Structural basis of glycerophosphodiester recognition by the *Mycobacterium tuberculosis* substrate-binding protein UgpB

Jonathan Fenn[a], Ridvan Nepravishta[b], Collette S Guy[a], James Harrison[a], Jesus Angulo[b], Alexander D Cameron[a], Elizabeth Fullam[a]*

[a] School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK
[b] School of Pharmacy, University of East Anglia, Norwich Research Park, Norwich, Norfolk NR4 7TJ, UK

*To whom correspondence should be addressed: Elizabeth Fullam, School of Life Sciences, University of Warwick, Coventry, CV4 7AL, United Kingdom;
e.fullam@warwick.ac.uk; Tel. +44 (0)2476 574239

Abstract

Mycobacterium tuberculosis (Mtb) is the causative agent of tuberculosis (TB) and has evolved an incredible ability to survive latently within the human host for decades. The *Mtb* pathogen encodes for a low number of ATP-binding cassette (ABC) importers for the acquisition of carbohydrates that may reflect the nutrient poor environment within the host macrophages. *Mtb* UgpB (Rv2833) is the substrate binding domain of the UgpABCE transporter that recognises glycerophosphocholine (GPC), indicating that this transporter has a role in recycling glycerophospholipid metabolites. By using a combination of saturation transfer difference (STD) NMR and X-ray crystallography we report the structural analysis of *Mtb* UgpB complexed with GPC and have identified that *Mtb* UgpB does not only recognise GPC but that it is promiscuous for a broad range of glycerophosphodiesters. Complementary biochemical analyses and site-directed mutagenesis precisely define the molecular basis and specificity of glycerophosphodiester recognition. Our results provide critical insights into the structural and functional role of the *Mtb* UgpB transporter and reveal that the specificity of this ABC-transporter is not limited to GPC therefore optimising the ability of *Mtb* to scavenge scarce nutrients and essential glycerophospholipid metabolites *via* a single transporter during intracellular infection.

Bacterial pathogens have evolved a wide range of strategies to survive and thrive within their host environment. The ability to assimilate nutrients is vital and pathogens have evolved diverse strategies to uptake and scavenge the scarce energy sources that are available to them. In the context of intracellular microbial infections there is growing evidence that in a nutrient limited environment the interplay between the host and the pathogen is important. This is manifested through the ability of bacterial pathogens to utilise discrete nutrient sources with dedicated transport machinery for import. Glycerophosphodiester metabolites that are released by the action of phospholipases on host phospholipids represent an important nutrient source for the supply of carbon and phosphate.

Mycobacterium tuberculosis (Mtb) is a major human pathogen and is now the leading cause of death from a single infectious agent worldwide, resulting in more deaths each year than HIV and malaria combined[1]. *Mtb* is a highly evolved pathogen that is able to persist and survive intracellularly within macrophages for decades[2]. However, the essential nutrients that are available to *Mtb* within the stringent environment of the human host and acquisition systems are poorly understood[3-4]. Understanding the molecular mechanisms that enable *Mtb* to survive within this niche environment and the nutrients that are assimilated is critical to understand this major global pathogen and for the development of new therapeutic approaches.

The sugars that are available within the nutrient-limited macrophage environment are unknown, however *Mtb* is equipped with five putative importers of carbohydrate substrates: four members of the ATP-binding cassette (ABC) transporter family and one belonging to the major facilitator superfamily[3-4]. Until recently the substrates for these transporters were unresolved, however, recent studies have demonstrated a role for the ABC-transporters in the recycling of components from the complex *Mtb* cell wall. Trehalose is recycled from the *Mtb* cell envelope glycolipid trehalose monomycolate and taken up by the LpqY-SugABC transporter, which plays a critical role in the virulence of the *Mtb* pathogen[5]. The *Mtb* UspABC transporter has been found to recognise amino-sugars with a potential role in the uptake of *Mtb* cell-wall peptidoglycan fragments[6].

The role of the UgpABCE ABC-transporter is less clear, however studies of its substrate binding domain *Mtb* UgpB (Rv2833c) indicate its importance for *Mtb* survival and pathogenesis and *in vivo Mtb* UgpB has been found to be upregulated during infection[7]. *Mtb* UgpB has been shown to bind the glycerophosphocholine (GPC) head group of the membrane phospholipid phosphatidylcholine and metabolomic profiling by NMR of intact lung tissue at various stages of *Mtb* infection has revealed that the GPC metabolite increases significantly as infection progresses, with a concomitant decrease in phosphatidylcholine[8]. However, despite the essential role of this *Mtb* transporter, the molecular mechanisms that dictate how GPC is recognised and whether other glycerophosphodiester metabolites are substrates for this ABC-transporter are currently unknown. The only crystal structure of *Mtb* UgpB is of the protein in an open conformation without substrate bound (PDB 4MFI)[9]. Some mechanistic understanding of substrate recognition can be obtained from the crystal structure of a homologue from *E. coli* with low sequence identity (25%) in complex with glycerol-3-phosphate (G3P) (PDB 4AQ4)[10]. However, *Mtb* UgpB does not bind G3P. Comparison of the closed G3P-bound *E.coli* UgpB with

the open *Mtb* UgpB in the absence of substrate (PDB 4MFI) reveals notable differences in the binding sites of these homologous proteins indicating that these UgpB ABC-transporters, belonging within the same structural classification (cluster D)[11], have diverged to have different substrate specificities. This may reflect the nutritional requirements of the specific organism within different host environments and also the ability of bacteria to produce G3P extracellularly through the action of secreted glycerophosphodiesterases that hydrolyse glycerophopshodiesters[12]. Other microorganisms that import GPC have evolved to use either permeases or proton symporters that belong to the major facilitator superfamily indicating that glycerophosphodiester uptake is not limited to ABC-transporters[13-14]. It is likely that the divergence of transport systems for the import of glycerophosphodiesters reflects the evolutionary divergence and intracellular life-style of the pathogen and the metabolites available within its niche environment.

In this study, we report a detailed functional and structural characterisation of the *Mtb* UgpB substrate binding domain of the ABC-transporter using a combination of biochemical and biophysical approaches. We report the first crystal structure of *Mtb* UgpB in complex with GPC and identify, in both solid and solution state, the molecular determinants of binding and critical features for glycerophosphodiester recognition. Structure guided-mutagenesis has revealed the crucial role of binding-site residues that underpin substrate binding and function. Moreover, we show that *Mtb* UgpB has a broad selectivity for glycerophosphodiesters which highlights that the *Mtb* UgpABCE transporter uptakes metabolites derived from various glycerophospholipids. Thus, *Mtb* has evolved to use a broad spectrum of nutrients *via* a single ABC-transporter that enables it to adapt and assimilate essential nutrients during intracellular infection.

RESULTS AND DISCUSSION

Production of *Mtb* UgpB. An *N*-terminal truncated *Mtb* UgpB, corresponding to removal of residues 1-34 predicted to form a trans-membrane anchor-helix, was cloned into the pYUB1062 vector with a *C*-terminal hexa-histidine affinity tag and expressed in *Mycobacterium smegmatis* mc^24517. Soluble *Mtb* UgpB protein was obtained and purified to apparent homogeneity using Co^{2+}-affinity, anion exchange and size-exclusion chromatography (Supplementary Figure S1). The identity of the *Mtb* UgpB protein was confirmed by using in-gel trypsin digestion and analysis of the peptides by mass spectrometry.

