Mapping the architecture of the initiating phosphoglycosyl transferase from *S. enterica* O-antigen biosynthesis in a liponanoparticle.

Greg J. Dodge,[a] Alyssa J. Anderson,[a] Yi He[b], Rosa Viner[b], Barbara Imperiali[a,*]

[a]Department of Biology and Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

[b]Thermo Fisher Scientific, San Jose CA 95134, USA.

*Corresponding author: imper@mit.edu

ORCID IDs:

BI: 0000-0002-5749-7869

RV: 0000-0003-0550-5545

YH: 0000-0002-5565-5579

AJA: 0000-0003-4871-4283

GJD: 0000-0002-6555-8350

Abstract:

Bacterial cell surface glycoconjugates are critical for cell survival and for interactions between bacteria and their hosts. Consequently, the pathways responsible for their biosynthesis have untapped potential as therapeutic targets. The localization of many glycoconjugate biosynthesis enzymes to the membrane represents a significant challenge for expressing, purifying, and characterizing these enzymes. Here, we leverage cutting-edge methods to stabilize, purify, and structurally characterize WbaP, a phosphoglycosyl transferase (PGT) from *Salmonella enterica* (LT2) O-antigen biosynthesis without extraction from the lipid bilayer. Results from these experiments establish WbaP as a homodimer, reveal the structural elements responsible for oligomerization, shed light on the role of a domain of unknown function embedded within WbaP, and identify conserved structural motifs between PGTs and functionally unrelated UDP-sugar dehydratases. The approach developed here is generalizable and provides a toolkit for studying small membrane proteins embedded in liponanoparticles beyond PGTs.

Introduction:

Glycoconjugates are heterogeneous biopolymers found on the cell surface and secreted from all cells. In prokaryotes,[1] glycoconjugates play important roles in survival,[2] antibiotic resistance,[3] immune evasion and immunogenicity,[4,5] and biofilm formation.[6] Despite the overwhelming compositional diversity of prokaryotic glycoconjugates, the underlying logic for their biosynthesis is often conserved and most commonly localized to cell membranes.[7] Genetic disruption or inhibition of the enzymes involved in glycoconjugate assembly can result in deleterious effects in many microorganisms and have been reported to attenuate virulence and pathogenicity.[8-10] Thus, structural and mechanistic studies of the

machinery for glycoconjugate assembly are of considerable interest although the membrane localization of glycoconjugate assembly enzymes poses a significant hurdle for in-depth characterization.

The Wzx/Wzy-dependent pathway is highly conserved and comprises a series of enzyme-catalyzed transfer reactions using nucleoside diphosphate sugar substrates and results in the production of a wide array of glycoconjugates.[11] O-antigen biosynthesis in *Salmonella enterica* serovar Typhimurium LT2 is a well-studied example of a Wzx/Wzy-dependent glycoconjugate biosynthesis pathway.[12-14] This pathway utilizes a phosphoglycosyl transferase (PGT) known as WbaP (formerly RfbP) to catalyze the transfer of phospho-galactose from UDP-galactose onto undecaprenol phosphate (UndP), forming UndPP-galactose.[15-19] This reaction represents the initial membrane-committed step in O-antigen repeat unit (RU) biosynthesis and is followed by a series of glycosyl transfer reactions to form the O-antigen RU. The tetrasaccharide RU is then flipped across the inner bacterial membrane, polymerized into full-length O-antigen, and appended to the lipid A core for display on the outer membrane (Fig. S1).[7,11,20-22]

PGTs belong to one of two superfamilies based on the topology of the catalytic core structure. Polytopic PGTs (polyPGTs), such as MraY, have a catalytic domain comprising multiple transmembrane helices (TMHs) and the catalytic core of monotopic PGTs (monoPGTs) contains a single re-entrant membrane helix (RMH).[23] Analysis of the monoPGT superfamily utilizing a sequence similarity network (SSN) recently classified several families containing variations around the core catalytic domain.[24] Small monoPGTs (Sm-PGTs) feature only the catalytic core, such as PglC from the *Campylobacter* protein glycosylation pathways.[25,26] Bifunctional monoPGTs (Bi-PGTs) include functional domains fused to either the N- or C-terminus of the monoPGT catalytic domain. Finally, large monoPGTs (Lg-PGTs), which are the most abundant, contain a putative four-transmembrane helix bundle and a domain of unknown function (DUF) in addition to the monoPGT core catalytic domain. The role of the accessory domains beyond the catalytic domain in the Lg-PGTs remains unclear. The *S. enterica* WbaP is a prototypic Lg-PGT that has been biochemically characterized.[15-19,27] Additionally, studies of a WbaP ortholog from *Streptococcus*

pneumoniae suggest that residues in the DUF may modulate PGT function or overall pathway flux.[28,29] Initial computational assignment of the topology of Lg-PGTs mis-assigned the topology of the reentrant membrane helix (RMH) as a transmembrane helix (TMH), confounding interpretations of biological data, as well as the subcellular localization of the DUF in these proteins.[29-31] Currently, although the structure of PglC, a sm-PGT from *Campylobacter concisus* has been determined by X-ray crystallography,[26] there is no experimental structure determination of a Lg-PGT to provide insight into the roles of the auxiliary domains or their interactions with the catalytic core. Furthermore, PglC was purified using detergent, potentially obscuring any interactions, which may depend on the placement of PglC within the phospholipid bilayer.

We have recently reported a robust pipeline for solubilizing and purifying Lg-PGTs from several organisms, including WbaP from *S. enterica*.[27] The strategy includes *in vivo* cleavage of an N-terminal SUMO tag via the protease Ulp1, application of the amphiphilic polymer SMALP-200 (formerly SMA30) for solubilization from *E. coli* membranes into styrene-maleic acid liponanoparticles (SMALPs), and purification via a dual-strep tag, which is exposed upon cleavage by Ulp1. The resulting protein is highly pure and stabilized in a native-like membrane environment by the SMALP. WbaP assembled in SMALP was vitrified on grids for CryoEM screening, and an initial dataset yielded 2D classes with unexpected symmetry. As these experiments were the first to study purified Lg-PGTs in a native-like lipid bilayer, the putative symmetry observed in SMALP prompted the investigation of the oligomeric state of Lg-PGTs in the liponanoparticles.

Here, we build upon the successful membrane protein solubilization techniques to further investigate the architecture of *S. enterica* WbaP in SMALP. We leverage AlphaFold modeling, crosslinking experiments, mass spectrometry, and CryoEM to map the structure of *S. enterica* WbaP in a native-like lipid bilayer environment. Insights from these experiments establish the oligomeric state of *S. enterica* WbaP as a dimer, facilitate the production of a soluble truncation construct to probe the role of the DUF, and highlight evolutionary relationships between Lg-PGTs and other enzymes from glycoconjugate

biosynthetic pathways. These findings demonstrate SMALP as a useful platform for capturing and studying native oligomerization and for CryoEM of small membrane proteins. These structural studies now allow us to develop a clearer picture of the initial steps of glycoconjugate biosynthesis and membrane positioning of Lg-PGTs. Finally, experiments to assess DUF binding assign this domain as a uridine nucleotide binder, implicating the DUF as a potential modulator of Lg-PGT function.

Materials and Methods:

Cloning and expression:

Expression of full-length Lg-PGTs

WbaP from *S. enterica* LT2 (Uniprot: P26406) was purified as previously described.[27] Briefly, C43 cells harboring pAM174[32] were transformed with a plasmid encoding *S. enterica* WbaP with an N-terminal SUMO tag – linker – dual-strep tag sequence. Protein was expressed using autoinduction[33] in 0.5 L terrific broth[34] supplemented with 150 µg/mL kanamycin and 25 µg/mL chloramphenicol. Cells were incubated at 37° C until the OD_{600} reached ~1.5, and then the temperature was adjusted to 18° C and 1 g solid (L)-arabinose was added to each culture. After 18-20 hr expression, cells were pelleted via centrifugation, pellets were transferred to 1 gal Ziplock® freezer bags, manually spread to a uniform thin layer, and frozen at -80° C.

WbaP DUF expression and purification

The region encoding residues 146-272 from *S. enterica* WbaP was ordered as a synthetic gene, and cloned into pMCSG7[35] using Gibson assembly.[36] The resulting plasmid was transformed into *E. coli* Bl21 (DE3), and protein was expressed using autoinduction[33] in 0.5 L cultures supplemented with 100 µg/mL ampicillin at 37 ° C. The temperature was adjusted to 18° C once the OD_{600} reached 1.5, and cultures were incubated

for 18 hr. After expression, cells were harvested via centrifugation, transferred to 1-gallon Ziplock® freezer bags, spread to a thin layer, and stored at -80 ° C.

Mutagenesis

S. enterica WbaP Cys-pair variants were generated using primers designed in the quikchange primer tool (Agilent). Successful mutations were confirmed by Sanger sequencing. Forward and reverse primers are shown below.

