Supporting information

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

Jonathan S 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

Table of Contents

General Procedures, Chemicals and reagents

All chemicals and reagents were purchased from Sigma-Aldrich, unless specified. Glycerophosphoinositol and glycerophosphoinositol-4-phosphate were purchased from Tebu-Bio. PCR and restriction enzymes were obtained from New England Biolabs. Double-distilled water was used throughout.

Plasmid construction. An *N*-terminal truncated form (codons 35-436) of the *Mtb ugpB* gene (Rv2833c) was amplified from *Mtb* genomic DNA by PCR using gene specific primers listed in Table S3. The PCR amplification (Q5 polymerase (NEB)) consisted of 30 cycles (95°C, 2 min; 95°C, 1 min; 60°C, 30 s; 72°C, 3 min), followed by an extension cycle (10 min at 72°C). The resulting PCR product was cloned into the vector pYUB1062 using the *NdeI* and *HindIII* restriction enzyme sites resulting in the construct *ugpB-pYUB1062*. Targeted single-site substitutions were introduced into *ugpB-pYUB1062* using the primers that are detailed in SI Appendix, Table S2, with Phusion HF polymerase and the PCR cycle (98°C, 30 s; 20 cycles of 98°C, 30 s; 60°C, 30 s; 72°C, 4 min; followed by 5 min at 72°C), followed by digestion with 1μL DpnI. All plasmid sequences were verified by DNA sequencing (GATC) and used for protein expression.

Heterologous overexpression of *Mtb* UgpB. *Mycobacterium smegmatis* mc²4517 competent cells were transformed with the appropriate *ugpB-pYUB1062* expression plasmid and grown at 37 °C to an optical density at 600 nm (OD_{600}) of 0.4 in LB media supplemented with 0.05% Tween-80, 0.2% glycerol, 100μg/mL hygromycin and 25μg/mL kanamycin. Protein production was induced with 0.2% acetamide and the culture was grown at 37 °C for 20 hours with shaking at 180 rpm. The cells were harvested and resuspended in lysis buffer (25 mM NaH_2PO_4, 500 mM NaCl pH 7.4 (Buffer A), supplemented with 0.1% Triton-X 100, DNAse and Complete Protease Inhibitor Cocktail (Pierce)). The cells were freeze-thawed and sonicated on ice. Following centrifugation (39,200 *g*, 40 min, 4°C) the supernatant was loaded onto a pre-equilibrated Co^{2+}-affinity resin (HisPure). The *Mtb* UgpB protein was eluted from the Co^{2+}-affinity column in buffer A with increasing concentrations of imidazole. Fractions containing the protein, as determined by SDS-PAGE, were pooled and dialysed against buffer B (25 mM HEPES, 150 mM NaCl, pH 7.0) at 4 °C for 16 hours. Following dialysis, the protein was loaded onto a pre-equilibrated QHP ion exchange column (GE Healthcare) and eluted with buffer B containing increasing concentrations of NaCl (0.1-1 M). Fractions containing *Mtb* UgpB were pooled and loaded onto a Superdex 75 pg HiLoad 16/600 gel filtration column (GE Healthcare) and eluted with buffer C (25 mM HEPES, 150 mM NaCl, 10% glycerol pH 7.0). The fractions that contained *Mtb* UgpB were combined and concentrated by ultrafiltration (10 kDa cut-off, Amicon Ultra) to ~ 5-10 mg/mL prior to storage at -80°C. The identity of the protein was confirmed by tryptic digest and nanoLC-ESI-MS/MS (WPH Proteomics facility, University of Warwick).

Circular Dichroism (CD) analysis. Purified *Mtb* UgpB proteins were diluted to 0.25 mg/mL and dialysed in the following buffer: 25 mM NaH_2PO_4, 100mM NaCl, 10% glycerol pH 7.0. The samples were transferred into a 1 mm path length quartz cuvette and analysed on Jasco J-810 DC spectrometer from 198-260nm. Spectra were acquired in triplicate and averaged after subtraction of the buffer background.

Methylation of *Mtb* UgpB. Purified *Mtb* UgpB was diluted to 1 mg/mL into buffer C and reductively methylated as described previously [1]. Briefly, dimethylborane amine (DMAB) and formaldehyde were added to *Mtb* UgpB and the mixture was left shaking at 100 rpm at 4 °C for two hours. This step was repeated two additional times. DMAB was then added to *Mtb* UgpB for a final incubation step (1 hour, 4 °C, shaking at 100 rpm) followed by the addition of Tris-HCl (final concentration 100 mM, pH 7.0) to remove any excess unreacted DMAB reagent and the sample was then dialysed at 4 °C for 16 hours against buffer C. *Mtb* UgpB was concentrated by ultrafiltration (10 kDa cut-off, Amicon Ultra) to 7 mg/mL.

Crystallization and structure determination. For co-crystallisation experiments methylated *Mtb* UgpB was incubated with 10mM glycerol-3-phosphocholine (GPC) and incubated at 4 °C for 30 min before crystallization. Crystals of *Mtb* UgpB in complex with GPC were grown by vapor diffusion in 96-well plates (Swiss-Ci) using a Mosquito liquid handling system (TTP LabTech) by mixing 1:1 volumes (150 nL) of concentrated methylated *Mtb* UgpB (7 mg/mL) with reservoir solution. *Mtb* UgpB crystals typically grew within three days at 22 °C in 0.2 M MgCl$_2$, 0.1 M Tris pH 8.5, 20% w/v PEG 8,000. The *Mtb* UgpB crystals were cryoprotected with 20% glycerol and flash frozen in liquid nitrogen prior to data collection.

