Repurposing of FDA-Approved drugs to predict new inhibitors against key regulatory genes in \textit{Mycobacterium tuberculosis}

XINJUN YANG$^1$; AFTAB ALAM$^2$; NAIYAR IQBAL$^3$; KHALID RAZA$^4$*

$^1$ Department of Pediatrics, 3201 Hospital, Hanzhong, China
$^2$ Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India
$^3$ Department of Computer Science and Information Technology, Maulana Azad National Urdu University, Hyderabad, India
$^4$ Department of Computer Science, Jamia Millia Islamia, New Delhi, India

Key words: Virtual screening, Molecular docking, MD simulation, PPI network

Abstract: Tuberculosis (TB) disease has become one of the major public health concerns globally, especially in developing countries. Numerous research studies have already been carried out for TB, but we are still struggling for a complete and quick cure for it. The progress of \textit{Mycobacterium tuberculosis} (MTB) strains resistant to existing drugs makes its cure and control very complicated. Therefore, it is the need of the hour to search for newer and effective drugs that can inhibit an increasing number of putative drug targets. We applied the drug repurposing concept to identify promising FDA-approved drugs against five key-regulatory genes (FurB, IdeR, KstR, MosR, and RegX3) of the MTB. The FDA drugs were virtually screened using a structure-based approach by GOLD versions 5.2, and subjected to rigid docking followed by an induced-fit docking algorithm to enhance the accuracy and prioritize drugs for repurposing. We found 11 candidate drugs (including ZINC03871613, ZINC03871614, ZINC03871615 as top scorer candidate drugs) that were frequently present within the top 20 GoldScore ranks and showed promising results. Furthermore, molecular dynamics simulation was performed to monitor the effect of the top scorer drugs on the structural stability of all the five targets, indicating that inhibitors preferentially bind to the active site of the targets. This work suggests that these known FDA-approved drugs open new application domains in the form of anti-tuberculosis agents.

Introduction

Tuberculosis (TB) is one of the oldest diseases with molecular evidence that dates back to 17,000 years ago. Despite advances in modalities for diagnosis and treatment of TB, unfortunately, people are still suffering, and it is among the top 10 killer communicable diseases worldwide, second only to HIV. TB is a communicable disease caused by \textit{Mycobacterium tuberculosis}. It typically affects the lungs (pulmonary TB) but can affect other sites as well (extrapulmonary TB) (Ahmad, 2011). The disease is spread in the air when people who are sick with pulmonary diseases expel bacteria during coughing. According to WHO, six countries hold 60% of all the TB burden of the entire world in which India is the leading country followed by Indonesia, China, Nigeria, Pakistan, and South Africa (shown in Fig. 1). The Central TB Division, Ministry of Health and Family Welfare, India has published a report on “India TB report 2018” which reported TB statistics for India for 2016 to give an estimated incidence figure of 2.79 million cases of TB for India and up to 4.23 lakh patients were estimated to have died during the year (India TB Report, 2018). Multidrug-resistant (MDR) tuberculosis (TB) is a serious global public health problem (Dua et al., 2018), and estimates indicate that unless the management of MDR TB changes radically, it will be one of the main drivers of antimicrobial resistance, which could kill more persons than cancer by 2050 (Ho et al., 2016). Major challenges to curb the TB in India include poor primary healthcare system in rural areas, unregulated private health care leading to the indiscriminate use of first-line and second-line anti-TB drugs, spreading HIV infection, poverty, and lack of administrative coordination among government bodies responsible for health and hygiene (Singh et al., 2017). Hence, there is an urgent need to design or developed nontoxic biocompatible drugs against TB (Peters et al., 2020; Weiner et al., 2020).

In our study, we have used the drug repurposing concept, which is a method for disease cure that accelerates the identification of new uses for existing drugs with minimum
side effects (Karaman and Sippl, 2019; Lanza et al., 2018; Savoia, 2016). Previous research has shown that the drug repurposing concept (a combination of virtual screening with docking and dynamics studies) can be successfully applied to treat new diseases with existing drugs. For example, the drug sildenafil, which was initially designed for the treatment of hypertension and ischemic heart disease, was later approved for the treatment of erectile dysfunction, representing a successful history of drug repurposing (Afzal et al., 2014; Karthick and Ramanathan, 2013; Li et al., 2001; Palos et al., 2017). Here, we use a computer-guided drug repurposing method and apply it to virtually screened FDA approved drugs against five key regulatory proteins from MTB; (a) FurB—as shown by the regulatory profiles, furB gene is up-regulated in macrophages-phagocytosed bacteria, fascinating that it might be involved in the regulation of intracellular gene expressions. The furB gene is co-transcribed with its upstream gene (Rv2358), which encodes another zinc-dependent regulator and responsible for repressing at least 32 genes, several of which have been implicated with zinc homeostasis. In addition, FurB controls five genes encoding ribosomal proteins three of which containing a putative zinc-binding motif. Based on comparative genomics data these ribosomal proteins have also been suggested to be involved in zinc homeostasis (Lucarelli et al., 2007; Panina et al., 2003). (b) IdeR—It is an iron-dependent regulator that maintains a moderate level of iron inside a bacterium by controlling the transcription of genes, including the expression of its own gene involved in iron uptake, transport, and storage (Banerjee et al., 2011). In M. tuberculosis regulation of the intracellular levels of iron are performed by the transcription factor IdeR. In the iron sufficient stage in MTB, iron binds and activates intracellular IdeR, which then represses the iron acquisition machinery of the pathogen and activates the synthesis of iron storage proteins. IdeR also plays a key role in protecting the cells against the oxidative stress generated by the host (Rohilla et al., 2017). (c) KstR—It is a highly conserved TetR family transcriptional repressor (TFR) that regulates 78 genes responsible for cholesterol catabolism in MTB (Kendall et al., 2010; Sanz et al., 2011). In MTB, a highly critical enzyme known as 3β-HSO(EC 1.1.1.145) has been primarily involved in the degradation of cholesterol, which transforms cholesterol into cholestenone. In fact, this enzyme is expressed in a cholesterol-independent manner suggesting its availability in the pathogen at any time (Garcia et al., 2012). (d) MosR—This is a highly connected protein in the transcriptional regulatory (TR) network of MTB, one of its functions is to regulate around 295 genes at the level of transcription. The main role of MosR is to up-regulate expression of rv1050 (a putative exported oxidoreductase that has not yet been characterized) in response to oxidants and propose that it is through this role that MosR contributes to the bacterium survival in the macrophage (Brugarolas et al., 2012). (e) RegX3—It is a self-regulatory protein and controls the expression of its own gene. Many RegX3-dependent genes have a role in the biosynthesis of DNA, RNA and other macromolecules (Pei et al., 2021). Whereas, a number of genes participate in the catabolism of fatty acids and are repressed by fadE6, fadE14 and accD2 genes in a direct way (Ahmad and Kumar, 2016; Carroll et al., 2011). This two component system is also responsible for cell wall biosynthesis, lipid biosynthesis and cell envelope protein synthesis (Almatar et al., 2017; Roberts et al., 2011).

