Bioinformatics Analysis to Uncover the Potential Drug Targets Responsible for *Mycobacterium tuberculosis* Peptidoglycan and Lysine Biosynthesis

Dian Ayu Eka Pitaloka[1,2], Afifah Izzati[1], Siti Rafa Amirah[1], Luqman Abdan Syakuran[3], Lalu Muhammad Irham[4,5], Athika Darumas Putri[6] and Wirawan Adikusuma[7,8]

[1]Department of Pharmacology and Clinical Pharmacy, Faculty of Pharmacy, Universitas Padjadjaran, Sumedang, Indonesia. [2]Center for Translational Biomarker Research, Universitas Padjadjaran, Sumedang, Indonesia. [3]Genetics and Molecular Laboratory, Faculty of Biology, Jenderal Soedirman University, Purwokerto, Indonesia. [4]Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia. [5]Research Center for Pharmaceutical Ingredienst and Traditional Medicine, National Research and Inovation Agency (BRIN), South Tangerang, Indonesia. [6]Semarang College of Pharmaceutical Sciences (STIFAR), Semarang, Indonesia. [7]Department of Pharmacy, Faculty of Health Science, Universitas Muhammadiyah Mataram, Mataram, Indonesia.[8]Research Center for Vaccine and Drugs, National Research and Inovation Agency (BRIN), South Tangerang, Indonesia.

ABSTRACT: Drug-resistant tuberculosis (TB), which results mainly from the selection of naturally resistant strains of *Mycobacterium tuberculosis* (MTB) due to mismanaged treatment, poses a severe challenge to the global control of TB. Therefore, screening novel and unique drug targets against this pathogen is urgently needed. The metabolic pathways of *Homo sapiens* and MTB were compared using the Kyoto Encyclopedia of Genes and Genomes tool, and further, the proteins that are involved in the metabolic pathways of MTB were subtracted and proceeded to protein-protein interaction network analysis, subcellular localization, drug ability testing, and gene ontology. The study aims to identify enzymes for the unique pathways for further screening to determine the feasibility of the therapeutic targets. The qualitative characteristics of 28 proteins identified as drug target candidates were studied. The results showed that 12 were cytoplasmic, 2 were extracellular, 12 were transmembrane, and 3 were unknown. Furthermore, druggability analysis revealed 14 druggable proteins, of which 12 were novel and responsible for MTB peptidoglycan and lysine biosynthesis. The novel targets obtained in this study are used to develop antimicrobial treatments against pathogenic bacteria. Future studies should further shed light on the clinical implementation to identify antimicrobial therapies against MTB.

KEYWORDS: In silico drug targets, metabolic pathways, protein-protein interaction, *Mycobacterium tuberculosis*

Introduction

Tuberculosis (TB), which ranges from asymptomatic infection to fatal disease, is an airborne infectious disease caused by *Mycobacterium tuberculosis* (MTB). Furthermore, it is estimated that these pathogenic bacteria have infected one-third of the world's population, and more than 250 people die of TB daily.[1] The emergence of multidrug-resistant tuberculosis (MDR-TB) has exacerbated the situation, making the disease a top priority to be resolved globally.

The conventional drug discovery and development approach uses expensive methods that are time-consuming, complex, and only uncover a small number of potential targets.[2] However, computational approaches leveraging Omics data analyses have been widely used in pharmaceuticals to identify and accelerate drug discovery with lower failure rates in clinical trials.[3,4] The search for potential pharmacological targets has become more accessible due to the development of sequenced human and pathogen genomes accessible in public databases.

The current methods for discovering therapeutic targets are greatly based on non-homologous enzymes, chokepoint enzymes, critical genes unique to a pathogen, and genes linked to resistance and virulence.[5] There is a previous report on MTB's metabolic pathways and protein-protein interaction (PPI).[6-8] However, a deep gene ontology (GO) analysis to identify the putative targets has not been performed. These actions are essential because GO investigation paves the way for identifying significant features, such as molecular function (MF) and cellular processes.[9]

