CrossMark

ORIGINAL ARTICLE

Genome scale identification, structural analysis, and classification of periplasmic binding proteins from *Mycobacterium tuberculosis*

Padmani Sandhu[1] · Monika Kumari[1] · Kamal Naini[1] · Yusuf Akhter[1]

Abstract Periplasmic-binding proteins occupy the periplasmic space of bacteria and are involved in binding and transport of various ions, siderophores, and other diverse types of solutes. These proteins may be associated with membrane transport systems or may help in activation of signal transducers. There is limited information available on *Mycobacterium tuberculosis* (*Mtb*) periplasm-inhabiting proteins. In the present study, we have performed genome-wide identification and functional annotation of periplasmic-binding proteins of *Mtb* on the basis of signature characteristics and their functional motifs. 37 putative periplasmic-binding proteins were identified in *Mtb* proteome and categorized into different classes mainly known for their association with membrane transport and signaling pathways. Conclusively, this study adds 11 completely novel proteins to the periplasmic binding proteome of *Mtb,* which were not annotated as PBPs earlier. This study provides an overview of the periplasmic binding proteome of *Mtb,* which may be involved in various important pathophysiological functions of the bacteria. These proteins may serve as novel drug targets, which may lead to better treatment strategies against this deadly pathogen.

Communicated by M. Kupiec.

✉ Yusuf Akhter
yusuf@cuhimachal.ac.in; yusuf@daad-alumni.de

1   Centre for Computational Biology and Bioinformatics, School of Life Sciences, Central University of Himachal Pradesh, Shahpur, District Kangra, Himachal Pradesh 176206, India

Introduction

Periplasm is a semi-liquid gel matrix filling the space between outer cell wall and cell membrane of the bacteria. The periplasm is similar to the cytoplasm in composition but it may contain different types of proteins according to its functional requirements. The environment of periplasmic space is dynamic as its composition varies constantly in different growth conditions and with changing external environment due to its close proximity with the bacterial outer membrane. For maintaining the shape of bacterium, the periplasm also contains some amount of peptidoglycans (murein) and other maintenance-related proteins (Vollmer and Seligman 2010). Other periplasmic proteins may be hydrolytic enzymes, chemoreceptors, detoxifying enzymes, and periplasmic binding proteins (PBP) (Ho et al. 2007). Periplasmic transporter proteins are termed as PBP in Gram-negative bacteria and substrate-binding proteins (SBP) in Gram-positive bacteria but, both type of proteins share structural similarity and perform similar functions (Berntsson et al. 2010). The main function of these proteins is to transport solutes or variety of substrates across inner and outer membranes. The PBPs of Gram-positive bacteria have diacylglycerol lipid anchor at the amino terminal with the help of which they remain

attached to the outer leaflet of the cytoplasmic membrane (Chu and Vogel 2011).

PBPs work in association with the transmembrane proteins to carry out the transport process. The PBP-dependent transport systems typically consist of a hydrophobic integral membrane protein, hydrophilic peripheral membrane protein, a periplasmic SBP (Chaudhary et al. 2016; Higgins 2001; Sandhu and Akhter 2015). ATP-binding cassette (ABC) permeases superfamily has been reported to contain many well-studied PBP-dependent transporters (Saurin and Dassa 1994). In addition to this, PBPs are also associated with tripartite ATP-independent periplasmic (TRAP)-transporters, many ligand-gated ion channels, G-protein-coupled receptors (GPCRs) and guanylate cyclase-atrial natriuretic peptide receptors which exhibit their functional diversity (Mulligan et al. 2011; Tasneem et al. 2004).

PBPs help in scavenging and sensing the nutrients and aiding in their transport from and toward the nutrient source and thus instigate the chemotaxis (Felder et al. 1999). PBPs structurally have three classes. Class I and II of PBPs have two or three inter-domain connections. The third class of PBPs contains two globular domains connected by a helix and includes proteins such as siderophore-binding PBPs and metal ions-binding PBPs such as ChuT (heme binding) and TroA (zinc and manganese binding) (Chu and Vogel 2011).

Mtb is a leading opportunistic human pathogen and is a causative agent of tuberculosis (TB). TB has high death toll irrespective of effective treatment strategies and have high risk of infection in the case of immuno-compromised patients suffering from human immuno deficiency virus (WHO 2015). The unique cell envelope of *Mtb* has a lipid-rich cell wall consisting of peptidoglycans, arabinogalactans, and mycolic acid-ligated arabinan chains forming a thick waxy characteristic outer coat. Cell wall surrounds the phospholipid-rich cell membrane and the space between the two envelopes is occupied by the periplasm (Kieser and Rubin 2014). The periplasmic space of *Mtb* is home to several PBPs specific to various physiological conditions; however, only few of them have been characterized till date. PBPs in *Mtb* have been documented to carry out versatile functions and serving to retain the viability of the organism under the stressed environmental conditions. The first reported PBP from mycobacterial genera was a maltose uptake carrying protein Mla; gene encoding this protein has been cloned and sequenced in the *M. bovis* and, sequence similarity searches has revealed its conservation and unique genetic arrangement across the species of *Mtb* complex. One homolog of this protein LpqY from *Mtb* has been reported to be a part of ABC transporter system LpqY-SugA-SugB-SugC and reported to facilitate the recycling of trehalose during the TB infection (Borich et al. 1999; Kalscheuer et al. 2010). PBPs have also

been shown to regulate the phosphate-limiting environment in *Mtb*. EPstS-1 (PDB ID: 1PC3) PBP from *Mtb* has been reported to have high affinity for phosphate binding and enabling the pathogen to survive under the phosphate-limiting conditions. This protein was shown to work in association with the transmembrane permease assembly of ABC transporter proteins, PstA and PstC (Peirs et al. 2005; Vyas et al. 2003). *Mtb* also possesses heme-binding PBP Rv0203 (PDB ID: 3MAY), which has been reported to bind heme and transfer it to transmembrane proteins MmpL3 and MmpL11 for its further transport into the cytoplasm (Tullius et al. 2011). PBPs in *Mtb* has also been reported to regulate transcription levels as periplasmic anti-sigma factor RsdA (PDB ID: 3VEP) has been reported and well characterized (Jaiswal et al. 2013). UgpB, another ABC transporters-related PBP, has also been identified and characterized from the periplasm of *Mtb*. This protein along with the UgpAEC proteins forms an ABC importer system for sn-glycerol 3-phosphate (G3P) uptake (Jiang et al. 2014). PBPs have also been reported to constitute the non-classical secretion system such as ESX-1 secretion, *Mtb*EccB1 (PDB ID: 4KK7), a single transmembrane helix containing PBP has been characterized. It was reported to be an essential part of the ESX-1 secretion system and also show ATPase activity as evident from the ATP docking studies (Wagner et al. 2016; Zhang et al. 2015). Recently, a sugar-binding periplasmic protein UspC (PDB ID: 5K2X) from *Mtb* has been characterized. It was also revealed that UspC forms an essential component of the trimeric assembly of ABC permeases transmembrane proteins involved in amino-sugar transport. This UspABC assembly serves to cope with the scarcity of carbohydrate by importing them from the carbohydrate-limited environment of phagosomes (Fullam et al. 2016). One more PBP (PDB ID: 4BEG) was characterized from *Mtb* and has been reported to share homology with the lipid-binding proteins; however, the molecular function of the protein is still unknown (Eulenburg et al. 2013). These reported PBPs from *Mtb* with diverse physiological functions and crucial for the survival of the pathogen in the host cells encouraged us to mine more PBPs from its proteome and which may be further characterized and could be targeted for drug discovery programs.

In the last few decades, the evolution of multidrug resistant and extensively drug-resistant strains has posed a new challenge to the eradication of TB infection (Cohen et al. 2016). Therefore, it is the need of the hour to explore further for novel drugs against *Mtb*. In the present study, we have performed a genome-wide identification of putative PBPs of *Mtb* which may serve as potential future drug targets against TB. We have screened the PBPs on the basis of presence of signatures like signal peptides, lipopeptide signal peptides, single N-terminal transmembrane helices, and their periplasmic localization (Fig. 1). All the putative PBPs have been annotated into different classes on the
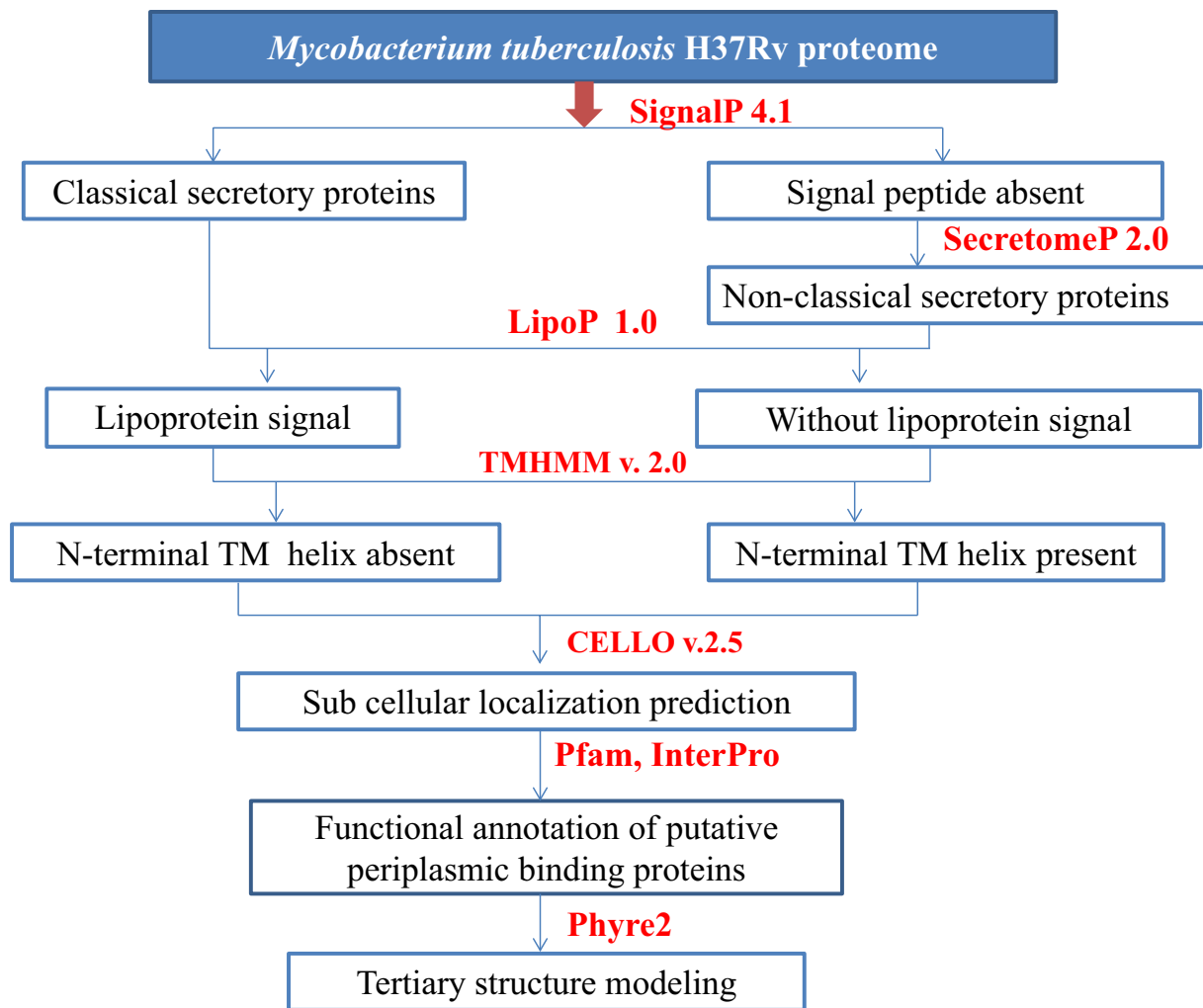
Fig. 1 Methodology for screening of putative periplasmic binding proteins across the *Mtb* proteome: Step-wise screening of putative PBPs. Whole proteome sequence was scanned by SignalP (Nielsen et al. 1997) program for the identification of classically secreted proteins. Rest of the proteins without classical signal peptides were analyzed using SecretomeP 2.0 (Bendtsen et al. 2004) for the non-classically secreted proteins identification. Classically and non-classically secreted proteins were analyzed individually for the identification of lipoprotein signal peptides using LipoP 1.0 (Juncker et al. 2003) and then single N-terminal TM-helix prediction was carried out using TMHMM v. 2.0 (Krogh et al. 2001). All the classically and non-classically secreted proteins were analyzed with or without lipoprotein signal and an N-terminal helix for their subcellular localization using CELLO v 2.5 (Yu et al. 2004). All periplasm localized proteins then subjected to the functional annotation to identify the families of PBPs using Pfam (Bateman et al. 2004) and InterPro (Hunter et al. 2009). Finally the modeled tertiary structures of putative PBPs were obtained using Phyre2 (Kelley et al. 2015). All the tools/softwares/databases used during this study have been listed in Table 2

basis of functional domains and motifs commonly found in well-known PBPs.

Results

Classically secreted proteins of *Mtb*

The PBPs are localized in periplasm and hence they must be targeted to periplasm from cytoplasm after being synthesized. To identify the whole secreted proteome, the complete proteome dataset of the *Mtb* H37Rv was subjected to signal peptides prediction using SignalP program (Nielsen et al. 1997). Out of 4027 protein sequences, we have obtained 256 proteins (Supplementary Table 1) with signal peptides and 3771 proteins were identified which were not containing any classical signal peptides.

Non-classically secreted proteins of *Mtb*

3771 proteins from *Mtb* proteome were lacking classical signal peptides and these may include some secretory

proteins which are transported as folded proteins or with the help of some other non-classical mechanism. Therefore, non-classically secreted proteins were identified using SecretomeP program (Bendtsen et al. 2004). 738 proteins (Supplementary Table 2) were identified with a threshold score >0.5 and were annotated as non-classically secreted proteins.