Materials and Methods

Preparation of drug target and drug-library
In our study, we considered five key regulatory genes from MTB; FurB (PDB ID: 2O03) (Lucarelli et al., 2007), IdeR (PDB ID: 1B1B) (Pohl et al., 1999), KstR (PDB ID: 3MNL) (Ho et al., 2016), MosR (PDB ID: 4FX0) (Brugarolas et al., 2012) and RegX3 (PDB ID: 2OQR) (King-Scott et al., 2007), details are given in Fig. 2. All the atomic coordinates (x-ray structure with good resolution, from 1.8Å to 2.7Å) of these proteins were downloaded from Protein Data Bank (Berman et al., 2000). The proteins were optimized for docking by
following standard procedures, which include removal of crystallographic water molecules, the addition of missing H-atoms, and minimizing the energy level of all the proteins using the Swiss-PDB viewer tool and this server provides an energy minimization facility using the GROMOS96 force field (Alam et al., 2016; Johansson et al., 2012). All the targets were further refined using SCWRL 4.0 tool to minimize side-chain to backbone clashes and the side-chain-to-side chain clashes (Wang et al., 2008). The drug library consists of 1985 drugs (FDA approved) which were downloaded from the ZINC-15 database (Irwin et al., 2012). CORINA Classic (the classic command-line version of CORINA) (https://www.mn-am.com/products/corina) was used to prepare the drug-library because it facilitates preparation of high quality, low energy, single, 3D structure with correct chirality.

**Virtual screening**

Virtual screening (VS) is a computational method used in drug designing and discovery to find out the series of (compound library) small molecules in order to identify the most potent molecules which are most likely to bind to a drug target (Krueger et al., 2016; Morrone Xavier et al., 2016; Rester, 2008; Rollinger et al., 2008). In our study, VS was performed by using GOLD versions 5.2 (Genetic Optimisation for Ligand Docking; Cambridge Crystallographic Data Centre) (Jones et al., 1995; Jones et al., 1997). GOLD uses genetic algorithms to dock ligand with the receptor. The scoring function “GoldScore” was tested individually for each protein. The molecular docking that utilizes the genetic algorithm (GA) was set to three different execution runs—(a) 50, (b) 100, and (c) 200 GA runs, so as to identify influential and high-affinity binding modes. The radius of the docking sphere was 10–12 Å centered at the position, a maximum of 100,000 operations were performed that deliver high prediction accuracy but are relatively slow (Heberlé and De Azevedo, 2011; Potemkin and Grishina, 2018). The information on the binding site for all the targets was observed from the available literature which was then used further for grid generation. For FurB and IdeR, we used its active repressor site (Lucarelli et al., 2007) for KstR and MosR, we used its DNA binding site (Brugarolas et al., 2012; Ho et al., 2016) and for RegX3, we used its conformational switch region (present ‘hotspot’ residues) (Ahmad et al., 2015; Khalid et al., 2018).

**Molecular dynamics study**

As we are aware of the remarkable biological functions of proteins and DNA and also their profound dynamic mechanisms which include switching between active and inactive states (Autiero et al., 2013), cooperative effects the intercalation of drugs into DNA (Chou and Mao, 1988), and allosteric transition (Ward et al., 2010) can be unveiled by studying their internal motions (Lin and Lapointe, 2013).

### FIGURE 2.

Key regulatory genes from MTB. FurB (PDB ID: 2O03), IdeR (PDB ID: 1B1B), KstR (PDB ID: 3MNL), MosR (PDB ID: 4FX0) and RegX3 (PDB ID: 2OQR).
Likewise, to understand the action mechanism of receptor-ligand binding, we should not consider only the static structures concerned, but also the dynamical information obtained by simulating their internal motions or dynamic processes (de Azevedo, Jr et al., 2001; de Azevedo, 2011; Qazi and Raza, 2021). To realize this, initially, MD simulations were performed for proteins only, e.g., FurB, IdeR, KstR, RegX3, and MosR at 300K to investigate the stability profile. We observed that drug ZINC03871613 was best docked with the three targets (FurB, KstR, and RegX3), and drugs ZINC03871615, ZINC03871614 were docked with MosR and IdeR targets, respectively. These are the best candidate drugs that exhibit high binding affinity with the targets. To observe conformational changes due to the presence of ligands in the active site, we performed MD simulations for protein complexes too.