The metabolic pathways of the pathogen and the host are compared to discover enzymes critical to the MTB's survival. The procedure started with identifying host and pathogen metabolic pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Subsequently, the specific pathways to the pathogen were compared by analyzing the distinctive pathways of the extracted enzymes and submitted to an online platform for identification. The discovered essential enzymes used as therapeutic targets are determined by the similarity of innovative drug targets (DT) using DrugBank, cellular localization, and GO analysis.
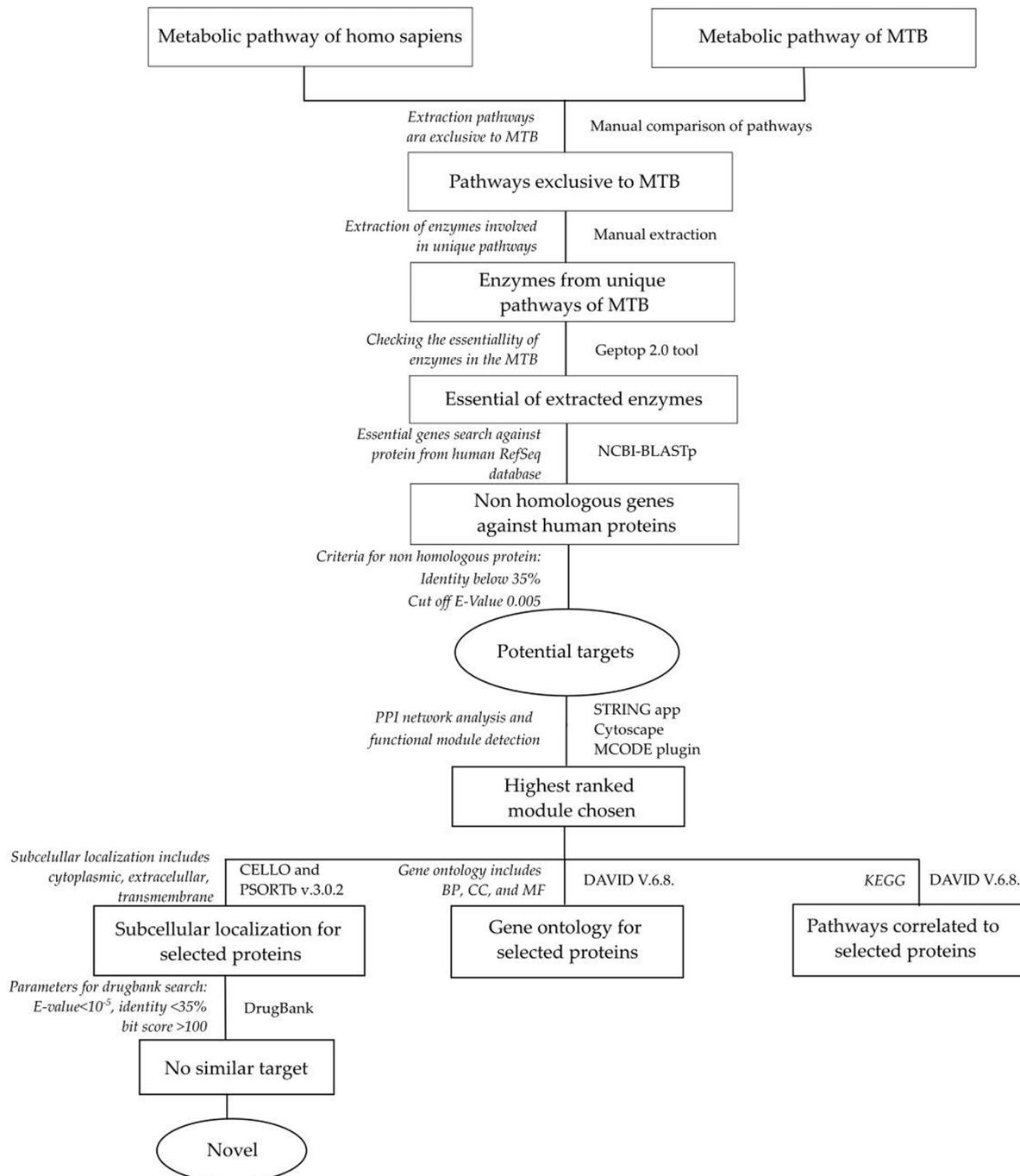
Figure 1. A schematic representation of the methodology.

