MEDICINAL CHEMISTRY RESEARCH

CrossMark

ORIGINAL RESEARCH

Computational identification of potential drug targets against *Mycobacterium leprae*

Reaz Uddin[1,2] · Syed Sikander Azam[3] · Abdul Wadood[4] · Waqasuddin Khan[1,5] · Umar Farooq[6] · Ajmal Khan[6]

Abstract Leprosy is caused by *Mycobacterium leprae* a major health concern in several countries of the world particularly in Asia and Africa. The preventive measurement has been adopted by the combined efforts of the leprosy burden countries and WHO. However, the situation is getting worse due to the emergence of the resistant strains of the *M. leprae*. The continuous efforts are underway to discover new chemical agent as a therapeutic to cure the diseases caused by the resistant pathogens of bacterial origins. The resistant pathogens are still growing on alarming rate. In order to overcome the resistant pathogens, a relatively newer approach has been applied since last decade. One of them involves the computational subtractive genomics, in which the complete proteome of the bacterial pathogen is step-wise reduced to few potential drug targets. The steps include the finding of non-host proteins, essentiality of the proteins to the pathogens and involvement of the shortlisted proteins in essential metabolic pathways of the pathogen, which are necessary for the bacterial survival. In the current study, we applied computational subtractive genomics on complete proteome of the *M. leprae* and ended up with 16 hypothetical proteins as potential drug targets against which new active molecules can be proposed to ameliorate the activity to cure the disease associated with them. The study is innovative and has a potential to improve the research directions in unraveling the novel cure of leprosy.

✉ Reaz Uddin
mriazuddin@iccs.edu

1   Lab 103 PCMD ext. Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi 75270, Pakistan

2   Prince of Wales Clinical School, Faculty of Medicine, UNSW, Sydney, Australia

3   National Centre for Bioinformatics, Quaid-i-Azam University, Islamabad, Pakistan

4   Department of Biochemistry, Abdul Wali Khan University Mardan, Mardan, Pakistan

5   Jamil-ur-Rahman Center for Genome Research, Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi, Pakistan

6   Department of Chemistry, COMSATS Institute of Information Technology, Abbottabad 22060, Pakistan

Introduction

The causative agent of leprosy, *Mycobacterium leprae* is an acid-fast bacterium. *M. leprae* is having structural resemblance with the *M. tuberculosis*. Leprosy is a major health concern in economically poor countries of Asia, Africa and Latin America. According to WHO official report from five WHO regions, the globally registered number of leprosy cases in the year of 2013 was 180,618, which is still high (http://www.who.int/mediacentre/factsheets/fs101/en/). The current recommended treatment for leprosy with multidrug therapy is designed to prevent the spread of drug-resistant *M. leprae*. Though there is no official definition of multidrug resistance (MDR) in leprosy, the term came into play when resistance to rifampin and one other drug of the standard regimen is observed. The drug-resistant strains have been

reported, since 1964 (Jacobson and Hastings 1976; Ji *et al.*, 1996; Pettit and Rees, 1964). Leprosy has two common forms: tuberculoid and lepromatous with similar symptoms; however, the lepromatous is much more severe.

The elucidation of *M. leprae* genome sequence has been a major achievement (Cole *et al.*, 2001). The most striking finding is the difference between *M. leprae* and its pathogenic relative, *M. tuberculosis*. Instead of the tightly packed 4000-gene chromosome of *M. tuberculosis*, the *M. leprae* sequence encodes only 1600 predicted open-reading frames. The complete genome of the *M. leprae* opens the new arena to understand the drug action and its resistance mechanism at genome level. The genomic information may provide the basis for the design of the new chemicals as therapeutics against *M. leprae*. There are 2770 genes within *M. leprae*. The genome codes for 1605 proteins and contains 1115 pseudogenes. Many of the pseudogenes were involved in catabolism. The biosynthetic pathways of the *M. leprae* tend to be well conserved, and only 49 % of the genome encodes for proteins.