Co-crystal structure of *Mtb* UgpB with GPC. Initial attempts to crystallize *Mtb* UgpB in the presence of GPC routinely resulted in crystals of UgpB in an open conformation with no ligand bound. Therefore, to overcome this we chemically modified the surface *Mtb* UgpB through reductively methylation and this resulted in crystals of UgpB in complex with GPC. The UgpB protein co-crystallized with GPC with four molecules in the asymmetric unit. Phases for the structure were determined by molecular replacement using each of the two domains from the *apo*-structure of *Mtb* UgpB (PDB 4MFI) as separate search models and the structure was refined at a resolution of 2.3 Å, to a R_{work} of 20.6 % and R_{free} of 25.6 %, Supplementary Table 1 for the data collection and refinement statistics. Structural superposition of each molecule of *Mtb* UgpB using PDBeFOLD[15] indicates that each molecule within the asymmetric unit is equivalent, aligning with r.m.s.d of 0.35 - 0.44 Å for 394-395 residues. The crystal packing and analysis of the packing interfaces using PDBePISA[16] does not suggest that *Mtb* UgpB forms dimers or higher oligomers and is consistent with our analytical gel filtration studies where the protein behaves as a monomer in solution with an apparent molecular weight of 44 kDa (Supplementary Figure S1D). It is therefore likely that the monomer is the biologically relevant unit, consistent with substrate binding domains of other ABC-transporters[17-18].

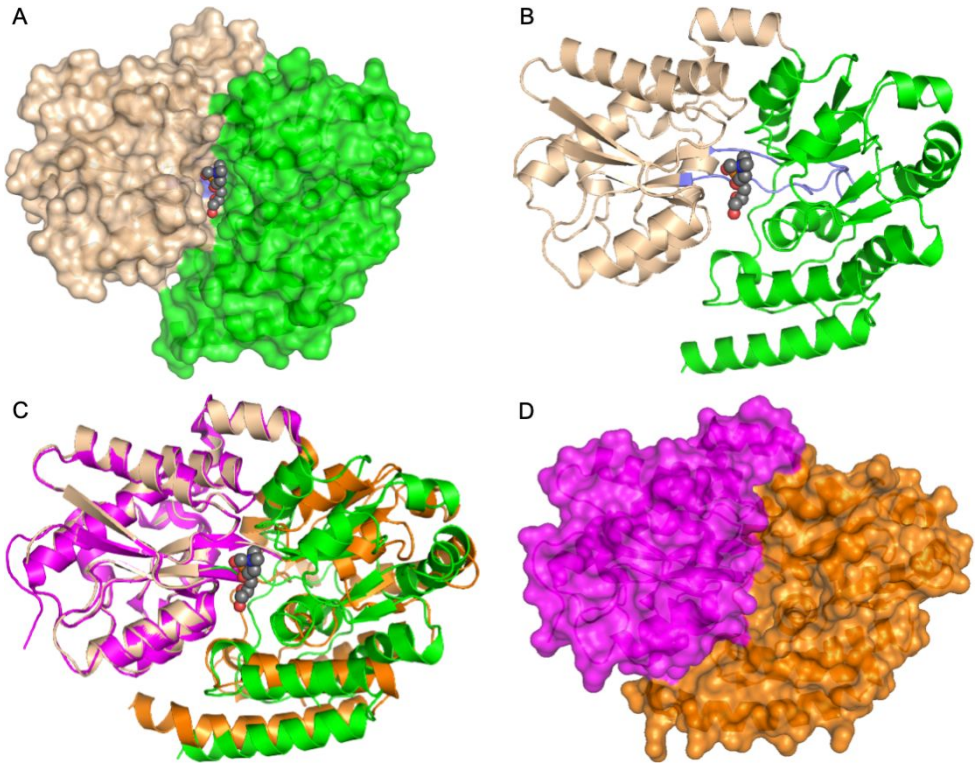

Figure 1. Crystal structure of *Mtb* UgpB. A) Surface representation of *Mtb* UgpB in complex with GPC. The two domains are highlighted, domain I (brown) and domain II (green). The GPC ligand is represented as spheres with dark gray carbon atoms. B) Cartoon representation of *Mtb* UgpB in complex with GPC identifying the secondary structure elements. Domain I (brown), domain II (green) and the two hinge regions are highlighted in blue. The GPC ligand is represented as spheres with dark gray carbon atoms C) Superposition of Domain I of GPC *Mtb* UgpB co-complex (brown/green) with Domain I of apo *Mtb* UgpB (PDB 4MFI) (magenta/orange). D) Surface representation of the unliganded *Mtb* UgpB (PDB 4MFI) with the two domains colored magenta (Domain I) and orange (Domain II).

Overall structure of the *Mtb* UgpB-GPC complex. *Mtb* UgpB comprises two α/β domains (Figure 1). Domain I (residues 1-154 and 307-365) consists of a five-stranded β-sheet surrounded by 11 α-helices and domain II (residues 155-306 and 366-436) of a four-stranded β-sheet enclosed by 9 α-helices. The two domains, or globular lobes, are connected *via* two flexible hinges that are formed between residues Arg152-Pro155 and Ala290-Ala307. Relative to the apo crystal structure there is a 22° rotation of domain I relative to domain II about the interdomain screw-axis with three hinge/binding regions identified from DynDom analysis[19] (residues 152-153, 304-306 and 362-372 (Supplementary Table S2)). This bending movement results in an almost two-fold reduction in the volume of the cavity from 1986 Å^3 to 791 Å^3, as determined by CAVER [20], which is in-line with the 'Venus Fly-trap mechanism' for other substrate-binding proteins[17-18] that close when substrate is bound. Interdomain bridging and stabilisation of this closed conformation of the protein is centred around Arg385, which forms interdomain hydrogen bonds with Asp102 from domain I and Gln381 from domain II. The individual domains of *Mtb* UgpB *apo*- and GPC co-complex structures align with r.m.s.ds of 0.57 Å and 0.75 Å for domains I and II respectively (over 178 atoms, Domain I and over 216 atoms, Domain II, PDBeFOLD[16]). In comparison, superposition of *Mtb* UgpB *apo*- and GPC co-complex structures align with a r.m.s.d. of 2.2 Å (over 385 residues) highlighting the importance of an interdomain conformational change mechanism for substrate recognition by *Mtb* UgpB.

The ligand-binding site of *Mtb* UgpB. Well defined electron density for the GPC ligand in all *Mtb* UgpB molecules within the crystal unit was observed enabling the GPC ligand to be modelled in the *Mtb* UgpB binding-site (Supplementary Figure S2A). The GPC ligand is found in an identical position and orientation in each subunit (Supplementary Figure S2B). Notably, the electrostatic surface shows that GPC is buried in the prominent, acidic interface that is formed between the two domains of UgpB and makes contact to both. The GPC is precisely orientated within the binding cleft such that the glycerol moiety is buried at the base of the cavity, in close proximity to the flexible hinge region centred around Arg385, whilst the choline moiety extends outwards towards the solvent exposed channel entrance (Figure 2).