Molecular dynamics (MD) simulations were performed using the GROMACS version 5.1 software (van der Spoel et al., 2005) with the standard GROMOS 53A6 force field. The proteins were soaked in a cubic box of water molecules with a dimension of 10 Å, i.e., setting the box edge 10 Å from the molecule perimeters, using the editconf module for making boundary conditions and genbox for solvation. For the solvation of proteins, the spc216 template was used. The charges of the complete system are neutralized by the addition of Na\(^+\) and Cl\(^-\) ions (if the system has a positive charge (e.g., +6), add 6 Cl\(^-\) ions and if the system has a negative charge (e.g., -6), add 6 Na\(^+\) ions) using the gmxgenion module to maintain neutrality. The system was then minimized using the 1500 steps of steepest descent at 300K during their equilibration period (100 ps) at a constant volume under periodic boundary conditions. Equilibration was performed in two phases: NVT ensemble (constant number of particles, volume, and temperature at 100 ps) and NPT ensemble (constant number of particles, pressure, and temperature at 100 ps). After the equilibration phase, the particle MeshEwald method (Cheatham et al., 1995) was used. Production phases were performed at 300K. The resulting trajectories were analyzed, using RMSD, RMSF, RG, and SASA by the utilities provided by GROMACS. For ligand topology, we used SWISSPARAM that provides topologies and parameters for small organic molecules, compatible with the CHARMM all atoms force field, for use with the CHARMM or GROMACS software (Bijlkermar et al., 2010; Zoete et al., 2011). The server is fully automatic, and we only need to provide the drug molecule in the mol2 format.

Protein–Protein interaction network

The protein-protein interaction (PPI) networks are much important for the system-level understanding of cellular processes (Cau et al., 2018). Such networks are used for screening and evaluation of functional genomics data and provide an instinctive platform for annotating structural, functional, and evolutionary properties of proteins. The five proteins (FurB, IdeR, KstR, RegX3, and MosR) were taken as seed proteins from which we obtained direct and indirect PPIs using the STRING 10.5 database (Snel et al., 2000). This database provides details on both predicted and experimentally verified interactions from different sources based on their neighborhood, co-occurrence, co-expression, gene fusions, experiments, and literature mining. We constructed a PPI network based on high confidence interaction which means that only interactions with a high level of confidence are considered as potential associations in the PPI network.

Results

Virtual screening

In the first step, the library of FDA approved drugs were screened using GOLD versions 5.2 server. Dock runs of all 1985 drugs on targets namely “FurB (binding cavity: His-81, His-83, Cys-76, Glu-78, Asp-24, Asp-62, etc.), IdeR (binding cavity: Glu-20, Asp-17, Arg-80, etc.), KstR (binding cavity: Arg-174, Glu-112, Pro-108, Trp-17, etc.), RegX3 (binding cavity: Pro-102, Tyr-103, Ser-104, etc.) and MosR (binding cavity: Thr-36, His-287, Glu-244, etc.)” using GOLD suite (with 50 GA runs were performed) resulted in a few best candidate drugs evaluated based on the Gold fitness score \( \geq 55 \) (The fitness score >50 is supposed to good binding).

We found a total of 828 drugs that were bound with all the targets like FurB, IdeR, KstR, MosR, and RegX3 that were docked with 111, 355, 193, 157, and 12 drugs out of 1985, respectively. In the second step, the genetic search algorithm was initially set to 100 then 200 GA runs for 828 drugs to each target and in most cases, similar results were obtained. The docking and scoring details are shown in Fig. 3 and Suppl. Tab. S1. In the third step, we again filtered out the drugs that frequently bounded with all the targets at set threshold among the top 20 GoldScore ranks. It was found that only 11 candidate drugs (designated with Zinc Database ids as ZINC03871612, ZINC03871613, ZINC03871614, ZINC03871615, ZINC01529323, ZINC03830426, ZINC03830428, ZINC03830430, ZINC03830635, ZINC03799072, and ZINC08551107) are consistently binding with all the targets.

The consistent binding idea was given preference for the selected 11 candidate drugs over top 20 (for each target) as a matter of feasibility towards in-vitro or in-vivo testing for these bindings. The 2D structures of these 11 candidate drugs can be obtained from the ZINC-15 database (https://zinc15.docking.org/). In this study, we have visualized only top-ranked candidate drugs which are best bound with the targets like FurB, KstR, and RegX3, are docked with ZINC03871613, while IdeR and MosR are docked with ZINC03871614, and ZINC03871615 with GoldScore 77.58, 75.02, 62.54, 81.69, and 79.72, respectively. While studying the docked conformation of the FurB complex, mainly the three residues, ASP-24, GLU-78, and HIS-81 were found to make H-bond with the ligand along with 14 surrounding residues. As in the case of KstR and RegX3 complexes, ASP-27, THR-36, ASN-37, THR-38, ARG-75 (also forms salt bridge interaction), and ASN-76 while in the IdeR complex, the residues ARG-13,
ARG 33, ARG-80 (also make π–π interaction) were seemed to make H-bonds with the ligands along with six and 17 surrounding residues, respectively. The complete details are given in Fig. 4 and Tab. 1. All the top scorrer candidate drugs are purine derivatives in the form of triphosphate and the basic difference among them is in the cyclic pyranose which are isomers of each other due to the different spatial orientation of the hydroxyl group.