The X-ray diffraction data for the ligand bound *Mtb* UgpB crystals were collected at the I04 beamline of Diamond Light Source. The diffraction data were indexed, integrated and scaled with XDS [2] through the XIA2 pipeline and the CCP4 suite of programmes [3]. Initial phases were determined by molecular replacement using PHASER [4] and the separate domains (Domain I residues 34-153 and 305-378/ Domain II residues 154-304 and 379-424) of the *apo-Mtb* UgpB structure as two ensembles as a search model (PDB 4MFI) specifying to search for 4 copies in the asymmetric unit. Autobuild [5] was initially used for model building followed by iterative cycles of alternating manual rebuilding in COOT [6] and reciprocal space crystallographic refinement with PHENIX-REFINE [7] assigning each domain as a separate TLS group. The coordinates for the glycerol-3-phosphocholine ligand were downloaded from the PDB and fitted into unoccupied electron density in all four chains of the asymmetric unit. The restraints for use in refinement were calculated using REEL [8]. Magnesium ions and glycerol molecules were also fitted into the unoccupied electron density as well as waters. Methylated lysine (MLZ) was fitted at position 161 in each chain.

The model of the ligand-bound structure comprises residues 36-428 in all chains (A-D), with an additional 1 residue in chains B and D and 2 residues in chain C. There is one disordered region between residues 355-366 in chains C and D and these residues were not modelled. No Ramachandran outliers were identified and structure validations were done by MolProbity [9]. Figures were prepared using Pymol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC), except for those showing electron density which were prepared using CCP4mg [10].

DynDom Analysis

DynDom (http://fizz.cmp.uea.ac.uk/dyndom/) [11] was used to determine dynamic domain and hinge regions comparing the refined ligand-bound structure and the previously solved apo structure (PDB 4mfi). Default parameters were used for the analysis: a window length of 5, minimum ratio of inter-domain to intradomain displacement of 1.0 and minimum domain size of 20 residues.

^{1}H STD NMR experiments. All the STD NMR experiments were performed in PBS D_2O buffer, pH 7.5. For the complex *Mtb* UgpB/GPC the protein concentration was 68 µM while the ligand concentration was 5 mM. STD NMR spectra were acquired on a Bruker Avance 500.13 MHz at 298 K. The on- and off-resonance spectra were acquired using a train of 50 ms Gaussian selective saturation pulses using a variable saturation time from 0.5 s to 4 s, and a relaxation delay (D1) of 4 seconds. The water signal was suppressed using the watergate technique [12] while the residual protein resonances were filtered using a $T_{1\rho}$-filter of 50 ms. All the spectra were acquired with a spectral width of 5 kHz and 32K data points using 128 scans. The on-resonance spectra were acquired by saturating at 0.77 or 6.78 ppm while the off-resonance spectra were acquired by saturating at 40 ppm. Instead, for the *Mtb* UgpB/GPIP4 complex, the protein concentration was 35 µM while the ligand concentration was 2.5 mM. STD NMR spectra were acquired on a Bruker Avance 800.23 MHz at 278 K. The on- and off-resonance spectra were acquired using a train of 50 ms Gaussian selective saturation pulses using a variable saturation time from 0.5 s to 4 s and a relaxation delay (D1) of 5 seconds. The water signal was suppressed by using the excitation sculpting technique [13] while the residual protein resonances were filtered using a $T_{1\rho}$ -filter of 24 ms. All the spectra were acquired with a spectral width of 12.82 kHz and 32K data points using 64 scans. The on-resonance spectra were acquired by saturating at 0.7 or 6.67 ppm while the off-resonance spectra were acquired by saturating at 40 ppm. To get accurate structural information from the STD NMR data and in order to minimize the T_1 relaxation bias, the STD build up curves were fitted to the equation $STD(t_{sat}) = STD_{max}*(1-exp(-k_{sat}*t_{sat}))$ calculating the initial growth rate STD_0 factor as $STD_{max}*k_{sat} = STD_0$ and then normalizing all of them to the highest value [14].

CORCEMA-ST calculations. The CORCEMA-ST software was used to calculate the theoretical STD intensities from the crystallographic structure of the *Mtb* UgpB/GPC complex. The parameters used for the calculations were: saturation frequency range 0-1.1 ppm; protein correlation time 45 ns; K_d 0.005 mM; order parameter 0.85; ligand correlation time 0.3 ns; ρ-leak 0.35 s; τ_m 10 ps; cutoff 8 Å; $[L]_0$ 2.5 mM; $[E]_0$ 35 µM; field 500 MHz. While for the *Mtb* UgpB/GPIP4 complex model obtained from docking calculations, the theoretical STD intensities were calculated using the following parameters: saturation frequency range 0-0.9 ppm; protein correlation time 45 ns; K_d 1 mM; order parameter 0.85; ligand correlation time 0.3 ns; ρ-leak 0.1 s; τ_m 10 ps; cutoff 8 Å; $[L]_0$ 5 mM; $[E]_0$ 68 µM; field 800 MHz. The calculations were repeated in order to have the best fitting possible between the calculated and the experimental ^{1}H STD NMR intensities. For CH_2 protons showing the same chemical shift an averaged calculated ^{1}H STD NMR intensity was assumed. NOE factor [15] was used to evaluate the best fit to the experimental data.