The alarming emergence of the drug resistance strains among many bacterial diseases including the *M. leprae* poses a big challenge to find the effective cures, since the existing drugs are no more active (Matsuoka *et al.*, 2010). In most of the instances, the scientists are engaged in finding the therapeutic compounds against the deadly infections including the resistant ones. However, finding the new drugs against already developed drug targets is not the solution since the pathogen may find the alternative mechanisms to bypass the drug action. One of the recently applied methods to overcome the resistance is to find the new and unique drug targets out of the complete proteomes of the bacteria. Several applications have been reported in the literature focusing on exploring the new target sites (Barh *et al.*, 2011; Uddin and Saeed, 2014; Uddin *et al.*, 2015). Within this context, the computational subtractive genomics is the most applicable method in order to find the novel drug targets. In the current study, we applied a computational subtractive genomics method to shortlist few unique and novel drug targets against *M. leprae*. There are few literature reports available which described the comparative and computational identification of the potential drug targets in mycobacterial species including *M. leprae* (Barh *et al.*, 2011; Cole 2002; Crowther *et al.*, 2010; Marri *et al.*, 2006; Sarker *et al.*, 2013; Singh *et al.*, 2014). However, the current study described the identification of potential drug targets within the hypothetical proteins pool of *M. leprae* which is largely ignored previously. We shortlisted at least 16 hypothetical proteins out of the 1604-sized proteome of the *M. leprae*. Those newly signified functions to the 16 hypothetical proteins may lead to discover a novel drug target against which the new chemical could be proposed in future as drugs. Since the proposed drug targets are non-

homologous to the human host therefore, we expect that there should not be any side effects associated with inhibiting their activities by active constituents.

Materials and methods

NCBI BLAST + standalone version 2.2.26 (Altschul *et al.*, 1990) was used for the study. The overall scheme of the current study is shown in Fig. 1.

Complete proteome retrieval

We obtained the complete proteome of the *M. leprae* from NCBI. The complete proteome of *H. sapiens* was retrieved from the HAMAP on ExPASy as UniProtKB FASTA format.

Determining non-paralogous sequences

CD-HIT (Li and Godzik, 2006) was sued for the identification of the paralogous or duplicate protein sequences with sequence identity cutoff 0.8 (i.e., 80 %). Paralogous sequences were filtered from complete proteome of *M. leprae* resulted in non-paralogous sequences only.

Determination of non-homologous protein sequences to the human proteome

BLASTp was used on the non-paralogous sequences of the *M. leprae* against *Homo sapiens* using threshold expectation value (e value 10^{-3}). The resultant sequences consisted of homologous sequences (significant similarity with Human host) and non-homologous sequences (no hits found). Sequences which showed significant similarity with the human host were removed leaving only the non-homologous sequences for subsequent analysis.

Identification of non-homologous essential proteins in *M. leprae*

Database of essential gene (DEG) (Zhang *et al.*, 2004) was downloaded from the DEG website (http://www.essentialgene.org/). The non-homologous sequences were passed through the BLASTp search using DEG as database with e value 10^{-5}. The filtered sequences with the significant similarity with the DEG database are represented as essential proteins for the pathogens.

KEGG metabolic pathway analyses

KEGG is Kyoto Encyclopedia of Genes and Genomes and contains the complete metabolic pathways of the organism. The KEGG can be accessed interactively via KEGG

Automated Annotation Server (KAAS) (Moriya *et al.,* 2007). The KAAS server was used to predict the involvement of the protein sequences within different metabolic pathways of the pathogens.

Prediction of subcellular localization

All non-homologous essential proteins were subjected to the prediction of subcellular localization by using PSORTb version 3.0 (Nancy *et al.,* 2010). The main principle is to use SubCellular Localization BLAST (SCL BLAST) which takes all non-homologous essential protein sequences and runs BLASTp against database of proteins of known subcellular localization. PSORTb defines prediction results for different subcellular localization, and it may include cytoplasm, cytoplasmic membrane, cell wall and extracellular and unknown.

Functional family prediction of all non-homologous, essential and hypothetical proteins

The SVMProt server is a method of choice to predict the functional family classification of the proteins particularly the hypothetical proteins for which there is no functional information available. The hypothetical protein sequences were subjected to the SVMProt server to predict the functional family classes of non-homologous hypothetical protein sequences. SVMProt is a server for the classification of a protein into functional class from its primary sequence including all major classes of enzymes, receptors, transporters, channels, DNA-binding proteins and RNA-binding proteins (Cai *et al.,* 2003).

Druggability potential of shortlisted sequences

The screening of all non-homologous, essential and hypothetical proteins was assessed by BLASTp comparison against DrugBank database (Knox *et al.,* 2011) which contains number of protein targets with respect to the drug IDs approved by FDA. In order to reach the novel drug targets, default parameter values with e value 10^{-3} were used in BLASTp search against the DrugBank database.