Classically and non-classically secreted lipoproteins as PBPs

Many lipoproteins also work as PBPs in archea and Gram-positive bacteria (Hutchings et al. 2009), therefore all classically and non-classically secreted proteins were subjected to lipoprotein signal peptides prediction using LipoP 1.0 program (Juncker et al. 2003). Out of classically secreted 256 proteins, 51 proteins were found to have lipoprotein signal peptides, whereas in the case of 738 non-classically secreted proteins, we have obtained only 11 proteins with lipoprotein signal peptides.

N-terminal single transmembrane helix containing PBPs

After lipoprotein signal peptides' prediction, all classically and non-classically secreted proteins were subjected to a single N-terminal TM helix prediction to identify the proteins which may be potentially anchored to the inner membrane of the bacteria. We have used TMHMM program for the prediction of the TM helix (Krogh et al. 2001). In the case of 51 classically secreted lipoproteins, we have identified 10 proteins with an N-terminal TM helix and 41 lipoproteins were observed without any N-terminal TM-helix. Out of 205 classically secreted proteins without lipoprotein signal peptides, 75 proteins were carrying an N-terminal helix, whereas rest of the 130 proteins were lacking an N-terminal helix.

In the case of non-classically secreted 11 lipoproteins, one protein was observed to carry an N-terminal TM helix and the rest of the 10 lipoproteins were lacking an N-terminal helix. 70 non-classically secreted proteins without lipoprotein signal peptides were identified with an N-terminal TM helix. Remaining 657 non-classically secreted proteins were identified in *Mtb* proteome, which were devoid of any type of membrane anchor.

Subcellular localization of classically and non-classically secreted proteins

We have found PBPs in *Mtb* proteome and these proteins are localized in periplasm. To investigate the subcellular localization of the screened proteins with and without signal peptides, lipoprotein signal peptides, and an N-terminal

TM helix, we have analyzed all classically and non-classically secreted proteins using CELLO subcellular localization program (Yu et al. 2004). From all classically secreted proteins, we have predicted five proteins localized in periplasm carrying lipoprotein signal peptides as well as an N-terminal transmembrane helix (Supplementary Table 3). 22 classically secreted periplasmic proteins were identified only with lipoprotein signal peptides (Supplementary Table 4). 55 proteins with an N-terminal TM-helix were found putatively localized in periplasmic space (Supplementary Table 5). 51 putative periplasmic proteins were classically secreted and any N-terminal membrane anchor was absent in their protein sequences (Supplementary Table 6).

Non-classically secreted proteins were containing one putative periplasmic protein with lipoprotein signal peptide and an N-terminal TM helix (Supplementary Table 3). Seven proteins were containing only lipoprotein signals found to be putatively localized in the periplasm (Supplementary Table 7). 39 non-classically secreted proteins with only single N-terminal TM helix were observed to be localized in periplasm (Supplementary Table 8). 215 non-classically secreted proteins were showing periplasmic localization and lacking lipoprotein signals or an N-terminal TM helix (Supplementary Table 9).

Functional domains and structural features of putative PBPs

All the potential classical and non-classically secreted protein sequences with periplasmic localization, with or without lipoprotein signal peptides and an N-terminal TM helix, were analyzed for the presence of functional domains and motifs with the help of Inter pro (Hunter et al. 2009) and Pfam (Bateman et al. 2004). We have identified 37 putative PBPs related to 15 different known classes with divergent solute-binding properties (Table 1) (Supplementary Fig. 1). These proteins from each family were modeled to obtain their tertiary structure using Phyre2 (Supplementary Tables 10, 11). The description of these families is as follows:

1. *MlaD family* The members of this protein family are involved in Mla pathway. They were named on the basis of their functions, i.e., maintenance of lipid asymmetry. MlaD is one of the reported membrane attached PBP with uncleaved signal peptides and involved in Mla pathway (Linton and Higgins 1998; Malinverni and Silhavy 2009). We have obtained ten putative PBPs from *Mtb* proteome out of which two proteins, Rv1969 and Rv0592, were classically secreted proteins with single N-terminal TM-helix. The rest of the eight proteins (Rv3494c, Rv1966, Rv1968, Rv0591, Rv0594,

Table 1 List of the identified periplasmic-binding proteins in *Mtb* proteome with their annotation descriptors

S. no.	GenBank sequence ID/Rv number	Gene name	Pfam (domain ID/domain name)	Interpro (domain ID/domain name)	Description	Mycobacterial orthologues	Functional annotations/localization/PDB ID	References
Classically secreted PBPs with lipoprotein peptide signal and N-terminal helix								
1.	Rv3666c (KBJ24735.1)	*dppA*	PF00496 (SBP bac 5)	IPR030678 (peptide/nickel binding protein)	Bacterial extracellular solute-binding proteins, family 5 middle	MT3767 (*Mtb* CDC1551), Mb3690c (*M. bovis*), ML2293 (*M. leprae*), MMAR_5154 (*M. maritnum*)	Dipeptide transporters	Braibant et al. (2000)
Classically secreted PBPs with lipoprotein peptide signal and without N-terminal helix								
2.	Rv3044 (KBJ29088.1)	*FecB*	PF01497 (Peripla BP 2)	IPR002491 (ABC transporter periplasmic binding protein)	Periplasmic binding protein	MT3129 (*Mtb* CDC1551), Mb3070 (*M. bovis*), ML1729 (*M. leprae*), MMAR_1650 (*M. marinum*), MSMEG_2319 (*M. smegmatis*)	Probable FEIII-dicitrate-binding periplasmic lipoprotein/exported	Målen et al. (2007), Prakash et al. (2005)
3.	Rv2864c (KBJ30253.1)	*Rv2864c*	PF05223 (MecA N), PF00905 (transpeptidase)	IPR007887 (Nuclear transport), IPR001460 (factor 2-like N-terminal transpeptidase)	NTF2-like N-terminal transpeptidase domain. Penicillin binding protein transpeptidase domain	MT2933 (*Mtb* CDC1551), Mb2889c (*M. bovis*), ML1577 (*M. leprae*)	Unknown function	Identified in this study
4.	Rv1166 (KBJ36413.1)	*Rv1166*	PF00496 (SBP bac 5)	IPR000914 (SBP bac 5)	Bacterial extracellular solute-binding proteins, family 5 middle	MT1203 (*Mtb* CDC1551), Mb1198 (*M. bovis*), ML1497 (*M. leprae*), MMAR_4288 (*M. marinum*) MSMEG_5130 (*M. smegmatis*)	Unknown function/exported	Målen et al. (2007)
5.	Rv0934 (KBJ38696.1)	*pstS1*	PF12849 (PBP like 2)	IPR005673 (ABC transporter, phosphate-binding protein)	PBP superfamily domain	MT0961 (*Mtb* CDC1551), Mb0959 (*M. bovis*), ML0740 (*M. leprae*)	Exported/PDB ID:1PC3	Målen et al. (2007), Vyas et al. (2003)
Classically secreted PBPs with N-terminal helix								
6.	Rv3682 (KBJ24753.1)	*ponA2*	PF00912 (transgly), PF00905 (trans-peptidase), PF03793 (PASTA)	IPR023346 (lysozyme like domain), IPR001264 (glycosyl trasferase), IPR012338 (beta-lactamase/transpeptidase-like), IPR001460 (penicillin-binding protein and serine/threonine kinase associated domain, IPR005543 (PASTA domain)	Transglycosylase. Penicillin binding protein transpeptidase domain. PASTA domain	MT3784 (*Mtb* CDC1551), Mb3707 (*M. bovis*), ML2308 (*M. leprae*), MMAR_5171 (*M. marinum*) MSMEG_6201 (*M. smegmatis*)	Penicillin binding protein/exported	Billman-Jacobe et al. (1999), Målen et al. (2007)
7.	Rv3194c (KBJ28233.1)	*Rv3194c*	PF13180 (PDZ 2), PF05362 (Lon C)	IPR027065 (Lon protease (S 16 peptidase), IPR001478 (PDZ domain), IPR020568 (ribosomal protein), IPR014721 (S5 domain 2-type fold), IPR008269 (peptidase S16)	PDZ domain. Lon protease (S16) C-terminal proteolytic domain	MT3286 (*Mtb* CDC1551), Mb3216c (*M. bovis*), ML0643 (*M. leprae*), MMAR_1370 (*M. marinum*) MSMEG_1958 (*M. smegmatis*)	Unknown function/exported	Målen et al. (2007)
8.	Rv1969 (KBJ33109.1)	*mce3D*	PF02470 (MlaD)	IPR005693 (virulence factor Mce protein), IPR003399 (MlaD protein from actinomycocetes)	MlaD protein	MT2021 (*Mtb* CDC1551), MMAR_2885 (*M. marinum*) MSMEG_0349 (*M. smegmatis*)	Unknown function	Identified in this study

Table 1 continued

S. no.	GenBank sequence ID/Rv number	Gene name	Pfam (domain ID/domain name)	Interpro (domain ID/domain name)	Description	Mycobacterial orthologues	Functional annotations/localization/PDB ID	References
9.	Rv0592 (KBJ39214.1)	mce2D	PF02470 (MlaD)	IPR005693 (virulence factor Mce protein), IPR003399 (MlaD protein from actinomycocetes)	MlaD protein	MT0622 (*Mtb* CDC1551), Mb0607, Mb0608 (*M. bovis*), MSMEG_0137 (*M. smegmatis*)	Unknown function	Identified in this study
10.	Rv0016c (KBJ41119.1)	pbpA	PF00905 (transpeptidase)	IPR012338 (beta-lactamase/transpeptidase-like), IPR001460 (penicillin-binding protein, transpeptidase)	Penicillin binding protein transpeptidase domain	MT0019 (*Mtb* CDC1551), Mb0016c (*M. bovis*), ML0018 (*M. leprae*), MMAR_0018 (*M. marinum*), MSMEG_0031 (*M. smegmatis*)	Probable penicillin binding protein/PDB ID: 3LO7	Fedarovich et al. (2010)
11.	Rv0203 (KBJ41312.1)	Rv0203	PF16525 (MHB)	IPR030937 (haemophore Rv0203), IPR032407 (haempophore, haem binding)	Haemophore, haem-binding. haemophore Rv0203	MT0213 (*Mtb* CDC1551), Mb0209 (*M. bovis*), MMAR_0443 (*M. marinum*), MSMEG_0242 (*M. smegmatis*)	Exported/PDB ID: 3MAY	Mâlen et al. (2007), Tullius et al. (2011)
Classically secreted PBPs without lipoprotein signal peptides and N-terminal helix								
12.	Rv2041c (KBJ33186.1)	Rv2041c	PF01547 (SBP bac 1)	IPR006059 (solute-binding family-1), IPR006311 (twin-arginine translocation pathway)	Bacterial extracellular solute-binding protein	MT2101 (*Mtb* CDC1551), Mb2067c (*M. bovis*), ML1427 (*M. leprae*), MMAR_3014 (*M. marinum*)	Probable sugar binding lipoprotein	Titgemeyer et al. (2007)
13.	Rv1235 (KBJ36248.1)	lpqY	PF01547 (SBP bac 1)	IPR006059 (solute-binding family 1)	Bacterial extracellular solute-binding protein	MT1273 (*Mtb* CDC1551), Mb1267 (*M. bovis*), ML1086 (*M. leprae*), MMAR_4206 (*M. marinum*), MSMEG_5061 (*M. smegmatis*)	Probable sugar binding lipoprotein	Titgemeyer et al. (2007)
Non-classically secreted PBPs with lipoprotein signal peptide								
14.	Rv1911c (KBJ34145.1)	lppC	PF01161 (PEBP)	IPR008914 (phosphatidylethanolamine-binding protein), IPR005247 (YbhB/YbcL)	Phosphatidylethanolamine-binding protein	MT1962 (*Mtb* CDC1551) Mb1946c (*M. bovis*) MSMEG_6851 (*M. smegmatis*)	Probable lipoprotein/exported	Mâlen et al. (2007)
Non-classically secreted PBPs with N-terminal helix								
15.	Rv1910c (KBJ34144.1)	Rv1910c	PF01161 (PEBP)	IPR008914 (phosphatidylethanolamine-binding protein)	Phosphatidylethanolamine-binding protein	MT1961 (*Mtb* CDC1551) Mb1945c (*M. bovis*) MSMEG_6851 (*M. smegmatis*)	Unknown function/exported	Mâlen et al. (2007)

Table 1 continued

S. no.	GenBank sequence ID/Rv number	Gene name	Pfam (domain ID/domain name)	Interpro (domain ID/domain name)	Description	Mycobacterial orthologues	Functional annotations/localization/PDB ID	References
16.	Rv2163c (KBJ32732.1)	phpB	PF03717 (PBP_dimer), PF00905 (transpeptidase)	IPR005311 (penicillin-binding protein, dimerisation domain), IPR012338 (beta-lactamase/transpeptidase-like)	Penicillin-binding protein dimerisation domain. Penicillin binding protein transpeptidase domain	MT2221 (Mtb CDC1551) Mb2187c (M. bovis) ML0908 (M. leprae) MMAR_3200 (M. marinum) MSMEG_4233 (M. smegmatis)	Probable penicillin-binding membrane protein	Datta et al. (2006)
17.	Rv2318 (KBJ32435.1)	UspC	PF01547 (SBP bac 1)	IPR006059 (solute-binding family 1)	Bacterial extracellular solute-binding protein	MT2381 (Mtb CDC1551) Mb2345 (M. bovis) ML1770 (M. leprae) MMAR_3619 (M. marinum), MSMEG_4468 (M. smegmatis)	Probable periplasmic sugar-binding lipoprotein UspC/PDB ID: 5K2X	Fullam et al. (2016)
18.	Rv3494c (KBJ26634.1)	mce4F	PF02470 (MlaD)	IPR005693 (virulence factor)	MlaD protein	MT3598 (Mtb CDC1551) Mb3524c (M. bovis) MMAR_4982 (M. marinum) MSMEG_5895 (M. smegmatis)	Unknown function	Identified in this study
19.	Rv1966 (KBJ33106.1)	mce3A	PF02470 (MlaD), PF11887 (Mce4 CUP1)	IPR005693 (virulence factor Mce protein), IPR003399 (Mce/MlaD)	MlaD protein. Cholesterol uptake porter CUP1 of Mce4, putative	MT2018 (Mtb CDC1551) MMAR_2882 (M. marinum) MSMEG_0346 (M. smegmatis)	Unknown function	Identified in this study
20.	Rv1968 (KBJ33108.1)	mce3C	PF02470 (MlaD)	IPR005693 (virulence factor Mce protein), IPR003399 (Mce/MlaD)	MlaD protein	MT2020 (Mtb CDC1551) MSMEG_0348 (M. smegmatis)	Unknown function	Identified in this study
21.	Rv0591 (KBJ39213.1)	mce2c	PF02470 (MlaD)	IPR005693 (virulence factor Mce protein), IPR003399 (Mce/MlaD)	MlaD protein	MT0621 (Mtb CDC1551) Mb0606 (M. bovis)	Unknown function	Identified in this study
22.	Rv0594 (KBJ39216.1)	mce2F	PF02470 (MlaD), PF11887 (Mce4 CUP1)	IPR005693 (virulence factor Mce protein), IPR003399 (Mce/MlaD)	MlaD protein. Cholesterol uptake porter CUP1 of Mce4, putative	MT0624 (Mtb CDC1551) Mb0610 (M. bovis)	Unknown function	Identified in this study