Autodock Vina Docking calculations

Autodock tools [16] was used to prepare for docking both the ligand GPIP4 and the *Mtb* UgpB protein. The calculations were performed by positioning a grid of 20 x 24 x 22 Å in the center of the binding site of *Mtb* UgpB, which was maintained rigid while the ligand was considered flexible. The calculations were performed using Autodock Vina [17]

DEEP-STD NMR . DEEP-STD factors were obtained as previously described [18]. Briefly, frequencies derived from shiftx2 [19] for aliphatic and aromatic residues present in the binding site of *Mtb* UgpB were used for the position of the saturating selective pulse. In order to perform the DEEP-STD NMR experiment two consecutive experiments were acquired where the protein was saturated with a low power selective saturation pulse for 0.5 seconds. The on-saturation pulse was positioned in the aliphatic region at 0.77 ppm while the off-saturation pulse was positioned at 40 ppm for GPC. The on-saturation pulse was positioned in the aromatic region at 6.78 while the off-saturation pulse was positioned at 40 ppm for GP14P. The DEEP-STD factor for each ligand proton (ΔSTDi) is calculated using the following equation:

$$\Delta STD_i = \frac{STD_{i,1}}{STD_{i,2}} - \frac{1}{n}\sum_{i}^{n}\left(\frac{STD_{i,1}}{STD_{i,2}}\right)$$

The STD intensities from experiment 1 and experiment 2 were used, were the $STD_{i,1}$ are the STD intensities obtained from experiment 1 (on-saturation pulse positioned at 0.77 ppm), while the $STD_{i,2}$ are the STD intensities obtained from experiment 2 (on-saturation pulse positioned at 6.78 ppm). In this way, positive DEEP-STD factors reveal which protons of the ligands are oriented toward the aliphatic residues of the protein while DEEP-STD negative factors reveal those protons of the ligand oriented toward aromatic residues in the binding site.

Affinity studies with Microscale thermophoresis (MST). *Mtb*-UgpB protein was labelled using Monolith His-tag labelling kit RED-Tris-NTA, PBS (PBS supplemented with 0.05 % Tween 20 and a constant concentration of UgpB (50 nM) was used. The compounds were prepared in PBS in the concentration range 0-0.5 M. The samples were loaded into the MonoLith NT.115 standard treated capillaries and incubated for 10 min before analysis using the Monolith NT.115 instrument (NanoTemper Technologies) at 21°C using medium laser power and 40 % LED power. The binding affinities were calculated using a single-site binding model with GraphPad Prism software (version 7.0). All experiments were carried out in triplicate.

Thermal shift assay. The transition unfolding temperature T_m of the *Mtb* UgpB protein (22 µM) was determined in the presence or the absence of ligands. The screen used a single ligand concentration of 100 mM. Reactions were performed in a total volume of 20 µL using Rotor-Gene Q Detection System (Qiagen), setting the excitation wavelength to 470 nm and detecting emission at 557 nm of the SYPRO Orange protein gel stain, 15 × final concentration (Invitrogen). The cycle used was a melt ramp from 30 to 95°C, increasing temperature in 1°C steps and time intervals of 5 s. Fluorescence intensity was plotted as a function of

temperature. The T_m was determined using the Rotor-Gene Q software and the Analysis Melt functionality. All experiments were performed in triplicate.

Enzymatic synthesis of glycerophosphoethanolamine (GPE) and glycerophosphoserine (GPS). Phospholipase A1 from *Aspergillus oryzae* was dialysed into PBS overnight at 4°C prior to use. The enzymatic reaction contained either 50 mg of 1,2-Dipalmitoyl-sn-glycero-3-phosphoethanolamine or 25 mg of 1,2-Diacyl-sn-glycero-3-phospho-L-serine in an organic-aqueous media (800 µL hexane, 138 µL H_2O: 5.8:1 ratio) and heated at 50°C for 10 mins prior to the addition of Phospholipase A1 from *Aspergillus oryzae* (1 µL Phospholipase A1 per 1 mg of phospholipid). The reaction mixtures were heated at 50°C and stirred at 300 rpm for 48 hours. The solvent was removed *in vacuo* and the reaction mixture redissolved in water (5 mL) and extracted with chloroform (3 x 25 mL). The aqueous phase was separated and the phospholipase A1 enzyme removed using a centrifugal filter unit (Amicon, 10 kDa molecular weight cut off) . The collected filtrate was concentrated *in vacuo* to give the products as a colourless oil (9.4 mg GPE) or yellow oil (3 mg GPS). GPE: ^1H NMR (400MHz, D_2O) δ_{ppm} 3.99 – 4.08 (2H, m, POC$\underline{H}_2$CH$_2$N), 3.75- 3.91 (3H, m, POC$\underline{H}_2$C$\underline{H}$CH$_2$), 3.49 – 3.63 (2H, m, POCH$_2$CHC$\underline{H}_2$), 3.20 (2H, t, J = 5.0 Hz, POCH$_2$C$\underline{H}_2$N). ^{13}C NMR (100MHz, D_2O) δ_{ppm} 70.7 ($\underline{C}$H), 66.5 (O$\underline{C}$H$_2$), 62.0 (O$\underline{C}$H$_2$), 61.8 (O$\underline{C}$H$_2$), 40.0 (N$\underline{C}$H$_2$). ^{31}P NMR (161 MHz, D_2O) δ_{ppm} 0.42. GPS: ^1H NMR (400MHz, D_2O) δ_{ppm} 4.17 – 4.26 (2H, m, POC$\underline{H}_2$CHN), 4.01 – 4.08 (1H, m, POCH$_2$C$\underline{H}$N), 3.71 – 3.93 (3H, m, 3H, m, POC$\underline{H}_2$CHCH$_2$), 3.48-3.65 (2H, m, POCH$_2$CHC$\underline{H}_2$). ^{13}C NMR (100MHz, D_2O) δ_{ppm} 178.5 ($\underline{C}$=O), 70.6 (CH$_2$$\underline{C}$H(OH)CH$_2$), 66.5 (O$\underline{C}H_2$), 63.7 (PO$\underline{C}H_2$CHN), 62.0 (O$\underline{C}H_2$), 54.4 (POCH$_2$$\underline{C}$HN). ^{31}P NMR (161 MHz, D_2O) δ_{ppm} 0.08.