Results and discussion

The major objective of the current study was to find the potential drug targets against *M. leprae*. The proposed drug targets should fulfill the druggability criteria, which include the non-homologous to human host, essential to the pathogen (*M. leprae*) and playing important role in major metabolic pathway of the pathogen. Here, we applied a computational routine which has been cited in the literature earlier for effective identification of the new and novel drug targets against multiple bacterial pathogens (Uddin and Saeed, 2014). Fig. 1 shows the complete workflow of
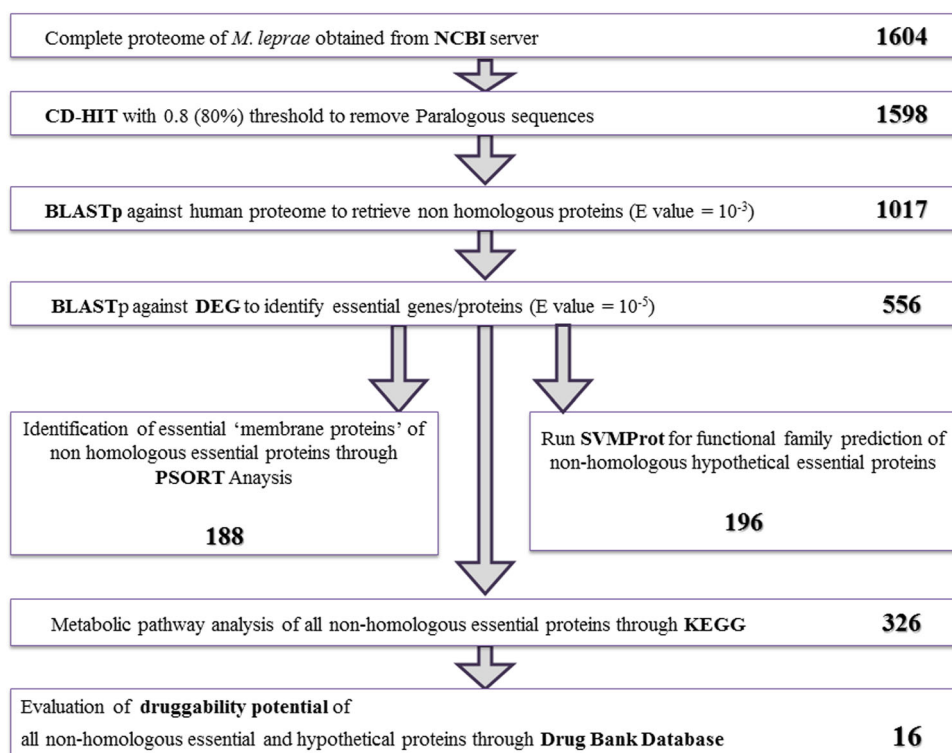
Fig. 1 Overall workflow chart

Complete proteome of *M. leprae* obtained from **NCBI** server | **1604**

CD-HIT with 0.8 (80%) threshold to remove Paralogous sequences | **1598**

BLASTp against human proteome to retrieve non homologous proteins (E value = 10^{-3}) | **1017**

BLASTp against **DEG** to identify essential genes/proteins (E value = 10^{-5}) | **556**

Identification of essential 'membrane proteins' of non homologous essential proteins through **PSORT** Analysis | **188**

Run **SVMProt** for functional family prediction of non-homologous hypothetical essential proteins | **196**

Metabolic pathway analysis of all non-homologous essential proteins through **KEGG** | **326**

Evaluation of **druggability potential** of all non-homologous essential and hypothetical proteins through **Drug Bank Database** | **16**

Table 1 Subtractive genomics steps and corresponding number of retrieved sequences

S. no.	Steps	*Mycobacterium leprae*
1	Total number of proteins	1604
2	Removed paralogous (>80 % identical) in CD-HIT	1598
3	Number of proteins against *H. sapiens* using BLASTp (E value 10^{-3})	1017
4	Essential proteins in DEG (*E* value 10^{-5})	556
5	Essential proteins involved in metabolic pathways (KEGG)	326
7	Number of essential cell membrane proteins (PSORT)	188
8	Number of non-homologous hypothetical protein as essential proteins (SVMProt)	196
9	Essential drug target proteins (DBD)	16

the current study, and Table 1 shows each step and the respective outcomes of the number of sequences.