Mutant	Primers (5'-3')
T239C	5'-GCGAAATGACGGAACTACGCAGACGGAGCGACAGTGATG-3' 5'-CATCACTGTCGCTCCGTCTGCGTAGTTCCGTCATTTCGC-3'
V240C	5'-GCGAAATGACGGAACGCATGTGACGGAGCGACAGTGATG-3' 5'-CATCACTGTCGCTCCGTCACATGCGTTCCGTCATTTCGC-3'
L265C	5'-CAGGTTATTCTGAATGCGGCACAACATCACCTCGTGGGAG-3' 5'-CTCCCACGAGGTGATGTTGTGCCGCATTCAGAATAACCTG-3'
R266C	5'-GGTTATTCTGAATGCAAAGCAACATCACCTCGTGG-3' 5'-CCACGAGGTGATGTTGCTTTGCATTCAGAATAACC-3'

Protein Purification:

WbaP in SMALP

WbaP was purified in SMALP200 liponanoparticles as previously described.[27] Frozen cells were resuspended in buffer A (50 mM HEPES pH 8.0, 300 mM NaCl) at 4 mL per g pellet. Resuspended pellets were supplemented with 2 mM $MgCl_2$, 0.06 mg/mL lysozyme (RPI), and 0.5 mg/mL DNase I (Millipore Sigma) and incubated on ice for 30 minutes. Cells were disrupted via sonication (2x 90 s, 50% amplitude,

1 s on 2 s off), and cell membrane was isolated via differential centrifugation (Ti-45 rotor, 9,000g 45 min, reserve supernatant, 140,000g 65 min). Isolated membranes were diluted to 50 mg/mL using buffer A (assessed by UV absorbance at 280 nm), flash frozen in liquid N_2, and stored at -80° C. Isolated membranes were thawed on ice, and mixed 1:1 (v/v) with a 2% stock solution of SMALP200 (Polyscope) or buffer A and rotated at room temperature for 1 hour. Soluble liponanoparticles were isolated by centrifugation (Ti45 rotor, 160,000 xg 65 min). After centrifugation, pellets were discarded, and the supernatant was flowed over 1 mL StreptactinXT 4flow resin (IBA Biosciences) pre-equilibrated with buffer A. The flowthrough was re-run over the column bed, and the column was washed with 5 mL buffer A. Protein was eluted using 3 mL Buffer A + 50 mM Biotin (buffer B). Protein-containing fractions were identified by UV-vis (Nanodrop) and pooled. Biotin was removed using 3x tandem 5 mL HiTrap desalting columns (Cytiva) equilibrated with 25 mM HEPES pH 8.0, 150 mM NaCl (buffer C). Protein purity was assessed via SDS-PAGE, and protein concentration was determined via BCA assay (Pierce). Protein was concentrated to 4 mg/mL, and flash frozen in liquid N_2, then stored at -80° C.

Soluble WbaP truncation:

All purification steps were performed on ice. Frozen cell pellets were resuspended in buffer D (50 mM HEPES pH 7.5, 300 mM NaCl, 20 mM imidazole pH 7.5, 5 % glycerol) at 4 mL per g cell pellet. Resuspended pellets were brought to 2 mM $MgCl_2$, 0.06 mg/mL lysozyme (RPI), and 0.5 mg/mL DNase I (Millipore Sigma). Cells were incubated on ice for 30 minutes. Cells were then disrupted via sonication (2x 90 s, 50% amplitude, 1 s on 2 s off). Insoluble material was removed via centrifugation (Ti45 rotor, 42,000 RPM 60 min). The supernatant was sterile filtered and loaded onto a 5 mL NiNTA His-Trap column using an Akta FPLC pre-equilibrated with buffer D. Protein was eluted using a linear gradient from 0 – 100 % buffer E (50 mM HEPES pH 7.5, 300 mM NaCl, 400 mM imidazole pH 7.5, 5 % glycerol) over 5 column volumes. Peak fractions were pooled, supplemented with TEV protease at a 1:7 TEV:protein ratio, and dialyzed overnight at 4°C against 3 L buffer F (25 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol). Inverse NiNTA

purification was used to remove the TEV protease and the cleaved 6xHis tag. Flowthrough and low-imidazole wash from inverse NiNTA purification were pooled, concentrated to 5 mL, and injected on a Superdex S200 column pre-equilibrated with buffer F (25 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol). Peak fractions were analyzed via SDS-PAGE, and fractions containing the *S. enterica* WbaP DUF were pooled, concentrated to 2 mg/mL, flash frozen in liquid N_2, and stored at -80° C.

Characterization of Lg-PGTs in SMALP

Size exclusion chromatography

S. enterica WbaP in SMALP was analyzed to assess monodispersity via size exclusion chromatography (SEC) using an Enrich S650 column (Biorad) pre-equilibrated with buffer C. A 500 µL sample was injected, and peak fractions were pooled and concentrated to 10 mg/mL.

Size exclusion chromatography with multi-angle light scattering (SEC-MALS)

A 50 µL aliquot of 0.5 mg/mL *S. enterica* WbaP in SMALP was injected onto a WTC-030 fused silica column pre-equilibrated with HEPES buffered saline pH 7.5 (HBS). Eluate was flowed through a DAWN MALS detector (Wyatt) and an Optilab differential refractive index detector (Wyatt). Data were analyzed using ASTRA software (Wyatt), and the system was pre-calibrated with a BSA standard.

Mass photometry

The *S. enterica* WbaP in SMALP was diluted to 40 nM in buffer C to a final volume of 50 µL. A CultureWell™ gasket (Grace Bio-Labs) was attached to a 24 mm x 50 mm glass coverslip (Electron Microscopy Sciences), and 10 µL buffer C was dispensed into a well to acquire focus. A 10 µL aliquot of 40 nM WbaP solution was added to the buffer well to achieve a final concentration of 20 nM WbaP. Following this 120 s data was collected, and the data was analyzed using DiscoverMP (Refyn Ltd). A standard curve was generated using NativeMark™ unstained protein ladder (Thermo Fisher Scientific).

Crosslinking experiments:

Crosslinking Panel:

SMALPs containing S. enterica WbaP were reacted with a panel of lysine-reactive crosslinkers with variable chemistries, solubilities, and spacer lengths (Table S1). Stock solutions were made of each crosslinker as follows: 50 mM dimethyl adipimidate (DMA) in water, 50 mM dimethyl suberimidate (DMS) in water, 25 mM bis[2-(succinimidyloxycarbonyloxy)ethyl]sulfone (BSOCOES) in dry DMSO, and 10 mM 3,3'-dithiobis(sulfosuccinimidyl propionate) (DTSSP) in water. Crosslinkers were immediately added to 15 µM WbaP in buffer C at a 20-40 fold molar excess. After incubation at room temperature for 30 min, the reaction was quenched by the addition of 1 µL of 1.5 M Tris-HCl, pH 8. Samples were analyzed by Western blot to detect the presence of crosslinked oligomers. Crosslinking efficiency was annotated as high (>30 % total protein crosslinked to dimer), moderate (<30 % total protein crosslinked to dimer), or none (no protein crosslinked to dimer).

Dithiobis(succinimidylpropionate) (DSP) Crosslinking:

DSP was dissolved in dry DMSO to a stock concentration of 25 mM and immediately added to a final concentration of 0.1 – 5 mM to 20 µL of WbaP in SMALP. After incubation at room temperature for 30 min, the reaction was quenched by the addition of 1 µL of 1.5 M Tris-HCl, pH 8. The sample was then divided into two aliquots and reducing or nonreducing loading dye was added to each sample. Crosslinking was analyzed by SDS-PAGE followed by anti-Strep Western blot analysis.

Cys-Pair variant Crosslinking:

A 2.5 mM stock solution of dibromobimane (bBBr) in 20% acetonitrile was prepared. bBBr was added to 15 µM WbaP to a final concentration of either 75 or 300 µM. After incubation at room temperature for 20 min, loading dye was added and SDS-PAGE was performed to assess the WbaP oligomerization state.

A 10 mM 1,10-phenanthroline DMSO stock and a 5 mM $CuSO_4$ aqueous stock were prepared. $CuSO_4$ and 1,10 phenanthroline were added to 20 µL of 15 µM WbaP to final concentrations of 50 and 100 µM or 200 and 400 µM, respectively. The reactions were allowed to proceed at room temperature for 20 min, followed by the addition of non-reducing loading dye and SDS-PAGE analysis.

Structural biology experiments

Electron Microscopy Details

Data were collected at the MIT Characterization.nano facility. Optimal freezing conditions were screened using the Vitrobot system (Thermo Scientific). Grids were clipped and screened on a Talos Arctica G2 (Thermo Scientific) equipped with a Falcon 3EC camera (Thermo Scientific). Optimal particles were observed in grid GD5-3, containing 2 mg/mL *S. enterica* WbaP. Based on these preliminary results, 4,481 movies of SMA30-solubilized *S. enterica* WbaP were collected on a Titan Krios G3i (Thermo Scientific, Table S2) equipped with a K3 camera (Gatan, Inc). Motion correction, CTF estimation, particle extraction, *ab-initio* reconstruction, and all refinement were conducted using Cryosparc V4.[37,38]

AlphaFold Modelling

Lg-PGTs from *T. thermophilus* (Uniprot: A0A510HWX9), *A. hydrophila* (Uniprot: B3FN88), *E. coli* (Uniprot: P71241), and *S. pneumoniae* (Uniprot: Q9ZII5) were selected for structure prediction. AlphaFold models were generated as described previously.[27] Sequences containing two chains of each Lg-PGT were used as input for a local installation of colabfold,[39] and predictions were generated using the flags --amber, --templates, --num-recycle 3, --use-gpu-relax activated.

Structure Refinement and modeling

PHENIX version 1.20.1 was used to process the AlphaFold prediction, dock, rebuild, and real-space refine the processed model, and validate the resulting structural model of *S. enterica* WbaP.[40,41] UCSF ChimeraX

was used to visualize CryoEM volumes.[42] Structural figures were generated using PyMOL.[43] The protein topology cartoon was generated using Pro-origami.[44] The EMBL PISA server was used to analyze the dimer interface.[45]

Mass spectrometry

Cross-linking, protein digestion and IMAC enrichment.