Fig. S1. SDS-PAGE analysis of the purification of *Mtb* UgpB from *M. smegmatis*. A) Elution of His$_6$-tagged *Mtb* UgpB from a Co^{2+} IMAC-column. M = molecular weight marker in kDa, IS = insoluble fraction, S = soluble lysate, FT = flow through, numbers 5 – 1000 refer to the imidazole concentration in the elution buffer (units of mM). B) QHP anion exchange chromatography of *Mtb* UgpB following the Co^{2+} IMAC step. L = protein after dialysis, FT1 = first flow through, FT2 = second flow through, numbers 150 – 1000 refer to NaCl concentration in the elution buffer (units of mM). C) Size exclusion chromatography of *Mtb* UgpB following anion exchange chromatography with the volumes shown as corresponding to D. D) Size exclusion trace of *Mtb* UgpB. See Materials and Methods for buffer compositions.
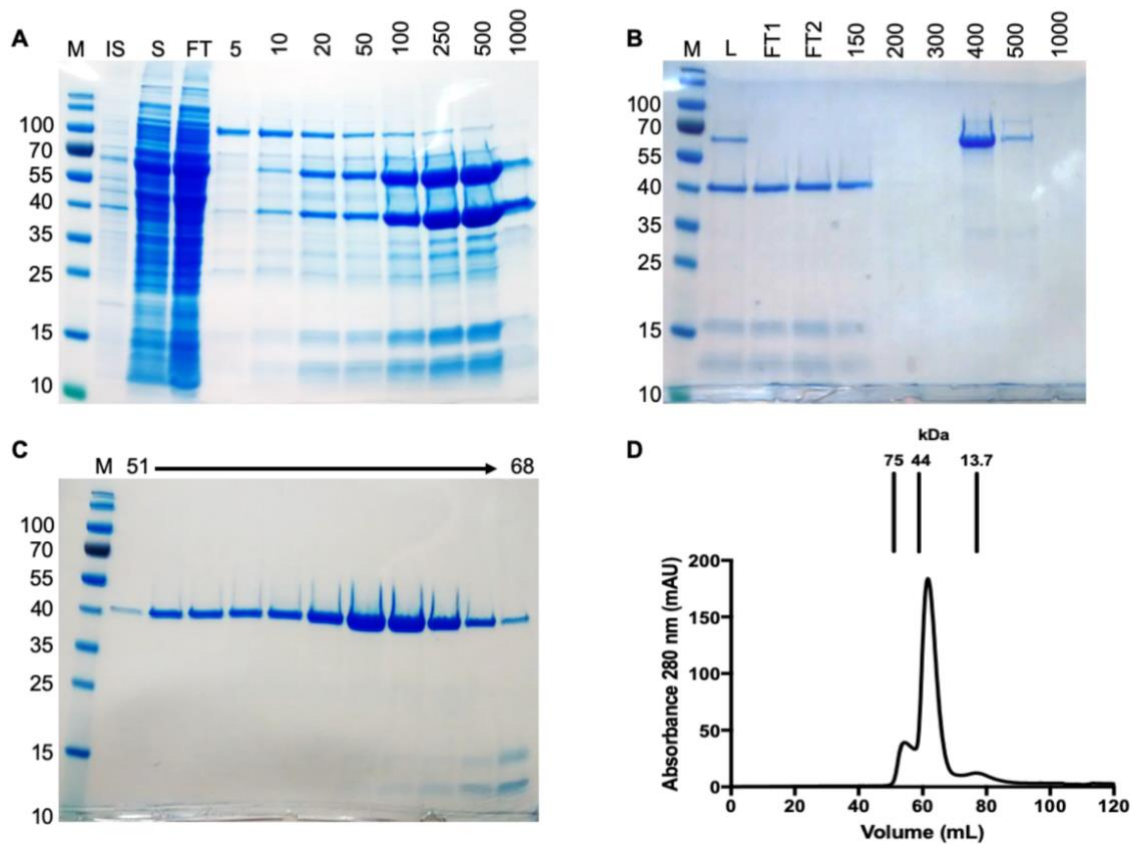
Fig S2: GPC binding

A) Electron density for the GPC substrate. Electron density map contoured at 0.38electrons/$\mathring{A}^3$. Carbon atoms are shown in green, oxygen atoms are shown in red, nitrogen atoms in blue and phosphate atom in purple. The figure was prepared using CCP4mg. The .mtz file was loaded directly with the default settings and clipped to select for the GPC atoms.