Table 1 continued

| S. no.	GenBank sequence ID/Rv number	Gene name	Pfam (domain ID/domain name)	Interpro (domain ID/domain name)	Description	Mycobacterial orthologues	Functional annotations/localization/PDB ID	References
23.	Rv0171 (KBJ41278.1)	*mce1C*	PF02470 (MlaD)	IPR005693 (virulence factor Mce protein), IPR003399 (Mce/MlaD)	MlaD protein	MT0180 (*Mtb* CDC1551) Mb0177 (*M. bovis*), ML2591 (*M. leprae*) MMAR_0414 (*M. marinum*) MSMEG_0136 (*M. smegmatis*)	Unknown function	Identified in this study
24.	Rv0172 (KBJ41279.1)	*mce1D*	PF02470 (MlaD), PF11887 (Mce4 CUP1)	IPR005693 (virulence factor Mce protein), IPR024516 (mammalian cell entry, C-terminal)	MlaD protein. Cholesterol uptake porter CUP1 of Mce4, putative	MT0181 (*Mtb* CDC1551) Mb0178 (*M. bovis*), ML2592 (*M. leprae*) MMAR_0415 (*M. marinum*), MSMEG_0137 (*M. smegmatis*)	Unknown function/exported	Målen et al. (2007)
25.	Rv0174 (KBJ41281.1)	*mce1F*	PF02470 (MlaD)	IPR005693 (virulence factor Mce protein), IPR003399 (Mce/MlaD)	MlaD protein	MT0183 (*Mtb* CDC1551) Mb0180 (*M. bovis*) ML2594 (*M. leprae*) MMAR_0417 (*M. marinum*) MSMEG_0139 (*M. smegmatis*)	Unknown function/exported	Målen et al. (2007)
26.	Rv1782 (KBJ34015.1)	*eccB5*	PF05108 (T7SS ESX1 EccB)	IPR007795 (type VII secretion system EccB)	Type VII secretion system ESX-1, transport TM domain B	MT1832 (*Mtb* CDC1551) Mb1811 (*M. bovis*) ML1544c (*M. leprae*) MMAR_2664 (*M. marinum*)	Type VII secretion system protein	Gey Van Pittius et al. (2001)
27.	Rv3869 (KBJ24947.1)	*eccB1*	PF05108 (T7SS ESX1 EccB)	IPR007795 (type VII secretion system EccB)	Type VII secretion system ESX-1, transport TM domain B	MT3982 (*Mtb* CDC1551) Mb3899 (*M. bovis*) ML0054c (*M. leprae*) MMAR_5444 (*M. marinum*) MSMEG_0060 (*M. smegmatis*)	Type VII secretion system protein/PDB: 4KK7	Gey Van Pittius et al. (2001), Zhang et al. (2015)

Table 1 continued

Non-classically secreted PBPs without lipoprotein signal peptides and N-terminal helices

S. no.	GenBank sequence ID/Rv number	Gene name	Pfam (domain ID/domain name)	Interpro (domain ID/domain name)	Description	Mycobacterial orthologues	Functional annotations/localization/PDB ID	References
28.	Rv2400c (KBJ30724.1)	subI	PF13531 (SBP bac 11)	IPR005669 (thiosulphate/Sulfate-binding protein)	Bacterial extracellular solute-binding protein	MT2471 (*Mtb* CDC1551) Mb2422c (*M. bovis*) ML0615 (*M. leprae*) MMAR_3718 (*M. marinum*), MSMEG_4533 (*M. smegmatis*)	Probable sulfate-binding lipoprotein	Wooff et al. (2002)
29.	Rv2140c (KBJ32708.1)	TB18.6	PF01161 (PEBP)	IPR008914 (PEBP), IPR005247 (YbhB/Ybc)	Phosphatidylethanolamine-binding protein	MT2198 (*Mtb* CDC1551) Mb2164c (*M. bovis*) ML1289 (*M. leprae*), MMAR_3122 (*M. marinum*), MSMEG_4199 (*M. smegmatis*)	Exported/PDB: 4BEG	Eulenburg et al. (2013), Målen et al. (2007)
30.	Rv0050 (KBJ41153.1)	ponA1	PF00912 (transgly), PF00905 (transpepti-dase)	IPR023346 (lysozyme-like domain), IPR001264 (glycosyl transferase, family 51), IPR012338 (beta-lacta-mase/transpeptidase-like), IPR001460 (penicillin-binding proteins)	Transglycosylase. Penicillin binding protein transpeptidase domain	MT0056 (*Mtb* CDC1551) Mb0051 (*M. bovis*) ML2688 (*M. leprae*) MMAR_0069 (*M. marinum*), MSMEG_6900 (*M. smegmatis*)	Probable bifunctional penicillin-binding protein	Bhakta and Basu (2002)
31.	Rv0054 (KBJ41157.1)	ssb	PF00436 (SSB)	IPR000424 (primosome PriB/single strand DNA-binding), IPR012340 (nucleic acid-binding, OB-fold)	Single-strand binding protein family	MT0060 (*Mtb* CDC1551) Mb0055 (*M. bovis*) ML2684 (*M. leprae*), MMAR_0073 (*M. marinum*), MSMEG_6896 (*M. smegmatis*)	Exported/PDB ID: 1UE1	Målen et al. (2007), Sai-krishnan et al. (2003)
32.	Rv3537 (KBJ24605.1)	kstD	PF00890 (FAD bind-ing 2)	IPR023753 (FAD/NAD(P)-binding domain), IPR003953 (FAD binding domain), IPR027477 (succinate dehy-drogenase catalytic domain)	FAD dependent oxidoreductase 2	MT3641 (*Mtb* CDC1551) Mb3567 (*M. bovis*), MSMEG_5941 (*M. smegmatis*)	Probable dehy-drogenase	Kendall et al. (2007)

Table 1 continued

S. no.	GenBank sequence ID/Rv number	Gene name	Pfam (domain ID/domain name)	Interpro (domain ID/domain name)	Description	Mycobacterial orthologues	Functional annotations/localization/PDB ID	References
33.	Rv1817 (KBJ34050.1)	*Rv1817*	PF00890 (FAD binding 2)	IPR023753 (FAD/NAD(P)-binding domain), IPR003953 (FAD binding domain), IPR027477 (succinate dehdrogenase catalytic domain)	FAD binding domain	MT1865 (*Mtb* CDC1551) Mb1848 (*M. bovis*), ML0934 (*M. leprae*), MMAR_2694 (*M. marinum*), MSMEG_3657 (*M. smegmatis*)	Unknown function	Identified in this study
34.	Rv3716c (KBJ24789.1)	*Rv3716c*	PF02575 (YbaB DNA bd)	IPR004401 (nucleoid-associated protein Yba/EbfC family)	YbaB/EbfC DNA-binding family	MT3819 (*Mtb* CDC1551) Mb3743c (*M. bovis*) ML2330 (*M. leprae*) MMAR_5232 (*M. marinum*), MSMEG_6280 (*M. smegmatis*)	Unknown function	Identified in this study
35.	Rv3505* (KBJ26646.1)	*fadE27*	PF00501 (AMP binding)	IPR030310 (long-chain-fatty-acid-CoA ligase FadD17), IPR020845 (AMP binding, conserved site)	AMP-binding enzyme. Long-chain-fatty-acid—coA ligase FaD17	MT3609 (*Mtb* CDC1551) Mb3535 (*M. bovis*) ML0356 (*M. leprae*) MMAR_4993 (*M. marinum*), MSMEG_5907 (*M. smegmatis*)	Probable lipid metabolism associated protein	Kendall et al. (2007)
36.	Rv3409c (KBJ26950.1)	*choD*	PF13450 (NAD binding 8)	IPR023753 (FAD/NAD(P)-binding domain, glucose-methanol-choline oxidoreductase	NAD(P)-binding Rossmann-like domain	MT3517 (*Mtb* CDC1551) Mb3443c (*M. bovis*) ML0389 (*M. leprae*) MMAR_1140 (*M. marinum*), MSMEG_1604 (*M. smegmatis*)	Cholesterol-O2 oxidoreductase	Brzostek et al. (2007)
37.	Rv3413c (KBJ26954.1)	*Rv3413c*	PF16751 (RsdA SigD bd)	IPR031928 (anti-sigma-D factor RsdA, sigma factor-binding region)	Anti-sigma-D factor RsdA to sigma factor binding region	MT3522 (*Mtb* CDC1551) Mb3447c (*M. bovis*) ML0385 (*M. leprae*) MMAR_1135 (*M. marinum*), MSMEG_1600 (*M. smegmatis*)	Exported/PDB ID: 3VEP	Jaiswal et al. (2013), Målen et al. (2007)

* KBJ26646.1 (Rv3505) (fadE27) share homology with the very long chain acyl-CoA synthetase protein (GI: BAA23644) of humans (*Homo sapiens*) with an expect-value of (3e-37)
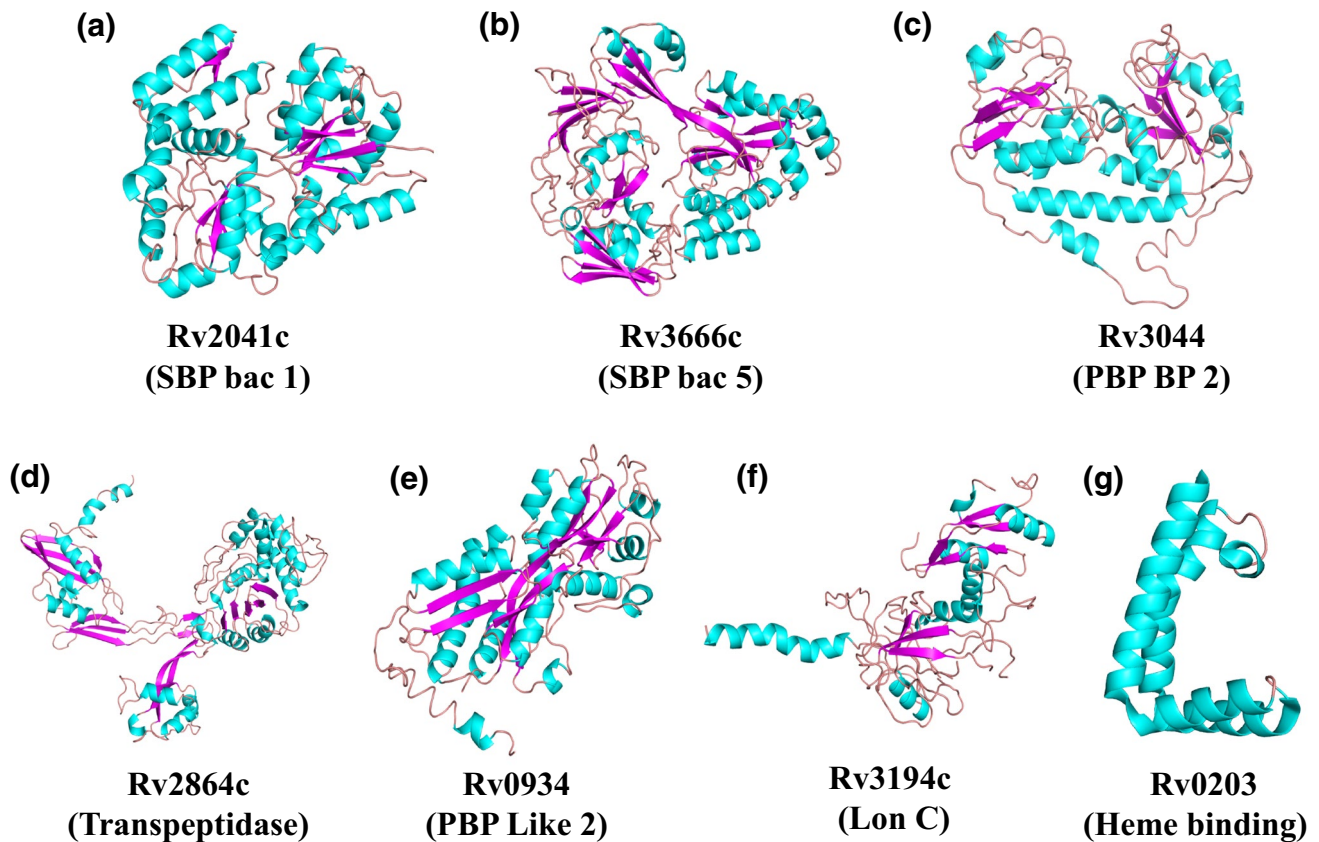
Fig. 2 Representative tertiary structures of classically secreted putative PBPs of *Mtb*: modeled tertiary and available PDB structures of classically secreted putative PBPs shown in figure. All the proteins are shown with *cartoon model* and *colored* according to the secondary structural elements. **a** Rv2041c protein of SBP bac 1 family, **b** Rv3666c protein of family SBP bac 5, **c** Rv3044 protein of PBP 2 family, **d** Rv2864c of transpeptidase family, **e** Rv0934 protein of PBP like 2 protein family, **f** Rv3194c of Lon C protease family, **g** Rv0203 protein of periplasmic heme-binding protein family (PDB ID: 3MAY)

Rv0171, Rv0172, and Rv0174) were non-classically secreted with N-terminal TM-helices. These proteins were homology modeled on the basis of the templates, namely coiled coil methyl accepting chemotaxis protein (PDB ID: 3G67) from *Thermotoga maritime* (Pollard et al. 2009) and serine chemotaxis receptor from *E. coli* (Kim et al. 1999) (Supplementary Tables 10, 11).