S. enterica WbaP in SMALP nanoparticles was prepared in 20 mM HEPES buffer pH 7.0 at a concentration of 0.6 mg/mL. 5 mM disuccinimidyl suberate (DSS) or 2 mM tert-butyl disuccinimidyl phenyl phosphonate (tBu-PhoX)(Thermo Fisher Scientific) in DMSO was added to the solution and incubated for 1 hr at room temperature. Reactions were quenched with 20 mM of Tris-HCl, pH 8.0, for 15 minutes. N-dodecyl β-D-maltoside (DDM,10% stock, Thermo Fisher Scientific) was added to the sample to a final concentration of 1% and the mixture was incubated on ice for 30 min. To remove SMALP, $MgCl_2$ (50 mM stock) was added to a final concentration of 4 mM, and the resulting solution incubated at 4 °C for 1 hr, and then centrifuged at 21,000 ×g at 4 °C for 1 hr. The supernatants were transferred to fresh microfuge tubes and diluted 1:1 with 0.1% SDS, 25 mM DTT and incubated at 50 °C for 1 hr. Chloroacetamide was added to 25 mM and incubated at RT for 30 min in the dark before acetone precipitation overnight at -20 ºC. The samples were washed twice with 90% acetone and the pellet was vortexed with 25 mM ammonium bicarbonate until re-solubilized. Enzymatic digestion was carried out with either trypsin in 0.1% Rapigest (Waters) (1:20 ratio) or pepsin (1:50 ratio). The trypsin digestion was stopped after 16 h with 1% formic acid (FA). Cross-linked peptides were desalted using Pierce peptide desalting spin column (Thermo Fisher Scientific) and dried. The tBu-PhoX cross-linked peptides were enriched as described previously .[46]

LC-MS data acquisition

Samples were separated by reverse phase-HPLC using a Thermo Scientific™ Vanquish™ Neo system connected to an EASY-Spray™ PepMap™ 75 µm x 25 cm column over a 60 min 3-65% gradient (A: water,

0.1% formic acid; B: 80% acetonitrile, 0.1% formic acid) at 300 nL/min flow rate. The crosslinked samples were analyzed on the Orbitrap Eclipse™ Tribrid™ mass spectrometer with Instrument Control Software version 4.0. 0. Cross-linked samples were analyzed using an HCD-MS2 acquisition strategy with 30% normalized collision energy (NCE). MS1 and MS2 scans were acquired in the orbitrap with a respective mass resolution of 60,000 and 30,000. MS1 scan range was set to m/z 375 – 1400 at 100% AGC target, 118 ms maximum injection time and 60 s dynamic exclusion. MS2 scans were set to an AGC target of 200%, 70 ms injection time, isolation window 1.2 m/z. Only cross-linked precursors at charged states +3 to +8 were subjected to MS2.

Data analysis

Spectral raw data files were analyzed using Proteome Discoverer 3.0 software (Thermo Fisher Scientific) with XlinkX node 3.0 using the non-cleavable or non-cleavable fast search algorithms for cross-linked peptides and SEQUEST HT search engine for unmodified peptides and loop-links/mono-links. MS1 ion mass tolerance: 10 ppm; MS2 ion mass tolerance: 20 ppm. Maximal number of missed cleavages: 2; minimum peptide length: 6; max. modifications: 4; peptide mass: 500 – 8,000 Da. Carbamidomethylation (+57.021 Da) of cysteines was used as a static modification. PhoX or DSS cross-linked mass modifications for lysine, the protein N-terminus and methionine oxidation (+15.995 Da) were used as variable modifications. Data were searched for cross-links against a protein database generated from protein identifications using WbaP and the *E. coli* proteome retrieved from Uniprot as search space. The false discovery rate (FDR) was set to 1% at CSM and cross-link levels. The maximum XlinkX score was set to be greater or equal to 40. Post-processing and visualization were carried out using XMAS plug-in for ChimeraX.[42,47]

Nano differential scanning fluorimetry (nDSF) of truncated S. enterica WbaP:

Data were collected using a Prometheus Panta nDSF instrument (Nanotemper). An initial panel of 20 nucleotides was chosen to screen for potential DUF ligands (Table S3). Truncated DUF at 1 mg/mL was diluted 1:1 with buffer C supplemented with 400 µM ligand, resulting in a final protein concentration of 0.5 mg/mL and 200 µM ligand. Data were processed using Pr.Analysis software (Nanotemper).

WbaP inhibition assay

WbaP activity assays were carried out as described previously,[27] with the following modifications. Prior to the addition of [^{3}H]-labeled UDP-Gal, reactions were incubated at room temperature with UMP for 5 minutes. A 2-fold serial dilution of UMP concentrations from 100 µM to 0.781 µM was utilized to assess UMP inhibition compared to a control reaction containing no additional UMP. Reactions were quenched after ten minutes. Data were collected in triplicate, normalized to percent activity of the control, and fit using the [inhibitor] vs normalized response –variable slope function in GraphPad Prism Version 9.5.1 for Windows, GraphPad Software, San Diego California, USA, www.graphpad.com.

Figures were generated using BioRender.

Data availability

CryoEM density has been deposited in the Electron Microscopy Data Bank with the accession code EMD-41042. The corresponding coordinates have been deposited in the Protein Data Bank with the accession code 8T53. Mass spectrometry data, full-length WbaP dimer model generated by AlphaFold, and Foldseek output are available on Mendeley: doi:10.17632/zddxv2k83x.1.

Results

Production of Lg-PGTs in Styrene maleic acid liponanoparticles (SMALP):

Characterization of S. enterica WbaP in SMALP:

We took a multifaceted approach to investigate the putative oligomerization of *S. enterica* in liponanoparticles (Figs. 1, S2). As SMALPs often display high heterogeneity we turned to size-exclusion chromatography (SEC) to remove aggregated material or free SMA polymer that had carried over through initial purification. A small peak was observed corresponding to the void volume of the column, but most of the material eluted in a sharp peak with a retention time of 11.09 min corresponding to a molecular weight of 464 kDa (Fig. S2). The predicted molecular weight of the expression construct of WbaP is 61.2 kDa. However, as the SMA and lipid accompanying WbaP in the SMALP may influence separation on SEC relative to globular standards, we sought alternative methods to accurately determine the mass of the WbaP SMALPs in solution. Both size exclusion chromatography with multi-angle light scattering (SEC-MALS) and mass photometry were utilized to assess the molar mass of WbaP in SMALP. In agreement with the initial SEC experiments, WbaP eluted in a sharp peak on an analytical SEC column in line with MALS (Fig. 1A). The polydispersity of this peak was calculated to 1.001, and the molar mass was calculated to 255 ± 9 kDa. Accurate assignment of mass of a species analyzed by SEC-MALS requires a precise assignment of the refractive index increment (dn/dc) of each component of the analyte.[48] Despite the low polydispersity and narrow mass range assigned to this peak, the refractive index increment (dn/dc) for SMA polymers is under debate in the literature,[49] obscuring the true mass calculation for WbaP in SMALP. To this end we turned to mass photometry, a technique that can accurately determine the mass of molecules in solution regardless of their composition.[50] When analyzed by mass photometry, WbaP-containing SMALPs were determined to have a molecular weight of 239 ± 28 kDa (Fig. 1B). As the monomer molecular weight of the WbaP construct was calculated to 61.2 kDa, and the SMA along with encapsulated lipids can add a variable amount of mass depending on the solubilized protein, we reasoned that WbaP could exist as a dimer or trimer in SMALP.

To further define the oligomeric state of WbaP in SMALP, we utilized amine-reactive crosslinking probes. By selecting crosslinkers with a defined radius, we reasoned that we could covalently trap oligomers and

distinguish dimer from trimer using non-reducing sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. S3A). A small panel of commonly used crosslinkers was selected to screen against WbaP in SMALP (Fig S3B, Table S1). Of these, the crosslinker dithiobis(succinimidylpropionate) (DSP), with a spacer length of 12.0 Å, readily crosslinked WbaP (Fig. 1C, Table S1). Crosslinking efficiency was dependent on DSP concentration, with complete crosslinking observed with 5 mM DSP after a five-minute reaction. Crosslinking could be reversed by the addition of reducing agent. Fully crosslinked WbaP migrated at ~2x the molecular weight of the non-crosslinked species, confirming a WbaP dimer in SMALP.

CryoEM of S. enterica WbaP in SMALP

With high confidence in the monodispersity of our sample and the oligomeric state of *S. enterica* WbaP in SMALP, we continued with CryoEM structure analysis. Prior to this work, only a small screening dataset had been collected on a Talos Arctica instrument.[27] Grids were prepared using the highly purified eluent from SEC, movies were collected using a Titan Krios G3i microscope. 2D classes from extracted particles closely matched those observed earlier, and initial *ab-initio* 3D models displayed features consistent with a liponanoparticle, as well as two distinct lobes outside of the membrane (Fig. S4). However, various refinement strategies failed to yield even moderate-resolution reconstructions. Reasoning that the auto-generated refinement mask may include SMA and lipid regions, we sought to generate a reasonable model to assist in manual masking during data processing.

Creation and validation of S. enterica WbaP dimer model

We turned to AlphaFold to generate models suitable for docking and refinement.[51] Although the prediction of *S. enterica* WbaP available from the AlphaFold protein structure database has high confidence throughout most of the model (Figs. 2A, S5), we noted a region of moderate confidence between the DUF and the PGT catalytic domain. In particular, the placement of a predicted β-hairpin motif from Asn251 to Gln268 stood out as anomalous, as this motif was flipped away from the remainder of the

protein, approximately 25 Å from the DUF, and ~37 Å from the PGT domain (Fig. 2A). As AlphaFold multimer routine has been reported to successfully predict oligomeric complexes,[39,52] we assessed whether a dimer model of *S. enterica* WbaP would represent a better starting point for docking and refinement into the CryoEM reconstruction. The dimer was also predicted with high overall confidence (Figs. 2B, S5). Unexpectedly, the β-hairpin motif that had poor placement in the monomer model was predicted to mediate domain swapping within the dimer model,[53,54] with the β-hairpin from chain A continuing a β-sheet structure present in the DUF of chain B (Figs. 2B, 2C). The predicted local distance difference test (pLDDT) score assigned to each amino acid by AlphaFold can be used to judge the quality of a predicted model on a per-residue basis.[51,52] Amino acids with pLDDT scores > 70 are considered to have high confidence.[55] For the β-hairpin domain swap region, the average pLDDT score is 78, indicating high confidence in the placement of this motif (Fig. S5). Significantly, the prediction of structures of other Lg-PGTs, reveals that this β-hairpin domain swap within a homodimer was highly conserved (Fig. S6).