**MD simulation analysis**

We performed MD simulation at 300K of the protein-ligand complex with top-ranked drugs to observe the effect on the stability of the trajectory.

**Stability of the trajectory**

a) Root-mean-square deviations (RMSD) from the native structure coordinates during the MD simulation were used to assess the stability and the reliability of the simulation. In our study, we noticed that FurB underwent structural fluctuations starting from the first 5 ns of the trajectory, although the amplitude of the fluctuation was smaller for the FurB complex at 6.44 ns, a slight transition of 0.325 nm was spotted and the system got optimized after 8 ns at 0.65 nm. This significant drift in RMSD of 0.325 nm is verified by the protein-ligand interactions within the complex. However, the simulation ends with a total drift of 0.28 nm from the native structure. A significant drift in RMSD of FurB and FurB complex indicate the conformational change in the receptor that has occurred to fit the ligand in the active site of the FurB.

b) In the case of the IdeR, the trajectory (RMSD) remains stable throughout the simulation with small drifts, but the fluctuations were smaller too for the IdeR complex, up to 7.15 ns. At the three points, a gradual transition of 0.083, 0.096, and 0.15 nm were observed at 0.24, 2.54 and 5.91 ns, respectively and after 11 ns, the trajectory was optimized. These significant drifts are verified by the protein-ligand interactions within the complex.

c) In the case of the RegX3, the trajectory (RMSD) drastically fluctuated with large drifts, but the fluctuations were smaller for the RegX3 complex. A gradual transition of 1.056 nm at 4.06 ns was observed in the native structure, but the system is optimized after 12 ns. A significant difference in RMSD of RegX3 and complex strongly indicates a conformational change in the protein due to ligand.

d) In the case of the MosR, the trajectory (RMSD) drastically fluctuated with large drifts, but the fluctuations were smaller for the MosR complex. A drastic transition of 0.5064 nm at 1.07 ns was observed in the native structure, but later system is optimized after 3.5 ns. However, the simulation ends with a total drift of 0.249 nm from the native structure, and this difference in RMSD indicates the conformational changes in the protein due to ligand.

e) In the case of the KstR, the trajectory (RMSD) remains stable throughout the simulation with small drifts of 0.103 nm, but the fluctuation pattern was slightly different from the kstR complex, a drastic transition of 0.287 nm is observed but the system is optimized after 8.6 ns. The average total drift in RMSD at the end of both simulations was 0.192 nm. This difference in RMSD of kstR and kstR complex indicates a conformational change in the protein due to ligand.

**The root mean square fluctuation (RMSF)**

a) We have performed RMSF from its time-averaged position to set up the conformational fluctuations of the FurB and FurB complex at the residual level. The average residual fluctuation of all the residues in the case of FurB complex is lower than FurB. These results suggest that the α-helical structure of FurB is more specifically maintained due to the binding of the ligand.

b) RMSF analysis for the IdeR revealed that fluctuation of amino acids was largest in the region from 33–41 residues in IdeR, but in the case of IdeR complex, residues V56, A57, G58, and D59 have a greater RMSF score. The region 61–80 residues slightly fluctuated up and down in both cases (complex and protein), which confirms the important role being played by these residues in the ligand-binding process.

![FIGURE 3. The plot represents the Gold Score of compounds that have scored equal, or above the threshold value of 55.0. (A) IdeR bind with 355 compounds (B) KstR bind with 193 compounds (C) MosR bind with 157 compounds (D) FurB bind with 111 compounds (E) RegX3 bind with 12 compounds.](image)
FIGURE 4. (continued)
c) On analyzing the fluctuations in the RegX3 and RegX3 complex, it was observed that fluctuation of amino acids was the largest in the two regions, 80–89 and 91–114 residues in both cases due to the presence of the active site in this region, which confirms the important role being played by these residues in ligand binding.
d) In RMSF analysis of the MosR, we observed that fluctuation of amino acids was largest in the two regions from 14–40 and 44–110 residues, in both MosR and mosR complex. The first fluctuation (between 14–40 residues) was started due to the presence of an active site where the ligand binds, and the second fluctuation occurred due to residual conformational changes in MosR protein.

e) The fluctuations of KstR are less than kstR complex; but at two points (77–92 and 149–172 residues) fluctuations are larger due to the presence of binding regions. In the presence of ligand with KstR protein, at two points, lower transitions were observed, and the system’s overall compactness shows an increase after 11 ns. The difference in RMSF value (for both) to the endpoint of the simulation is 0.07 nm. This significant difference in the RMSF value of the native structure indicates that the ligand fits into the binding site.

Radius of gyration (Rg)

a) The radius of gyration (Rg) analysis of the FurB was carried out to determine the compactness of the protein system during the simulations. In the presence of ligand, the average Rg score was drastically reduced from 1.80 rg (native structure of FurB) to 1.59 nm. This complex underwent two drastic transitions which were observed from 1.60 nm to 1.40 nm and 1.40 nm to 1.55 nm, then the system was optimized after 15.4 ns at 1.5 rg (nm). The decrease in Rg score from the native structure points out that ligand binds into the binding cavity of the FurB.

b) In the presence of ligand with IdeR protein, the average Rg score was 0.06 rg decreased. A slight transition of 0.12 nm was spotted and after 12.6 ns, the system was optimized at 1.45 rg and the average difference in Rg score at the end of both simulations was 0.11 nm. This significant difference in Rg score from the native structure points out that ligand binds into the binding site of the FurB.