2. *Mce4 CUP1 family* We have identified three non-classically secreted putative PBPs (Rv1966, Rv0594 and Rv0172) with single N-terminal helices in *Mtb* proteome. These three proteins were the members of Mce4 CUP1 family and were also carrying MlaD domain (PF11887). These proteins were also homology modeled and have similar structural fold as that of MlaD family proteins (Supplementary Tables 10, 11).

3. *Solute-binding proteins family* We have obtained total six putative PBPs of solute-binding family. Three proteins (Rv2041c, Rv1235, and Rv2318) have SBP bac 1 (PF01547) domain. Two proteins, Rv1166 and

Rv3666c, have SBP bac 5 (PF00496) and one protein, Rv2400c, has SBP bac 11 (PF13531) domain.

One of the identified SBP bac 1 protein, Rv2318, is a well-studied and characterized solute-binding protein, UspC (PDB ID: 5K2X) of *Mtb* (Fullam et al. 2016). Rest of the two proteins, Rv1166 and Rv3666c, of SBP bac 1 family (Fig. 2a) is shown to have homology with the ABC transporter solute-binding protein (PDB ID: 5CI5) (Payne et al. 2016) and maltose binding protein (PDB ID: 1EU8) (Diez et al. 2001) (Supplementary Tables 10, 11).

One SBP bac 5 protein has an N-terminal TM-helix along with the classical signal and lipoprotein signal peptides and another possesses only lipoprotein signal peptide. SBP bac 5 proteins were modeled using template proteins, periplasmic murein tripeptide binding protein, mppA (PDB ID: 3O9P) of *E. coli* (Maqbool et al. 2011) (Fig. 2b) and LpqW (PDB ID: 2GRV) (Marland et al. 2006) lipoprotein from *M. smegmatis* (Supplementary Tables 10, 11).
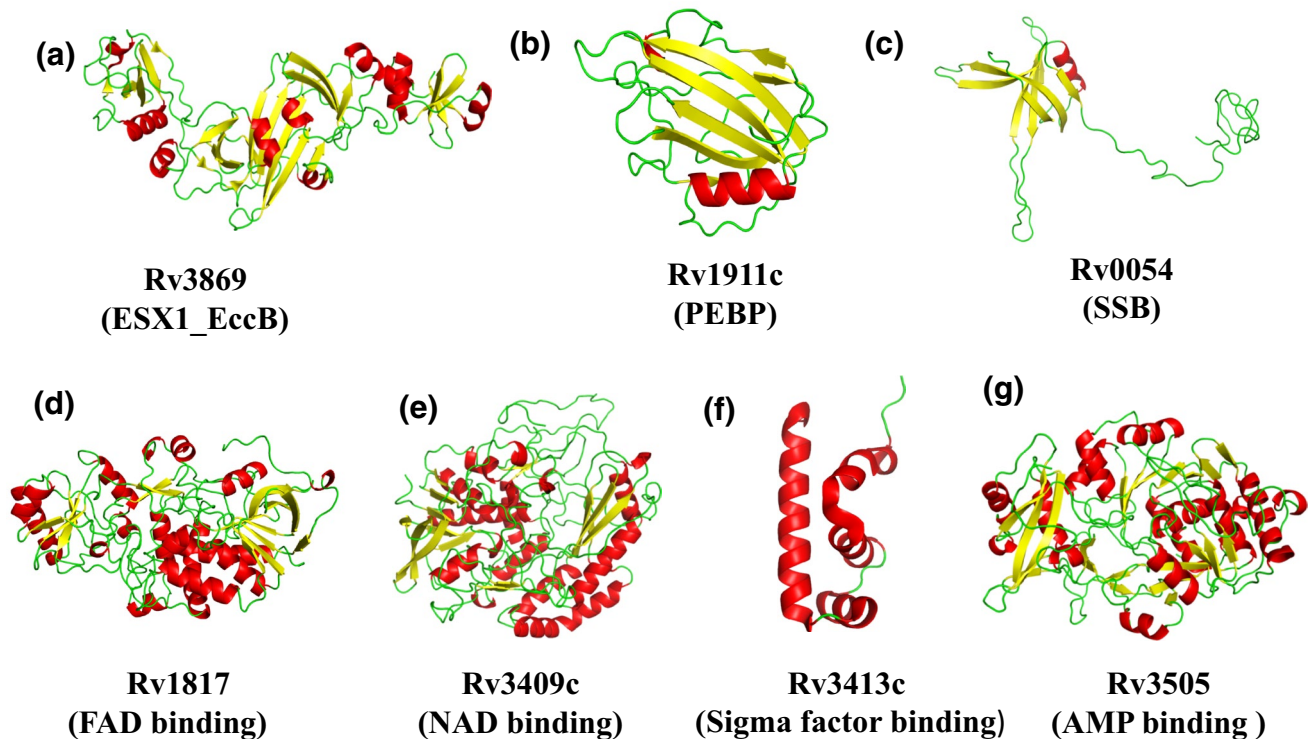
Fig. 3 Representative tertiary structures of non-classically secreted putative PBPs of *Mtb*: modeled tertiary and available PDB structures of non-classically secreted putative PBPs are shown in figure. All the proteins are shown with *cartoon model* and *colored* according to secondary structural elements. **a** Rv3869 protein ESX1_EccB (PDB ID: 4KK7) of ESX1 family, **b** Rv1911c protein of PEBP family, **c** Rv0054 protein of SSB family, **d** Rv1817 of FAD binding family, **e** Rv3409c of NAD-binding family, **f** Rv3413c (PDB ID: 3VEP), anti-sigma factor-binding protein RsdA (**g**) Rv3505 protein of AMP-binding protein family

One protein contained SBP bac 11 functional domain. This protein was non-classically secreted and both lipoprotein signal and N-terminal helix were absent in its sequence. This putative solute-binding protein was modeled using sulfate-binding protein (PDB ID: 1SBP) from *Salmonella enterica* (He and Quiocho 1993) (Supplementary Table 11).

4. *PBP 2 family* This family contains type III PBPs and SBPs. Type III PBPs include periplasmic ferric (Fe^{+3})-dicitrate transporter, FecB from *E. coli* and form substrate-binding part of the ABC transporters involved in Fe^{+3} uptake (Banerjee et al. 2016, Staudenmaier et al. 1989). We have identified one classically secreted putative PBP, Rv3044 of this family. The predicted protein has classical signal peptide along with the lipoprotein signal like all other known members of the PBP 2 family. The proteins share functional domain PF01497 homologous to ABC transporter PBPs (Table 1). The tertiary structure of the protein (Fig. 2c) was modeled using HtsA (PDB ID: 3EIW) lipoprotein of *E.coli* (Beasley et al. 2009) as template (Supplementary Table 10).

5. *Penicillin-binding protein or transpeptidase family and PBP dimer family* This family has proteins containing penicillin-binding domain, hence referred to as penicillin-binding proteins. The topological arrangement of periplasmic-binding proteins shows a cytoplasmic tail, a transmembrane helix with the help of which these proteins remain anchored in the membrane, and the two domains are joined by a linker and are exposed into the periplasm (Sauvage et al. 2008). These proteins have been reported to carry a lipoprotein signal along with a signal peptide which forms a water soluble form of this enzyme (Bowler and Spratt 1989). We have obtained three classically secreted proteins (Rv2864c, Rv3682, and Rv0016c) and two non-classically secreted proteins, Rv0050 and Rv2163c, of this family. These proteins have N-terminal transmembrane helix and possess PBP dimer (PF03717) and transpeptidase (PF00905) domains. Rv0016c protein is characterized by a penicillin-binding protein of the *Mtb* (PDB ID: 3LO7) (Fedarovich et al. 2010). The remaining four proteins were modeled using homology modeling. Tertiary structure of these (Fig. 2d) indicated the homology with
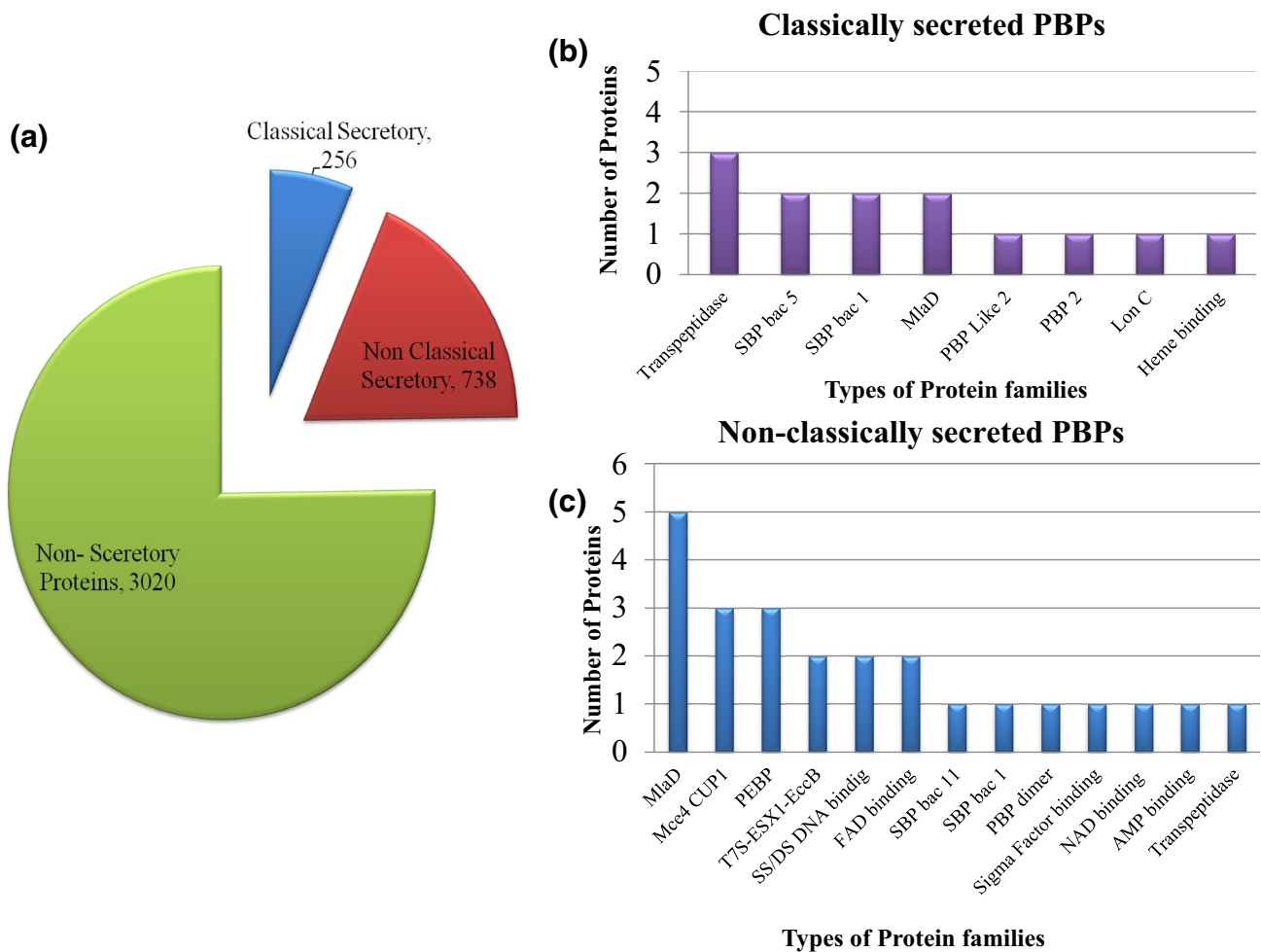
Fig. 4 Proteome-wide distribution of PBPs in *Mtb*: **a** classically secreted 256 proteins, non-classically secreted 738 proteins make the total secretome of *Mtb*. **b** 13 putative classically secreted PBPs were identified in classically secreted periplasmic proteins with 8 different classes. Transpeptidase or penicillin binding proteins have largest share with three members, while SBP bac 5, SBP bac 2 and MlaD families have two members each. PBP like 2, Periplasmic BP 2, Lon C, heme-binding proteins have one member each. **c** 24 Proteins were identified as periplasmic proteins without classical secretory signal peptides. Proteins were distributed among 12 different families of PBPs. MlaD proteins were identified related to mammalian cell entry and virulence followed by phosphatidylethanolamine (lipid) binding protein with three members. Type-VII secretion system (T7S), DNA-binding proteins and FAD binding proteins had two members each. Solute binding protein families (SBP bac 11, SBP bac 1), PBP dimer, Lon C, extracellular sigma factor-binding protein and NAD-binding protein families have one member each identified in *Mtb* proteome

the methicillin acyl-penicillin-binding protein (PDB ID: 1MWU) from *Staphylococcus aureus* (Lim and Strynadka 2002), bifunctional peptidoglycans transferase (PDB ID: 3DWK) from *Staphylococcus aureus* (Lovering et al. 2008) and penicillin-binding protein (PDB ID: from *Pseudomonas aeruginosa* (Sainsbury et al. 2011) (Supplementary Tables 10, 11).

6. *PBP like 2 family* Proteins of this family possess functional domain PF12849 which is designated for proteins which have PBPs like structural and functional features. In *Mtb* proteome, we have identified one classical secretory protein, Rv0934, of this family

(Fig. 2e) which also carries a lipoprotein signal peptide. This protein is a characterized PBP, PstS-1 from *Mtb* (PDB ID: 1PC3) (Supplementary Table 10).