The glycerol moiety is located between the side chains of Leu205 and Trp208 from domain II (Figure 2). The ring system of Trp208 lies approximately parallel to the C1, C2 and 2-hydroxy group of the glycerol moiety enabling π-stacking interactions, whilst Leu205 is orientated perpendicular to this plane and provides additional stabilisation. There is an important network of hydrogen bonding interactions that anchors GPC in the binding-pocket. The side chain of Asp102, from domain I, is orientated to enable direct hydrogen bonding to both the 1- and 2-hydroxy groups of the glycerol moiety. Two residues that comprise the flexible hinge-linkages are able to directly interact with GPC through the formation of additional hydrogen bond interactions between the side chain of Arg385 and the 1-hydroxy group and the backbone amide nitrogen atom of Gly306 with the 2-hydroxy group respectively. The direct interaction of these flexible-hinge linkages with the GPC ligand may help to stabilise the UgpB-GPC complex in the closed conformation. The phosphate group of GPC is stabilised through hydrogen bond interactions with the side chains of Tyr78 and Tyr345 (domain I), Ser153 (domain I), Ser272 (domain II) and the backbone amide of Gly306. It is striking that there are no direct or charged interactions between *Mtb* UgpB and the positively charged choline moiety, though this moiety is well-defined in the electron density.

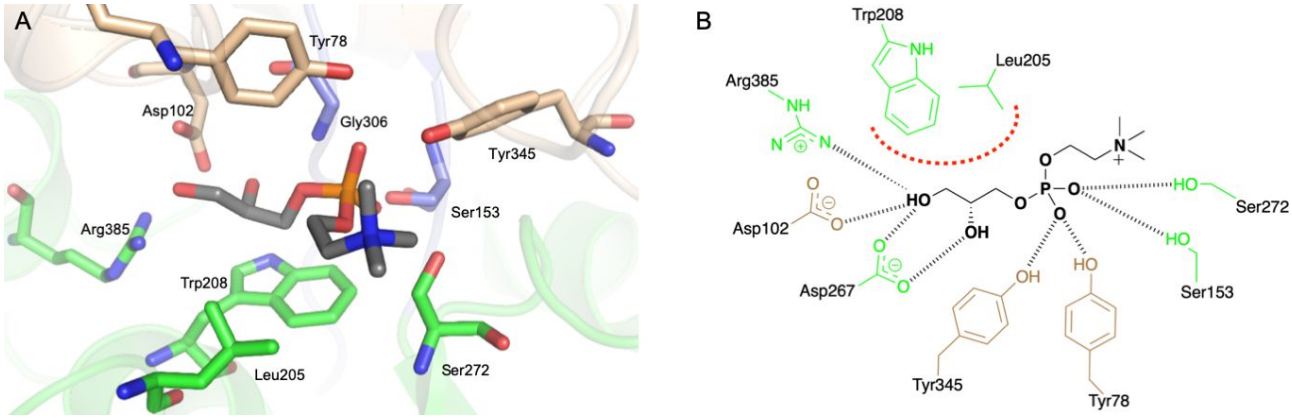

Figure 2. The GPC binding site in *Mtb* UgpB. A) Illustration showing GPC with dark gray carbon atoms and selected *Mtb* UgpB amino acid residues in stick representation (coloured brown for residues within Domain I, and green for residues with Domain II. B) Schematic diagram of the interactions of *Mtb* UgpB with GPC. Dashed lines (black) represent hydrogen bonding, thick dotted line (red) represents hydrophobic interactions

Comparison with the binding site of *E. coli* UgpB. Comparison with UgpB from *E coli*[10] indicates that the overall architecture of these two periplasmic binding proteins in complex with substrate is similar, with a r.m.s.d. of 2.1 Å (PDBeFold[15], target residues: 394, sequence identity 25 % (Supplementary Figure S3), PDB code 4AQ4), Fig. 3. Whilst *Mtb* was crystallised with GPC, the *E coli* protein was crystallised with G3P that we, as well as previous studies[10], show does not bind to *Mtb* UgpB. It is interesting to note that the binding mode of the G3P core of GPC resembles the situation found in the *E. coli* UgpB-G3P complex[10], even though *Mtb* UgpB is unable to bind or recognise this smaller G3P ligand (Fig. 3B). However, whilst the substrate binding pocket of *Mtb* UpgB resembles that of *E. coli* UgpB there are several important differences. Notably, there are substitutions of critical residues involved in substrate binding. Leu205 is specific to *Mtb* and is replaced by a larger indole-side chain from a tryptophan residue (Trp169) in *E. coli* UgpB. In addition, *Mtb* UgpB Asp102 is replaced in *E. coli* UgpB by a glutamic acid residue (Glu66) (Fig. 3C). In this instance, the difference in the length of these acidic side-chains may influence substrate selectivity between the different organisms. Intriguingly, whilst the interaction with an arginine residue is conserved between *Mtb* and *E. coli* the arginine residues in the two proteins originate from different regions of the protein indicating an evolutionary divergence of these substrate-binding proteins. In addition, a narrowing of the *E. coli* UgpB binding cleft results from two different loop regions. One loop region (Gly221-Asp230) in domain II of *E. coli* UgpB linking α-helices 10 and 11 narrows the substrate binding cavity as a result of a 5 Å translational shift. The difference in position of a second loop comprised of residues His8-Gly12 results in the translation of the first α-helix of *E. coli* UgpB (residues 12-30) located in domain I by approximately 6 Å towards α-helix 11 of domain II which further narrows the *E. coli* UgpB substrate binding channel (Fig. 3D/E). Comparison of the region at the entrance to the binding cleft reveals an expanded pocket for *Mtb* UgpB. It is of interest to note that in chain B of *Mtb* UgpB we observe an additional glycerol molecule located in this expanded pocket that is within 4 Å of the choline moiety of GPC (Supplementary Figure S4). A glycerol molecule is also present in the *E. coli* UgpB-G3P complex, though at a different position, indicating that for both proteins the binding pockets are larger than the recognised GPC substrate[10]. This may be functionally significant in substrate

recognition and have an important role in the accommodation and binding of alternative phosphodiester substrates.