Based on the earlier success with crosslinking *S. enterica* WbaP in SMALP, we further investigated the predicted domain swapping using a targeted crosslinking strategy. A set of Cys-pair variants of WbaP were constructed such that thiol-reactive crosslinkers with limited crosslinking radii could be used to "staple" the WbaP dimer at the predicted domain swap interface (Figs. 3A,B). For these experiments, we employed two different crosslinking strategies. For the first Cys-crosslinking experiment, we utilized Cu phenanthroline to catalyze the direct oxidation of Cys residues to form a disulfide bond between Cys residues within 2.05 Å. For the second strategy, we utilized dibromobimane (bBBr) to form a covalent crosslink between Cys residues within 4.88 Å. We observed concentration-dependent crosslinking for two discrete sets of crosslinking pairs along the predicted domain swap interface using either Cu phenanthroline or bBBr (Figs. 3D, E). No crosslinking was observed for wild-type protein using either crosslinking strategy, demonstrating that there are no native cysteines within crosslinking distance, (Fig. 3E) as predicted by the AlphaFold dimer model.

As a final validation of the predicted dimer model, we used crosslinking mass spectrometry (XLMS). Both disuccinimidyl suberate (DSS) and tert-butyl disuccinimidyl phenyl phosphonate (tBu-PhoX) were chosen for crosslinking, as DSS is commonly used for XLMS,[56,57] and tBu-PhoX has membrane-penetrating properties (Fig. 4A).[46] After crosslinking, WbaP was exchanged into DDM micelles,[49] and subjected to protease digestion. In-solution digestion resulted in fragments covering 93.86% *S. enterica* WbaP, and 49 unique crosslinking sites were identified (Figs. 4B, S7). The majority of the DSS crosslinks localized to the solvent-accessible areas of the PGT domain, while several of the tBu-PhoX crosslinks localized to the membrane-adjacent regions at the putative dimer interface (Fig 4C). Inter vs intra crosslinkers were assigned by XMAS{Lagerwaard, 2022 #43} based on distance constraints for each crosslinker (Fig. 4A). The Cα−Cα distances bridged by either crosslinker should not exceed 35 Å. The observed inter-chain crosslinks agreed with the predicted dimer interface, with Lys273 proximal to Lys146, Lys148, and Lys81. In addition, Lys81 and Lys451 were within crosslinking range (Fig. 4D). With targeted crosslinking validating the predicted domain swapping and XLMS validating the dimeric oligomerization state, we proceeded with the AlphaFold dimer model for CryoEM data processing and refinement.

S. enterica WbaP Structure:

A soft mask was generated using the AlphaFold dimer model of WbaP to mask around phospholipid and SMA density in the best unmasked CryoEM reconstruction. Subsequent masked refinement and cleaning of particle stacks led to a 4.1 Å reconstruction from 196,663 particles (Figs. 5A,B). Docking and real-space refinement led to the placement of 796 out of 1056 residues of the dimer (Figs. 5A,B). Superposition between the refined model and the AlphaFold dimer prediction results in an RMSD of 2.83 Å, demonstrating close agreement between the two models (Fig. 5C).

The overall architecture of the WbaP dimer is T-shaped, with the two four-TMH bundles forming the central stalk, each capped by a DUF. The dimer interface comprises ~780 $Å^2$ of buried surface area,

involving 7.1% of the residues of the protein. A modest (176 $Å^2$) interface is found along the four TMH bundles, with the C-terminal half of helix 3 mediating contacts. While the C-terminal regions of helix 3 are within interacting distance, the N-termini (Pro82) are ~32 Å apart, creating a large void in the center of the dimer, presumably occupied by phospholipid. The remainder of the dimer interface occurs at the DUF domains, mediated almost entirely by the β-hairpin domain crossover discussed above. Based on previous experiments establishing the topology of monoPGT catalytic domains, the DUF domains cap the cytosolic side of the four TMH bundles.[26,30,31]

The catalytic domains of the Lg-PGT protrude away from the dimer interface, placing the two active sites 77 Å apart. Despite the large distance between catalytic sites, the interdigitating β-hairpin motif at the dimer interface between the DUF domains places Ser259, found at the apex of the turn of the β-hairpin, only ~23 Å from the active site of the opposite chain. Although the β-hairpin is poorly ordered in the experimental map, the targeted Cys-pair crosslinking confirms the placement of this motif found in the AlphaFold model. A predicted helix-turn-helix motif from residues 337-369 of the catalytic domain is also disordered in the experimental map (Fig. 5C). This region is analogous to a mobile loop found in *C. concisus* PglC, which has been implicated in formation of the substrate binding pocket.[58] Although both the mobile loop in PglC and the helix-turn-helix found in WbaP are positioned proximal to the active site, the length, sequence conservation, and structural motifs of this region differ greatly between these two PGTs (Fig. 6). XLMS experiments demonstrate that lysines within the predicted helix-turn-helix of WbaP can crosslink with lysines across the soluble core of the catalytic domain (Fig. 4C), giving high confidence to the placement of this region in the AlphaFold model, despite the lack of density in the experimental map.

During the course of this work, a structural description of a WbaP ortholog from *E. coli* (NCBI:txid562) by CryoEM was published.[59] This protein was purified using both detergent and amphipols, and was complexed with a nanobody fusion to facilitate particle identification and alignment during 3D-reconstruction. Although the maps and models associated with this data are not yet publicly available,

visual comparison between *S. enterica* WbaP and published images of the *E. coli* WbaP reveals an overall similar architecture, including disorder in the helix-turn-helix region. However, the handedness of the *E. coli* WbaP reconstruction is inverted compared to *S. enterica* WbaP as well as the predicted AlphaFold model. Notably, AlphaFold multimer was not available at the time of publication of the *E. coli* structure and the handedness may be ambiguous at the observed resolution without a model for guidance.

With a clear idea of the overall architecture of WbaP, we sought to better understand the role of the DUF beyond stabilization of the dimer interface. To this end, we used both the CryoEM reconstruction and AlphaFold structural models to guide generation of truncated constructs of WbaP encoding only the DUF. A construct encoding residues 146 to 272 resulted in the production of soluble protein (Fig. S8), which was then used to screen for potential nucleotide ligands by nano differential scanning fluorimetry (nDSF). This screen identified UMP, UDP, and UTP as potential ligands, and demonstrated that the UDP-Gal substrate of the catalytic domain destabilized the truncated DUF construct. These results imply the DUF may play a role in allosteric regulation by sensing nucleotides and nucleotide-sugars. Indeed, full-length WbaP was found to be strongly inhibited by UMP, with an apparent IC_{50} of 4.1 µM (Fig. 7). The unusual profile of the UMP inhibition curve may reflect UMP binding both competitively to UDP-Gal at the active site, as well as to an allosteric site within the DUF.

Relationship Between Lg-PGTs and PglF-like dehydratases:

The N-terminal domains found in WbaP are highly conserved across the Lg-PGT family,[24] however, a search of the PDB for structural homologs of the domain alone returned only nucleotide binding domains with limited homology to the DUF. To sample a larger set of structural models, we utilized the recently described Foldseek tool to parse the AlphaFold model database (AFDB proteome).[60] In this analysis, although most of the aligned models were other PGTs, the N-terminal domains of WbaP exhibited structural homology to models generated from a class of NDP-sugar dehydratases also involved in

glycoconjugate biosynthesis (Fig. S10). Homologs identified by Foldseek include PglF from *Campylobacter* and Cap5D from *Staphylococcus*. This family of DHs is well characterized both structurally and mechanistically, however, constructs used for most structural studies truncate the N-terminus for practical reasons.[61,62] As these two classes of enzymes are functionally unrelated, the conservation of their N-terminal domains is puzzling, and warrants re-examination of full-length constructs of PglF-like DHs.

Discussion

Detailed studies of monoPGTs are hampered by the perennial problems associated with the expression, purification, and structural characterization of membrane proteins. However, the essential role of monoPGTs in the first membrane-committed step of the biosynthesis of diverse classes of prokaryotic glycoconjugates makes structural and functional analysis of this prokaryote-specific enzyme superfamily imperative. Here, we build on our recent advances in the development of an optimized solubilization and purification system for the abundant Lg-PGT members of the monoPGT superfamily and present the structural characterization of WbaP, the initiating PGT from O-antigen biosynthesis in *S. enterica*. We anticipate that the strategy applied here will be broadly applicable to other bacterial membrane proteins and provide an efficient route to map the structure of target proteins in a near-native environment, even in the absence of high-resolution structural data. In addition, these methods will be well suited to the study of membrane protein complexes or membrane-bound metabolons,[63,64] a long-standing goal in the study of glycoconjugate biosynthesis.

The structure presented represents the first monotopic PGT characterized in a native-like liponanoparticle. We show that in contrast to the Sm-PGT PglC from *Campylobacter*, which is functional as a monomer, the *S. enterica* Lg-PGT, WbaP, is a tightly associated dimer, with the majority of inter-chain contacts occurring in the highly conserved but previously uncharacterized DUF. Thermal denaturation binding experiments implicate the DUF in nucleotide and nucleotide-sugar binding and biochemical assay

studies show that UMP is a potent inhibitor of full-length WbaP activity (Fig. 6B). The profile of the inhibition curve suggests additional UMP interactions beyond the active site in the full-length protein. Taken together, these results point toward the DUF serving both a *structural and functional* role in stabilizing the dimer and allosterically regulating the catalytic domain upon UMP binding.