Identification of paralogous, non-homologous and essential proteins

The obtained complete proteome of the *M. leprae* strain Br4923 from NCBI consisted of 1604 protein sequences. The complete proteome of *M. leprae* was initially subjected to the CD-HIT with the sequence identity cutoff of 0.8 (80 % threshold). The CD-HIT step was performed to remove the paralogous sequences from complete proteome of the *M. leprae*. The CD-HIT resulted in the identification of at least six duplicate sequences with the corresponding cluster similarities from 92 to 100 %. Consequently, the six duplicated sequences were removed and hence resulted in the 1598 sequences. Table 2 shows the GI numbers of the paralogous sequences. The next step was to identify the non-host proteins. Since the major limitation of any drug is its side effects via cross-reaction with the host proteins, the drug target from the pathogen should be unique and non-homologous to any host protein to refrain from any cross-reactivity of the drug with the host proteins. In order to find out the non-homologous proteins from the *M. leprae* proteome, we ran a BLASTp standalone script where the queries were the complete non-redundant proteins of the *M. leprae* and the database was complete human proteome obtained from UniProt (e value 10^{-3}). This process resulted in the identification of only those proteins in *M. leprae,* which were absent in the human host (i.e., no hits found in BLASTp run). As many as 581 proteins of the *M. leprae* were found to have at least one of the human homolog and therefore not suitable to be employed as potential drug targets. We removed those 581 proteins from *M. leprae* proteome, which left us with 1017 of proteins for which there were no corresponding human homologs and therefore ideal to be a best candidate as potential drug targets. Other important criterion for considering a druggable

protein is the essentiality of the protein for the survival of the pathogen. The essentiality of any protein can establish its druggable characteristics. The database of essential genes (DEG) is the source from where we can find the essentiality of any protein sequence by making a comparison with the essential proteins present in DEG. We ran the BLASTp using the non-homologous (non-host proteins) as queries while the DEG as database with an e value of 10^{-5}. The step resulted in as many as 556 proteins from *M. leprae* as essential for the survival of the bacteria. The process ensured that the shortlisted sequences were essential for the survival of the pathogen and hence could be considered safely as potential drug targets in order to find cure for the infections caused by *M. leprae*.

Subcellular localization of the non-homologous and essential proteins

An important prerequisite of a protein to show its function is its compartmentalization or localization. In order to perform its optimum function, the protein needs to be in a specific location. There are methods available which can predict the subcellular localization of the proteins by comparing the sequences only. One of the best subcellular localization prediction methods is PSORTb as most reliable method (Nancy *et al.,* 2010). We subjected the non-homologous and essential proteins to the PSORTb which revealed that the majority of the sequences ($\sim$50 %) belonged to the cytoplasmic region of the cell (Fig. 2). The next big fractions of the sequences were located at the cytoplasmic membrane. These cytoplasmic membrane proteins could be the potential vaccine targets. Some of the cytoplasmic membrane proteins from the *M. leprae* were penicillin-binding protein (gi = 221229359), D-alanyl-D-alanine carboxypeptidase (gi = 221229769), sec-independent translocase (gi = 221230006), etc. A complete list is provided as supplementary information (File S1).

Table 2 CD-HIT identified paralogous sequences

Cluster	Size	Protein NCBI GI	%Threshold
Cluster# 0			99.00
0	498aa,	gi_221230151_ref_YP_002503567.1	
1	505aa,	gi_221230152_ref_YP_002503568.1	
Cluster# 1			92.00
0	438aa,	gi_221229434_ref_YP_002502850.1	
1	435aa,	gi_221229437_ref_YP_002502853.1	
Cluster# 2			100.00
0	369aa,	gi_221229599_ref_YP_002503015.1	
1	369aa,	gi_221230932_ref_YP_002504348.1	
Cluster# 3			100.00
0	100aa,	gi_221229989_ref_YP_002503405.1	
1	100aa,	gi_221230075_ref_YP_002503491.1	
Cluster# 4			100.00
0	99aa,	gi_221229988_ref_YP_002503404.1	
1	99aa,	gi_221230077_ref_YP_002503493.1	
Cluster# 5			100.00
0	95aa,	gi_221229990_ref_YP_002503406.1	
1	95aa,	gi_221230074_ref_YP_002503490.1	

Functional family classification of the non-homologous and essential but hypothetical proteins

The hypothetical proteins are those for which the sequence is available; however, the function is not known yet. There are thousands of hypothetical proteins in pathogenic bacteria which present the challenging job of curating their functions by any means. One of the methods to predict the function of the hypothetical proteins is to classify them in functional families based on the sequence similarity. One of the best methods reported in the literature to classify the proteins in functional groups/families is SVMProt method. There were at least 196 hypothetical proteins in the non-homologous and essential proteins of the *M. leprae*. We

Fig. 2 Subcellular localization of non-homologous essential proteins of *M. leprae* reflected in percentage

used the SVMProt method to predict the functional families of all the hypothetical proteins of *M. leprae*. The resulted frequency distribution of the different functional classes is shown in Fig. 3. The most prevalent families were transmembranes, zinc-binding proteins, lipid-binding proteins, transferases and iron-binding proteins. The supporting file S2 contains the detailed report of the SVMProt step.