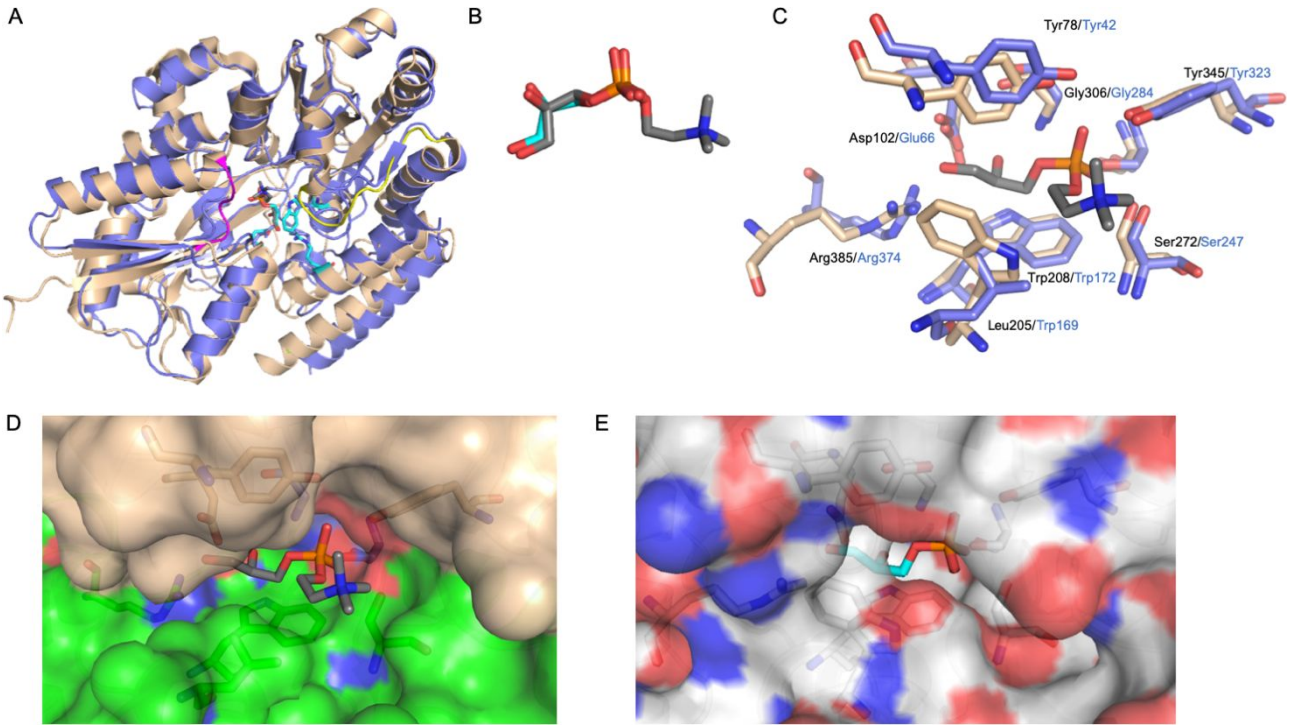

Figure 3. Comparison of *Mtb* UgpB with *E. coli* UgpB. A) Superposition of the *Mtb* UgpB GPC complex structure (blue) with *E. coli* UgpB in complex with G3P (PDB 4AQ4) (brown). Loop regions that differ are highlighted in yellow and magenta. B) Close-up illustration showing the binding orientation of the GPC ligand and G3P ligand in stick representation (dark gray carbon atoms, GPC, cyan carbon atoms G3P) C) Close-up of the overlay of the binding-sites of GPC (*Mtb*) and G3P (*E. coli*). Selected residues are shown as sticks (*Mtb* blue, *E. coli* brown) and the font labelled in black (*Mtb*) and blue (*E. coli*). D) Surface representation of the *Mtb* UgpB GPC binding pocket with the GPC ligand in stick representation. E) Surface representation of the *E. coli* UgpB G3P binding pocket in the same orientation as D with the G3P ligand in stick representation.

Solution saturation transfer difference (STD) NMR of *Mtb* UgpB with glycerophosphocholine. Given the apparent discrepancy between the lack of interactions formed between the choline moiety and its importance in binding, given that G3P lacking the choline moiety does not bind, we investigated binding in the solution state. We employed saturation transfer difference (STD) NMR to obtain quantitative maps of the ligand-protein complex in solution (Fig. 4)[21]. Binding was detected for GPC and binding epitope mapping was obtained and analysed as described in the methods section[22]. The STD NMR signals and the GPC binding epitope and maps obtained are shown in Fig. 4. From the epitope map, the glycerol moiety of GPC is identified as the main recognition element showing the highest STD normalized values. In particular, the highest STD intensity values were observed for the protons in position 1 and 2 (H1G and H2G) of the glycerol moiety (Fig. 4A), with slightly lower intensity values for the protons in position 3 (H3G). The STD values decrease from the glycerol moiety to the choline group, indicating that the ligand-protein contacts are closer with the glycerol group than with choline. Intermediate and low STD NMR intensity values were observed for the protons in position 1 and 2 (H1C and H2C) while low intensity values were observed for the methyl groups from the choline moiety. A quantitative comparison of the NMR solution data with the X-ray structure of the complex

was carried out using CORCEMA-ST calculations[23], as well as the newly developed method DEEP-STD NMR[24] and the results are summarised in Fig. 4. An NOE R-factor[25] of 0.25 was obtained when comparing the CORCEMA-ST calculated STD NMR intensities using the crystal structure with the experimentally obtained solution data. This indicates a very good agreement of the complex in solution state with the crystal structure. In order to probe for additional structural information in the solution state we then utilised differential epitope mapping by STD NMR (DEEP-STD NMR). This methodology allows us to gain information about the orientation of the ligand within the architecture of the binding site and indirectly gives information about the type of amino acids (aromatic, polar or apolar residues) surrounding the ligand in the bound state[26]. The DEEP-STD NMR factors clearly identified that the protons in position 3 of the glycerol moiety of GPC are orientated towards aliphatic amino acids whilst the protons in position 1 in the choline moiety are oriented toward aromatic residues (Fig. 4C). Based on the crystal structure of *Mtb* UgpB these residues can be mapped to Leu205, Tyr78 and Tyr345 respectively (Fig. 2). Notably, our data shows strong correlation for the molecular determinants of GPC ligand binding to *Mtb* UgpB to GPC in both solution and solid state.

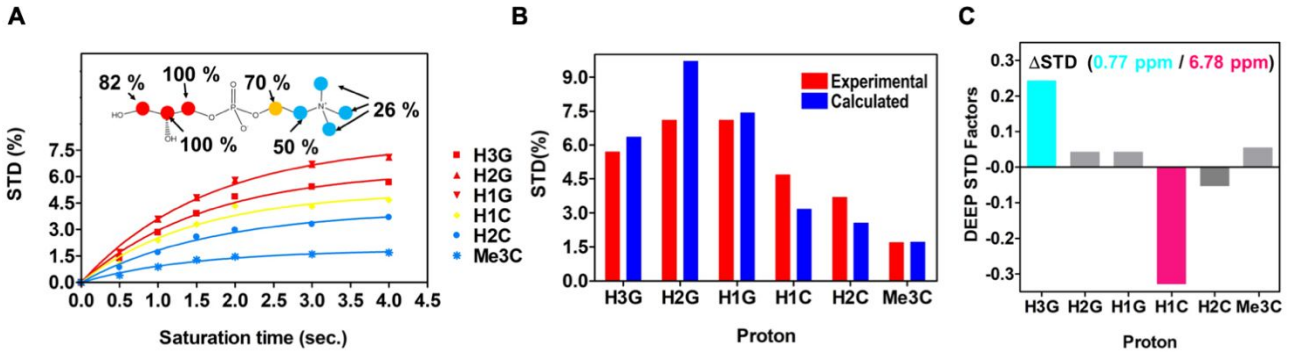

Figure 4. STD-NMR for *Mtb* with GPC. A) Experimental STD build up curves for the GPC/*Mtb* UgpB complex and the obtained epitope map of GPC/*Mtb* UgpB. B) STD in red bars obtained with a 4s saturation time while in blue bars the CORCEMA-ST calculated STD from the 3D crystallographic structure of the *Mtb* UgpB/GPC complex obtained for the same saturation time. RNOE factor 0.25. C) Differential epitope (DEEP)-STD factors showing the type of amino acid that the protons of the GPC ligand are orientated towards. Protons orientated towards aliphatic residues are highlighted in blue and protons orientated towards aromatic residues are highlighted in magenta.