Intriguingly, we observe a surprising structural homology between the N-terminal domains of Lg-PGTs and PglF-like dehydratases, which are common UDP-sugar modifying enzymes involved in the biogenesis of unusual carbohydrates for glycoconjugate biosynthesis. In this case, it is tempting to hypothesize that the DUFs may play a larger role in the modulation of the flux of glycoconjugate biosynthesis.

Although PglC and WbaP are predicted to catalyze similar chemical transformations, these PGT enzymes exhibit different substrate specificity. PglC shows high specificity for UDP-diNAcBac, while WbaP is specific for UDP-Gal. Recently, structural and sequence elements were identified that facilitated the bioinformatic assignment of substrate specificity to Sm-PGTs that use the same UDP-diNAcBac substrate as PglC.[58] Interestingly, the major structural deviations between PglC and WbaP occur at these regions (Fig. 6). An extended loop containing a conserved "GLLP" motif in diNAcBac-utilizing PGTs is replaced with a predicted helix-turn-helix motif in WbaP. This region is poorly resolved in the experimental CryoEM reconstruction of WbaP, however, lysines in this putative structural motif crosslink with the main domain of the WbaP dimer. Given the apparent high degree of motion in the vitrified WbaP particles at this region, and the proximity of the helix-turn-helix motif to the active site, we envision a mechanism in which the active site "closes" upon the helix-turn-helix motion towards the center of the dimer, and the active site is "open" when the motif moves away, as observed in the AlphaFold model. This resembles structural conformers observed in both crystal structures and molecular dynamics simulations of *C. concisus* PglC.[65] Another structural element in the diNAcBac-specific PGTs that is predictive of substrate specificity is an aromatic box motif comprising residues in the solvent-accessible region of the protein and a conserved Phe near the C-terminus of this subset of Sm-PGTs. In WbaP, the C-terminus of the catalytic domain is ~14 amino

acids shorter than in the diNAcBac Sm-PGTs, and the aromatic box motif is absent. These subtle changes surrounding the conserved Asp-Glu catalytic dyad are likely to drive the specificity of WbaP to UDP-Gal. As additional UDP-Gal utilizing PGTs are characterized, we envision the creation of a bioinformatic pipeline to identify the specific residues involved in substrate selection, akin to the method applied to the diNAcBac PGTs.

The liponanoparticle-based approach for purification of *S. enterica* WbaP provides a convenient way to add contrast to vitrified particles in CryoEM by way of additional lipid and SMA mass, similar to the rationale for utilization of nanobody fusion strategies in the study of other small membrane proteins.[59,66] Future efforts to improve the overall resolution of Lg-PGT CryoEM reconstructions may benefit from a combination of these approaches whereby mass from a liponanoparticle facilitates particle identification and picking, while bound nanobody or nanobody complexes/fusions reduce structural motion of the bound Lg-PGT and add distinct features to aid in masking and 3D refinement.

Bibliography:

1. Tytgat, H.L. & Lebeer, S. The sweet tooth of bacteria: common themes in bacterial glycoconjugates. *Microbiology and Molecular Biology Reviews* **78**, 372-417 (2014).
2. Chen, J., Lee, S.M. & Mao, Y. Protective effect of exopolysaccharide colanic acid of *Escherichia coli* O157: H7 to osmotic and oxidative stress. *International journal of food microbiology* **93**, 281-286 (2004).
3. Campos, M.A. et al. Capsule polysaccharide mediates bacterial resistance to antimicrobial peptides. *Infection and immunity* **72**, 7107-7114 (2004).
4. Tan, Y.H., Chen, Y., Chu, W.H., Sham, L.T. & Gan, Y.H. Cell envelope defects of different capsule-null mutants in K1 hypervirulent *Klebsiella pneumoniae* can affect bacterial pathogenesis. *Molecular microbiology* **113**, 889-905 (2020).
5. Bentley, S.D. et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS genetics* **2**, e31 (2006).
6. Vu, B., Chen, M., Crawford, R.J. & Ivanova, E.P. Bacterial extracellular polysaccharides involved in biofilm formation. *Molecules* **14**, 2535-2554 (2009).
7. Whitfield, C., Wear, S.S. & Sande, C. Assembly of bacterial capsular polysaccharides and exopolysaccharides. *Annual Review of Microbiology* **74**, 521-543 (2020).

8. Chua, W.-Z. et al. High-Throughput Mutagenesis and Cross-Complementation Experiments Reveal Substrate Preference and Critical Residues of the Capsule Transporters in *Streptococcus pneumoniae*. *Mbio* **12**, e02615-21 (2021).
9. Hong, Y. & Reeves, P.R. Model for the controlled synthesis of O-antigen repeat units involving the WaaL ligase. *Msphere* **1**, e00074-15 (2016).
10. Yethon, J.A., Vinogradov, E., Perry, M.B. & Whitfield, C. Mutation of the lipopolysaccharide core glycosyltransferase encoded by *waa*G destabilizes the outer membrane of *Escherichia coli* by interfering with core phosphorylation. *Journal of bacteriology* **182**, 5620-5623 (2000).
11. Islam, S.T. & Lam, J.S. Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway. *Canadian journal of microbiology* **60**, 697-716 (2014).
12. Jiang, X.M. et al. Structure and sequence of the rfb (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). *Molecular microbiology* **5**, 695-713 (1991).
13. Samuel, G. & Reeves, P. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydrate research* **338**, 2503-2519 (2003).
14. Kalynych, S., Morona, R. & Cygler, M. Progress in understanding the assembly process of bacterial O-antigen. *FEMS microbiology reviews* **38**, 1048-1065 (2014).
15. Wang, L., Liu, D. & Reeves, P.R. C-terminal half of *Salmonella enterica* WbaP (RfbP) is the galactosyl-1-phosphate transferase domain catalyzing the first step of O-antigen synthesis. *Journal of Bacteriology* **178**, 2598-2604 (1996).
16. Wang, L. & Reeves, P.R. Involvement of the galactosyl-1-phosphate transferase encoded by the *Salmonella enterica* rfbP gene in O-antigen subunit processing. *Journal of bacteriology* **176**, 4348-4356 (1994).
17. Patel, K.B., Ciepichal, E., Swiezewska, E. & Valvano, M.A. The C-terminal domain of the *Salmonella enterica* WbaP (UDP-galactose: Und-P galactose-1-phosphate transferase) is sufficient for catalytic activity and specificity for undecaprenyl monophosphate. *Glycobiology* **22**, 116-122 (2012).
18. Patel, K.B., Furlong, S.E. & Valvano, M.A. Functional analysis of the C-terminal domain of the WbaP protein that mediates initiation of O antigen synthesis in *Salmonella enterica*. *Glycobiology* **20**, 1389-1401 (2010).
19. Saldías, M.S. et al. Distinct functional domains of the *Salmonella enterica* WbaP transferase that is involved in the initiation reaction for synthesis of the O antigen subunit. *Microbiology* **154**, 440-453 (2008).
20. Whitfield, C., Williams, D.M. & Kelly, S.D. Lipopolysaccharide O-antigens—bacterial glycans made to measure. *Journal of Biological Chemistry* **295**, 10593-10609 (2020).
21. Woodward, R. et al. In vitro bacterial polysaccharide biosynthesis: defining the functions of Wzy and Wzz. *Nature chemical biology* **6**, 418-423 (2010).
22. Liu, D., Cole, R.A. & Reeves, P.R. An O-antigen processing function for Wzx (RfbX): a promising candidate for O-unit flippase. *Journal of bacteriology* **178**, 2102-2107 (1996).
23. O'Toole, K.H., Bernstein, H.M., Allen, K.N. & Imperiali, B. The surprising structural and mechanistic dichotomy of membrane-associated phosphoglycosyl transferases. *Biochemical Society Transactions* **49**, 1189-1203 (2021).
24. O'Toole, K.H., Imperiali, B. & Allen, K.N. Glycoconjugate pathway connections revealed by sequence similarity network analysis of the monotopic phosphoglycosyl transferases. *Proceedings of the National Academy of Sciences* **118**, e2018289118 (2021).
25. Lukose, V. et al. Conservation and covariance in small bacterial phosphoglycosyltransferases identify the functional catalytic core. *Biochemistry* **54**, 7326-7334 (2015).
26. Ray, L.C. et al. Membrane association of monotopic phosphoglycosyl transferase underpins function. *Nature chemical biology* **14**, 538-541 (2018).