c) In the case of RegX3, we observed structural fluctuations throughout the simulation, but in the case of RegX3 complex the amplitude of the fluctuations was smaller. The decrease in the transition of 0.487 nm occurred at 3.7 ns, but the system was optimized after 14.6 ns at 1.9 nm. However, the simulation finished with a total drift of 0.69 nm from the native structure and this difference indicates that the ligand binds into the RegX3 pocket.

d) In the presence of ligand with MosR, the average Rg score drastically decreases from 1.73 nm (native structure) to 1.35 n. This complex underwent structural fluctuations starting from first 0.17 ns to 5.29 ns and the compactness of the system increasing after 5.5 ns at 1.37 nm. This
difference indicates that the ligand binds to the active site of MosR.

e) In presence of ligand with KstR protein, at two points, lower transitions were observed, and the system’s overall compactness witnessed an increase after 11 ns. The difference in rg score (for both) to the endpoint of the simulation is 0.07 nm. This significant difference in Rg score from the native structure strongly indicates that the ligand binds to the active site of KstR.

**Solvent accessible surface area**

a) Solvent accessible surface area (SASA) remained relatively constant for the FurB protein at 73.68 nm², but in the case of FurB complex the pattern is different, SASA decreases from 73.20 nm² to 71 nm² and average fluctuation occurred near 70 nm² up to 10 ns. The system is optimized after 68.5 nm² at the 11 ns up to the endpoint of the simulation. The decrease in SASA caused by the change in protein conformation during its folding must be accompanied by the corresponding increase in the number of native contacts with the ligand, clearly indicating that ligand (ZINC03871613) binds into the active site of the FurB.

b) In the case of IdeR, we observed the structural fluctuations in SASA throughout the simulation, but these fluctuations diminish later on and the overall system is optimized after 79.5 but in the case of IdeR complex the pattern is different, the average surface area is drastically decreased from 71.78 nm² (native structure of IdeR) to 77.78 nm². The decrease in SASA strongly indicates that inhibitor (ZINC03871614) preferentially fits into the active site of IdeR.

c) In the case of RegX3, SASA remained stable up to 127.434 nm² at 8.5 ns, but the slight transition was observed at 7.185 nm², but in the case of RegX3 complex the pattern is different, the average surface area is slightly decreased from 123.75 nm² (native structure of RegX3) to 120.014 nm² and system is optimized after 6.68 ns at 119 nm². However, the simulation finished with a total drift off 10.209 nm² from the native structure and this difference in SASA indicates the ligand binding into the RegX3 pocket.

d) The SASA remained relatively constant for the MosR at 70.25 nm², but in the case of FurB complex the pattern is different; the SASA is decreasing from 69.59 nm² to below 62 nm², and therefore, systems compactness increases. The decrease in SASA strongly indicates that inhibitor (ZINC03871615) preferentially fits into the active site of MosR.

e) In the case of KstR protein and complex, gradual upper transition and lower transition were observed, respectively, but SASA remained relatively constant, and the overall system is optimized after 7.5 ns at 103 nm².

**Network analysis and nodes characterization**

The Protein-protein interaction (PPI) network was constructed using the five seed genes from the STRING database. In the present study, we considered only the immediate interacting partners (not more than 5 interactors) that abided the medium confidence score of 0.400. We have considered occurrence across the genome (suggesting that protein has a similar function or an occurrence in the same metabolic pathway further which to be expressed together and hold similar phylogenetic profile), neighborhood (similar genomic context in different species suggesting similar functions of the proteins), putative homologs, gene fusions (that means proteins that are fused in some genomes are the most likely to be functionally linked), co-expression. We observed that FurB is present in the neighborhood of MRA_2382 and GlyS and IdeR, KstR, Regx3, and MosR are in the neighborhood of SigB, FadE34, SenX3, and Rv1050 respectively, while RegX3 is fused with PhoR, MtrB, and MtubH3_010100022685. In the light, being a broad-spectrum antibiotics concept (i.e., designing multi-target drugs), it can be inferred from Tab. 2 that the interacting partners of RegX3 make it most pliable target to any drug design for it. As homologs of RegX3’s interacting partners are found co-expressed in other species of bacteria, it means it may be closely linked to MTB. In order to rank the targets based on their network interactions, KstR and RegX3 showcase highly influential interactions followed by MosR, IdeR, and FurB. The complete details of five seed genes and their interactions with other important genes are demonstrated in Fig. 6 and Tab. 2, which illustrates the PPI neighborhood of the 5 receptors (FurB, IdeR, KstR, RegX3, and MosR). The nodes of the graph represent proteins and the connections illustrate their known or predicted direct and indirect interactions.

**Discussion**

Till now TB has claimed the lives of millions of people across the world, despite our continuous efforts it is emerging as the next challenge for us to eradicate this ailment. Here we have tried to identify new scaffolds for drug discovery by repurposing FDA-approved drugs against M. tuberculosis. Almost all drugs repositioning concepts require at least some experimental validation of assessment for its efficacy, therefore we have adopted an *in-silico* approach to support experimental work in developing effective drugs in a shorter duration. Our systematic study has highlighted several important observations; molecular docking analysis of a set of 828 drugs with the best GoldScore (≥ 55) were selected and out of which 11 drugs were frequently present within the top 20 GoldScore ranks and ZINC03871613, ZINC03871614, ZINC03871615 are top scorer drugs. Hence, we consider all these 11 drugs as promising drugs as lead drugs for further research towards repurposing against TB.