Substrate specificity of *Mtb* UgpB. To establish the importance of both the polar head group and the glycerol moiety for substrate recognition binding we analysed the binding interactions of *Mtb* UgpB with G3P, the preferred substrate of *E. coli* UgpB, and phosphocholine by thermal shift analysis and microscale thermophoresis. In contrast to GPC, no binding interactions were observed for these smaller derivatives. Taken together with our structural studies, these results indicate that whilst the glycerol moiety is the main recognition element for *Mtb* UgpB and that there are minimal interactions with the polar head group, the entire phosphodiester moiety is critical for substrate recognition and binding. The lack of recognition of G3P by *Mtb* UgpB is consistent with the intracellular location of two putative *Mtb* glycerophosphodiesterase enzymes (GlpQ1, Rv3842c; GlpQ2, Rv03127c) that are predicted to degrade glyercophosphodiesters to produce G3P and the corresponding alcohol [27-28]. In direct contrast *E. coli* secretes glycerophosphodiesterase enzymes to

enable the extracellular production of G3P and this is consistent with the ability of the periplasmic *E. coli* UgpB to recognise the G3P metabolite[12].

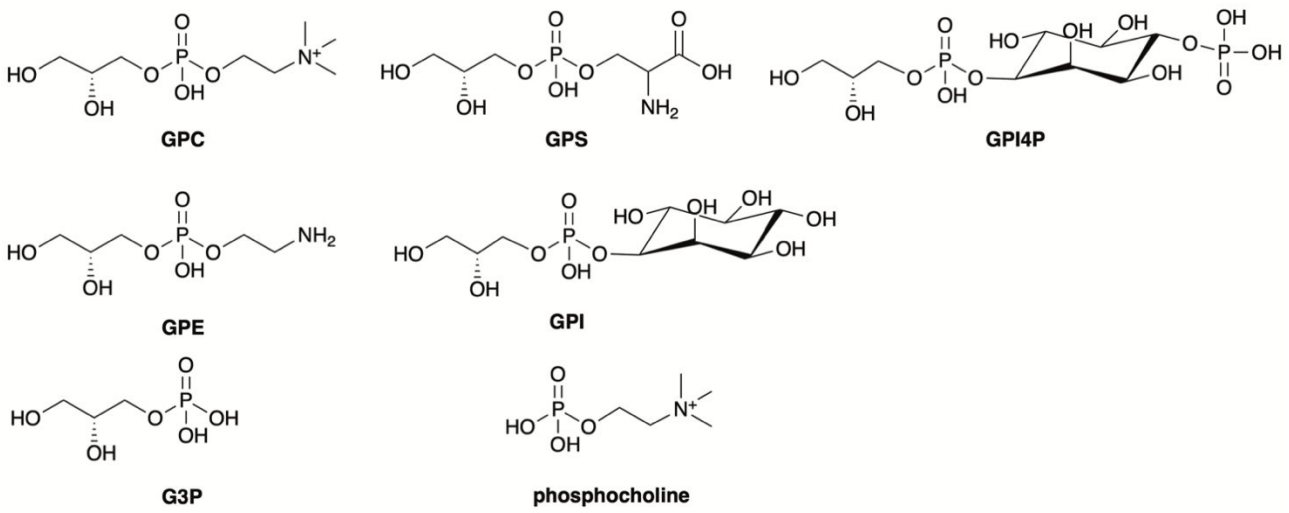

Figure 5. Structure of glycerophosphodiesters and derivatives probed in this study

Our structural studies in both the solid and solution state revealed that the GPC substrate interacts predominantly with *Mtb* UgpB through interactions with the glycerol backbone. The lack of specific interactions between the protein and the polar choline head group located at the entrance of the substrate binding pocket led us to speculate that *Mtb* UpgB may recognise alternative glycerophosphodiester analogues. To directly investigate the substrate specificity of *Mtb* UgpB we used microscale thermophoresis (MST) to analyse the binding interactions of other phosphodiester products formed from the lipolysis of membrane glycerophospholipids (Fig. 5). From the substrates tested, in each case we were able to detect binding for GPC, glycerophosphoserine (GPS), glycerophosphoethanolamine (GPE), glycerophosphoinositol (GPI) and glycerophosphoinositol-4-phosphate (GPI4P), (Table 1, Fig. 6). The measured K_d value for GPC was consistent with previous results obtained by isothermal titration calorimetry (ITC) [9]. Notably, *Mtb* UgpB also binds and recognises GPE, GPS, GPI and GPI4P glycerophosphodiesters with binding affinities in the micromolar range (Table 1) with a preference for positively charged polar head groups. Together, this suggests that *Mtb* has evolved to have a single ABC-transporter to scavenge a range of glycerophosphodiesters within its nutrient poor intracellular environment. The preference for GPC could suggest that as phosphatidylcholine is the main glycerophospholipid in human lung tissue[29] *Mtb* UgpB has evolved to recognise the most abundant glycerophosphodiester available within the host environment, with the potential to recognise and transport a spectrum of additional glycerophosphodiesters depending on the growth conditions and nutrient availability during intracellular infection that can subsequently be catabolised by *Mtb* pathways that are involved in polar head group recycling [27]. Notably, these glycerophospholipids are also major constituents of the *Mtb* cell envelope [30-31] and further experiments are underway to elucidate whether the glycerophosphodiesters are derived from host- or *Mtb*-lipids.