B) Alignment of the glycerophosphocholine ligand from each *Mtb* UgpB subunit. Superposition of GPC from each *Mtb* UgpB subunit in the asymmetric unit. The GPC ligand is shown with green carbon atoms (subunit A), light blue carbon atoms (subunit B), wheat carbon atoms (subunit C) and grey carbon atoms (subunit D). In all subunits oxygen atoms are coloured red and the phosphorous atom is coloured orange.
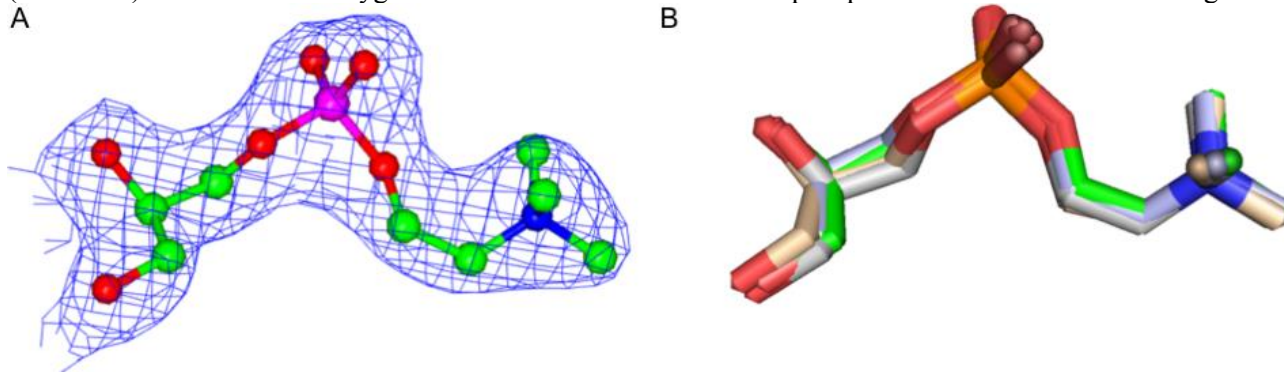
Fig. S3. Sequence alignment of UgpB from *Mycobacterium tuberculosis* with the UgpB homologue from *Escherichia coli.* The sequence alignment was generated using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and ESPript version 3. Identical residues are indicated by a red background and conserved residues by red characters. The secondary structure elements of *Mtb* UgpB are shown above the sequences and the secondary structure elements of *E. coli* UgpB (4AQ4) are shown below the sequences.

```
                                                           β1                    α1
                                                        ───────▶        ℓℓℓℓℓℓℓℓℓℓ
               1         10        20        30        40        50
Mtb    MDPLNRRQFLALAAAAAGVTAGCAGMGGGGSVKSGSGPIDFWSSHPGQSSAAERELIG
Ecoli  MKPLHY........TASALGLALMGNAQ....AVTTIPFWHSMEGELGKEVDSLAQ
                                                        ───────▶        ℓℓℓℓℓℓℓℓℓℓℓ
                                                           β1                    α1

                 β2              α2                β3        α3
            ───────▶     ℓℓℓℓℓℓℓℓℓℓ            ───────▶   ℓℓℓℓℓℓℓ
         60        70        80        90        100       110
Mtb    RFQDRFPTLSVKLIDAGKDYDEVAQKFNAALIGTDVPDVVLLDDRWWFHFALSGVLTA
Ecoli  RFNAENPDYKIVPTYKGNYEQNLSAGI.AAFRTGNAPAILQVYEVGTATMMASKAIKP
       ℓℓℓℓℓ   ───────▶     ℓℓℓℓℓℓℓℓℓ.ℓℓℓℓ      ───────▶  ℓℓℓℓℓℓℓ   TT───▶
                β2              α2                β3        η1  α3    β4

            α4           η1      α5     β4   β5     β6       β7      α6
         ℓℓℓℓℓℓℓ       ℓℓℓℓ    ℓℓℓℓℓ    ──▶TT──▶  ───▶   ───────▶ ℓℓℓℓℓℓℓ
         120       130       140       150       160       170
Mtb    LDDLFGQVGVD..TTDYVDSLLADY..EFNGRHYAVPYARSTPLFYYNKAAWQQAGLP
Ecoli  VYDVFKEAGIQFDESQFVPTVSGYYSDSKTGHLLSQPFNSSTPLYYYNKDAFKKAGLD
       ℓℓℓℓℓℓℓℓℓ     ℓℓℓ   ℓℓℓℓℓℓ    TTT      ───────▶ ───────▶  ℓℓℓℓℓℓ