7. *Lon C family* Lon C domain-containing proteins possess functional domain similar to protein-degrading enzyme Lon protease. We have identified one putative PBP, Rv3194c, with the Lon protease domain (PF05362) in *Mtb* proteome. This was identified as classically secreted with an N-terminal TM helix. The protein was modeled (Fig. 2f) using template protein Lon-like protease (PDB ID: 4FW9), an MtaLonC of *Meiothermus taiwanensis* (Supplementary Tables 10,
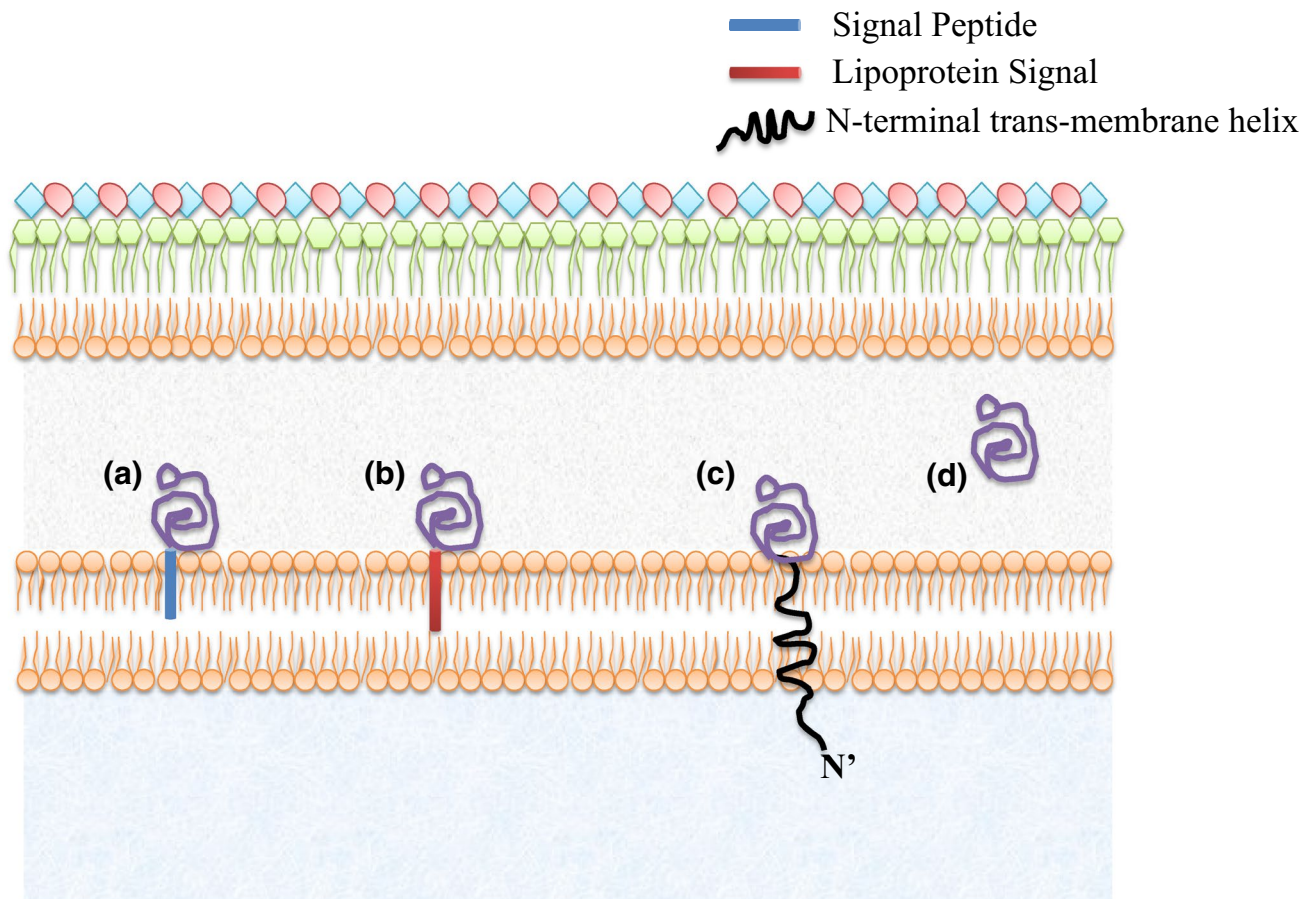
Fig. 5 Different possible modes of membrane anchoring in periplasmic binding proteins: PBPs remain anchored to the cell membrane to prevent their extracellular secretion from periplasm, therefore these proteins use different modes to remain attached to the plasma membrane. *a* PBPs attached to cell membrane via uncleaved signal peptides. *b* PBPs (lipoproteins) attached to membrane with the help of lipoprotein signal peptides. *c* PBPs attached to membrane via an N-terminal TM helices. *d* Proteins residing in periplasm without any membrane anchors

11). These proteins also possess PDZ_2 domain (PF13180) (Table 1).

8. *Heme-binding PBP family* Heme is one of the major iron sources for bacterial pathogens. There are many proteins involved in heme acquisition machinery of the bacteria: hemophores, heme receptors, heme importers, and cytoplasmic heme-utilizing proteins. The hemophore receptor proteins bind to the available heme from the host cell and transport it to the periplasm. In periplasm, periplasmic heme-binding proteins capture heme and transfer it to membrane spanned heme transporters which further transfer it to the bacterial cytoplasm for making it available to cytoplasmic heme utilizing proteins (Kaur et al. 2009). We have found one putative classically secreted periplasmic heme-binding protein, Rv0203, (Fig. 2f) with an N-terminal transmembrane helix in *Mtb* proteome. This protein possesses PF16525 domain which is a characteristic feature of mycobacteria heme-binding protein.

9. *T7SS ESX1 EccB family* Type-VII secretion system (T7SS) is an important non-classical secretory pathway. This pathway has been reported to transport virulence factors and folded proteins across the membrane (Serafini et al. 2013). We have identified two nonclassically secreted putative PBPs with an N-terminal TM helix and possessing T7SS ESX1 EccB domain (PF05108). One of the proteins, Rv3869, has been already reported as membrane-spanned periplasmic protein, $EccB_1$ of *Mtb* (PDB ID: 4KK7). The other protein, Rv1782, of this family was modeled using $EccB_1$ protein of type-VII secretion system (Fig. 3a) (Supplementary Table 11) (Wagner et al. 2016; Zhang et al. 2015).

10. *Phosphatidylethanolamine-binding protein (PEBP) family* These proteins bind lipid headgroups of

the membrane lipid molecules. We have identified three non-classically secreted putative PBPs related to PEBP family. One of the identified PEPB family proteins out of three has lipoprotein signals, one protein has an N-terminal TM-helix and another one was lacking any lipoprotein signals or an N-terminal helix. All the three proteins have PF01161 functional domain homologous to PEBP proteins. One identified member of this family in *Mtb*, Rv2140c, has been documented as a PEBP (PDB ID: 4BEG) (Fig. 3b) (Eulenburg et al. 2013). Rest two proteins were homologues of the Rv2140c protein as observed from their modeled tertiary structures (Supplementary Table 11).

11. *Single strand and double strand DNA-binding protein family* We have identified two non-classically secreted putative PBPs with DNA-binding domains, PF00436 and PF02575. In both these proteins, lipoprotein signal peptides as well as N-terminal helix are absent. One protein out of the two, Rv0054, is a periplasmic single strand binding protein (PDB ID: 1UE1) (Fig. 3c) of *Mtb* (Saikrishnan et al. 2003). The second protein Rv3716 was modeled using DNA-binding protein HP_0035 (PDB ID: 3F42) of *Helicobacter pylori* (Park et al. 2012) (Supplementary Table 11).

12. *FAD (flavin adenine nucleotide) binding protein family* In *Mtb* proteome, we have identified two non-classically secreted FAD-binding putative PBPs, Rv3537 and Rv1817. These proteins have FAD-binding domains (PF00890) and they lack lipoprotein signals as well as an N-terminal TM-helix. These putative FAD-binding proteins were modeled (Fig. 3d) using soluble flavocytochrome C3 fumarate reductase protein (PDB ID: 1QO8) from *Shewanella frigidimarina* (Bamford et al. 1999) as template protein structure (Supplementary Table 11).

13. *NAD-binding protein family* Periplasmic NAD(P)-binding proteins have been reported in some pathogenic bacteria and are required to fulfill NAD requirements of pathogens which lack NAD biosynthetic pathway. In *Mtb* proteome, we have identified one non-classically secreted protein with NAD-binding domain (PF13450) and without membrane anchor. This putative NAD-binding protein Rv3409c was modeled using alcohol oxidase, AOX1 (PDB ID: 5HSA) of *Pichia pastoris* (Supplementary Table 11) (Fig. 3e) (Koch et al. 2016).

14. *RsdA SigD binding family* We have identified one protein from RsdA SigD binding family Rv3413c, which is potentially a non-classically secreted protein without any secretory signals or TM-helix. This protein has a sigma factor (transcription regulatory proteins) binding domain similar to the known anti-sigma factor or sigma factor antagonist proteins (PF16751).

This putative PBP (Fig. 3f) is a well characterized membrane spanning anti-sigma factor RsdA (PDB ID: 3VEP) (Supplementary Table 11).

15. *AMP-binding family* We have identified one non-classically secreted putative AMP-binding protein Rv3505 in *Mtb* proteome. Proteins of this family possess functional domain PF00501 from known AMP-binding enzymes. This protein in *Mtb* lacks any lipoprotein signals and an N-terminal helix anchor. The protein was modeled (Fig. 3g) using non-ribosomal peptide synthetase LgrA (PDB ID: 5ES8) from *Breviabacillus parabrevis* (Reimer et al. 2016) and ATP-binding acetyl-coA synthetase (PDB ID: 1PG4) from *Salmonella enterica* (Gulick et al. 2003) as template structures (Supplementary Table 11).

Conservation of putative PBPs among the mycobacterial genera and non-homology with the human host

The orthologues search for the putative PBPs has shown that all identified putative PBPs of *Mtb* show conservation across the species of mycobacterial genera. The candidature of these proteins as drug targets was further tested by looking for their non-homology with the human host. The BLASTP searches have resulted in "no hits" or no homologous sequence for any of the putative PBPs except for the one, Rv3505 protein which share homology with the very long chain acyl-CoA synthetase protein (GI: BAA23644) of humans (Table 1).

Discussion

The PBPs or SBPs capture and bind ions, solutes, and various other substrates from the periplasm and transfer them to related membrane transporters for their export/import. Sometimes substrate binding by periplasmic proteins is related to signaling mechanism of the bacteria and in few cases with the virulence of the bacteria when these proteins are associated with the peptide transport system (Tam and Saier 1993). With the help of these proteins periplasm act as selective check point for the transfer of material inside and outside the bacterial cells.

In the present study, we have identified putative PBPs in the *Mtb* proteome. Out of 4027 proteins we have obtained 256 classical secretory and 738 non-classical secretory proteins (Fig. 4a). Further, we have shortlisted 133 classically secreted periplasmic proteins and 262 non-classically secreted periplasmic proteins on the basis of their periplasmic localization. Out of 133 classically secreted putative periplasmic proteins, we have screened 13 putative PBPs. From these proteins, one protein was identified to be

carrying both lipoprotein signal peptides and an N-terminal helix, four proteins with lipoprotein signal peptides, six proteins with the single N-terminal helix and two proteins were completely lacking any N-terminal anchor sequences (Fig. 4b).

From 262 non-classically secreted proteins, we have identified 24 putative PBPs. This set of proteins comprises one protein with the lipoprotein signal, 14 proteins with the single N-terminal helix, and 10 proteins were completely lacking any kind of lipoprotein signal peptides or an N-terminal helix, and therefore are devoid of any membrane anchors (Fig. 4c). All of these proteins were categorized into 15 different families on the basis of functional domains. These families include proteins involved in cell signaling, solute binding, and membrane transport by serving as periplasmic piece of ABC transporters, heme-binding and lipid-binding activities (Table 1).

Most of the PBPs or SBPs have requirements to stay at the interface between plasma membrane and the cell wall, so to restrict their secretion into outside environment, these proteins require a kind of anchoring attachment in the form of retained uncleaved classical signal peptides, lipoprotein signal peptides or a membrane spanning N-terminal helix (Fig. 5) (Tjalsma et al. 2000). Mycobacteria are surrounded by double membranous structure comprising plasma membrane, periplasmic space, and an outer membrane similar to Gram-negative bacteria (Daffé and Reyrat 2008). While phylogenetically based on 16s ribosomal RNA sequence analysis, these were reported to be closer to Gram-positive bacteria (Hett and Rubin 2008). To identify the membrane anchored PBPs classical signal peptides, lipoprotein signal peptides, and N-terminal TM helices were predicted. For the prediction of lipoproteins from both the Gram-positive as well as Gram-negative bacteria, LipoP program reportedly has a high accuracy of above 95% (Juncker et al. 2003; Rahman et al. 2008); hence this program was used during the present study.

Substrates across the cell membrane are transported by means of inner-membrane transporters and channels (Chaudhary et al. 2016; Sandhu and Akhter 2016), while the transport across the outer membrane is mediated by outer membrane proteins. There is one more part of transporter proteins involved in periplasmic translocation of various molecules known as adaptor protein or periplasmic efflux proteins. We have identified 8 classically secreted and 12 non-classically secreted PBPs related to ABC transport systems. Ten proteins were related to MlaD/Mce4 CUP1 family (Table 1) (Berks et al. 2000; Kumar et al. 2003). The MlaD domain containing proteins have been reported to constitute the periplasmic part of ABC importer system which carries out the function of phospholipids asymmetry maintenance at the outer membrane (OM) of Gram-negative bacteria by importing surface exposed phospholipids during the stressed conditions and thereby restoring the membrane impermeability (Linton and Higgins 1998; Malinverni and Silhavy 2009). Mammalian cell entry (Mce) CUP1 domain containing proteins constitute parts of ABC transporter systems involved mainly in the steroid uptake as reported in the case of *Rhodococcus jostii* (Mohn et al. 2008). Mce proteins have been reported to facilitate entry of pathogen inside the host cells and are important from virulence perspective. They are encoded by four mce operons in *Mtb*, *mce1-mce4*. *mce4* is mostly observed to express at the later stage of *Mtb* infection. They were reported to have a possible role in long-term persistence of *Mtb* pathogen inside the host tissues as reported for the infected mice (Kumar et al. 2003; Zhang and Xie 2011).