Materials and Methods

Comparative analysis of host and pathogen metabolic pathways

The metabolic pathways for the host (*Homo sapiens*, KEGG ID: T01001) and the pathogen (MTB H37Rv-sensitive strain to anti-tuberculosis drugs, KEGG ID: T00015) were collected from the KEGG database (accessed on April 8, 2022) (https://www.genome.jp/kegg/).[10] The pathways were compared to identify the unique routes only present in MTB. Furthermore,

the enzymes in shared and unique pathways of MTB were extracted from KEGG, and the protein sequences were retrieved in FASTA format from the NCBI (National Center for Biotechnology Information) database (Figure 1).

Identification of non–homologous essential proteins

The protein sequences were submitted to Geptop tool 2.0 (http://guolab.whu.edu.cn/geptop/) to identify their essentiality in MTB.[11] The webserver discovers essential genes of MTB

by comparing their orthology and phylogeny with the crucial gene database and differential expression gene. These essential genes were searched against proteins from the human RefSeq protein database for non-homology using NCBI-BLASTP (https://blast.ncbi.nlm.nih.gov/).[12] Proteins with identity below 35% and an E-value cut-off of 0.005 were selected as non-hosts.[13]

PPI network analysis

The PPI network of non-homologous proteins was analyzed using string analysis (https://string-db.org/) in Cytoscape v.3.9.0 (https://cytoscape.org/).[14] The interaction of network data was examined by the network analyzer module.[15] Furthermore, the functional modules of non-homologous proteins were detected using the Cytoscape plugin MCODE. The scores and parameters include the degree cut-off of 2, maximum depth of 100, k-core of 2, and node score cut-off of 0.2.[16] The uppermost hierarchical module was selected as the possible metabolic functional association of the interacting proteins and was assigned for further analysis.

Subcellular localization and identification of novel drug targets through PPI network analysis

Subcellular localization of the essential non-human proteins selected from network analysis was predicted by PSORTb v3.0.2 (https://www.psort.org/psortb/)[17] and CELLO v2.518 (http://cello.life.nctu.edu.tw/cello2go/).[18] Transmembrane proteins were identified by TMHMM-2.0 (https://services.health-tech.dtu.dk/service.php?TMHMM-2.0) based on the hidden Markov model.[19] The most probable topology of a membrane protein was determined using the N-best algorithm. Proteins with transmembrane helices estimated to have less than 50 amino acid residues from the N terminus were extracted as possible candidates for signal peptides. Furthermore, when a cleavage site is predicted to be >0.5, the signal peptide was cleaved off, and the prediction was redone.[19] The proteins selected as novel drug targets were cytoplasmic and transmembrane.[20] The DrugBank database (https://go.drugbank.com/) was used to identify novel targets with an E-value of less than 10^{-5}, sequence identity more significant than 35%, and score slightly greater than 100.[21]

Functional enrichment analysis

The Database for Annotation, Visualization, and Integrated Discovery (DAVID) v.6.8 (https://david.ncifcrf.gov/)[22] was used to perform GO and KEGG enrichment analysis to investigate functional annotation and pathways involved in novel drug targets. The complete list of all selected proteins was sent to GO enrichment analysis under the headings of the cellular compartment (CC), biological process (BP), MF, and KEGG. Finally, the significance threshold was set at a *P* value <.05.

Results

Metabolic pathway analysis and identification of essential proteins

A total of 345 and 131 metabolic pathways of *H sapiens* (S1) and MTB (S2), respectively, were extracted from the KEGG pathway database. Furthermore, 43 of the 131 pathways of MTB were unique and comprised 548 enzymes. The essentiality of these enzymes for the pathogen was analyzed using Geptop 2.0, and it was discovered that 313 (S3) of them were predicted as essential genes. NCBI-Blastp was performed to identify the homology of the enzyme with human proteins. Of the 313 essential proteins, 197 proteins (S4) identified as human non-homologous protein.

PPI analysis

According to STRING analysis, 192 nodes and 1154 interaction lines were analyzed from 197 essential proteins of MTB (S5). The PPI data originating from STRING into Cytoscape were further analyzed to explore the significance of proteins in the protein networks and the primary cluster using the MCODE plugin (S6). This was conducted due to the complexity and originality of the network. The highest cluster with the lowest *P* value comprised 29 nodes and 194 edges and was selected as the possible metabolic functional association between identified proteins (Figure 2).