            α4           η2      α5          β5       β6       α6

                    α7              β8          α8        α9
            ℓℓℓℓℓℓℓℓℓℓℓℓℓℓ        ───▶       ℓℓℓℓ      ℓℓℓℓℓ
         180       190       200       210       220
Mtb    .DRGPQSWSEFDEWGPELQRVVGAGRSAHGWA...NADL.ISWT...FQGPNWAFGGA
Ecoli  PEQPPKTWQDLADYAAKLKASGMKCGYASGWQGWIQLENFSAWNGLPFASKNNGFDGT
       TT      ℓℓℓℓℓℓℓℓℓℓℓℓℓℓ      ───────▶    ℓℓℓℓℓℓℓ ℓℓℓℓℓ    ℓℓℓℓ  T
                    α7              β7      α8      α9       α10

         β9   β10           α10             β11     α11        β12    η2
         ──▶TT──▶     ℓℓℓℓℓℓℓℓℓℓℓ..ℓℓ     ───▶   ℓℓℓℓℓ      ───▶  ℓℓℓℓ
              230       240       250       260       270
Mtb    YSDKWTLTLTEPATIAAGNFYRN..SIHGKGYAAVANDIANEFATGILASAVASTGSL
Ecoli  ...DAVLEFNKPEQVKHIAMLEEMNKKGDFSYVGRKDESTEKFYNGDCAMTTASSGSL
       ...T        ℓℓℓℓℓℓℓℓℓℓℓ            ℓℓℓℓℓ      ───▶  ℓℓℓℓ
                    α11                     α12     β8  η3

         α12        β13      η3            β14     TT     α13
         ℓℓℓℓℓℓ    ───────▶  ℓℓℓ        ───────▶      ℓℓℓℓℓℓℓℓℓℓℓℓℓ
         280       290       300       310       320       330
Mtb    AGITASARFDFGAAPLPTGPDAAP...ACPTGGAGLAIPAKLSEERKVNALKFIAFVT
Ecoli  ANIREYAKFNYGVGMMPYDADAKDAPQNAIIGGASLWVMQGKDKETYTGVAKFLDFLA
       ℓℓℓℓℓℓℓ    ───────▶    TT TT        ───────▶      ℓℓℓℓℓℓℓℓℓℓ
            α13        β9                     β10           α14

            α14              η4      α15        α16      η5        α17
         ℓℓℓℓℓℓℓℓℓℓℓℓ        ℓℓℓℓ   ℓℓℓℓℓℓℓ   ℓℓℓℓℓℓℓℓℓℓ    ℓℓℓℓℓ
              340       350       360       370       380
Mtb    NPTNTAYFSQQTGYLPVRKSAVDDASERHYLADNPRARVALDQLPHTRTQDYARVFL.
Ecoli  KPENAAEWHQKTGYLPITKAAYDLTREQGFYEKNPGADTATRQMLNKPPLPFTKGLRL
       ℓℓℓℓℓℓℓℓℓ     ℓℓℓℓℓℓℓℓℓℓℓ ℓℓℓℓ ℓℓℓℓℓℓℓℓℓℓ   TT
            α15           α16        α17 η4    α18

                    α18            α19              η6
         ℓℓℓℓℓℓℓℓℓℓℓℓℓ        ℓℓℓℓℓℓℓℓℓℓℓℓ     ℓℓℓℓ
         390       400       410       420       430
Mtb    ...PGGDRIISAGLESIGLRGADVTKTFTNIQKRLQVILDRQIMRKLAGHG..
Ecoli  GNMPQIRVIVDEELESVWTGKKTPQQALDTA.....VERGNQLLRRFEKSTKS
       ℓℓℓℓℓℓℓℓℓℓℓℓ        ℓℓℓℓℓℓℓℓℓ....ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ
            α19                    α20
```
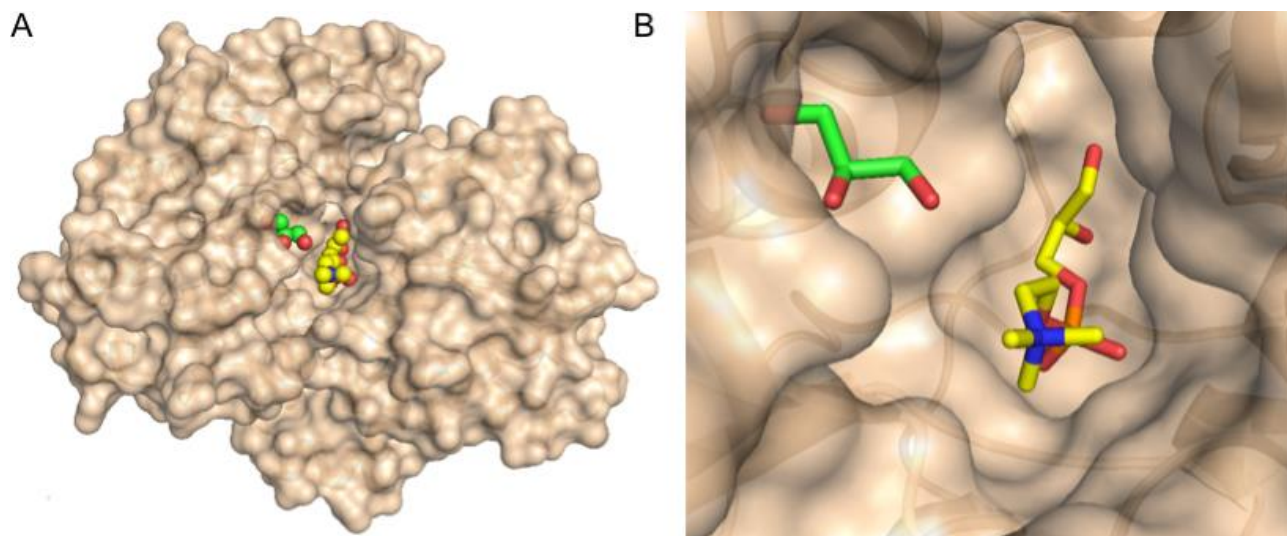
10