Two other putative PBPs from PBP 2 and PBP like 2 families have ABC transporter periplasmic-binding proteins like domain and may act as periplasmic part of ABC transporter system (Table 1). An identified PBP like 2 protein Rv0934 or PstS-1 has been reported to work as extracellular phosphate receptor of ABC transporter assembly and activate the ESX-5 secretion system in repose to phosphate limiting conditions (Elliott and Tischler 2016). Earlier this protein has been designated as secreted protein antigen-B of *Mtb* and used for the delayed type hypersensitivity test of TB (Vyas et al. 2003).

Six proteins were reported to be related to three different periplasmic solute-binding protein families (SBP bac 1, SBP bac 5, and SBP bac 11) (Table 1). We have identified two classically secreted proteins out of which one protein contains lipoprotein signal peptide and another contains N-terminal TM helix. Both of these are members of SBP bac 5 family. SBP bac 5 proteins have been reported to form periplasmic solute-binding part of the ABC transport systems in many Gram-negative bacteria or reported as membrane-anchored lipoproteins (Saurin and Dassa 1994; Singh and Röhm 2008), therefore we assume that they may be playing similar roles in the case of *Mtb*. These bacterial extracellular solute-binding proteins were reported to be involved in the transport of solutes in periplasm. They were reported to act as the chemoreceptors or recognition components of many ABC transporter systems involved in solute transport across the cellular membrane. These proteins may serve as a trigger to translocation of solutes through the membrane by binding to the outer leaflet of membrane and also perform an important role in the sensory transduction pathways. In the case of Gram-negative bacteria, these proteins are periplasmic soluble proteins but in the case of archea and Gram-positive bacteria, where periplasm is absent in most of the cases, some of the cell membrane-anchored lipoproteins are designated as solute-binding proteins (Saurin and Dassa 1994). These proteins have been classified into eight families on the basis of the nature of the solute they bind (Tam and Saier 1993).

Table 2 List of program/servers/databases used for the genome-wide identification of putative PBPs in *Mtb* proteome

S. no.	Program	Web link	Algorithm	Function	References
1.	SignalP 4.1	http://www.cbs.dtu.dk/services/TMHMM/	HMM	Prediction of classical secretory proteins	Nielsen et al. (1997)
2.	SecretomeP 2.0	http://www.cbs.dtu.dk/services/SecretomeP/	Neural network	Prediction of non-classical secretory proteins	Bendtsen et al. (2004)
3.	LipoP	http://www.cbs.dtu.dk/services/LipoP/	HMM	Prediction of Lipoproteins	Juncker et al. (2003)
4.	TMHMM2.0	http://www.cbs.dtu.dk/services/TMHMM/	HMM	Prediction of single N-terminal TM-helix containing proteins	Krogh et al. (2001)
5.	CELLO v.2.5	http://cello.life.nctu.edu.tw/	SVM	Subcellular localization prediction	Yu et al. (2004)
6.	InterPro	https://www.ebi.ac.uk/interpro/		Prediction of functional domains and families	Hunter et al. (2009)
7.	Pfam	http://pfam.xfam.org/	HMM and ADDA	Prediction of protein domain families	Bateman et al. (2004)
8.	PHYRE2	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index	HMM	Tertiary structure Modeling	Kelley et al. (2015)
9.	NCBI	http://www.ncbi.nlm.nih.gov/	–	Retrieval of proteome data set	NCBI (2013)
10.	PDB	http://www.rcsb.org/	–	Retrieval of crystal structure of characterized PBPs	Bernstein et al. (1977)

Two putative PBPs related to T7S secretion pathway (Table 1) were identified in *Mtb* proteome. Proteins of this family are component of a specialized secretion pathway, type-VII secretion systems (T7S) and carry out the export of virulence factors like ESAT-6 or EsxA (early secreted antigen target) proteins, EsxB (CFP-10) and their partners in mycobacteria (Brodin et al. 2006; Poulsen et al. 2014; Zhang et al. 2015). This secretions system is localized in the RD1 region of *M. tuberculosis* and is absent in avirulent candidate vaccine strains, *M. bovis* BCG and *M. microti*. These proteins constitute a multi-component secretion system and provide a speedy transport of virulence factors across the complex cell envelope of mycobacteria during the infection (Houben et al. 2012).

Six transpeptidase domain containing proteins or penicillin-binding proteins were identified in *Mtb* proteome (Table 1). These proteins play an important role in peptidoglycan polymerization, a principal component of bacterial cell wall, and its insertion into the pre-existing cell wall (Yoshida et al. 2012). Penicillin-binding proteins are important from the virulence prospective of the pathogen as they resist against the β-lactam antibiotics and play an important role in maintaining cell morphology (Guinane et al. 2006). One known penicillin-binding protein PbpA (PDB ID: 3LO7) from *Mtb* was also identified among the putative PBPs. This protein has been reported to be important for the cell division in *M. smegmatis* (Fedarovich et al. 2010).

Two putative periplasmic DNA-binding proteins were identified in the *Mtb* proteome (Table 1). DNA uptake is involved in natural transformation process and is an important part of horizontal gene transfer. Different proteins involved in various steps of DNA uptake in extracellular environment, periplasm, membrane, and cytoplasm have been reported in *Bacillus subtilis*, *Streptococcus pneumonia,* and *Hemophilus influenzeae.* Periplasmic DNA receptor protein has also been reported in *Nisseria gonorrhoeae,* which is involved in DNA transformation process (Jeon and Zhang 2007). However, the DNA transformation process is difficult in *Mycobacterium* genus but, *Mycobcterium smegmatis* has been reported to be capable of plasmid transformation and support their intracellular replication which makes it a model organism to study the DNA transformation process in this genus (Panas et al. 2014). These two PBPs may have a role in DNA transformation and uptake process of the *Mtb.*

Three putative PBPs of PEBP family and one lipid-binding protein were identified in *Mtb* proteome (Table 1). This family has proteins with PEBP (PF01161) domain. The name indicates their affinity for binding lipids (Hengst et al. 2001). The crystal structure of one of the characterized protein of this family from humans (hpEBP) showed a ligand-binding site which has the capability to accommodate the phosphate head group of phospholipids, with the help of which they bind to inner membrane rich in phosphatidylethanolamine. It is assumed that this kind of binding conformation makes it helpful to relay signal across the cytoplasm from the membrane (Serre et al. 1998). These proteins exhibit a wide-range of biological activities including lipid binding, interaction with the cell-signaling machinery, and odorant effectors (Banfield et al. 1998; Książkiewicz et al. 2016). The putative PBP proteins of

PEBP may have cell membrane remodeling-related function in *Mtb*.

RsdA sigma factor-binding PBPs were also identified in *Mtb* (Table 1). Anti-sigma proteins are membrane-spanning proteins with the soluble extracytoplasmic and extraperiplasmic part. These proteins bind the sigma factors to regulate extra cytoplasmic functions which in turn modulate the intracellular transcription in response to diverse environmental signals from outside the bacteria cells (Mettrick and Lamont 2009). These extracytoplasmic sigma factors were reported to be important for bacterial virulence and pathogenesis (Bashyam and Hasnain 2004). The identified PBP of RsdA sigma factor-binding protein (PDB ID: 3VEP) has been reported to play an important role in regulation of the expression profile of *Mtb* in response to external environmental stimuli (Jaiswal et al. 2013). This protein has also been reported to bind periplasmic sigma factors and thereby regulate intracellular transcription (Miot and Betton 2004).

One putative PBP with Lon C domain was also identified (Table 1). Lon C domain-containing proteins are ATP-dependent proteases and utilize serine in their catalytic activity. They are mainly functioning in the selective degradation of abnormal and mutant proteins thus playing a crucial role in cellular cleansing. These proteins are required in certain pathogenic bacteria for rendering pathogenicity and host infectivity (Lee et al. 2006). In *E. coli*, many periplasmic proteases have been reported related to degradation of misfolded proteins in periplasm and thus helping in quality control of bacterial periplasm (Miot and Betton 2004). The *Mtb* putative Lon proteases may be serving the similar function of misfolded protein degradation in periplasmic space. This protein also has PDZ_2 domain (PF13180), which have been reported to anchor receptor proteins to the cytoskeleton of eukaryotic cells (Brakeman et al. 1997).

One AMP-binding protein was also identified in putative PBP (Table 1). The function of this protein in *Mtb* could not be predicted due to unavailability of any AMP-binding periplasmic protein yet, but the AMP-binding domain is shared by many other nucleotide-binding periplasmic proteins. These proteins act as receptors and have been reported to be involved in cell signaling (Devi and McCurdy 1984).

PBPs of *Mtb* also had two non-classically secreted FAD-binding proteins (Table 1). FAD-binding proteins carry out mainly cellular functions and involve in the transfer of electrons to the acceptor molecules. Most of the FAD-binding proteins are membrane-associated cytoplasmic proteins like tricarballylate dehydrogenase (TcuA) and transmembrane proteins like succinate dehydrogenase, but bacterial periplasm was also reported to contain some FAD-binding proteins which are meant to prevent the membrane translocation of substrates before the reduction. Some of these proteins are fumarate reductase enzyme from *Wolinella succinogenes*, methaycrylate reductase from *Geobacter sulfurreducens*, periplasmic aldehyde oxidoreductase YagS in *E. coli,* and membrane-associated FAD-binding periplasmic protein ApbE from *Salmonella enteric* (Boyd et al. 2011; Neumann et al. 2009).

One NAD-binding putative PBP was also identified in *Mtb* proteome (Table 1). Periplasmic NAD(P) binding and degrading protein NadN has been reported in human pathogen *Hemeophilus influenza*e. This protein is specialized in exploiting host NAD(P) by capturing it and then subsequently degrading it into adenosine and NR (nicotinamide riboside). NR is reported to be then transported into cytoplasmic space by specific membrane transporters. NAD is then synthesized in the cytoplasm using this recycled NR (Garavaglia et al. 2012) and in this way the intracellular requirement of NAD(P) can be accomplished by the pathogens at the expense of the host resources.

Classically secreted characterized heme-binding protein Rv0203 (PDB ID: 3MAY) was also identified among *Mtb* putative PBPs (Table 1). These proteins possess a unique heme-binding fold. In mycobacteria, a specialized heme acquisition system has been reported involving secreted Rv0203 (heme carrier protein) protein along with MmpL11 and MmpL3 as transmembrane proteins associated with transport of heme across the inner membrane. Cytoplasmic protein, MhuD (PDB ID: 3HX9), was reported to release iron by breaking down heme in the cytoplasm of *Mtb*. This periplasmic heme-binding protein was reported to capture periplasmic heme in *Mtb* (Tullius et al. 2011). Iron is reported to be crucial for the pathogenesis of *Mtb*, therefore this protein could be essential for the survival of *Mtb* in the host cell (Banerjee et al. 2011).

Bacterial species have reported with diverse kind of heme-binding proteins; but in the present study, we have attempted to identify the substrate-binding proteins in *Mtb* which are potentially localized in the periplasm. The *Mtb* proteome contains approximately 256 classically secreted and 738 non-classically secreted proteins. Out of these proteins, many are inner and outer membrane associated or may be secreted to the extracellular environment. We have obtained 395 potentially periplasm-localized proteins out of all the secreted proteome. This number makes about 10% of the total proteome of the *Mtb*. These periplasm localized proteins may or may not have some substrate-binding role. From this small number of proteins, it is justified to obtain only one or two protein from a particular family of PBPs. In addition to this, presence of high number conserved hypothetical and species-specific proteins is also a drawback for functional annotation of proteins in *Mtb* (Cole 1999). This may be further due to the availability of only few characterized periplasmic heme-binding proteins in the PDB and lack of divergent known structural folds.

All the identified PBPs of *Mtb* share sequence conservation across the species of *Mycobacterium* genera. From the evolutionary studies of drug targets, it has been observed that changing evolutionary rates of drug targets is low in higher percentage of orthologous genes than the non-target genes (Lv et al. 2016). 36 of these 37 identified putative PBPs were found to be non-homologous and unique when compared with human proteome as observed from the BLASTP searches. Thus, the sequence conservation of these PBPs among the mycobacterial species and non-homology with the human proteins is evidence of their eligibility as suitable potential drug targets for designing novel drugs against TB (Melak and Gakkhar 2014).

Out of 37 classically and non-classically secreted putative PBPs identified in the present manuscript, 13 proteins have been already identified as exported proteins by the experimental analysis (Table 1). Four proteins (Rv3044, Rv0934, Rv3682, Rv0054) out of these exported proteins were already annotated as substrate-binding proteins (Målen et al. 2007; Wiker et al. 2000). The identified putative PBPs include seven proteins which have been already well characterized and reported to bind with different substrates in the periplasm (Table 1). Our in silico approach-based genome-wide screening has added 11 new proteins in the list of periplasmic binding proteome with unknown function and potential cellular localization (Table 1). Fourteen proteins with probable functions or experimentally verified exported proteins were reassigned functional classes on the basis of domain identification. Inclusion of all characterized PBPs in the list of identified PBPs validates the accuracy of our in silico approach.

All of the identified putative PBPs have a variety of physiologically important functions as indicated by the domain information and have tertiary structures similar to known PBPs. These proteins form an essential substrate capturing part of many ABC transporters which may be involved in ion-import, heme-import, and may putatively transport many of the biologically important substrates. As many of the PBPs are non-homologous to proteins from the humans, they could be used as potential drug targets (Counago et al. 2012). Some of these proteins are also involved in the sensory system, regulation of the cell wall symmetry or contributing to the virulence of the bacteria (Felder et al. 1999, Rana et al. 2015). These functions can be altered or inhibited by targeting the specific PBPs and may be aimed to target against the *Mtb* in integrative drugs. Lipid-rich thick cellular envelope of *Mtb* makes it difficult to isolate, purify, and crystallize mycobacterial membrane-associated proteins. Due to harsh chemical treatment required to lyse bacterial cell wall and membrane, possibilities of mixing of cell wall components is also a major hurdle in characterizing bacterial cell membrane-associated proteins. It is further to state that recombinant expression and purification for the membrane-associated proteins is technically challenging using the current state-of-the-art technologies. The bioinformatics analyses like presented here may cut down the costly and time consuming hit and trial assays involved in the experimental characterization of these proteins. This study provides a blue print overview of the putative PBPs of *Mtb* and may prove to be an initial step for the in vitro analyses to check the substrate specificity and their roles in the viability of pathogenic mycobacteria.