GO enrichment analysis

Because 1 protein-coding gene (uppP) is not mappable in the MCODE, we decided to use 28 proteins for further analysis. Gene ontology enrichment analysis was performed on these 28 proteins (S7) to explore their underlying mechanisms in MTB using the DAVID tool (Figure 3). There were 13 GO enrichment terms for BP. The top 5 enriched terms are cell cycle, cell division, shape regulation, peptidoglycan biosynthetic process, and cell wall organization (false discovery rate [FDR] = 0.0001). Only 2 CC items were obtained from GO enrichment cytoplasmic and an integral membrane component. There were 5 GO terms for MF enrichment, and the most enriched (FDR < 0.0001) were transferase activity, transferring glycosyl group, and arabinosyltransferase activity. Moreover, KEGG analysis revealed 3 pathways correlated with the respected proteins: vancomycin resistance, lysine biosynthesis, and peptidoglycan biosynthesis.

Prediction of subcellular localization and identification of novel drug targets

The subcellular localization of 28 proteins revealed that 12 were cytoplasmic, 2 were extracellular, and 13 were transmembrane, with the exclusion of 1 protein that did not fulfill the requirements (number of predicted transmembrane helix <50 amino acid residues and total probability of peptide cleavage
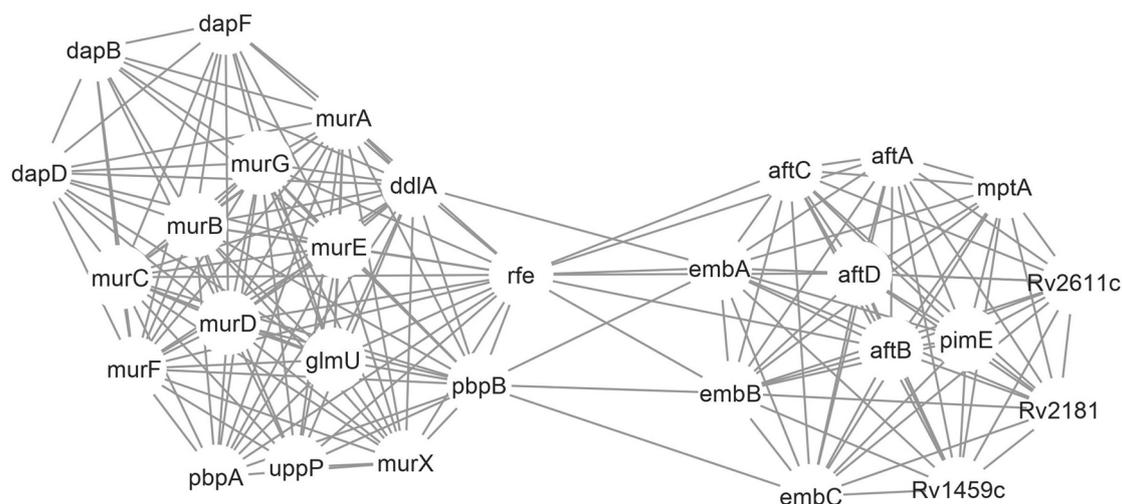
Figure 2. Protein-protein interaction of 29 proteins from non-host essential proteins from *Mycobacterium tuberculosis*.

>0.5).[19] Subsequently, the novel targets were queried against the DrugBank database. Proteins showing no matching hits against the DrugBank database at the threshold were nominated as novel drug targets. The results showed 12 proteins were uniquely involved in pathogen-specific unique pathways, and peptidoglycan and lysine biosynthesis (Table 1).

Discussion

This study focused on subtractive genome analysis, which resulted in identifying proteins that could serve as prospective drug targets against the pathogenicity of MTB. Developing a drug, particularly for non-homologous targets, does not affect the host's biology and has a specific effect on the pathogen. Furthermore, 12 unique proteins from MTB were proposed, and 4 of them are cytoplasmic and 8 are transmembrane unique. Furthermore, 12 non-homologous proteins were discovered in different pathways, such as arabinogalactan biosynthesis, lipoarabinomannan (LAM) biosynthesis, lysine biosynthesis pathway, and O-antigen nucleotide sugar biosynthesis.