27. Dodge, G.J., Bernstein, H.M. & Imperiali, B. A generalizable protocol for expression and purification of membrane-bound bacterial phosphoglycosyl transferases in liponanoparticles. *Protein Expr Purif* **207**, 106273 (2023).
28. Cartee, R.T., Forsee, W.T., Bender, M.H., Ambrose, K.D. & Yother, J. CpsE from type 2 *Streptococcus pneumoniae* catalyzes the reversible addition of glucose-1-phosphate to a polyprenyl phosphate acceptor, initiating type 2 capsule repeat unit formation. *Journal of bacteriology* **187**, 7425-7433 (2005).
29. James, D.B., Gupta, K., Hauser, J.R. & Yother, J. Biochemical activities of *Streptococcus pneumoniae* serotype 2 capsular glycosyltransferases and significance of suppressor mutations affecting the initiating glycosyltransferase Cps2E. *Journal of bacteriology* **195**, 5469-5478 (2013).
30. Furlong, S.E., Ford, A., Albarnez-Rodriguez, L. & Valvano, M.A. Topological analysis of the *Escherichia coli* WcaJ protein reveals a new conserved configuration for the polyisoprenyl-phosphate hexose-1-phosphate transferase family. *Scientific reports* **5**, 1-9 (2015).
31. Entova, S., Billod, J.-M., Swiecicki, J.-M., Martin-Santamaria, S. & Imperiali, B. Insights into the key determinants of membrane protein topology enable the identification of new monotopic folds. *Elife* **7**, e40889 (2018).
32. Sjodt, M. et al. Structure of the peptidoglycan polymerase RodA resolved by evolutionary coupling analysis. *Nature* **556**, 118-121 (2018).
33. Studier, F.W. Protein production by auto-induction in high-density shaking cultures. *Protein expression and purification* **41**, 207-234 (2005).
34. Tartof, K. Improved media for growing plasmid and cosmid clones. *Bethesda Res Lab Focus* **9**, 12 (1987).
35. Eschenfeldt, W.H., Lucy, S., Millard, C.S., Joachimiak, A. & Mark, I.D. A family of LIC vectors for high-throughput cloning and purification of proteins. *High Throughput Protein Expression and Purification: Methods and Protocols*, 105-115 (2009).
36. Gibson, D.G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**, 343-345 (2009).
37. Punjani, A. Algorithmic advances in single particle cryo-EM data processing using CryoSPARC. *Microscopy and Microanalysis* **26**, 2322-2323 (2020).
38. Punjani, A., Rubinstein, J.L., Fleet, D.J. & Brubaker, M.A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature methods* **14**, 290-296 (2017).
39. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nature methods* **19**, 679-682 (2022).
40. Adams, P.D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* **66**, 213-221 (2010).
41. Afonine, P.V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallographica Section D: Structural Biology* **74**, 531-544 (2018).
42. Pettersen, E.F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30**, 70-82 (2021).
43. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).
44. Stivala, A., Wybrow, M., Wirth, A., Whisstock, J.C. & Stuckey, P.J. Automatic generation of protein structure cartoons with Pro-origami. *Bioinformatics* **27**, 3315-3316 (2011).
45. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* **372**, 774-797 (2007).
46. Jiang, P.L. et al. A Membrane-Permeable and Immobilized Metal Affinity Chromatography (IMAC) Enrichable Cross-Linking Reagent to Advance In Vivo Cross-Linking Mass Spectrometry. *Angewandte Chemie International Edition* **61**, e202113937 (2022).

47. Lagerwaard, I.M., Albanese, P., Jankevics, A. & Scheltema, R.A. Xlink mapping and analySis (XMAS)-smooth integrative modeling in chimeraX. *bioRxiv*, 2022.04. 21.489026 (2022).
48. Thomsen, M. Determination of the Molecular Mass of Membrane Proteins Using Size-Exclusion Chromatography with Multiangle Laser Light Scattering (SEC-MALLS). *Methods in Molecular Biology (Clifton, NJ)* **2168**, 263-269 (2020).
49. Hesketh, S.J. et al. Styrene maleic-acid lipid particles (SMALPs) into detergent or amphipols: an exchange protocol for membrane protein characterisation. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1862**, 183192 (2020).
50. Olerinyova, A. et al. Mass photometry of membrane proteins. *Chem* **7**, 224-236 (2021).
51. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
52. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. *BioRxiv*, 2021.10. 04.463034 (2021).
53. Bennett, M., Choe, S. & Eisenberg, D. Domain swapping: entangling alliances between proteins. *Proceedings of the National Academy of Sciences* **91**, 3127-3131 (1994).
54. Liu, Y. & Eisenberg, D. 3D domain swapping: as domains continue to swap. *Protein science* **11**, 1285-1299 (2002).
55. Akdel, M. et al. A structural biology community assessment of AlphaFold2 applications. *Nature Structural & Molecular Biology*, 1-12 (2022).
56. Chavez, J.D. & Bruce, J.E. Chemical cross-linking with mass spectrometry: a tool for systems structural biology. *Current opinion in chemical biology* **48**, 8-18 (2019).
57. Piersimoni, L., Kastritis, P.L., Arlt, C. & Sinz, A. Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein–Protein Interactions– A Method for All Seasons. *Chemical Reviews* **122**, 7500-7531 (2021).
58. Anderson, A.J., Dodge, G.J., Allen, K.N. & Imperiali, B. Co-conserved sequence motifs are predictive of substrate specificity in a family of monotopic phosphoglycosyl transferases. *Protein Science*, e4646 (2023).
59. Uchański, T. et al. Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. *Nature methods* **18**, 60-68 (2021).
60. van Kempen, M. et al. Foldseek: fast and accurate protein structure search. *Biorxiv*, 2022.02. 07.479398 (2022).
61. Riegert, A.S. et al. Structural and biochemical investigation of PglF from *Campylobacter jejuni* reveals a new mechanism for a member of the short chain dehydrogenase/reductase superfamily. *Biochemistry* **56**, 6030-6040 (2017).
62. Olivier, N.B., Chen, M.M., Behr, J.R. & Imperiali, B. In vitro biosynthesis of UDP-N,N'-diacetylbacillosamine by enzymes of the *Campylobacter jejuni* general protein glycosylation system. *Biochemistry* **45**, 13659-13669 (2006).
63. Møller, B.L. Dynamic metabolons. *Science* **330**, 1328-1329 (2010).
64. Bassard, J.-E. & Laursen, T. Molecular snapshots of dynamic membrane-bound metabolons. in *Methods in Enzymology*, Vol. 617 1-27 (Elsevier, 2019).
65. Majumder, A. et al. Synergistic computational and experimental studies of a phosphoglycosyltransferase membrane/ligand ensemble. *bioRxiv*, 2023.05. 07.539694 (2023).
66. Wu, X. & Rapoport, T.A. Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies). *Proceedings of the National Academy of Sciences* **118**, e2115001118 (2021).

Acknowledgements

We thank Ed Brignole and Christopher Borsa for their support with electron microscopy and Karen Allen for thoughtful comments on the manuscript.

Funding

This work was funded by NIH grants R01 GM131627 and GM039334 (to B.I) and F32 GM134576 (to G.J.D).

Author contribution:

GJD: Conceived and carried out experiments, generated AlphaFold prediction, purified protein and collected data for electron microscopy, processed microscopy data, built and refined experimental model, assisted in XLMS protocol development, generated figures, drafted, reviewed, and edited manuscript. **AA**: Conceived and carried out experiments, designed crosslinking panel for WT protein, designed and purified variants for Cys-pair crosslinking, drafted, reviewed, and edited manuscript. **YH** and **RV**: Developed XLMS protocol, ran XLMS experiments, processed XLMS data, reviewed and edited manuscript. **BI**: Conceived experiments, review and editing of manuscript, funding.

Authors

Greg J. Dodge – Department of Biology and Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

orcid.org/0000-0002-6555-8350

Alyssa J. Anderson – Department of Biology and Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

orcid.org/0000-0003-4871-4283

Yi He – Thermo Fisher Scientific, 355 River Oaks Pkwy, San Jose, CA 95134 USA

Rosa Viner – Thermo Fisher Scientific, 355 River Oaks Pkwy, San Jose, CA 95134 USA

orcid.org/0000-0003-0550-5545

*Barbara Imperiali - Department of Biology and Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

imper@mit.edu

orcid.org/0000-0002-5749-7869

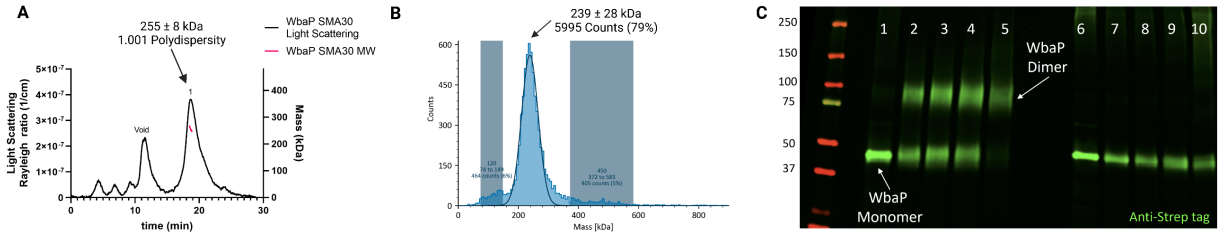


Figure 1 Characterization of *S. enterica* WbaP in SMALP. **A**. WbaP in SMALP analyzed by SEC-MALS. A small peak is observed at the void volume of the column, while the main peak has a calculated molecular weight of 255 ± 8 kDa, with a polydispersity value of 1.001. **B**: WbaP in SMALP analyzed by mass photometry. The sample is monodisperse, and the main species has an apparent molecular weight of 239 ± 28 kDa. **C**. Western blot analysis of Lysine-reactive dithiobis(succinimidyl propionate) (DSP) crosslinker reacted with WbaP in SMALP. Lanes: 1: Control, 2: 0.1 mM DSP, 3: 0.25 mM DSP, 4: 1 mM DSP, 5: 5 mM DSP, 7: 0.1 mM DSP reduced, 8: 0.25 mM DSP reduced, 9: 1 mM DSP reduced, 10: 5 mM DSP reduced