Materials and methods

Annotated proteome data set of *Mtb* H37Rv

The latest release of complete proteome of *Mtb* H37Rv in the form of a FASTA file (GCA_000667805.1) containing 4027 protein sequences was retrieved from the ftp repository of National Centre for Biotechnology Information (Table 2) (NCBI 2013).

Signal peptide prediction

The signal peptides containing sequences were identified using SignalP 4.1 (Nielsen et al. 1997) (Table 2). We have analyzed the whole proteome of *Mtb* H37Rv using SignalP program which provides batch search for multiple sequences. It is one of the most used signal peptides prediction server currently and predicts signal peptidase I cleavage site in bacterial and eukaryotic protein sequences. The prediction is carried out using different predictors based on neural network. The tool has been trained on positive and negative signal peptides data set from Gram-negative, Gram-positive bacteria, and eukaryotes (Nielsen et al. 1997).

Prediction of non-classical secretory proteins

All the proteins which were not carrying any signal peptide sequence were subjected to SecretomeP 2.0 (Table 2) for the identification of non-classically secreted proteins. These non-classically secreted proteins lack the classical signal peptides which is required for the proteins to be transported across the cell membrane by major Sec transport pathway. Many non-classical transport pathways have been reported for the transport of these proteins. SecretomeP uses artificial neural network constructed on the basis of features present in known non-classically secreted proteins and different from classically secreted proteins. It can distinguish between the two classes (Bendtsen et al. 2004).

Prediction of lipoprotein signal containing sequences

All signal peptides containing and non-containing proteins were subjected to the prediction of lipoprotein signal peptides. LipoP (Table 2) was used for the screening of lipoprotein signal peptides containing sequences. LipoP is based upon hidden Markov model (HMM) based predictor and is able to distinguish between signal peptidase I cleavable and signal peptidase II cleavable proteins. LipoP was reported to identify lipoproteins with over 95% accuracy in Gram-positive as well as in Gram-negative bacteria (Juncker et al. 2003; Rahman et al. 2008).

N-terminal transmembrane helix prediction

All the protein sequences with or without signal peptidase I and signal peptidase II cleavage site were subjected to the N-terminal single transmembrane helix prediction. TMHMM2.0 (Table 2) was used for the prediction of transmembrane helices using HMM (hidden Markov model) algorithm. HMM incorporate hydrophobicity, helix length, and electrical charge into a single model which efficiently results in transmembrane helix prediction. The program efficiently differentiates between soluble and membrane-embedded proteins. The transmembrane helix predictions by TMHMM are reported to have 97–98% accuracy (Krogh et al. 2001).

Subcellular localization prediction

To predict the subcellular localization of targeted proteins, we have used the CELLO v.2.5 subcellular localization prediction server (Table 2). CELLO uses multi class Simple Vector Machine (SVM) based classification system for the prediction of putative periplasmic localization of proteins. The tool takes less computational time and the prediction is carried out on the basis of amino acid composition, the di-peptide composition, the partitioned amino acid composition and the sequence composition based on the physicochemical properties of individual amino acids. The accuracy of the tool is better than all other available tools for the subcellular localization prediction (Yu et al. 2004).

Functional domain prediction

We have predicted the functional domains in the putative periplasmic proteins. Two programs were used for the functional domain predictions, InterPro (Hunter et al. 2009) and Pfam (Bateman et al. 2004) (Table 2). InterPro is a database of signatures related to protein domains, families, and functional sites (Hunter et al. 2009), whereas Pfam is a consortium of protein domains and families (Bateman et al. 2004).

Pathogenic specificity and drug target suitability identification

Orthologues of all putative PBPs were searched across the genomes of species from mycobacterial genera using the tuberculist database (Lew et al. 2011). To predict the suitability of putative PBPs as drug target, the identified 37 putative PBPs were scanned for the homologs in human (*Homo sapiens*) proteome (taxid: 9606). The proteins were subjected individually to BLASTP (Mahram and Herbordt 2010) search using a cut-off value of 0.0001. The criteria of "no hits found" was set to consider the proteins as potential drug targets (Vetrivel et al. 2011).

Tertiary structure modeling

The tertiary structure modeling of putative PBPs from *Mtb* was carried out using homology modeling program PHYRE2 (Protein Homology/Analogy Recognition Engine) (Table 2). It is a homology modeling method which compares the query sequences using multiple alignments with the protein sequences of already known tertiary structures. The program is able to model the protein sequences with remote homology and sequence identity >15% with the known sequences from database (Kelley et al. 2015).

Compliance with ethical standards

References

Bamford V, Dobbin PS, Richardson DJ, Hemmings AM (1999) Open conformation of a flavocytochrome c3 fumarate reductase. Nat Struct Mol Biol 6:1104–1107

Banerjee S, Farhana A, Ehtesham NZ, Hasnain SE (2011) Iron acquisition, assimilation and regulation in mycobacteria. Infect Genet Evol 11:825–838

Banerjee S, Paul S, Nguyen LT, Chu BC, Vogel HJ (2016) FecB, a periplasmic ferric-citrate transporter from *E. coli*, can bind different forms of ferric-citrate as well as a wide variety of metal-free and metal-loaded tricarboxylic acids. Metallomics Integr Biometal Sci 8:125–133

Banfield MJ, Barker JJ, Perry AC, Brady RL (1998) Function from structure? The crystal structure of human phosphatidylethanolamine-binding protein suggests a role in membrane signal transduction. Structure 6:1245–1254

Bashyam MD, Hasnain SE (2004) The extracytoplasmic function sigma factors: role in bacterial pathogenesis. Infect Genet Evol 4:301–308

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL (2004) The Pfam protein families database. Nucleic Acids Res 32:D138–D141

Beasley FC, Vinés ED, Grigg JC, Zheng Q, Liu S, Lajoie GA, Murphy ME, Heinrichs DE (2009) Characterization of staphyloferrin A biosynthetic and transport mutants in *Staphylococcus aureus*. Mol Microbiol 72:947–963

Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S (2004) Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Design Sel PEDS 17:349–356

Berks BC, Sargent F, Palmer T (2000) The Tat protein export pathway. Mol Microbiol 35:260–274

Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) The protein data bank. Eur J Biochem 80:319–324

Berntsson RP-A, Smits SH, Schmitt L, Slotboom D-J, Poolman B (2010) A structural classification of substrate-binding proteins. FEBS Lett 584:2606–2617

Bhakta S, Basu J (2002) Overexpression, purification and biochemical characterization of a class A high-molecular-mass penicillin-binding protein (PBP), PBP1* and its soluble derivative from *Mycobacterium tuberculosis*. Biochem J 361:635–639

Billman-Jacobe H, Haites RE, Coppel RL (1999) Characterization of a *Mycobacterium smegmatis* mutant lacking penicillin binding protein 1. Antimicrob Agents Chemother 43:3011–3013

Borich S, Murray A, Gormley E (1999) Genomic arrangement of a putative operon involved in maltose transport in the *Mycobacterium tuberculosis* complex and *Mycobacterium leprae*. Microbios 102:7–15

Bowler LD, Spratt BG (1989) Membrane topology of penicillin-binding protein 3 of *Escherichia coli*. Mol Microbiol 3:1277–1286

Boyd JM, Endrizzi JA, Hamilton TL, Christopherson MR, Mulder DW, Downs DM, Peters JW (2011) FAD binding by ApbE protein from *Salmonella enterica*: a new class of FAD-binding proteins. J Bacteriol 193:887–895

Braibant M, Gilot P, Content J (2000) The ATP binding cassette (ABC) transport systems of *Mycobacterium tuberculosis*. FEMS Microbiol Rev 24:449–467

Brakeman PR, Lanahan AA, O'Brien R, Roche K, Barnes CA, Huganir RL, Worley PF (1997) Homer: a protein that selectively binds metabotropic glutamate receptors. Nature 386:284–288

Brodin P, Majlessi L, Marsollier L, de Jonge MI, Bottai D, Demangel C, Hinds J, Neyrolles O, Butcher PD, Leclerc C (2006) Dissection of ESAT-6 system 1 of *Mycobacterium tuberculosis* and impact on immunogenicity and virulence. Infect Immun 74:88–98

Brzostek A, Dziadek B, Rumijowska-Galewicz A, Pawelczyk J, Dziadek J (2007) Cholesterol oxidase is required for virulence of *Mycobacterium tuberculosis*. FEMS Microbiol Lett 275:106–112

Chaudhary N, Kumari I, Sandhu P, Ahmed M, Akhter Y (2016) Proteome scale census of major facilitator superfamily transporters in *Trichoderma reesei* using protein sequence and structure based classification enhanced ranking. Gene 585:166–176

Chu BC, Vogel HJ (2011) A structural and functional analysis of type III periplasmic and substrate binding proteins: their role in bacterial siderophore and heme transport. Biol Chem 392:39–52

Cohen KA, El-Hay T, Wyres KL, Weissbrod O, Munsamy V, Yanover C, Aharonov R, Shaham O, Conway TC, Goldschmidt Y, Bishai WR, Pym AS (2016) Paradoxical hypersusceptibility of drug-resistant *Mycobacterium tuberculosis* to beta-lactam antibiotics. EBioMedicine 9:170–179

Cole ST (1999) Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv. FEBS Lett 452:7–10

Counago RM, McDevitt CA, Ween MP, Kobe B (2012) Prokaryotic substrate-binding proteins as targets for antimicrobial therapies. Curr Drug Targets 13:1400–1410

Daffé M, Reyrat J-M (2008) The mycobacterial cell envelope. ASM Press, Washington

Datta P, Dasgupta A, Singh AK, Mukherjee P, Kundu M, Basu J (2006) Interaction between FtsW and penicillin-binding protein 3 (PBP3) directs PBP3 to mid-cell, controls cell septation and mediates the formation of a trimeric complex involving FtsZ, FtsW and PBP3 in mycobacteria. Mol Microbiol 62:1655–1673

Devi A, McCurdy H (1984) Cyclic GMP and cyclic AMP binding proteins in *Myxococcus xanthus*. Microbiology 130:1845–1849

Diez J, Diederichs K, Greller G, Horlacher R, Boos W, Welte W (2001) The crystal structure of a liganded trehalose/maltose-binding protein from the hyperthermophilic archaeon *Thermococcus litoralis* at 1.85 Å. J Mol Biol 305:905–915

Elliott SR, Tischler AD (2016) Phosphate responsive regulation provides insights for ESX-5 function in *Mycobacterium tuberculosis*. Curr Genet 1–5

Eulenburg G, Higman VA, Diehl A, Wilmanns M, Holton SJ (2013) Structural and biochemical characterization of Rv2140c, a phosphatidylethanolamine-binding protein from *Mycobacterium tuberculosis*. FEBS Lett 587:2936–2942

Fedarovich A, Nicholas RA, Davies C (2010) Unusual conformation of the SxN motif in the crystal structure of penicillin-binding protein A from *Mycobacterium tuberculosis*. J Mol Biol 398:54–65

Felder CB, Graul RC, Lee AY, Merkle H-P, Sadee W (1999) The Venus flytrap of periplasmic binding proteins: an ancient protein module present in multiple drug receptors. AAps Pharmsci 1:7–26

Fullam E, Prokes I, Fütterer K, Besra GS (2016) Structural and functional analysis of the solute-binding protein UspC from *Mycobacterium tuberculosis* that is specific for amino sugars. Open Biol 6:160105

Garavaglia S, Bruzzone S, Cassani C, Canella L, Allegrone G, Sturla L, Mannino E, Millo E, De Flora A, Rizzi M (2012) The high-resolution crystal structure of periplasmic *Haemophilus influenzae* NAD nucleotidase reveals a novel enzymatic function of human CD73 related to NAD metabolism. Biochem J 441:131–141

Gey Van Pittius NC, Gamieldien J, Hide W, Brown GD, Siezen RJ, Beyers AD (2001) The ESAT-6 gene cluster of *Mycobacterium tuberculosis* and other high G + C Gram-positive bacteria. Genome Biol 2: RESEARCH0044

Guinane CM, Cotter PD, Ross RP, Hill C (2006) Contribution of penicillin-binding protein homologs to antibiotic resistance, cell morphology, and virulence of Listeria monocytogenes EGDe. Antimicrob Agents Chemother 50:2824–2828

Gulick AM, Starai VJ, Horswill AR, Homick KM, Escalante-Semerena JC (2003) The 1.75 Å crystal structure of acetyl-CoA synthetase bound to adenosine-5′-propylphosphate and coenzyme A. Biochemistry 42:2866–2873

He JJ, Quiocho FA (1993) Dominant role of local dipoles in stabilizing uncompensated charges on a sulfate sequestered in a periplasmic active transport protein. Protein Sci 2:1643–1647

Hengst U, Albrecht H, Hess D, Monard D (2001) The phosphatidylethanolamine-binding protein is the prototype of a novel family of serine protease inhibitors. J Biol Chem 276:535–540

Hett EC, Rubin EJ (2008) Bacterial growth and cell division: a mycobacterial perspective. Microbiol Mol Biol Rev 72:126–156

Higgins CF (2001) ABC transporters: physiology, structure and mechanism—an overview. Res Microbiol 152:205–210

Ho WW, Li H, Eakanunkul S, Tong Y, Wilks A, Guo M, Poulos TL (2007) Holo- and apo-bound structures of bacterial periplasmic heme-binding proteins. J Biol Chem 282:35796–35802

Houben EN, Bestebroer J, Ummels R, Wilson L, Piersma SR, Jiménez CR, Ottenhoff TH, Luirink J, Bitter W (2012) Composition of the type VII secretion system membrane complex. Mol Microbiol 86:472–484

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L (2009) InterPro: the integrative protein signature database. Nucleic Acids Res 37:D211–D215