Kushwaha et al used a similar strategy to identify therapeutic targets in MTB. Subsequently, 18 prospective drug targets were identified using metabolic pathway and chokepoint analysis.[7] However, this study is more refined as GO and non-homology analysis, as well as druggability, functionality, essentiality, and cellular localization, were included. These provided detailed information, such as their location in the cell, about the drug targets.

Phosphatidylinositol mannoside acyltransferase (patA) is an essential enzyme involved in the biosynthesis of phosphatidyl-myo-inositol mannosides (PIMs), which are vital components of glycolipids/glycoglycans of the mycobacterial cell envelope.[23] Phosphatidyl-myo-inositol mannosides are an important virulence factor during MTB infection and have been shown to be an important enzyme in both in vitro and in vivo growth.[24,25] Defects in these proteins do not directly affect the life of the

pathogenic bacteria but reduce the integrity of the cell wall. Therefore, these results suggest that patA is a promising drug target candidate.[25,26]

Galactan 5-O-arabinofuranosyltransferase, terminal beta-(1→2) arabinofuranosyltransferase, and decaprenylphosphate *N*-acetylglucosamine phosphotransferase are involved in arabinogalactan and LAM biosynthesis, which are essential components of the MTB cell wall. The mycobacterial cell wall is the most frequently adopted target for anti-TB drugs due to the fundamental nature of its synthesis and assembly.[26] This intricating structure, which consists of 3 separate layers of peptidoglycan, arabinogalactan, and mycolic acid, enhances cell proliferation, virulence, and antibiotic resistance.[27] Targeting the enzymes for synthesizing and assembling the arabinan domains of arabinogalactan and LAM presents opportunities for new therapies.

Diaminopimelate epimerase is an important enzyme for lysine biosynthesis, a significant component in the bacterial peptidoglycan cell wall.[28] It plays an essential role in converting LL-DAP into meso-DAP in the lysine biosynthesis pathway in bacteria by converting LL-DAP into meso-DAP. The products of this pathway, namely, meso-DAP and L-lysine, are involved in cross-linking of the peptidoglycan cell wall of gram-negative and gram-positive bacteria.[15] The L-lysine biosynthesis pathway is considered an attractive target for anti-TB drugs due to its unavailability in animals.[29]