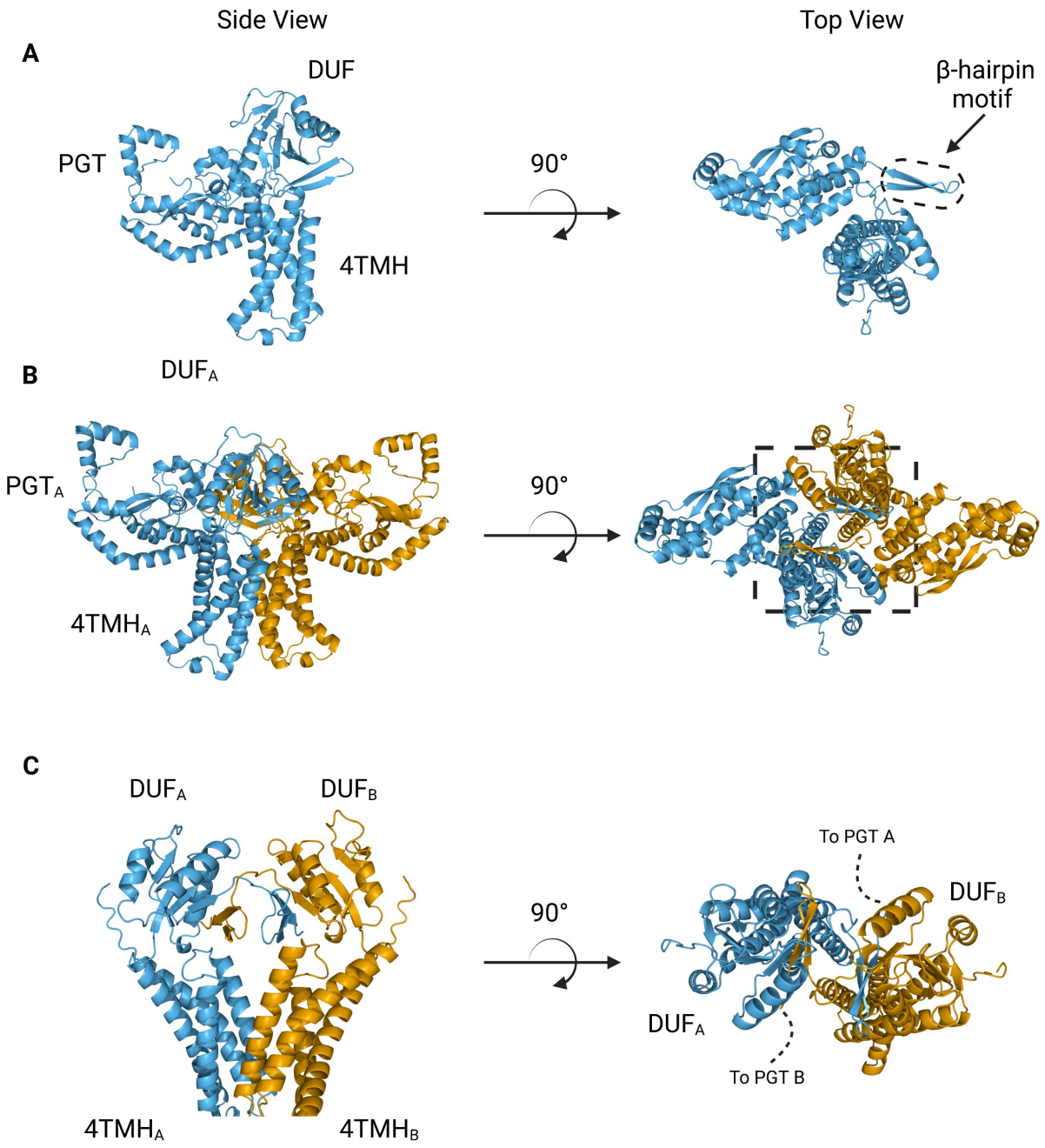


Figure 2 AlphaFold predictions of *S. enterica* WbaP monomer and dimer. **A**: WbaP monomer prediction. **B**: WbaP dimer prediction. **C**: Close-up view of predicted dimer interface showing the interdigitating β-hairpin motifs between DUF_A and DUF_B. PGT domains hidden for clarity.

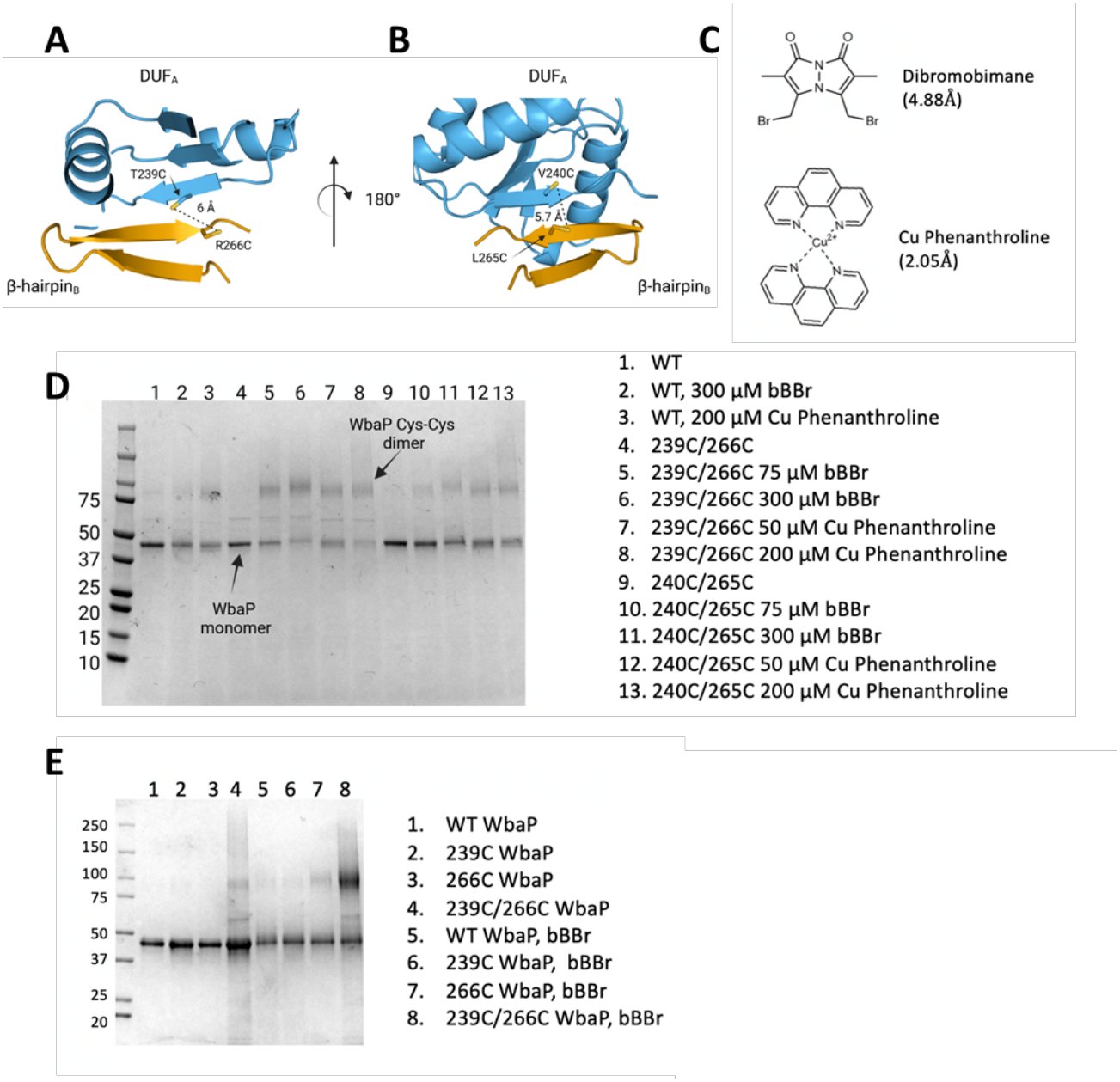
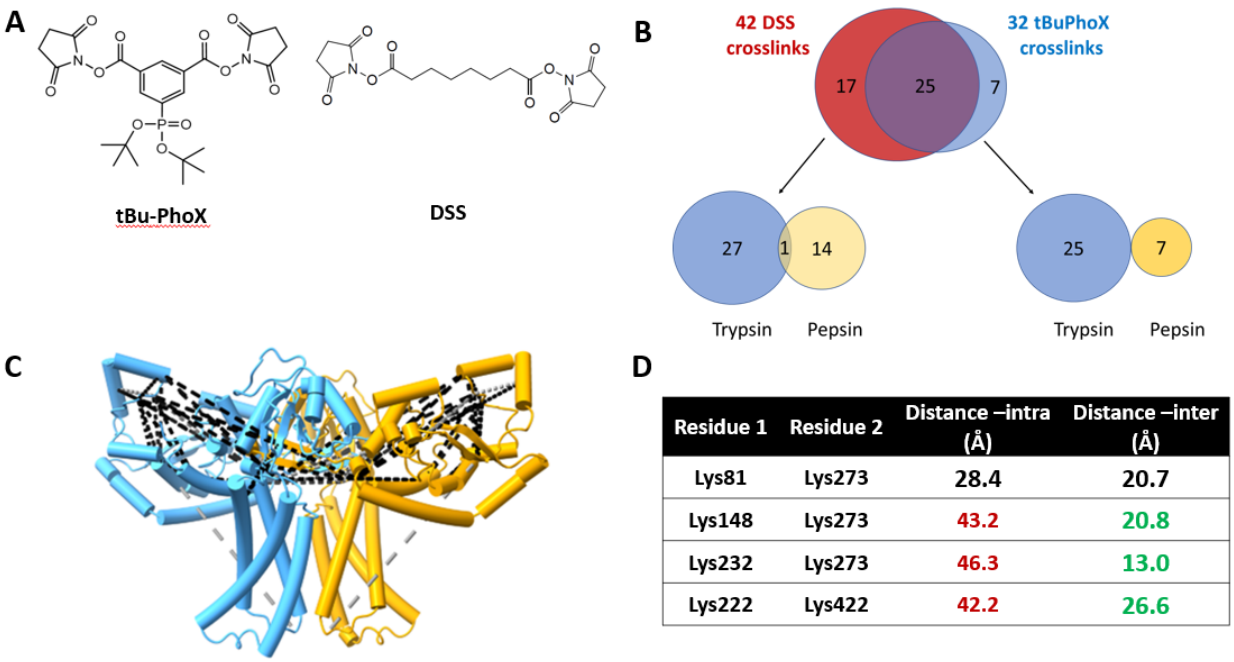


Figure 3 Probing the predicted interface of the WbaP dimer. **A.** and **B.** AlphaFold prediction with Cys pair variants modeled as sticks. The β-hairpin of one monomer is shown in sky blue and the continuing β- strand of the other monomer is shown in orange. **C.** Structure and crosslinking radii of thiol reactive crosslinkers. **D**. Crosslinking of Cys pair WbaP variants in SMALP using Dibromobimane (bBBr) and Cu Phenanthroline. **E.** bBBr crosslinking of single and double WbaP Cys variants in SMALP.