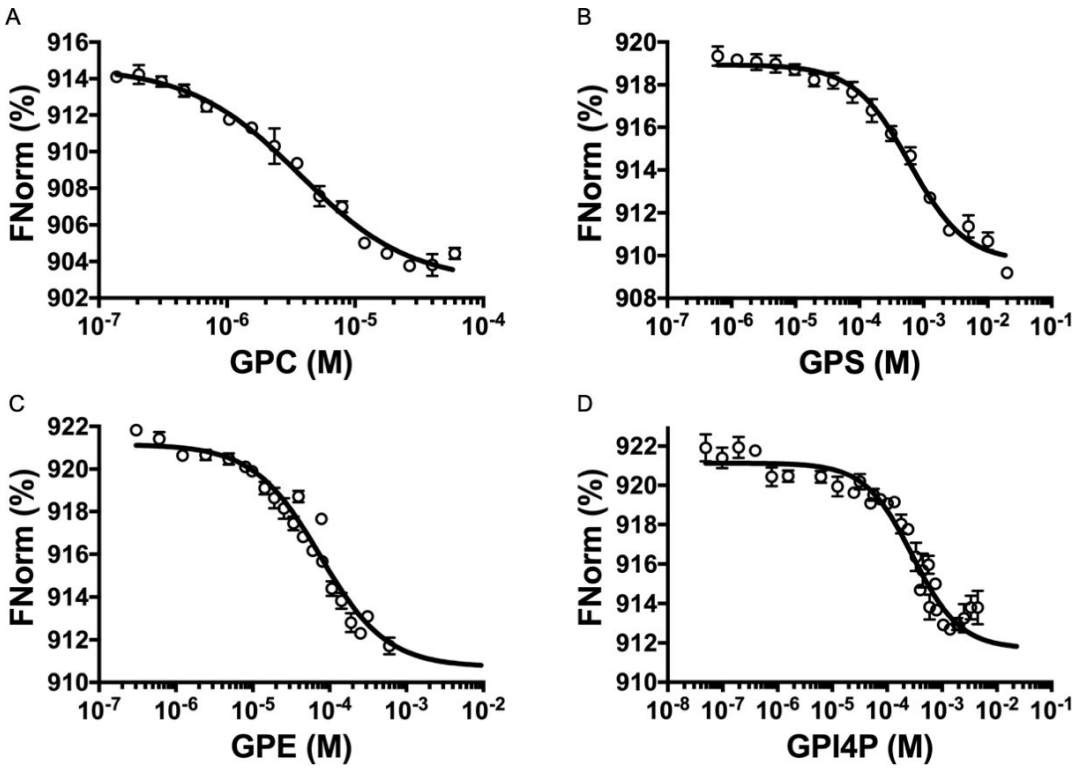

Figure 6. Binding affinities for *Mtb* UgpB Binding of A) GPC, B) GPS, C) GPE and D) GPI4P to *Mtb* UgpB measured by microscale thermophoresis (MST). FNorm (%) is the normalized fluorescence signal of the change in MST signal. Error bars represent standard deviations from at least three independent experiments.

As a final evaluation for potential substrate promiscuity we screened a panel of carbohydrates and amino acids using a thermal shift assay and assessed the binding of putative ligands that resulted in a change in the melting temperature (T_m) of *Mtb* UgpB which can be indicative of binding. In total 37 potential substrates were probed, including trehalose which is known to be a substrate of the *Mtb* LpqY-SugABC ABC-transporter[5], and we found that none of the ligands that were screened influenced the melting temperature (Supplementary Figure S5). It appears that although *Mtb* encodes for only five putative carbohydrate importers, each transport system has a defined substrate preference. Interestingly, these data indicate that the substrate binding-pocket of *Mtb* UgpB can efficiently accommodate glycerophosphodiesters but that it is not able to recognise other carbohydrates or amino acids.

Table 1: Binding data for *Mtb* UgpB

Enzyme	Substrate	K_d (μM)	Reference
Mtb UgpB	GPC	3.6 ± 0.5	This study
Mtb UgpB	GPS	14.9 ± 1.6	This study
Mtb UgpB	GPE	74.7 ± 13.9	This study
Mtb UgpB	GPI	1053.2 ± 313.4	This study
Mtb UgpB	GPI4P	289.8 ± 54.1	This study
Mtb UgpB	G3P	-	This study
Mtb UgpB	phosphocholine	-	This study
Mtb UgpB Y78A	GPC	-	This study
Mtb UgpB Y78A	GPS	-	This study
Mtb UgpB Y78A	GPE	-	This study
Mtb UgpB D102A	GPC	-	This study
Mtb UgpB D102A	GPS	-	This study
Mtb UgpB D102A	GPE	-	This study
Mtb UgpB Ser153Ala	GPC	309.8 ± 56.1	This study
Mtb UgpB S153A	GPS	102.5 ± 16.4	This study
Mtb UgpB S153A	GPE	-	This study
Mtb UgpB L205A	GPC	161.7 ± 15.9	This study
Mtb UgpB L205A	GPE	1360 ± 210	
Mtb UgpB W208A	GPC	-	This study
Mtb UgpB S272A	GPC	-	This study
Mtb UgpB Y345A	GPC	-	This study
Mtb UgpB R385A	GPC	-	This study
Mtb UgpB	GPC	27.3 ± 2.0	[9]
Mtb UgpB	G3P	-	[9]
Mtb UgpB	Maltose	-	[9]
Mtb UgpB L205W	GPC	-	[9]
Mtb UgpB L205W	G3P	-	[9]
E. coli UgpB	GPC	5.1 ± 0.3	[10]
E. coli UgpB	G3P	0.68 ± 0.02	[10]

(-) = no binding detected, standard deviations from at least three independent experiments
GPC: glycerophosphocholine, GPS, glycerophosphoserine, GPE: glycerophosphoethanolamine, GPI Glycerophosphoinositol, GPI4P:glycerophosphoinositol-4-phosphate.

STD NMR of *Mtb* UgpB with GPI4P . Next, to validate some of the MST-binding data we used STD NMR spectroscopy for a more in-depth investigation of GPI4P binding to *Mtb* UgpB. Again the glycerol moiety of GPI4P was the main recognition element with close contacts to *Mtb* UgpB. High STD NMR intensity values were also observed for the H1 and H2 protons of the inositol ring with intermediate STD-NMR values for H3 and H4 protons and low values for H5 and H6 protons (Fig 7A, B). This differs from the situation of the choline head group of GPC where instead low STD intensities were observed. Furthermore, the DEEP-STD NMR maps reveal a slight modification in the binding orientation of the glycerol tail of GPI4P compared to GPC as protons in position 3 orientated towards aromatic residues this time. To gain 3D structural insights about this interaction we carried out docking calculations using Autodock Vina[32] followed by validation using CORCEMA-ST calculations. An NOE R-factor of 0.31 was obtained by comparing the CORCEMA-ST calculated STD intensities from the best scored docked structure of GPIP4 bound to *Mtb* UgpB and the

experimental STD values. This indicates a good agreement of the proposed docking structure of the *Mtb* UgpB/GPIP4 complex with the experimental STD NMR data. From Fig. 7 we can observe that the protons in position 3 (H3G) are oriented toward the aromatic residues, which was also determined from DEEP-STD factors analysis. Further, also the protons of inositol-phosphate moiety are in line with the observed orientation from DEEP-STD factor analysis. In fact protons H4I, H1G, H2G are oriented toward aliphatic residue Leu205, while protons H1I, H3G, H6I, H5I are oriented toward the aromatic residues Tyr78 and Tyr345, validating the proposed model structure with the experimental STD and DEEP-STD NMR data. These studies indicate that the size and charge of the glycerophosphodiester head group is critical in defining substrate selectivity and the binding orientation of the glycerol tail.