KEGG metabolic pathways analysis

The non-homologous and essential proteins were passed through the online server KAAS. The KAAS server allowed us to find the involvement of the submitted protein sequences in various essential metabolic pathways present in bacteria. The results obtained by the KASS server are shown in Fig. 4. Total 556 protein sequences were passed through the KAAS server. The names of various metabolic pathways in which the *M. leprae* take part are: *carbon metabolism* including 2-oxocarboxylic acid metabolism, fatty acid metabolism, biosynthesis of amino acids; *carbohydrate metabolism* including glycolysis/gluconeogenesis, citrate cycle (TCA), pentose phosphate pathway, fructose and mannose metabolism, galactose metabolism, starch and sucrose metabolism, amino sugar and nucleotide metabolism, pyruvate metabolism, glyoxylate and dicarboxylate metabolism, butanoate metabolism, C5-branched dibasic acid metabolism, inositol metabolism; *energy metabolism* including oxidative phosphorylation, photosynthesis, carbon fixation in photosynthetic organisms,
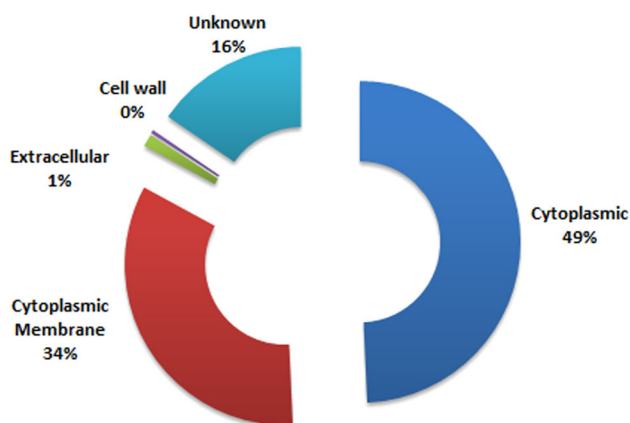
Fig. 3 Functional family predictions of non-homologous, hypothetical, essential proteins of *M. leprae*. Frequency of each functional family is reported in *Y*-axis
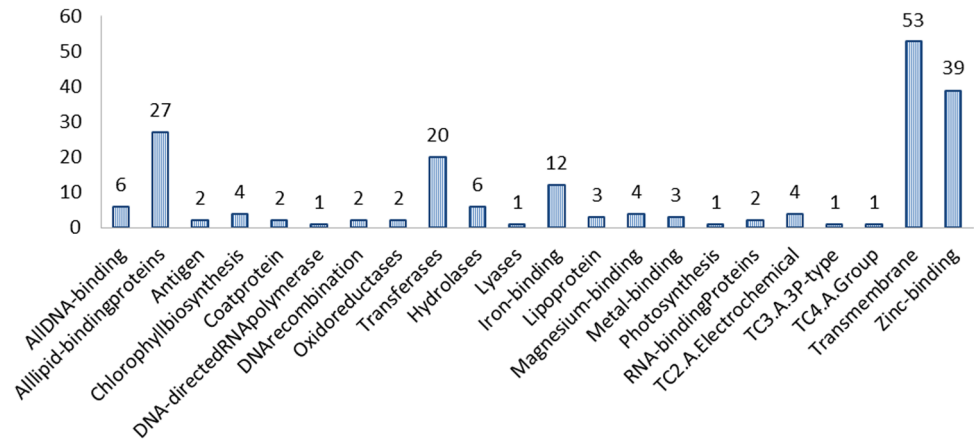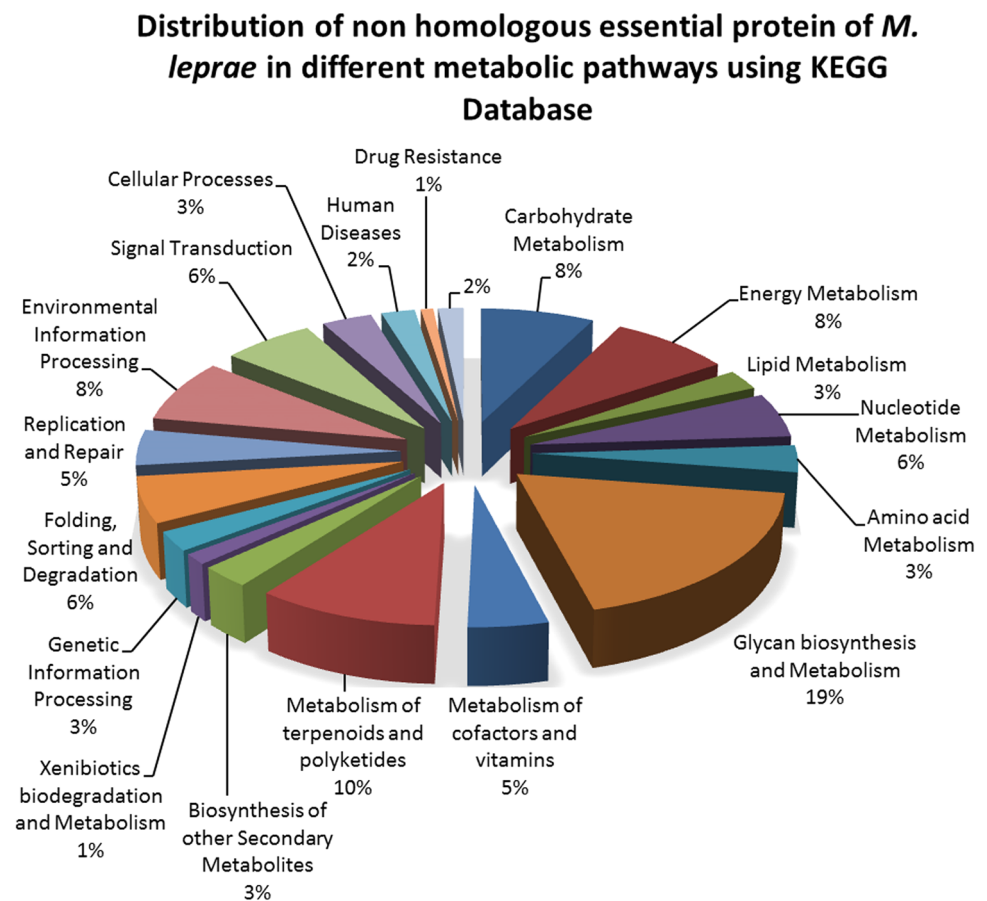