N-acetylglucosamine-1-phosphate uridyltransferase (GlmU) is a bifunctional enzyme with uridyltransferase and acetyltransferase activities catalyzed by the N-terminal and C-terminal domains, respectively.[30] It plays a crucial role in synthesizing UDP-*N*-acetylglucosamine, a fundamental precursor of the cell wall peptidoglycan of MTB.[31,32] A study by Soni et al showed that GlmU depletion led to decreased MTB survival.[30] TPSA, a GlmU inhibitor, was reported to impair the cell wall and membrane integrity of MTB.[33] Therefore, this bifunctional enzyme could be a promising target for new TB drugs.[31]
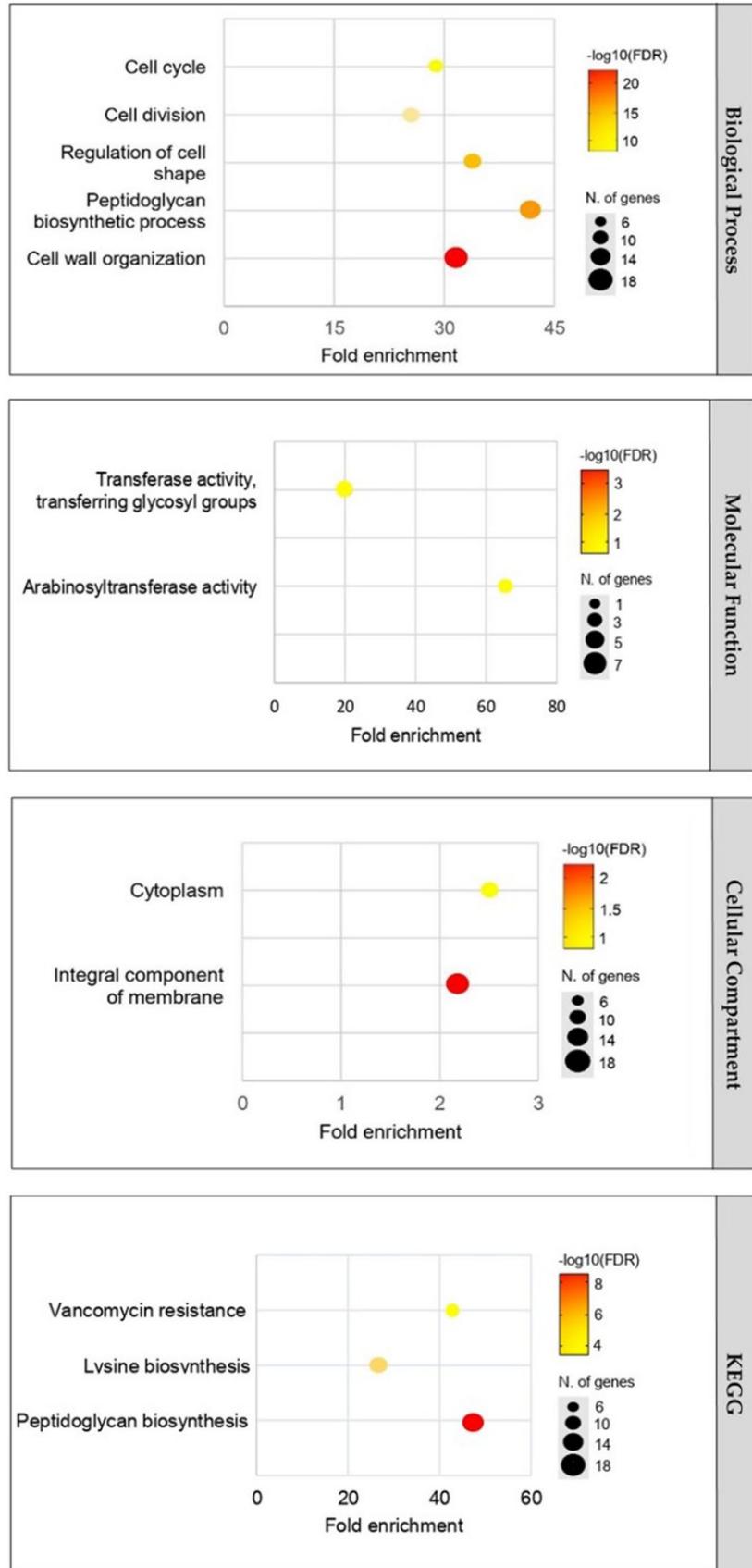
Figure 3. Gene ontology analysis of 28 proteins from non-host essential proteins from MTB. Analysis of the proteins under the headings of biological process, molecular function, cellular compartment, and pathways. The *x*-axis shows significantly enriched categories of the proteins, and the *y*-axis shows the terms (*P* < .0001).
KEGG, Kyoto Encyclopedia of Genes and Genomes; MTB, *Mycobacterium tuberculosis*.

Table 1. List of proteins selected as novel drug targets.

ACCESSION NO.	PROTEIN NAMES	SUBCELLULAR LOCALIZATION
NP_217127.1	Phosphatidylinositol mannoside acyltransferase (Rv2611c)	Cytoplasm
NP_217242.1	Diaminopimelate epimerase (dapF)	Cytoplasm
NP_215534.1	Bifunctional protein GlmU	Cytoplasm
NP_215717.1	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate *N*-succinyltransferase (dapD)	Cytoplasm
NP_215675.1	Polyprenol-phosphate-mannose-dependent alpha-(1-2) phosphatidylinositol pentamannoside (pimE)	Transmembrane
NP_214750.1	Alpha-(1→3)-arabinofuranosyltransferase (aftD)	Transmembrane
NP_216697.1	Polyprenol-phosphate-mannose-dependent alpha-(1-2)-phosphatidylinositol mannoside (Rv2181)	Transmembrane
NP_216690.1	Alpha-(1→6)-mannopyranosyltransferase A (mptA)	Transmembrane
NP_215975.1	Alpha-(1→6)-mannopyranosyltransferase Rv1459c	Transmembrane
NP_218309.1	Galactan 5-O-arabinofuranosyltransferase (aftA)	Transmembrane
NP_218322.1	Terminal beta-(1→2) arabinofuranosyltransferase (aftB)	Transmembrane
NP_215818.1	Decaprenyl-phosphate *N*-acetylglucosaminephosphotransferase (rfe)	Transmembrane

Abbreviation: GlmU, *N*-acetylglucosamine-1-phosphate uridyltransferase.