Fig S4: Location of additional glycerol moiety in the *Mtb* UgpB binding pocket. A) Surface representation of *Mtb* UgpB. The GPC ligand is represented by yellow spheres and a solvent glycerol moiety as green spheres and the. B) Close-up of the *Mtb* UgpB binding pocket with the GPC ligand and glycerol moiety shown in stick representation.

Fig S5. Thermal shift assay probing a panel of potential *Mtb* UgpB ligands. Bar graphs illustrating shifts of ΔT_m for the series of potential ligands. Thirty seven different ligands were probed for binding at a final concentration of 100mM. Data are shown from three independent repeats represented as mean ± SD.
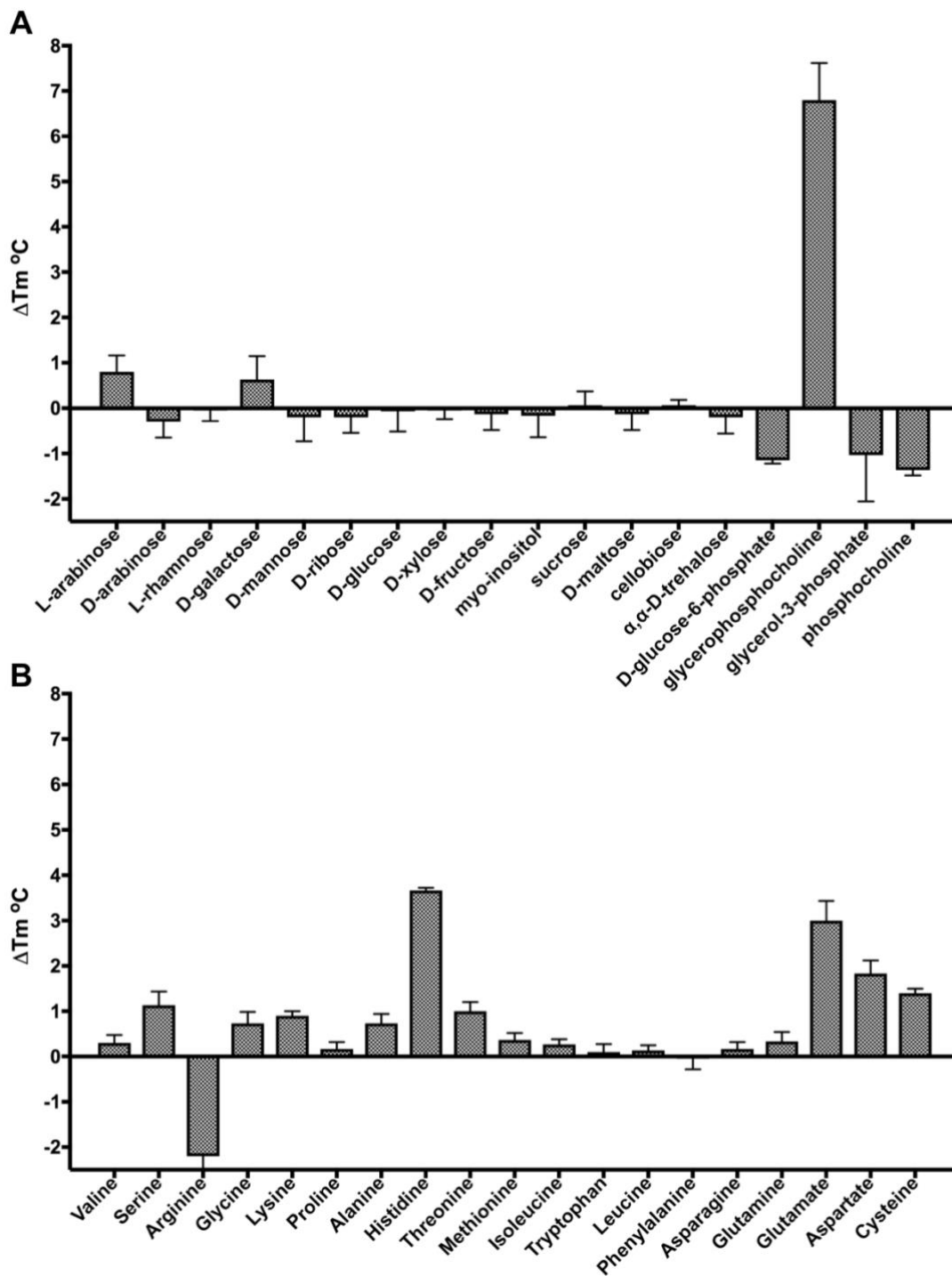
Fig S6. CD spectra of *Mtb* UgpB and site directed mutant proteins. CD spectra of *Mtb* UgpB (red)), *Mtb* UgpB Tyr78Ala (green), *Mtb* UgpB Asp102Ala (cyan), *Mtb* UgpB Ser153Ala (purple), *Mtb* UgpB Leu205Ala (magenta), *Mtb* UgpB Trp208Ala (brown), *Mtb* UgpB Ser272Ala (orange), Tyr345Ala (blue), *Mtb* UgpB Arg385Ala (yellow).
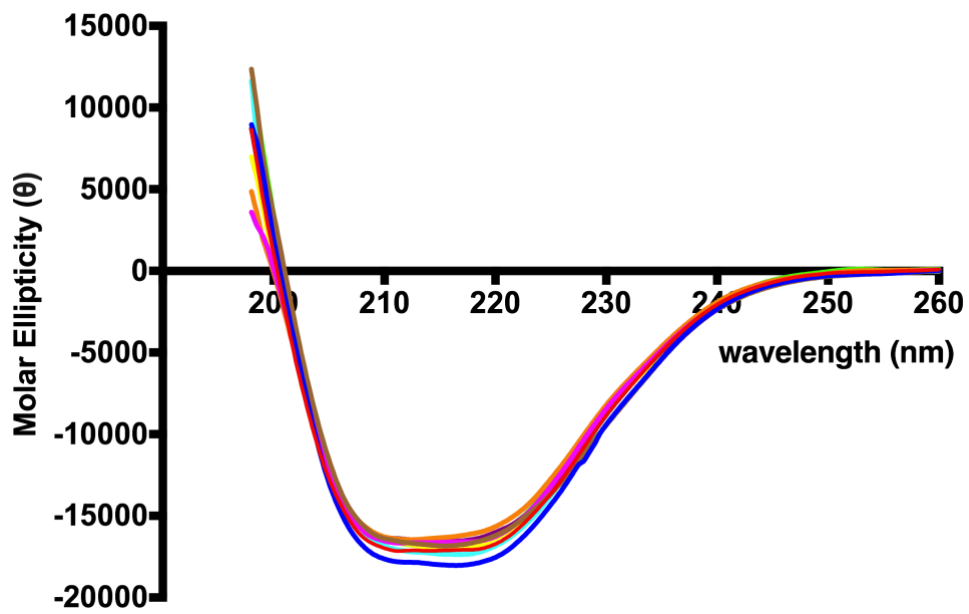
Table S1. Crystallographic parameters for *Mtb* UgpB in complex with GPC

PDB ID	UgpB-GPC 6R1B
Data collection	
Beam line	Diamond I04-1
Wavelength (Å)	0.92
Space group	P 2_1 2_1 2_1
Unit cell parameters	
a (Å)	169.9
b (Å)	213.3
c (Å)	46.1
α	90
β	90
γ	90
Molecules in ASU	4
Resolution (Å)	38-2.27
(Outer shell)[a]	(2.33-2.27)
Unique reflections	78,213 (5,626)
Multiplicity	7.8 (6.7)
CC$_{1/2}$	0.997 (0.476)
Completeness (%)[a]	99.2 (97.5)
R$_{merge}$ (%)[a]	13.4 (5.1)
Mean I/σ(I)[a]	10.3 (1.7)
Refinement	
R work (%)	20.6
R free (%)	25.6
r.m.s.d	
Bond lengths (Å)	0.006
Bond angles (degrees)	0.66
No. of non-hydrogen atoms	
Protein atoms	11,953
Ligand/Ions	13
Solvent waters	259
Average B factors (Å^2)	
Overall	54.2
Protein	54.5
Ligand/Ions	42.7
Solvent	45.0
Ramachandran plot[b]	
Favoured region (%)	96.3
Allowed region (%)	3.72
Outer region (%)	0

[a]Numbers in parentheses refer to the highest-resolution shell.
[b]Ramachandran plot statistics were calculated by MolProbity.