Fig. 4 Metabolic pathway (KEGG) analysis of *M. leprae*. The distribution of non-homologous essential proteins of *M. leprae* in different metabolic pathways using KEGG database

Distribution of non homologous essential protein of *M. leprae* in different metabolic pathways using KEGG Database

carbon fixation pathways in prokaryotes, methane metabolism, nitrogen metabolism, sulfur metabolism, lipid metabolism, including fatty acid biosynthesis, glycerolipid metabolism, glycerophospholipid metabolism, biosynthesis of unsaturated fatty acids; *nucleotide metabolism* including purine metabolism and pyrimidine metabolism; *amino acid metabolism* including alanine, aspartate and glutamate metabolism, glycine, serine and threonine metabolism,

cysteine and methionine metabolism, valine, leucine and isoleucine biosynthesis, lysine biosynthesis, arginine and proline metabolism, histidine metabolism, tyrosine metabolism, phenylalanine, tyrosine and tryptophan biosynthesis, lipopolysaccharide biosynthesis, peptidoglycan biosynthesis; *metabolism of cofactors and vitamins* including thiamine metabolism, riboflavin metabolism, vitamin B6 metabolism, nicotinate and nicotinamide

Table 3 Unique metabolic pathways of *M. leprae* from KEGG

S. no.	Names of unique metabolic pathway of *M. leprae*
1	C5-branched dibasic acid metabolism
2	Methane metabolism
3	Nitrogen metabolism
4	Pantothenate and CoA biosynthesis
5	Limonene and pinene degradation
7	Streptomycin biosynthesis
8	Novobiocin biosynthesis
9	Benzoate degradation
10	Aminobenzoate degradation
11	Ethylbenzene degradation
12	Bacterial secretion system
13	Two-component system
14	Beta-lactam resistance
15	Vancomycin resistance

metabolism, pantothenate and CoA biosynthesis, biotin metabolism, folate biosynthesis, porphyrin and chlorophyll metabolism, ubiquinone and other terpenoid-quinone biosynthesis; *metabolism of terpenoids and polyketides* including terpenoid backbone biosynthesis, limonene and pinene degradation, polyketide sugar unit biosynthesis, biosynthesis of siderophore group nonribosomal peptides; *biosynthesis of other secondary metabolites* including monobactam biosynthesis, streptomycin biosynthesis, novobiocin biosynthesis; *xenobiotic biodegradation and metabolism* including benzoate degradation, aminobenzoate degradation, ethylbenzene degradation; *genetic information processing* including RNA polymerase, ribosome, aminoacyl t-RNA biosynthesis; *folding, storing and degradation* including protein export, sulfur relay system, proteasome, RNA degradation. *Replication and repair* including DNA replication, base excision repair, nucleotide excision repair, mismatch repair, homologous recombination; *membrane transport* including ABC transporters and bacterial secretion system; *signal transduction* including two-component system; *cellular processes* including peroxisome and *Cell* motility. *Cell growth and death* include cell cycle caulobacter; *drug resistance* including beta-lactam resistance and vancomycin resistance. Out of all of the above-mentioned metabolic pathways, few of them belonged to the unique metabolic pathways of the pathogen; for example, those pathways were absent in the human host. The unique metabolic pathways are shown in Table 3. The proteins belonging to those pathways are the most likely potential drug targets since there are no competing pathways in the human host so there is no possibility of the side effects. The supporting file S3 contains the detailed description of each of the protein sequence involved in metabolic pathways obtained by KAAS.