There are some limitations based on this study. Several drug targets were identified, but not all were in pharmacological activity; hence, they could potentially miss the medication target (undruggable). In addition, functional studies and clinical trials are still required to confirm the safety and efficacy of the drugs.

Conclusions

The availability of complete genome sequences and computer-aided analysis to discover potential anti-TB drug targets has become a new trend. This study performed comparative metabolic pathways to identify the probable anti-TB targets. These results highlight an innovative method to discover therapeutic targets for treating MTB infection. Twelve novel drug targets were reported in this research and are involved in different pathways, including arabinogalactan biosynthesis, LAM biosynthesis, lysine biosynthesis pathway, and O-antigen nucleotide sugar biosynthesis. These results can be further exploited for rational drug design for MTB.

Acknowledgements

Author Contributions

D.A.E.P. was involved in conceptualization, methodology, writing the original draft, and supervision. A.I., S.R.A., and L.A.S. collected data, analyzed, and wrote the original draft. L.M.I., A.D.P., and W.A. wrote, revised, and edited the manuscript. All authors contributed to manuscript revision, and read and approved the submitted version.

SUPPLEMENTAL MATERIAL

Supplemental material for this article is available online.

REFERENCES

1. Natarajan A, Beena PM, Devnikar AV, Mali S. A systemic review on tuberculosis. *Indian J Tuberc*. 2020;67:295-311. doi:10.1016/j.ijtb.2020.02.005
2. Dickson M, Gagnon JP. Key factors in the rising cost of new drug discovery and development. *Nat Rev Drug Discov*. 2004;3:417-429. doi:10.1038/nrd1382
3. Chen L. Computational systems biology for omics data analysis. *J Mol Cell Biol*. 2019;11:631-632. doi:10.1093/jmcb/mjz095
4. Pitaloka DAE, Syamsunarno MRAAA, Abdulah R, Chaidir L. Omics biomarkers for monitoring tuberculosis treatment: a mini-review of recent insights and future approaches. *Infect Drug Resist*. 2022;15:2703-2711. doi:10.2147/IDR. S366580
5. Shanmugham B, Pan A. Identification and characterization of potential therapeutic candidates in emerging human pathogen Mycobacterium abscessus: a novel hierarchical in silico approach. *PLoS ONE*. 2013;8:e59126. doi:10.1371/journal.pone.0059126
6. Cloete R, Oppon E, Murungi E, Schubert WD, Christoffels A. Resistance related metabolic pathways for drug target identification in Mycobacterium tuberculosis. *BMC Bioinformatics*. 2016;17:75. doi:10.1186/s12859-016-0898-8
7. Kushwaha SK, Shakya M. Protein interaction network analysis—approach for potential drug target identification in Mycobacterium tuberculosis. *J Theor Biol*. 2010;262:284-294. doi:10.1016/j.jtbi.2009.09.029
8. Asif SM, Asad A, Faizan A, et al. Dataset of potential targets for Mycobacterium tuberculosis H37Rv through comparative genome analysis. *Bioinformation*. 2009;4:245-248. doi:10.6026/97320630004245
9. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. *Nat Genet*. 2000;25:25-29. doi:10.1038/75556
10. Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. *Nucleic Acids Res*. 2002;30:42-46. doi:10.1093/nar/30.1.42
11. Wei W, Ning LW, Ye YN, Guo FB. Geptop: a gene essentiality prediction tool for sequenced bacterial genomes based on orthology and phylogeny. *PLoS ONE*. 2013;8:e72343. doi:10.1371/journal.pone.0072343
12. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403-410. doi:10.1016/S0022-2836(05)80360-2
13. Narad P, Himashu Bansal H. Computational identification of essential enzymes as potential drug targets in Shigella flexneri Pathogenesis using metabolic pathway analysis and epitope mapping. *J Microbiol Biotechnol*. 2021;31:621-629. doi:10.4014/jmb.2007.07006

14. Szklarczyk D, Gable AL, Nastou KC, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. 2020;49:605-612. doi:10.1093/nar/gkaa1074