Residue 1	Residue 2	Distance –intra (Å)	Distance –inter (Å)
Lys81	Lys273	28.4	20.7
Lys148	Lys273	43.2	20.8
Lys232	Lys273	46.3	13.0
Lys222	Lys422	42.2	26.6

Figure 4 XLMS of WbaP in SMALP. **A.** Structures of amino reactive crosslinkers used for XLMS analysis. **B.** Overview of identified WbaP crosslinks from different protease treatments and crosslinkers. **C.** Identified crosslinks mapped onto S. enterica WbaP AlphaFold dimer model. DSS crosslinks shown in black, tBu-PhoX crosslinks shown in gray. **D.** Residues linked by intermolecular crosslinks. The distance between these residues within a single monomer and between monomers is shown.

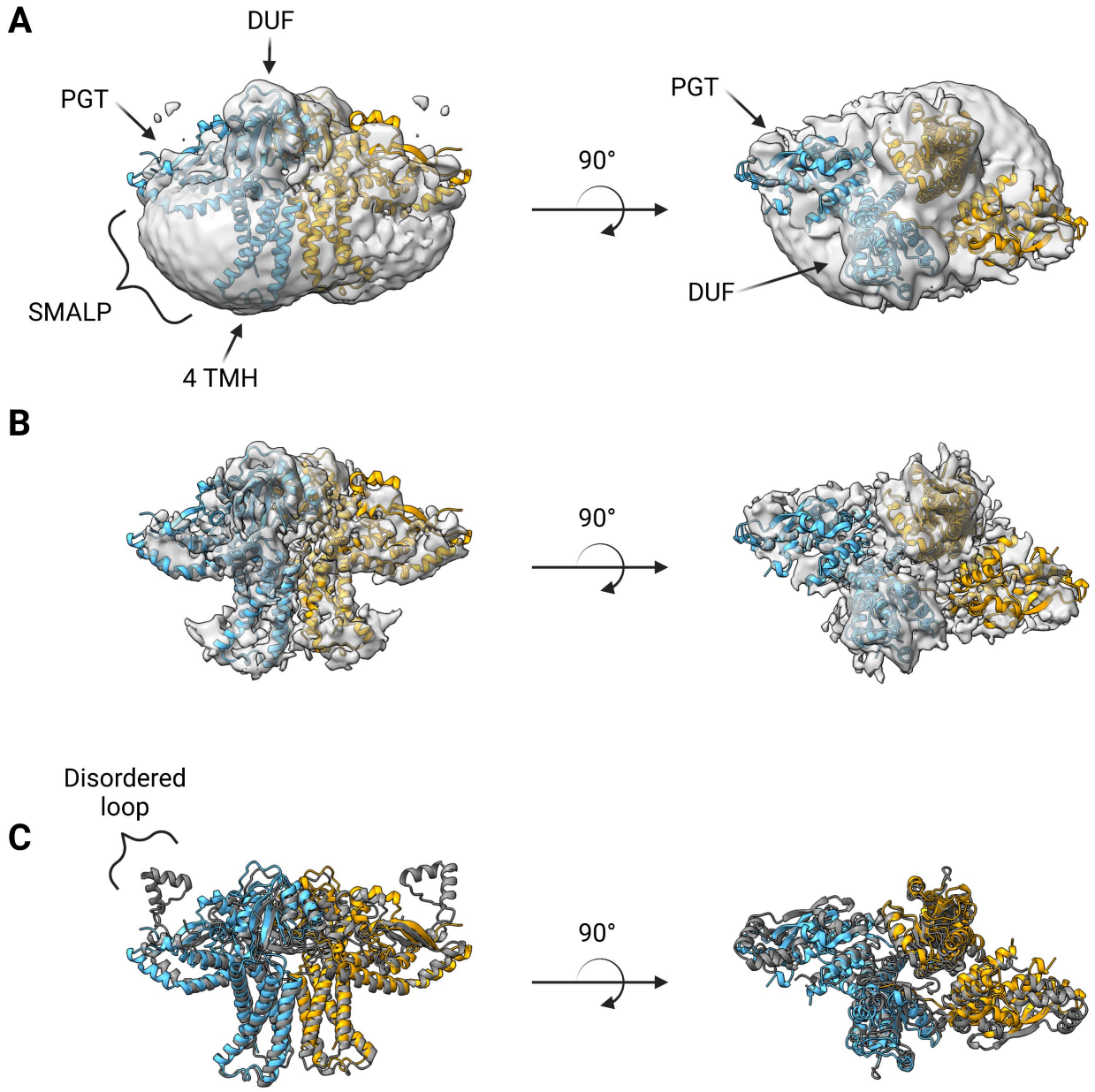


Figure 5 Structure of *S. enterica* WbaP in SMALP at 4.1 Å. WbaP dimer colored in blue and orange. **A.** Unsharpened map displaying clear density for stabilizing liponanoparticle. Unsharpened density colored light gray. **B**. Sharpened map after masked local refinement. Sharpened density colored light gray **C.** Superposition of refined WbaP model and full-length AlphaFold prediction, RMSD: 2.83 Å. AlphaFold dimer colored gray. A predicted helix-turn-helix motif lacks density in the experimental map.

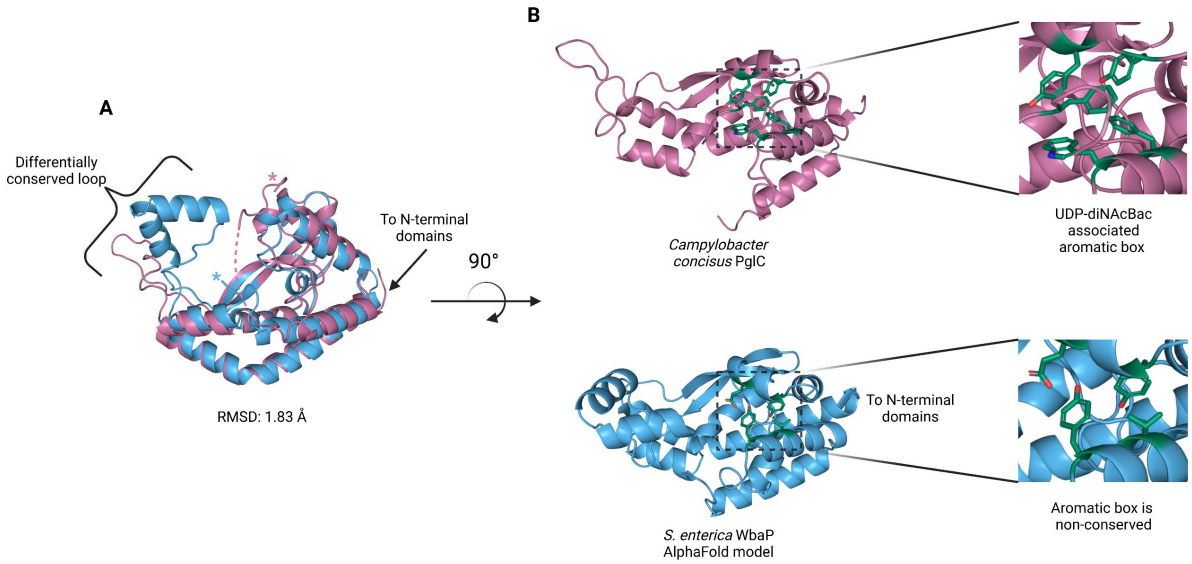


Figure 6 Comparison of *S. enterica* WbaP PGT domain and PglC from *Campylobacter concisus* (PDBid 8G1N). WbaP shown in light blue, PglC in magenta. **A**. Superposition of WbaP and PglC. While the overall RMSD is 1.83 Å, regions in PglC which can be used to computationally assign substrate are significantly different in WbaP. A loop region in PglC is replaced with a helix-turn-helix motif in WbaP, and the C-terminus of WbaP is significantly shorter than that of PglC. **B**. Top: an aromatic box motif in PglC is conserved among PGTs that utilize UDP-diNAcBac. Aromatic box residues are shown as dark green sticks. Bottom: Residues found in S. enterica WbaP at homologous positions to aromatic box residues in PglC shown as dark green sticks. Aromatic box residues are non-conserved in WbaP.

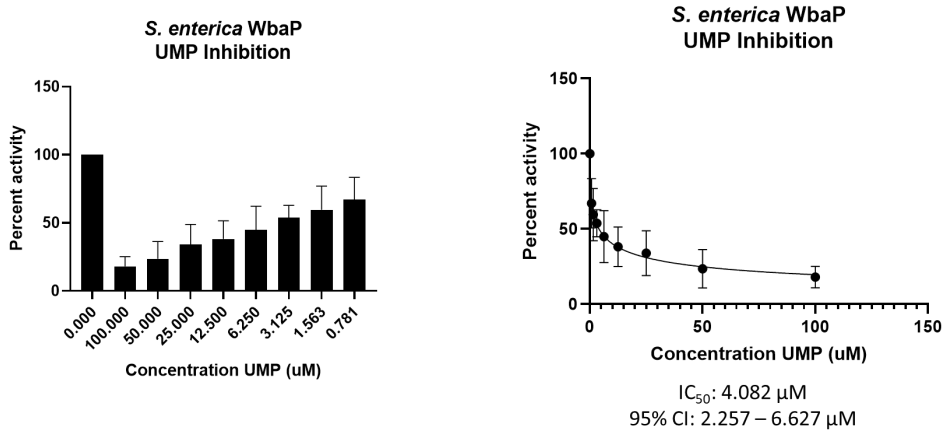


Figure 7 WbaP UMP titration assay. Left: Percent activity vs UMP concentration. A UMP-dependent decrease in WbaP activity is observed. Right: Curve-fitting of UMP inhibition data yields an IC50 apparent of 4.1 µM.