Table S2. DynDom analysis of *Mtb* UgpB (pdb 4MFI) and *Mtb* UgpB in complex with GPC

Backbone RMSD (Å)	Bending region	Rotation angle	Translation (Å)	Closure (%)
0.47 (Domain I)	152-153	21.8°	0.8	98.7
0.54 (Domain II)	304-306			
	362-372			

Domain I comprises residues 38-152 and 306-365 and Domain II comprises residues 153-305 and 366-426

Table S3. Sequence of primers for cloning and site-directed mutagenesis

Restriction recognition sites are in italics. The codon encoding the amino acid mutation is indicated in bold type.

Name	Use	Sequence (5'-3')
UgpB_T_pYUB_5	Clone truncated *Mtb* UgpB pYUB1062	aaaaaa*catat*gggttccggcccaatcgacttctgg
UgpB_T_pYUB_3	Clone truncated *Mtb* UgpB pYUB1062	aaaaaa*aagctt*gccatgccccgccagcttccg
Tyr78Ala_F	Mutate *Mtb* UgpB residue Tyr78Ala	ggcaaggac**gcc**gacgaggtg
Tyr78Ala_R	Mutate *Mtb* UgpB residue Tyr78Ala	cacctcgtc**ggc**gtccttgcc
Asp102Ala_F	Mutate *Mtb* UgpB residue Asp102Ala	cgttttgctcgac**gcc**cgatggtggttcc
Asp102Ala_R	Mutate *Mtb* UgpB residue Asp102Ala	ggaaccaccatcg**ggc**gtcgagcaaaacg
Ser153Ala_F	Mutate *Mtb* UgpB residue Ser153Ala	ccgtatgctcgc**gcg**acgccgctgttc
Ser153Ala_R	Mutate *Mtb* UgpB residue Ser153Ala	gaacagcggcgt**cgc**gcgagcatacgg
Leu205Ala_F	Mutate *Mtb* UgpB residue Leu205Ala	gctaacgccgac**gcc**atctcgtggacg
Leu205Ala_R	Mutate *Mtb* UgpB residue Leu205Ala	cgtccacgagat**ggc**gtcggcgttagc
Trp208Ala_F	Mutate *Mtb* UgpB residue Trp208Ala	ccgacctcatctc**ggc**gacgtttcagggacc
Trp208Ala_R	Mutate *Mtb* UgpB residue Trp208Ala	ggtccctgaaacgtc**gcc**gagatgaggtcgg
Ser272Ala_F	Mutate *Mtb* UgpB residue Ser272Ala	gccgtggca**gcc**accggctcg
Ser272Ala_R	Mutate *Mtb* UgpB residue Ser272Ala	cgagccggt**ggc**tgccacggc
Tyr345Ala_F	Mutate *Mtb* UgpB residue Tyr345Ala	cagcaaaccggc**gct**ctgccggtgcgcaag
Tyr345Ala_R	Mutate *Mtb* UgpB residue Tyr345Ala	cttgcgcaccggcag**agc**gccggtttgctg
Arg385Ala_F	Mutate *Mtb* UgpB residue Arg385Ala	cacaagactacgca**gcg**gttttcctgcc
Arg385Ala_R	Mutate *Mtb* UgpB residue Arg385Ala	ggcaggaaaac**cgc**tgcgtagtcttgtg

References

1. Shaw, N., Cheng, C., and Liu, Z.-J. (2007) Procedure for reductive methylation of protein to improve crystallizability, *Nature Protocols*.
2. Kabsch, W. (2010) XDS, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 125-132.
3. Collaborative Computational Project, N. (1994) The CCP4 suite: programs for protein crystallography, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 50*, 760-763.
4. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J. Appl. Crystallogr. 40*, 658-674.
5. Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 64*, 61-69.
6. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 2126-2132.
7. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine, *Acta Crystallogr., Sect. D: Biol. Crystallogr.68*, 352-367.
8. Moriarty, N. W., Draizen, E. J., and Adams, P. D. (2017) An editor for the generation and customization of geometry restraints, *Acta Crystallogr., Sect. D: Struct. Biol. 73*, 123-130.
9. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr., Sect. D: Biol. Crystallogr.66*, 12-21.
10. McNicholas, S., Potterton, E., Wilson, K. S., and Noble, M. E. (2011) Presenting your structures: the CCP4mg molecular-graphics software, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 67*, 386-394.
11. Hayward, S., and Lee, R. A. (2002) Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50, *J. Mol. Graphics Modell. 21*, 181-183.
12. Piotto, M., Saudek, V., and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions, *J. Biomol. NMR 2*, 661-665.
13. Hwang, T. L., and Shaka, A. J. (1995) Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients, *Journal of Magnetic Resonance, Series A 112*, 275-279.
14. Mayer, M., and James, T. L. (2004) NMR-based characterization of phenothiazines as a RNA binding scaffold, *J. Am. Chem. Soc. 126*, 4453-4460.
15. Xu, Y., Sugar, I. P., and Krishna, N. R. (1995) 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 5*, 37-48.
16. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, *J. Comput. Chem. 30*, 2785-2791.
17. Trott, O., and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J. Comput. Chem. 31*, 455-461.
18. Monaco, S., Tailford, L. E., Juge, N., and Angulo, J. (2017) Differential Epitope Mapping by STD NMR Spectroscopy To Reveal the Nature of Protein-Ligand Contacts, *Angew Chem Int Ed Engl 56*, 15289-15293.
19. Han, B., Liu, Y., Ginzinger, S. W., and Wishart, D. S. (2011) SHIFTX2: significantly improved protein chemical shift prediction, *J. Biomol. NMR 50*, 43-57.