Druggability potential of the shortlisted hypothetical sequences

We further looked at the druggability potential of the shortlisted protein sequences in earlier steps. There were 196 hypothetical proteins from essential and non-homologous proteins. In this step, we found the possible homolog of these 196 hypothetical proteins in DrugBank database. We shortlisted significant homologs of 16 hypothetical proteins from the established drug targets present in the DrugBank. The details are shown in Table 4. The 16 DrugBank homologs of the hypothetical proteins belonged to sugar phosphatase YbiV, exopolyphosphatase, FEZ-1 protein, putative ribosome biogenesis, laccase domain protein YfiH, uracil-DNA glycosylase, periplasmic oligopeptide-binding protein, streptogramin A acetyltransferase, response regulator PleD, ubiquitin-like modifier activating enzyme, aminomethyltransferase, cyclopropane mycolic acid synthase MmA2, exopolyphosphatase, peroxisomal multifunctional enzyme and putative serine/threonine protein kinase. All of the 16 hypothetical proteins can be considered as potential drug targets since all have one homolog in DrugBank with at least 25 % sequence identity. We proposed to explore all of these 16 hypothetical proteins by structure-based methods, for example, homology modeling and molecular docking, etc., which can further enlighten us about the potential role of these proteins for the survival of *M. leprae*. In the following, we explain one of the 16 proteins with the GI# 221229528. The protein with GI# 221229528 showed 39 % match with the DrugBank target exopolyphosphatase (DB03382). The exopolyphosphatase have been studied as drug target against bacteria, and inhibition of this enzyme can specifically block the cleavage of the chain of phosphates. The

Table 4 DrugBank targets of the essential, non-homologous, hypothetical proteins

NCBI GI	Name of the protein in *M. leprae*	DrugBank target name and DrugBank ID
221229403	Hypothetical protein MLBr_00089	Sugar phosphatase YbiV (DB04156)
221229528	Hypothetical protein MLBr_00258	Exopolyphosphatase (DB03382)
221229531	Hypothetical protein MLBr_00270	FEZ-1 protein (DB02032)
221229821	Hypothetical protein MLBr_00791	Putative ribosome biogenesis GTPase RsgA(DB04315)
221229915	Hypothetical protein MLBr_00918	Laccase domain protein YfiH (DB04272)
221230022	Hypothetical protein MLBr_01105	Uracil-DNA glycosylase (DB03419)
221230265	Hypothetical protein MLBr_01497	Periplasmic oligopeptide-binding protein.(DB07365)
221230398	Hypothetical protein MLBr_01719	Streptogramin A acetyltransferase (DB01669)
221230413	Hypothetical protein MLBr_01750	Response regulator PleD (DB01972)
221230414	Hypothetical protein MLBr_01751	Ubiquitin-like modifier activating enzyme (DB04119)
221230658	Hypothetical protein MLBr_02203	Aminomethyltransferase (DB03256; DB03760)
221230713	Hypothetical protein MLBr_02304	Uncharacterized protein (DB02078)
221230734	Hypothetical protein MLBr_02334	Cyclopropane mycolic acid synthase MmaA2 (DB01718)
221230792	Hypothetical protein MLBr_02434	Exopolyphosphatase (DB03382)
221230856	Hypothetical protein MLBr_02566	Peroxisomal multifunctional enzyme type DB00157
221230937	Hypothetical protein MLBr_02700	Putative serine/threonine protein kinase (DB05318)

SVMProt identified the functional family class of this query protein as manganese-binding protein. It is reported that for the optimal activity the enzyme exopolyphosphatase required manganese (Wurst and Kornberg, 1994). Because of the ubiquitous nature of the exopolyphosphatase, we additionally checked whether there is any sequence similarity of this query protein with any of the human protein. The BLASTp resulted in the 'no significant similarity' with any of the protein of the human. The above-mentioned characteristics turned exopolyphosphatase as one of the promising drug targets against *M. leprae*.