The ROC curve applied to the retrospective analysis of a virtual screening experiment is a plot of the true positive fractions (TPR, y-axis) versus false positive fractions (FPR, x-axis) for all drugs in a ranked dataset. The ROC curve analysis reveals either perfect discrimination, i.e., passes through the upper left corner (100% sensitivity, 100% specificity), or defective discrimination. Therefore, the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test. The area under the ROC curve (ROC AUC) outlines the overall performance thus higher the AUC, the higher is the accuracy. In this case, we
performed former experiment in order to identify the interaction capability of the targets with other potential based on the fitness value calculated using various parameters, e.g., interaction entropy. We found fitness values give a perfect discrimination, i.e., highly accurate, while interaction entropy contributes to the larger part of this perfection in each case (for both active and inactive ligands) as shown in ROC curve analysis of active and inactive ligands against targets in Fig. 7.

The protein-ligand complex with top-ranked drugs was used for the dynamics simulation study in order to see the effect of top scorer drugs on the structural stability of all five targets. It was found that these top scorer drugs preferentially fit into the active site of the targets. RMSD of all the protein’s initial structures and protein-ligand complexes were calculated. Throughout the simulation period, significant drifts were observed in the backbone of FurB, IdeR, KstR, RegX3, and MosR and protein-ligand complexes, implying that binding of drug candidates at the active sites of the targets is not only preferentially fitting, but also disturbs the structure of proteins, i.e., the structure became more compact that is also proved by RG-score analysis (The lowest radius of

| Table 2 |
|---|

Protein-Protein interaction: The evidence suggesting a functional link of proteins, including (a) Co-occurrence (b) Neighbourhood, (c) Putative homologs, (d) Gene fusions, (e) Co-expression

| Protein | Co-occurrence (Score) | Intergenic distance (BP) | Putative Homologous (Score) | Neighbourhood (Score) | Not neighbourhood in the genome but homologous in other species (Score) | Gene Fusion (Score) | Co-expression (Score) | Putative homologs are co-expressed in other species (Score) |
|---|---|---|---|---|---|---|---|---|
| FurB | NA | 0 bp | NA | 0.859 | NA | NA | NA | NA |
| GlyS | NA | 103524 bp | 0.406 | 0.579 | NA | NA | NA | NA |
| MRA_1920 | 0.611 | NA | NA | NA | NA | NA | NA | NA |
| MRA_2383 | 0.782 | NA | NA | NA | NA | NA | NA | NA |
| MRA_0086 | NA | NA | 0.612 | NA | NA | NA | NA | NA |
| MRA_2739 | 0.782 | NA | NA | NA | NA | NA | NA | NA |
| MtubH3_10100009904 | 0.651 | NA | NA | 0.047 | NA | NA | NA | NA |
| sigB | NA | 134 bp | NA | 0.550 | 0.128 | NA | NA | NA |
| IdeR | NA | NA | 0.612 | NA | NA | NA | NA | NA |
| Rv3521 | 0.779 | NA | NA | NA | 0.047 | NA | NA | 0.050 |
| Rv3542c | 0.752 | NA | NA | NA | 0.047 | NA | NA | 0.058 |
| fade34 | 0.630 | 80.35 k bp | NA | 0.541 | 0.120 | NA | NA | 0.051 |
| kshA | 0.745 | NA | NA | NA | NA | NA | NA | 0.058 |
| Rv0760c | 0.769 | NA | NA | NA | NA | NA | NA | NA |
| KstR | NA | NA | 0.612 | NA | NA | 0.047 | NA | NA |
| phoR | 0.719 | NA | 0.415 | NA | NA | 0.001 | 0.075 | 0.075 |
| senX3 | 0.764 | 229 bp | 0.415 | 0.480 | NA | NA | NA | 0.075 |
| mtrB | 0.662 | NA | 0.415 | NA | NA | 0.002 | 0.086 | 0.075 |
| prrB | 0.665 | NA | 0.415 | NA | NA | 0.002 | 0.083 | 0.132 |
| MtbH3_010100022685 | 0.765 | NA | 0.415 | NA | NA | 0.002 | 0.075 | 0.075 |
| RegX3 | NA | NA | 0.415 | NA | NA | 0.001 | 0.075 | 0.075 |
| MosR | NA | NA | 0.415 | NA | NA | 0.001 | 0.075 | 0.075 |
| narL [Rv0844c] | NA | NA | NA | NA | NA | NA | NA | 0.058 |
| Rv1364c | 0.304 | NA | NA | NA | NA | NA | NA | 0.058 |
| Rv0195 | NA | NA | NA | NA | NA | NA | NA | 0.058 |
| devR | NA | NA | NA | NA | NA | NA | NA | 0.058 |
| Rv1050 | NA | NA | NA | NA | NA | NA | NA | 0.058 |

Note: Probabilistic values (0 to 1), Not Available (NA).
gyration mean more compact structure). RMSF analysis suggests that the observed residual fluctuations were largest in active sites of proteins that confirm the important role being played by these residues in the ligand-binding process. The significant drifts in native structures and protein-ligand complexes were observed in SASA analysis. The decrease in SASA caused by the change in protein conformation during its folding must be accompanied by the corresponding increase in the number of native contacts with ligand, clearly indicating that ligands tightly fit into the active sites. We performed hydrogen bond analysis and extracted the complex at the particular interval (5000, 10,000, 15,000, 20,000 ps, respectively) in the PDB format using g_h bond utility of GROMACS. However, we have noticed that one or two residues (that formed H-bonds from ligands) from each protein are common in all 4 intervals of time (5, 10, 15, 20 ns) that indicate that ligands are tightly bound in the active sites. Apart from virtual screening and simulation