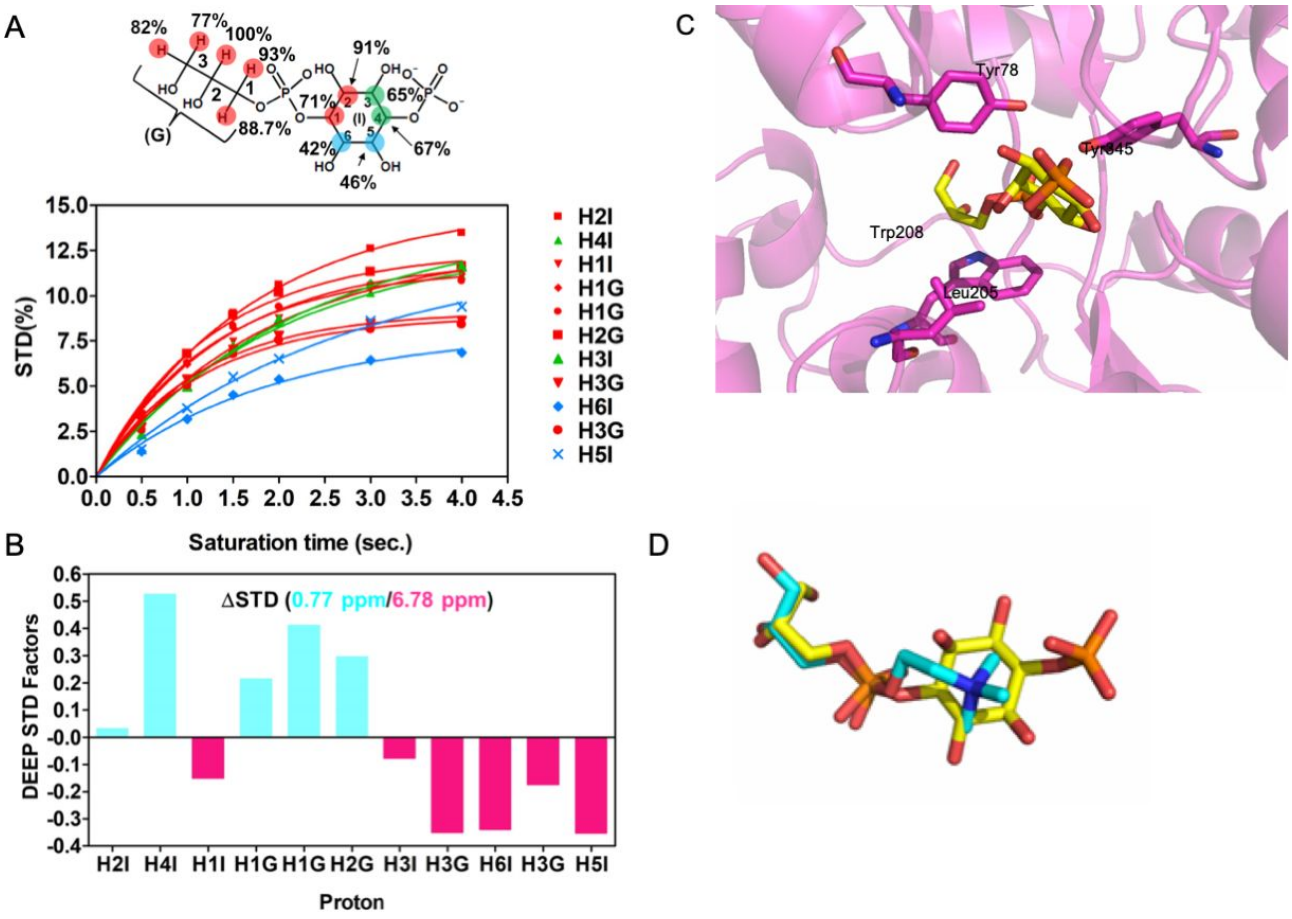

Figure 7. STD NMR of *Mtb* UgpB with GPI4P A) Experimental STD build up curve for the *Mtb* UgpB/GPIP4 complex and the obtained epitope map of GPI4P/*Mtb* UgpB. B) Differential epitope (DEEP)-STD factors showing the type of amino acid that the protons of the GPI4P ligand are orientated towards. Protons orientated towards aliphatic residues are highlighted in blue and protons orientated towards aromatic residues are highlighted in magenta. C) Docked structure of the GPIP4 in the binding site of *Mtb* UgpB. GPI4P is in stick representation with the carbon atoms in yellow. The binding orientation of GPC obtained from the crystal structure is shown in stick representation with orange carbon atoms. D) Close-up overlay of the binding orientations of GPC (cyan carbon atoms) with GPI4P (yellow carbon atoms).

Activity of sequence variants. In order to complement our structural studies in both the solution and solid state and assess the significance of individual amino acids that were identified to be important in molecular recognition and binding we introduced single point mutations in eight individual residues that were suggested

to interact with the glycerophosphodiester ligands. In each case, we confirmed that the substituted alanine mutation was not detrimental to the correct folding of the protein by circular dichroism spectroscopy (Supplementary Figure S6). MST was used to determine the binding affinities of the *Mtb* UpgB protein with GPC and complete abrogation of binding was observed when Tyr78, Asp102, Trp208, Ser272, Tyr345 and Arg385 were individually replaced by an alanine, confirming the significance of these residues in substrate selectivity and importance in binding recognition. In contrast, binding of GPC was still observed when Ser153 and Leu205 were replaced by alanine, with a corresponding 85- and 45-fold reduction in the K_d values respectively (Table 1), indicating that whilst these two individual residues are important for binding, they are not critical. Failure of these single-residue mutants to completely abolish binding reflects that multiple amino-acids are involved in the interaction with GPC, as observed from the crystal structure. Previous studies that mutated *Mtb* UgpB Leu205 to a tryptophan residue to mimic the situation found in *E. coli* UgpB were detrimental for binding of GPC, indicating that the bulky indole side-chain cannot be tolerated in *Mtb* UgpB[9] and did not enable recognition of G3P. The distinct glycerophosphodiester-recognition of *Mtb* UgpB compared with *E.coli* UgpB indicates that the mycobacterial UgpB transporter has evolved to have unique specificity and function that is distinct from other UgpB proteins.

In conclusion, to date, the nutrient requirements of *Mtb* during infection and the corresponding transport systems have not been fully elucidated. The structural and functional understanding of mycobacterial ABC-transporters that import essential nutrients is an important step to understanding the mechanisms that support intracellular survival. Importantly, we have identified that the essential *Mtb* UgpABCE importer is linked with glycerophosphodiester uptake with wide substrate selectivity. For the first time, we have established the molecular determinants of the distinct substrate selectivity of the UgpB substrate binding protein from the *Mtb* pathogen that has important structural and functional differences with *E. coli* UgpB. We therefore propose a new role for the *Mtb* UgpABCE transporter in the uptake of glycerophosphodiesters generated from the degradation of membrane phospholipids as a route to scavenge scarce nutrients during intracellular infection.

METHODS

Procedures for cloning, protein expression, crystallization, X-ray data collection and refinement, STD-NMR experiments, docking, micro-scale thermophoresis, thermal shift assays and enzymatic synthesis of substrates in this study are described in the Supporting Information

Accession codes

Coordinates and structure factors for *Mtb* UgpB have been deposited in the Protein Data Bank under accession code 6R1B.

Associated content

The Supporting Information is available free of charge.

Acknowledgements

We would like to thank M. Ahanger for technical assistance. We thank W. Jacobs (Albert Einstein College of Medicine, USA) for providing expression vector pYUB1062 and the *M. smegmatis* $mc^2$4517 expression system. We acknowledge the contribution of the WPH Proteomics Facility research technology platform, University of Warwick. We thank Diamond Light Source for access to synchrotron beamlines and staff for support during experiments. Equipment was supported through the Warwick Integrative Synthetic Biology (WISB) research technology platform (BB/M017982/1). This work was supported by a Sir Henry Dale Fellowship to EF jointly funded by the Wellcome Trust and Royal Society (104193/Z/14/Z), a research grant from the Royal Society (RG120405), the MRC for a studentship to JF (MR/J003964/1) and the EPSRC for funding an Integrate Early Career fellow, JH (EP/M027503/1). JA and RN acknowledge support from the BBSRC through a New Investigator grant awarded to JA (BB/P010660/1). We are grateful for the use of the University of East Anglia (UEA) Faculty of Science NMR facility.