Conclusion

We applied a comprehensive computational approach on the complete proteome of the *M. leprae* in a hope to propose new and potential drug targets against *M. leprae*. *M. leprae* proteome consisted of 1604 proteins which step-wise reduced to 16 proposed drug targets. The identified drug targets were hypothetical proteins which were short-listed by following the drug-target-like filtering criteria including the non-homology to the host, essential to the pathogen's survival and involvement in significant metabolic pathway during the pathogen's life cycle. Previous studies have focus on discovering the novel chemical candidates as new therapeutics against deadly infections including mycobacterial diseases. However, limited studies are available to propose novel and unique drug targets against which the new therapeutics could be discovered. This study reported an interesting application of the computational subtractive genomics approach to shortlist few potential drug targets which can be proposed to discover antibacterial candidates against *M. leprae*. We are quite optimist that the study will move forward the research in a new and fruitful directions to cure the deadly disease caused by the *M. leprae* (Leprosy).

Compliance with ethical standards

Conflict of interest The authors report no conflict of interest, and they are responsible for the content and writing of the paper.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Barh D, Tiwari S, Jain N, Ali A, Santos AR, Misra AN, Azevedo V, Kumar A (2011) In silico subtractive genomics for target identification in human bacterial pathogens. Drug Dev Res 72:162–177

Cai C, Han L, Ji ZL, Chen X, Chen YZ (2003) SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence. Nucleic Acids Res 31:3692–3697

Cole S (2002) Comparative mycobacterial genomics as a tool for drug target and antigen discovery. Eur Respir J Suppl 20:78s–86s

Cole S, Eiglmeier K, Parkhill J, James K, Thomson N, Wheeler P, Honore N, Garnier T, Churcher C, Harris D (2001) Massive gene decay in the leprosy bacillus. Nature 409:1007–1011

Crowther GJ, Shanmugam D, Carmona SJ, Doyle MA, Hertz-Fowler C, Berriman M, Nwaka S, Ralph SA, Roos DS, Van Voorhis WC (2010) Identification of attractive drug targets in neglected-disease pathogens using an in silico approach. PLoS Negl Trop Dis 4:e804

Jacobson R, Hastings R (1976) Rifampin-resistant leprosy. Lancet 308:1304–1305

Ji B, Perani EG, Petinom C, Grosset JH (1996) Bactericidal activities of combinations of new drugs against *Mycobacterium leprae* in nude mice. Antimicrob Agents Chemother 40:393–399

Knox C, Law V, Jewison T, Liu P, Ly S, Frolkis A, Pon A, Banco K, Mak C, Neveu V (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. Nucleic Acids Res 39:D1035–D1041

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659

Marri PR, Bannantine JP, Golding GB (2006) Comparative genomics of metabolic pathways in Mycobacterium species: gene duplication, gene decay and lateral gene transfer. FEMS Microbiol Rev 30:906–925

Matsuoka M, Suzuki Y, Garcia IE, Fafotis-Morris M, Vargas-Gonzalez A, Carreno-Martinez C, Fukushima Y, Nakajima C (2010) Possible mode of emergence for drug-resistant leprosy is revealed by an analysis of samples from Mexico. Jpn J Infect Dis 63:412–416

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 35:W182–W185

Nancy YY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608–1615

Pettit J, Rees R (1964) Sulphone resistance in leprosy: an experimental and clinical study. Lancet 284:673–674

Sarker M, Talcott C, Galande AK (2013) In silico systems biology approaches for the identification of antimicrobial targets. Methods Mol Biol 993:13–30

Singh Y, Kohli S, Sowpati DT, Rahman SA, Tyagi AK, Hasnain SE (2014) Gene cooption in Mycobacteria and search for virulence attributes: comparative proteomic analyses of *Mycobacterium tuberculosis*, *Mycobacterium indicus pranii* and other mycobacteria. Int J Med Microbiol 304:742–748

Uddin R, Saeed K (2014) Identification and characterization of potential drug targets by subtractive genome analyses of methicillin resistant *Staphylococcus aureus*. Comput Biol Chem 48:55–63

Uddin R, Saeed K, Khan W, Azam SS, Wadood A (2015) Metabolic pathway analysis approach: identification of novel therapeutic target against methicillin resistant *Staphylococcus aureus*. Gene 556:213–226

Wurst H, Kornberg A (1994) A soluble exopolyphosphatas e of *Saccharomyces cerevisiae*. purification and characterization. J Biol Chem 269:10996–11001

Zhang R, Ou HY, Zhang CT (2004) DEG: a database of essential genes. Nucleic Acids Res 32:D271–D272