**FIGURE 6.** Analysis of protein-protein interaction (PPI) using STRING 10.5 server. The figure illustrates the PPI neighborhood of the 5 receptors (FurB, IdeR, KstR, RegX3, and MosR), the nodes of the graph represent proteins and the connections illustrate their known or predicted direct and indirect interactions. The connection between any two protein nodes is based on the available information mined from relevant.
FIGURE 7. (continued)
In this paper, we executed a computational drug repurposing pipeline to identify promising FDA-approved drugs against five key-regulatory genes of Mtb, namely, FurB, IdeR, KstR, MosR, and RegX3. Our study found 11 drug candidates including ZINC03871613, ZINC03871614, and ZINC03871615 as the top three drug candidates. In order to validate these results, a molecular dynamics simulation study was performed which indicates that inhibitors preferentially bind to the active site of the targets. This finding opens a new application domain in the form of anti-tuberculosis agents. This study can be further extended to explore the receptor-ligand interactions experimentally and assessment of their biological activity against key targets of MTB.

**Availability of Data and Materials:** Available on request to the authors.

**Author Contribution:** The idea was conceived by XY and KR, an experiment was designed by AA and KR, result analysis by AA, and the manuscript written by XY, AA, and KR. All authors reviewed the results and approved the final version of the manuscript.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.
References

Afzal O, Kumar S, Kumar R, Firoz A, Jaggi M, Bawa S (2014). Docking based virtual screening and molecular dynamics study to identify potential monoacylglycerol lipase inhibitors. Bioorganic & Medicinal Chemistry Letters 24: 3986–3996.

Ahmad A, Cai Y, Chen X, Shuai J, Han A (2015). Conformational dynamics of response regulator RegX3 from Mycobacterium tuberculosis. PLoS One 10: e0133389.

Ahmad F, Kumar S (2016). Rewiring of interconnected hubs in Gene Interaction Network of Mycobacterium tuberculosis and its survival in human host. Journal of Applied Computing 1: 38–50.

Ahmad S (2011). Pathogenesis, immunology, and diagnosis of latent Mycobacterium tuberculosis infection. Clinical and Developmental Immunology 2011: 814943.

Alam A, Ali S, Ahamad S, Malik MZ, Ihsrat R (2016). New drugs for the treatment of Mycobacterium tuberculosis infection. Biomedical & Pharmacotherapy 91: 546–558.

Almatar M, Almandeal H, Var I, Kayar B, Köksal F (2017). New insights into the mechanism of interaction between trehalose-conjugated beta-sheet breaker peptides and Aβ (1-42) fibrils by molecular dynamics simulations. Molecular BioSystems 9: 2835–2841.

Banerjee S, Farhana A, Ehtesham NZ, Hasnain SE (2011). Iron drug intercalation. Biochemical and Biophysical Research Communications 287: 277–281.

De Azevedo Jr WF, Canduri F, Fadel V, Teodoro LG, Hial V et al. (2001). Molecular model for the binary complex of uropesin and pepstatin. Biochemical and Biophysical Research Communications 287: 277–281.

De Azevedo WF (2011). Molecular dynamics simulations of protein targets identified in Mycobacterium tuberculosis. Current Medicinal Chemistry 18: 1353–1366.

Dua K, Rapalli VK, Shukla SD, Singhv G, Shastri MD et al. (2018). Multi-drug resistant Mycobacterium tuberculosis & oxidative stress complexity: Emerging need for novel drug delivery approaches. Biomedicine & Pharmacotherapy 107: 1218–1229.

Garcia JL, Uhia I, Galán B (2012). Catabolism and biotechnological applications of cholesterol degrading bacteria. Microbial Biotechnology 5: 679–699.

Heberle G, De Azevedo WF (2011). Bio-inspired algorithms applied to molecular docking simulations. Current Medicinal Chemistry 18: 1339–1352.

Ho NAT, Dawes SS, Crowe AM, Casbon I, Gao C et al. (2016). The structure of the transcriptional repressor KstR in complex with CoA thioester cholesterol metabolites sheds light on the regulation of cholesterol catabolism in Mycobacterium tuberculosis. Journal of Biological Chemistry 291: 7256–7266.

India TB Report (2018). Ministry Of Health and Family Welfare.

Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG (2012). ZINC: A free tool to discover chemistry for biology. Journal of Chemical Information and Modeling 52: 1757–1768.

Johansson MU, Zoete V, Michielen O, Guex N (2012). Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. BMC Bioinformatics 13: 173.

Jones G, Willett P, Glen RC (1995). Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. Journal of Molecular Biology 245: 43–53.

Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997). Development and validation of a genetic algorithm for flexible docking. Journal of Molecular Biology 267: 727–748.

Karaman B, Sippl W (2019). Computational drug repurposing: Current trends. Current Medicinal Chemistry 26: 5389–5409.

Karthick V, Ramanathan K (2013). Virtual screening for oseltamivir-resistant H5N1 influenza neuraminidase from traditional Chinese medicine database: A combined molecular docking with molecular dynamics approach. Springerplus 2: 115.

Kendall SL, Burgess P, Balhana R, Withers M, Ten Bokum A et al. (2010). Cholesterol utilization in mycobacteria is controlled by two TetR-type transcriptional regulators: KstR and katR2. Microbiology 156: 1362.

Khalid Z, Ahmad S, Raza S, Azam SS (2018). Subtractive proteomics revealed plausible drug candidates in the proteome of multi-drug resistant Corynebacterium diphtheriae. Meta Gene 17: 34–42.

King-Scott J, Nowak E, Mylonas E, Panjikar S, Roessle M et al. (2007). The structure of a full-length response regulator from Mycobacterium tuberculosis in a stabilized three-dimensional domain-swapped, activated state. Journal of Biological Chemistry 282: 37717–37729.