References

1. WHO Global Tuberculosis Report. http://www.who.int/tb/publications/global_report/en/.
2. Getahun, H.; Matteelli, A.; Chaisson, R. E.; Raviglione, M., Latent *Mycobacterium tuberculosis* infection. *N Engl J Med* **2015,** *372* (22), 2127-35.
3. Titgemeyer, F.; Amon, J.; Parche, S.; Mahfoud, M.; Bail, J.; Schlicht, M.; Rehm, N.; Hillmann, D.; Stephan, J.; Walter, B.; Burkovski, A.; Niederweis, M., A genomic view of sugar transport in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. *J Bacteriol* **2007,** *189* (16), 5903-15.
4. Niederweis, M., Nutrient acquisition by mycobacteria. *Microbiology* **2008,** *154* (Pt 3), 679-92.
5. Kalscheuer, R.; Weinrick, B.; Veeraraghavan, U.; Besra, G. S.; Jacobs, W. R., Jr., Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* **2010,** *107* (50), 21761-6.
6. Fullam, E.; Prokes, I.; Futterer, K.; Besra, G. S., Structural and functional analysis of the solute-binding protein UspC from *Mycobacterium tuberculosis* that is specific for amino sugars. *Open Biol.* **2016,** *6* (6).
7. Kruh, N. A.; Troudt, J.; Izzo, A.; Prenni, J.; Dobos, K. M., Portrait of a pathogen: the *Mycobacterium tuberculosis* proteome in vivo. *PloS one* **2010,** *5* (11), e13938.
8. Somashekar, B. S.; Amin, A. G.; Rithner, C. D.; Troudt, J.; Basaraba, R.; Izzo, A.; Crick, D. C.; Chatterjee, D., Metabolic profiling of lung granuloma in *Mycobacterium tuberculosis* infected guinea pigs: ex vivo 1H magic angle spinning NMR studies. *J Proteome Res* **2011,** *10* (9), 4186-95.
9. Jiang, D.; Zhang, Q.; Zheng, Q.; Zhou, H.; Jin, J.; Zhou, W.; Bartlam, M.; Rao, Z., Structural analysis of *Mycobacterium tuberculosis* ATP-binding cassette transporter subunit UgpB reveals specificity for glycerophosphocholine. *FEBS J* **2014,** *281* (1), 331-41.
10. Wuttge, S.; Bommer, M.; Jager, F.; Martins, B. M.; Jacob, S.; Licht, A.; Scheffel, F.; Dobbek, H.; Schneider, E., Determinants of substrate specificity and biochemical properties of the sn-glycerol-3-phosphate ATP binding cassette transporter (UgpB-AEC2) of *Escherichia coli*. *Mol Microbiol* **2012,** *86* (4), 908-20.
11. Scheepers, G. H.; Lycklama, A. N. J. A.; Poolman, B., An updated structural classification of substrate-binding proteins. *FEBS Lett* **2016,** *590* (23), 4393-4401.
12. Larson, T. J.; Ehrmann, M.; Boos, W., Periplasmic glycerophosphodiester phosphodiesterase of *Escherichia coli*, a new enzyme of the glp regulon. *J Biol Chem* **1983,** *258* (9), 5428-32.
13. Grosshennig, S.; Schmidl, S. R.; Schmeisky, G.; Busse, J.; Stulke, J., Implication of glycerol and phospholipid transporters in Mycoplasma pneumoniae growth and virulence. *Infect Immun* **2013,** *81* (3), 896-904.
14. Patton-Vogt, J. L.; Henry, S. A., GIT1, a gene encoding a novel transporter for glycerophosphoinositol in *Saccharomyces cerevisiae*. *Genetics* **1998,** *149* (4), 1707-15.
15. Krissinel, E.; Henrick, K., Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **2004,** *60* (Pt 12 Pt 1), 2256-68.
16. Krissinel, E.; Henrick, K., Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **2007,** *372* (3), 774-97.
17. Locher, K. P., Structure and mechanism of ATP-binding cassette transporters. *Philos Trans R Soc Lond B Biol Sci* **2009,** *364* (1514), 239-45.
18. Berntsson, R. P.; Smits, S. H.; Schmitt, L.; Slotboom, D. J.; Poolman, B., A structural classification of substrate-binding proteins. *FEBS Lett* **2010,** *584* (12), 2606-17.
19. Hayward, S.; Lee, R. A., Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J Mol Graph Model* **2002,** *21* (3), 181-3.
20. Kozlikova, B.; Sebestova, E.; Sustr, V.; Brezovsky, J.; Strnad, O.; Daniel, L.; Bednar, D.; Pavelka, A.; Manak, M.; Bezdeka, M.; Benes, P.; Kotry, M.; Gora, A.; Damborsky, J.; Sochor, J., CAVER Analyst 1.0: graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics* **2014,** *30* (18), 2684-5.
21. Mayer, M.; Meyer, B., Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew Chem Int Ed Engl* **1999,** *38* (12), 1784-1788.
22. Mayer, M.; James, T. L., NMR-based characterization of phenothiazines as a RNA binding scaffold. *J Am Chem Soc* **2004,** *126* (13), 4453-60.
23. Jayalakshmi, V.; Krishna, N. R., Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes. *J Magn Reson* **2002,** *155* (1), 106-18.

24. Monaco, S.; Tailford, L. E.; Juge, N.; Angulo, J., Differential Epitope Mapping by STD NMR Spectroscopy To Reveal the Nature of Protein-Ligand Contacts. *Angew Chem Int Ed Engl* **2017,** *56* (48), 15289-15293.
25. Xu, Y.; Sugar, I. P.; Krishna, N. R., A variable target intensity-restrained global optimization (VARTIGO) procedure for determining three-dimensional structures of polypeptides from NOESY data: application to gramicidin-S. *J Biomol NMR* **1995,** *5* (1), 37-48.
26. Nepravishta, R.; Walpole, S.; Tailford, L.; Juge, N.; Angulo, J., Deriving Ligand Orientation in Weak Protein-Ligand Complexes by DEEP-STD NMR Spectroscopy in the Absence of Protein Chemical-Shift Assignment. *Chembiochem* **2018**.
27. Larrouy-Maumus, G.; Biswas, T.; Hunt, D. M.; Kelly, G.; Tsodikov, O. V.; de Carvalho, L. P., Discovery of a glycerol 3-phosphate phosphatase reveals glycerophospholipid polar head recycling in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* **2013,** *110* (28), 11320-5.
28. de Souza, G. A.; Leversen, N. A.; Malen, H.; Wiker, H. G., Bacterial proteins with cleaved or uncleaved signal peptides of the general secretory pathway. *J Proteomics* **2011,** *75* (2), 502-10.
29. Veldhuizen, R.; Nag, K.; Orgeig, S.; Possmayer, F., The role of lipids in pulmonary surfactant. *Biochim Biophys Acta* **1998,** *1408* (2-3), 90-108.
30. Brennan, P. J.; Nikaido, H., The envelope of mycobacteria. *Annu Rev Biochem* **1995,** *64*, 29-63.
31. Jackson, M., The mycobacterial cell envelope-lipids. *Cold Spring Harb Perspect Med* **2014,** *4* (10).
32. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **2010,** *31* (2), 455-61.