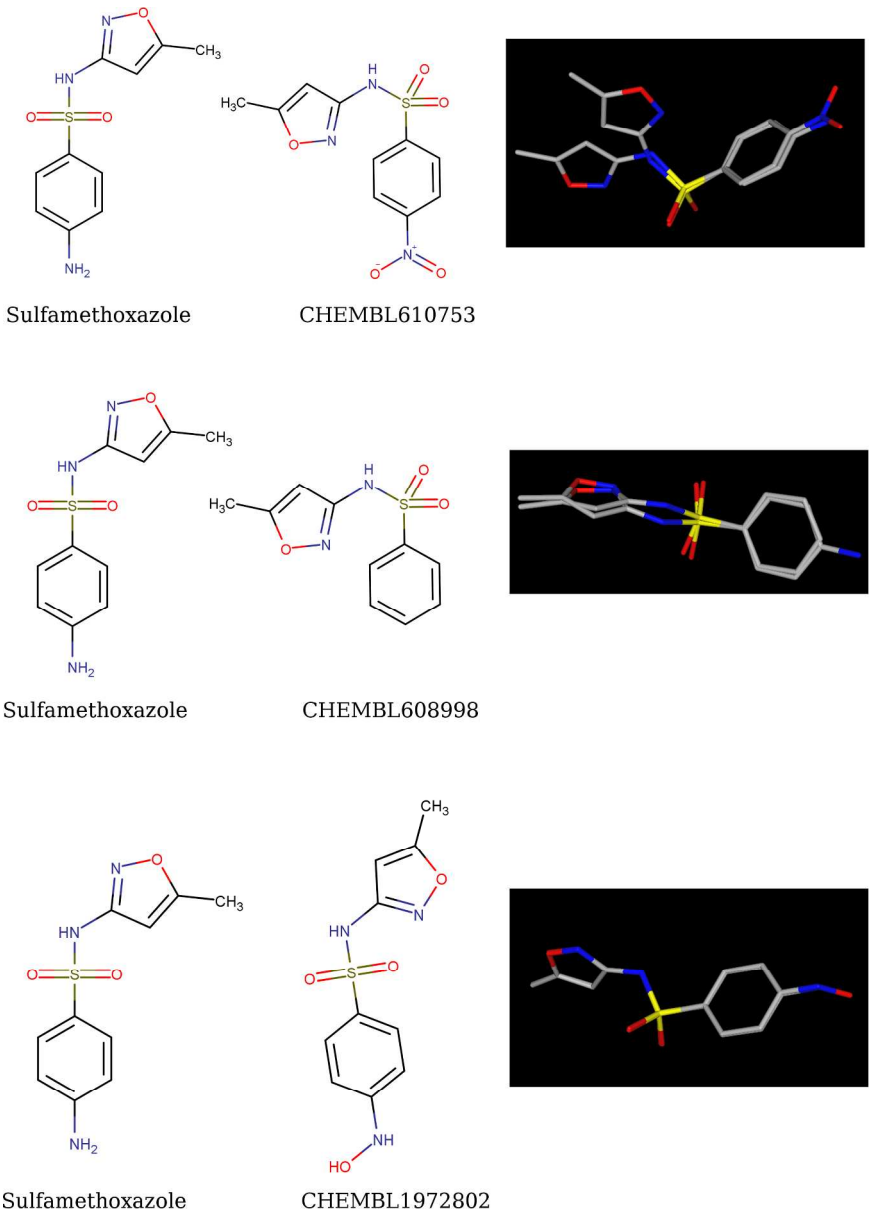


Figure 7. Sulfamethoxazole and its structurally similar ChEMBL compounds. 2D chemical structures of the drug sulfamethoxazole and the three structurally similar ChEMBL compounds (CHEMBL610753, CHEMBL608998, CHEMBL1972802) are depicted in the left panel of the figure, while the right panel depicts their respective pairwise 3D alignment. Visualization and alignment of chemical structures was pursued using Marvin 15.5.18, 2015, ChemAxon (http://www.chemaxon.com)
196x271mm (300 x 300 DPI)

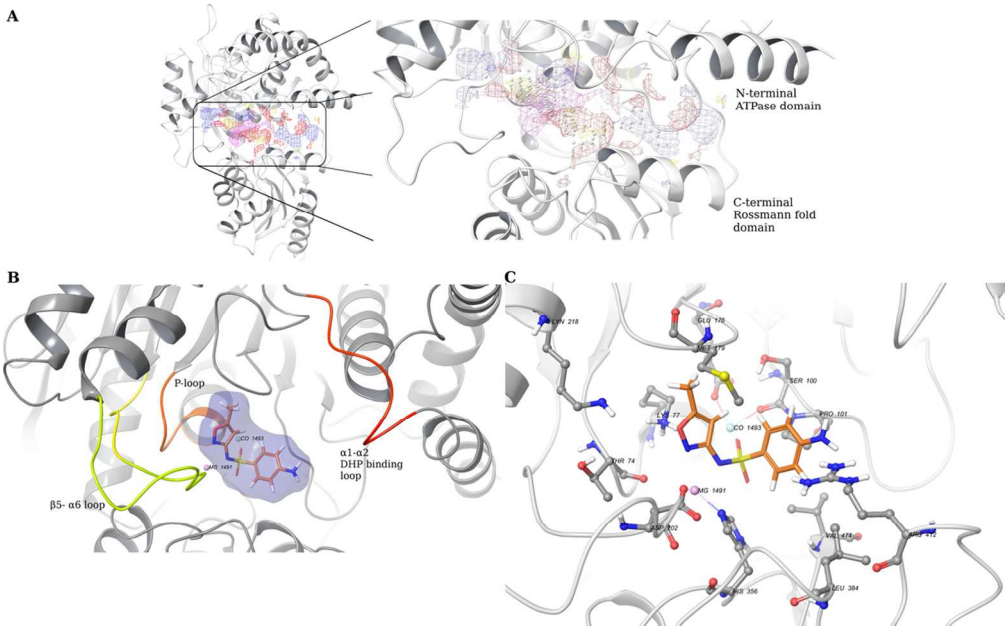


Figure 8. Characterizing the site and the mode of binding of sulfamethoxazole. A) The bifunctional enzyme folC of M. tuberculosis (PDB code: 2VOS) is shown in cartoon representation (grey) with a zoomed-in view of the high scoring binding site recognized using SiteMap. This high scoring binding site sandwiched between N-terminal ATPase domain and C-terminal Rossmann fold domain houses functionally important loops as highlighted in B). The proximity of the docked sulfamethoxazole (shown as sticks) to these loops is illustrated in B). Cobalt and magnesium ions are represented as blue and pink spheres, respectively. C) The probable residues influenced (ball and stick representation) upon binding of sulfamethoxazole are indicated. Note: The missing coordinates of the disordered residues of α1-α2 DHP-binding loop were filled, corrected and refined using Prime71, availed through Schrodinger suite. Figures were generated using Maestro72.

110x68mm (300 x 300 DPI)