Krueger J, Thiell P, Merelli I, Grunze K, Gesing S (2016). Portals and web-based resources for virtual screening. Current Drug Targets 17: 1649–1660.

Lanza V, Milardi D, Di Natale G, Pappalardo G (2018). Repurposing of copper (II)-chelating drugs for the treatment of neurodegenerative diseases. Current Medicinal Chemistry 25: 525–539.

Li MS, Monahan IM, Waddell SJ, Mangan JA, Martin SL et al. (2001). cDNA-RNA subtractive hybridization reveals increased expression of mycocerosic acid synthase in intracellular Mycobacterium bovis BCG. Microbiology 147: 2293–2305.
Lucarelli D, Russo S, Garman E, Milano A, Meyer-Klaucke W, Pohl E (2007). Crystal structure and function of the zinc uptake regulator FurB from Mycobacterium tuberculosis. *Journal of Biological Chemistry* **282**: 9914–9922.

Morrone Xavier M, Sehnem Heck G, Boff De Avila M, Maria Peters JS, Ismail N, Dippenaar A, Ma S, Sherman DR et al. (2020). RegX3-mediated regulation of methylcitrate cycle in Mycobacterium smegmatis. *Frontiers in Microbiology* **12**: 119.

Panina EM, Mironov AA, Gelfand MS (2003). Comparative genomics of bacterial zinc regulons: Enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 9912–9917.

Pei JF, Qi N, Li YX, Wo J, Ye BC (2021). RegX3-mediated regulation of methylcitrate cycle in *Mycobacterium smegmatis*. *Frontiers in Microbiology* **12**: 119.

Peters JS, Ismail N, Dippenaar A, Ma S, Sherman DR et al. (2020). Genetic diversity in Mycobacterium tuberculosis clinical isolates and resulting outcomes of tuberculosis infection and disease. *Annual Review of Genetics* **54**: 511–537.

Pohl E, Holmes RK, Hol WG (1999). Crystal structure of the iron-dependent regulator (IdeR) from *Mycobacterium tuberculosis* shows both metal binding sites fully occupied. *Journal of Molecular Biology* **285**: 1145–1156.

Potemkin V, Grishina M (2018). Grid-based technologies for in silico screening and drug design. *Current Medicinal Chemistry* **25**: 3526–3537.

Qazi S, Raza K (2021). Phytochemicals from Ayurvedic plants as potential medicaments for ovarian cancer: An in silico analysis. *Journal of Molecular Modeling* **27**: 114.

Rester U (2008). From virtuality to reality-Virtual screening in lead discovery and lead optimization: A medicinal chemistry perspective. *Current Opinion in Drug Discovery & Development* **11**: 559–568.

Roberts G, Vadrevu IS, Madiraju MV, Parish T (2011). Control of CylD and GltAI expression by the SenX3 RegX3 two component regulatory system of *Mycobacterium tuberculosis*. *PLoS One* **6**: e21090.

Robilla A, Khare G, Tyagi AK (2017). Virtual Screening, pharmacophore development and structure based similarity search to identify inhibitors against IdeR, a transcription factor of *Mycobacterium tuberculosis*. *Scientific Reports* **7**: 4653.

Rollinger JM, Stuppner H, Langer T (2008). Virtual screening for the discovery of bioactive natural products. *Natural Compounds as Drugs I*: 211–249.

Sanz J, Navarro J, Arbués A, Martin C, Marijuán PC et al. (2011). The transcriptional regulatory network of *Mycobacterium tuberculosis*. *PLoS One* **6**: e22178.

Savoia D (2016). New antimicrobial approaches: Reuse of old drugs. *Current Drug Targets* **17**: 731–738.

Singh P, Jaiyeola B, Kerru N, Ebenezer O, Bissessur A (2017). A review of recent advancements in anti-tubercular molecular hybrids. *Current Medicinal Chemistry* **24**: 4180–4212.

Snel B, Lehmann G, Bork P, Huynen MA (2000). STRING: A web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Research* **28**: 3442–3444.

van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **26**: 1701–1718.

Wang Q, Canutescu AA, Dunbrack RL Jr (2008). SCWRL and MolIDE: Computer programs for side-chain conformation prediction and homology modeling. *Nature Protocols* **3**: 1832.

Ward SK, Abomoeak B, Marcus S, Talaat AM (2010). Transcriptional profiling of *Mycobacterium tuberculosis* during infection: Lessons learned. *Frontiers in Microbiology* **1**: 121.

Weiner J, Domaszewska T, Donkor S, Kaufmann SH, Hill PC, Sutherland JS (2020). Changes in transcript, metabolite, and antibody reactivity during the early protective immune response in humans to *Mycobacterium tuberculosis* infection. *Clinical Infectious Diseases* **71**: 30–40.

Zoete V, Cuydert MA, Grosdidier A, Michielin O (2011). SwissParam: A fast force field generation tool for small organic molecules. *Journal of Computational Chemistry* **32**: 2359–2368.

**Supplementary Table S1**

| Docking Results | Set Threshold | Gold Score |
|-----------------|--------------|------------|
| FurB | 111 | 559–568 |
| IdeR | 355 | 559–568 |
| KstR | 193 | 559–568 |
| MosR | 157 | 559–568 |
| RegX3 | 12 | 559–568 |

List of docking results at the set threshold of 55 Gold score: FurB was docked with 111, IdeR was docked with 355, KstR was docked with 193, MosR was docked with 157, and RegX3 was docked with 12 out of 1985 compounds.