15. Assenov Y, Ramírez F, Schelhorn SE, Lengauer T, Albrecht M. Computing topological parameters of biological networks. *Bioinformatics*. 2008;24:282-284. doi:10.1093/bioinformatics/btm554

16. Bader GD, Hogue CWV. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*. 2003;4:2. doi:10.1186/1471-2105

17. Yu NY, Wagner JR, Laird MR, et al. PSORTb 3. 0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*. 2010;26:1608-1615. doi:10.1093/bioinformatics/btq249

18. Yu CS, Cheng CW, Su WC, et al. CELLO2GO: a web server for protein subCELLular LOcalization prediction with functional gene ontology annotation. *PLoS ONE*. 2014;9:e99368. doi:10.1371/journal.pone.0099368

19. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001;305:567-580. doi:10.1006/jmbi.2000.4315

20. Yang H, Qin C, Li YH, et al. Therapeutic target database update 2016: enriched resource for bench to clinical drug target and targeted pathway information. *Nucleic Acids Res*. 2016;44:1069-1074. doi:10.1093/nar/gkv1230

21. Wishart DS, Feunang YD, Guo AC, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. 2018;46:1074-1082. doi:10.1093/nar/gkx1037

22. Jiao X, Sherman BT, Huang DW, et al. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics*. 2012;28:1805-1806. doi:10.1093/bioinformatics/bts251

23. Albesa-Jové D, Svetlíková Z, Tersa M, et al. Structural basis for selective recognition of acyl chains by the membrane-associated acyltransferase PatA. *Nat Commun*. 2016;7:10906. doi:10.1038/ncomms10906

24. Anso I, Basso LGM, Wang L, et al. Molecular ruler mechanism and interfacial catalysis of the integral membrane acyltransferase PatA. *Sci Adv*. 2021;7:eabj4565. doi:10.1126/sciadv.abj4565

25. Tersa M, Raich L, Albesa-Jové D, et al. The molecular mechanism of substrate recognition and catalysis of the membrane acyltransferase PatA from mycobacteria. *ACS Chem Biol*. 2018;13:131-140. doi:10.1021/acschembio.7b00578

26. Bhat ZS, Rather MA, Maqbool M, Lah HU, Yousuf SK, Ahmad Z. Cell wall: a versatile fountain of drug targets in Mycobacterium tuberculosis. *Biomed Pharmacother*. 2017;95:1520-1534. doi:10.1016/j.biopha.2017.09.036

27. Abrahams KA, Besra GS. Mycobacterial cell wall biosynthesis: a multifaceted antibiotic target. *Parasitology*. 2018;145:116-133. doi:10.1017/S0031182016002377

28. Peverelli MG, Costa TPS, da Kirby N, Perugini MA. Dimerization of bacterial diaminopimelate decarboxylase is essential for catalysis. *J Biol Chem*. 2016;291:9785-9795. doi:10.1074/jbc.M115.696591

29. Sagong H-Y, Kim K-J. Structural basis for redox sensitivity in Corynebacterium glutamicum diaminopimelate epimerase: an enzyme involved in l-lysine biosynthesis. *Sci Rep*. 2017;7:42318. doi:10.1038/srep42318

30. Soni V, Upadhayay S, Suryadevara P, et al. Depletion of M. tuberculosis GlmU from infected murine lungs effects the clearance of the pathogen. *PLoS Pathog*. 2015;11:e1005235. doi:10.1371/journal.ppat.1005235

31. Zhou Y, Xin Y, Sha S, Ma Y. Kinetic properties of Mycobacterium tuberculosis bifunctional GlmU. *Arch Microbiol*. 2011;193:751-757. doi:10.1007/s00203-011-0715-8

32. Craggs PD, Mouilleron S, Rejzek M, et al. The mechanism of acetyl transfer catalyzed by mycobacterium tuberculosis GlmU. *Biochemistry*. 2018;57:3387-3401. doi:10.1021/acs.biochem.8b00121

33. Chen C, Han X, Yan Q, et al. The inhibitory effect of GlmU acetyltransferase inhibitor TPSA on mycobacterium tuberculosis may be affected due to its methylation by methyltransferase Rv0560c. *Front Cell Infect Microbiol*. 2019;9:251. doi:10.3389/fcimb.2019.00251