Hutchings MI, Palmer T, Harrington DJ, Sutcliffe IC (2009) Lipoprotein biogenesis in Gram-positive bacteria: knowing when to hold 'em, knowing when to fold 'em. Trends Microbiol 17:13–21

Jaiswal RK, Prabha TS, Manjeera G, Gopal B (2013) *Mycobacterium tuberculosis* RsdA provides a conformational rationale for selective regulation of sigma-factor activity by proteolysis. Nucleic Acids Res 41:3414–3423

Jeon B, Zhang Q (2007) Cj0011c, a periplasmic single-and double-stranded DNA-binding protein, contributes to natural transformation in *Campylobacter jejuni*. J Bacteriol 189:7399–7407

Jiang D, Zhang Q, Zheng Q, Zhou H, Jin J, Zhou W, Bartlam M, Rao Z (2014) Structural analysis of *Mycobacterium tuberculosis* ATP-binding cassette transporter subunit UgpB reveals specificity for glycerophosphocholine. FEBS J 281:331–341

Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H, Krogh A (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci 12:1652–1662

Kalscheuer R, Weinrick B, Veeraraghavan U, Besra GS, Jacobs WR (2010) Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of *Mycobacterium tuberculosis*. Proc Natl Acad Sci 107:21761–21766

Kaur AP, Lansky IB, Wilks A (2009) The role of the cytoplasmic heme-binding protein (PhuS) of *Pseudomonas aeruginosa* in intracellular heme trafficking and iron homeostasis. J Biol Chem 284:56–66

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858

Kendall SL, Withers M, Soffair CN, Moreland NJ, Gurcha S, Sidders B, Frita R, Ten Bokum A, Besra GS, Lott JS, Stoker NG (2007) A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. Mol Microbiol 65:684–699

Kieser KJ, Rubin EJ (2014) How sisters grow apart: mycobacterial growth and division. Nat Rev Microbiol 12:550–562

Kim KK, Yokota H, Kim S-H (1999) Four-helical-bundle structure of the cytoplasmic domain of a serine chemotaxis receptor. Nature 400:787–792

Koch C, Neumann P, Valerius O, Feussner I, Ficner R (2016) Crystal structure of alcohol oxidase from *Pichia pastoris*. PLoS One 11:e0149846

Krogh A, Larsson B, Von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580

Książkiewicz M, Rychel S, Nelson MN, Wyrwa K, Naganowska B, Wolko B (2016) Expansion of the phosphatidylethanolamine binding protein family in legumes: a case study of *Lupinus angustifolius* L. FLOWERING LOCUS T homologs, LanFTc1 and LanFTc2. BMC Genom 17:820

Kumar A, Bose M, Brahmachari V (2003) Analysis of expression profile of mammalian cell entry (mce) operons of *Mycobacterium tuberculosis*. Infect Immun 71:6083–6087

Lee I, Berdis AJ, Suzuki CK (2006) Recent developments in the mechanistic enzymology of the ATP-dependent Lon protease from *Escherichia coli*: highlights from kinetic studies. Mol Bio-Syst 2:477–483

Lew JM, Kapopoulou A, Jones LM, Cole ST (2011) TubercuList—10 years after. Tuberculosis 91:1–7

Lim D, Strynadka NC (2002) Structural basis for the β lactam resistance of PBP2a from methicillin-resistant *Staphylococcus aureus*. Nat Struct Mol Biol 9:870–876

Linton KJ, Higgins CF (1998) The *Escherichia coli* ATP-binding cassette (ABC) proteins. Mol Microbiol 28:5–13

Lovering AL, De Castro L, Strynadka NC (2008) Identification of dynamic structural motifs involved in peptidoglycan glycosyltransfer. J Mol Biol 383:167–177

Lv W, Xu Y, Guo Y, Yu Z, Feng G, Liu P, Luan M, Zhu H, Liu G, Zhang M (2016) The drug target genes show higher evolutionary conservation than non-target genes. Oncotarget 7:4961

Mahram A, Herbordt MC (2010) Fast and accurate NCBI BLASTP: acceleration with multiphase FPGA-based prefiltering Proceedings of the 24th ACM International Conference on Supercomputing. ACM, p 73–82

Målen H, Berven FS, Fladmark KE, Wiker HG (2007) Comprehensive analysis of exported proteins from *Mycobacterium tuberculosis* H37Rv. Proteomics 7:1702–1718

Malinverni JC, Silhavy TJ (2009) An ABC transport system that maintains lipid asymmetry in the Gram-negative outer membrane. Proc Natl Acad Sci 106:8009–8014

Maqbool A, Levdikov VM, Blagova EV, Hervé M, Horler RS, Wilkinson AJ, Thomas GH (2011) Compensating stereochemical changes allow murein tripeptide to be accommodated in a conventional peptide-binding protein. J Biol Chem 286:31512–31521

Marland Z, Beddoe T, Zaker-Tabrizi L, Lucet IS, Brammananth R, Whisstock JC, Wilce MC, Coppel RL, Crellin PK, Rossjohn J (2006) Hijacking of a substrate-binding protein scaffold for use in mycobacterial cell wall biosynthesis. J Mol Biol 359:983–997

Melak T, Gakkhar S (2014) Potential non homologous protein targets of *Mycobacterium tuberculosis* H37Rv identified from protein-protein interaction network. J Theor Biol 361:152–158

Mettrick KA, Lamont IL (2009) Different roles for anti-sigma factors in siderophore signalling pathways of *Pseudomonas aeruginosa*. Mol Microbiol 74:1257–1271

Miot M, Betton J-M (2004) Protein quality control in the bacterial periplasm. Microb Cell Fact 3:1

Mohn WW, Van Der Geize R, Stewart GR, Okamoto S, Liu J, Dijkhuizen L, Eltis LD (2008) The actinobacterial mce4 locus encodes a steroid transporter. J Biol Chem 283:35368–35374

Mulligan C, Fischer M, Thomas GH (2011) Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. FEMS Microbiol Rev 35:68–86

Ncbi RC (2013) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 41:D8

Neumann M, Mittelstädt G, Iobbi-Nivol C, Saggu M, Lendzian F, Hildebrandt P, Leimkühler S (2009) A periplasmic aldehyde oxidoreductase represents the first molybdopterin cytosine

dinucleotide cofactor containing molybdo-flavoenzyme from *Escherichia coli*. FEBS J 276:2762–2774

Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 10:1–6

Panas MW, Jain P, Yang H, Mitra S, Biswas D, Wattam AR, Letvin NL, Jacobs WR (2014) Noncanonical SMC protein in *Mycobacterium smegmatis* restricts maintenance of *Mycobacterium fortuitum* plasmids. Proc Natl Acad Sci 111:13264–13271

Park SJ, Son WS, Lee B-J (2012) Structural analysis of hypothetical proteins from *Helicobacter pylori*: an approach to estimate functions of unknown or hypothetical proteins. Int J Mol Sci 13:7109–7137

Payne TM, Yee EF, Dzikovski B, Crane BR (2016) Constraints on the radical cation center of cytochrome c peroxidase for electron transfer from cytochrome c. Biochemistry 55(34):4807–4822

Peirs P, Lefevre P, Boarbi S, Wang X-M, Denis O, Braibant M, Pethe K, Locht C, Huygen K, Content J (2005) *Mycobacterium tuberculosis* with disruption in genes encoding the phosphate binding proteins PstS1 and PstS2 is deficient in phosphate uptake and demonstrates reduced in vivo virulence. Infect Immun 73:1898–1902

Pollard AM, Bilwes AM, Crane BR (2009) The structure of a soluble chemoreceptor suggests a mechanism for propagating conformational signals. Biochemistry 48:1936–1944

Poulsen C, Panjikar S, Holton SJ, Wilmanns M, Song Y-H (2014) WXG100 protein superfamily consists of three subfamilies and exhibits an α-helical C-terminal conserved residue pattern. PLoS One 9:e89313

Prakash P, Yellaboina S, Ranjan A, Hasnain SE (2005) Computational prediction and experimental verification of novel IdeR binding sites in the upstream sequences of *Mycobacterium tuberculosis* open reading frames. Bioinformatics (Oxford, England) 21:2161–2166

Rahman O, Cummings SP, Harrington DJ, Sutcliffe IC (2008) Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria. World J Microbiol Biotechnol 24:2377–2382

Rana A, Ahmed M, Rub A, Akhter Y (2015) A tug-of-war between the host and the pathogen generates strategic hotspots for the development of novel therapeutic interventions against infectious diseases. Virulence 6:566–580

Reimer JM, Aloise MN, Harrison PM, Schmeing TM (2016) Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. Nature 529:239–242

Saikrishnan K, Jeyakanthan J, Venkatesh J, Acharya N, Sekar K, Varshney U, Vijayan M (2003) Structure of *Mycobacterium tuberculosis* single-stranded DNA-binding protein. Variability in quaternary structure and its implications. J Mol Biol 331:385–393

Sainsbury S, Bird L, Rao V, Shepherd SM, Stuart DI, Hunter WN, Owens RJ, Ren J (2011) Crystal structures of penicillin-binding protein 3 from *Pseudomonas aeruginosa*: comparison of native and antibiotic-bound forms. J Mol Biol 405:173–184

Sandhu P, Akhter Y (2015) The internal gene duplication and interrupted coding sequences in the MmpL genes of *Mycobacterium tuberculosis*: towards understanding the multidrug transport in an evolutionary perspective. Int J Med Microbiol IJMM 305:413–423

Sandhu P, Akhter Y (2016) The drug binding sites and transport mechanism of the RND pumps from *Mycobacterium tuberculosis*: insights from molecular dynamics simulations. Arch Biochem Biophys 592:38–49

Saurin W, Dassa E (1994) Sequence relationships between integral inner membrane proteins of binding protein-dependent transport systems: evolution by recurrent gene duplications. Protein Sci 3:325–344

Sauvage E, Kerff F, Terrak M, Ayala JA, Charlier P (2008) The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. FEMS Microbiol Rev 32:234–258

Serafini A, Pisu D, Palu G, Rodriguez GM, Manganelli R (2013) The ESX-3 secretion system is necessary for iron and zinc homeostasis in *Mycobacterium tuberculosis*. PLoS One 8:e78351

Serre L, Vallée B, Bureaud N, Schoentgen F, Zelwer C (1998) Crystal structure of the phosphatidylethanolamine-binding protein from bovine brain: a novel structural class of phospholipid-binding proteins. Structure 6:1255–1265

Singh B, Röhm K-H (2008) Characterization of a *Pseudomonas putida* ABC transporter (AatJMQP) required for acidic amino acid uptake: biochemical properties and regulation by the Aau two-component system. Microbiology 154:797–809

Staudenmaier H, Van Hove B, Yaraghi Z, Braun V (1989) Nucleotide sequences of the fecBCDE genes and locations of the proteins suggest a periplasmic-binding-protein-dependent transport mechanism for iron (III) dicitrate in *Escherichia coli*. J Bacteriol 171:2626–2633

Tam R, Saier M (1993) Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. Microbiol Rev 57:320–346

Tasneem A, Iyer LM, Jakobsson E, Aravind L (2004) Identification of the prokaryotic ligand-gated ion channels and their implications for the mechanisms and origins of animal Cys-loop ion channels. Genome Biol 6:1

Titgemeyer F, Amon J, Parche S, Mahfoud M, Bail J, Schlicht M, Rehm N, Hillmann D, Stephan J, Walter B, Burkovski A, Niederweis M (2007) A genomic view of sugar transport in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. J Bacteriol 189:5903–5915

Tjalsma H, Bolhuis A, Jongbloed JD, Bron S, van Dijl JM (2000) Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. Microbiol Mol Biol Rev 64:515–547

Tullius MV, Harmston CA, Owens CP, Chim N, Morse RP, McMath LM, Iniguez A, Kimmey JM, Sawaya MR, Whitelegge JP (2011) Discovery and characterization of a unique mycobacterial heme acquisition system. Proc Natl Acad Sci 108:5051–5056

Vetrivel U, Subramanian G, Dorairaj S (2011) A novel in silico approach to identify potential therapeutic targets in human bacterial pathogens. HUGO J 5:25

Vollmer W, Seligman SJ (2010) Architecture of peptidoglycan: more data and more models. Trends Microbiol 18:59–66

Vyas NK, Vyas MN, Quiocho FA (2003) Crystal structure of *M. tuberculosis* ABC phosphate transport receptor: specificity and charge compensation dominated by ion–dipole interactions. Structure 11:765–774

Wagner JM, Chan S, Evans TJ, Kahng S, Kim J, Arbing MA, Eisenberg D, Korotkov KV (2016) Structures of EccB 1 and EccD 1 from the core complex of the mycobacterial ESX-1 type VII secretion system. BMC Struct Biol 16:1

WHO (2015) Global tuberculosis report 2015. World Health Organization, Geneva

Wiker HG, Wilson MA, Schoolnik GK (2000) Extracytoplasmic proteins of *Mycobacterium tuberculosis*–mature secreted proteins often start with aspartic acid and proline. Microbiology 146:1525–1533

Wooff E, Michell SL, Gordon SV, Chambers MA, Bardarov S, Jacobs WR Jr, Hewinson RG, Wheeler PR (2002) Functional genomics reveals the sole sulphate transporter of the *Mycobacterium tuberculosis* complex and its relevance to the acquisition of sulphur in vivo. Mol Microbiol 43:653–663

Yoshida H, Kawai F, Obayashi E, Akashi S, Roper DI, Tame JR, Park S-Y (2012) Crystal structures of penicillin-binding

protein 3 (PBP3) from methicillin-resistant *Staphylococcus aureus* in the apo and cefotaxime-bound forms. J Mol Biol 423:351–364

Yu CS, Lin CJ, Hwang JK (2004) Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Sci 13:1402–1406

Zhang F, Xie JP (2011) Mammalian cell entry gene family of *Mycobacterium tuberculosis*. Mol Cell Biochem 352:1–10. doi:10.1007/s11010-011-0733-5

Zhang X-L, Li D-F, Fleming J, Wang L-W, Zhou Y, Wang D-C, Zhang X-E, Bi L-J (2015) Core component EccB1 of the *Mycobacterium tuberculosis* type VII secretion system is a periplasmic ATPase. FASEB J 29:4804–4814