OPEN BIOLOGY

royalsocietypublishing.org/rsob

Research

Cite this article: Weckener M *et al.* 2026 Structural insights into the polytopic *Escherichia coli* phosphoglycosyl transferase WbaP. *Open Biol.* **16**: 250369.
https://doi.org/10.1098/rsob.250369

Received: 29 September 2025
Accepted: 23 June 2026

Subject Areas:
biochemistry, biophysics

Keywords:
membrane protein, cell wall, oligosaccharides, phosphoglycosyl transferase, cryo-electron microscopy

Author for correspondence:
James H. Naismith
e-mail: james.naismith@strubi.ox.ac.uk

Structural insights into the polytopic *Escherichia coli* phosphoglycosyl transferase WbaP

Miriam Weckener[1,†], Audrey Le Bas[1,2,†], Philip N. Ward[1], Bradley R. Clarke[6], Peter J. Harrison[7,8], Stephen A. McMahon[9], Huanting Liu[9], Alice Eddershaw[1,3], Andrew M. Giltrap[1,4,5], Els Pardon[10,11], Abirami Lakshminarayanan[2], Jan Steyaert[10,11], Benjamin G. Davis[1,4,5], Chris Whitfield[6] and James H. Naismith[1,2]

[1]The Rosalind Franklin Institute, Harwell Campus, Didcot, UK
[2]Division of Structural Biology (STRUBI), [3]Nuffield Department of Clinical Neurosciences, [4]Department of Pharmacology, and [5]Department of Chemistry, University of Oxford, Oxford, UK
[6]Department of Molecular and Cellular Biology, University of Guelph, Guelph, Canada
[7]Membrane Protein Lab, Diamond Light Source and [8]Research Complex at Harwell, Harwell Science & Innovation Campus, Didcot, UK
[9]Biomedical Sciences Research Complex, University of St Andrews, St Andrews, UK
[10]Structural Biology Brussels, Vrije Universiteit Brussel (VUB), Brussels, Belgium
[11]VIB-VUB Center for Structural Biology, VIB, Brussels, Belgium

MW, 0000-0002-1419-9378; ALB, 0000-0002-1100-6833; PNW, 0000-0003-2546-3426; BRC, 0000-0003-2233-1268; PJH, 0000-0001-5167-0547; SAM, 0000-0003-4273-8364; HL, 0000-0001-5953-9367; AE, 0000-0002-8381-4135; AMG, 0000-0001-8960-553X; EP, 0000-0002-2466-0172; AL, 0000-0002-1688-3667; JS, 0000-0002-3825-874X; BGD, 0000-0002-5056-407X; CW, 0000-0002-5267-7027; JHN, 0000-0001-6744-5061

Host–pathogen interactions frequently depend on key components of the bacterial cell surface, such as capsules and lipopolysaccharides in Gram-negative bacteria. The first step in the synthesis of lipid-linked polysaccharides is the substitution of a uridine diphosphate (UDP)-sugar by a lipid monophosphate catalysed by a phosphoglycosyl transferase (PGT). We report the 3.0 Å cryo-electron microscopy apo-structure of the PGT enzyme WbaP from *Escherichia coli*, a UDP-galactose:undecaprenolphosphate galactose-1-phosphoryl transferase. The structure is a dimer with each monomer formed of four N-terminal transmembrane helices, a small α/β domain with a distinctive β-hairpin that inserts into the other monomer, and a catalytic domain, which sits perpendicular to the transmembrane domain. A complex of WbaP with the UDP-galactose substrate shows binding of the UDP moiety by R319 and R377. Mutations of R319 and R377, along with K331 and R401, highlighted the essential nature of these residues for the catalytic activity of the protein, as confirmed by an *in vivo* functional assay. Our results provide new insights into the PGT family of enzymes.

1. Introduction

Bacterial cell-surface polysaccharides exhibit remarkable diversity in their repeat-unit structures due to a wide range of component sugars, non-carbohydrate constituents and glycosidic linkage possibilities [1–4]. Despite the

[†]These authors contributed equally to the study.

Supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.8588522.

THE ROYAL SOCIETY PUBLISHING

structural variations, the chemical strategies for their synthesis are limited; most use a lipid polyprenol monophosphate to attach and elaborate the sugar component. Individual carbohydrate repeat units, linked to the lipid, are synthesized at the cytosolic face of the membrane. The resulting compound is translocated across the cytoplasmic membrane by a multidrug and toxin extrusion (MATE)-family flippase or ABC transporter [5]. Most systems use one of two polymerization strategies. In the Wzy-dependent systems, once the lipid-linked sugar reaches the periplasm, the polymerase (Wzy) catalyses block-wise chain extension [6]. In non-Wzy-dependent pathways, glycosyltransferases build the sugar polymer in the periplasm by sequential sugar addition [7]. The activity of Wzy is regulated by a protein co-polymerase (PCP), which, depending on the pathway, belongs to the Wzz or Wzc family [8,9]. The full-length polymers from Wzy-dependent synthesis are found in the capsules of Gram-positive and Gram-negative bacteria [10,11], in the lipopolysaccharide O-antigens in Gram-negative bacteria [12,13] and in N-linked glycans found in glycoproteins in Gram-positive and Gram-negative bacteria as well as Archaea [14].

Despite the diversity in synthesis pathways, they share an initiating step involving the transfer of hexose-1-phosphate or 2-acetamido-2-deoxyhexose-1-phosphate from sugar nucleotide donors to polyprenol-phosphate acceptors. These reactions are catalysed by phosphoglycosyl transferase (PGT) enzymes and result in polyprenol-pyrophosphate-sugar products. There are two families of PGTs, monotopic (monoPGTs) and polytopic (polyPGTs) [15]. PolyPGTs have 10–11 transmembrane helices which contribute to a complex catalytic site [16–18]. MonoPGTs possess a catalytic core within a peripheral membrane structure and a single transmembrane helix [19–23]. PolyPGTs are found in prokaryotes and eukaryotes, while monoPGTs appear to be confined to prokaryotes and thus have potential as new antibiotic targets. MonoPGTs are divided into three sub-families [24]. Small monoPGTs (SmPGTs) possess a glycosyltransferase catalytic domain exemplified by PglC, an undecaprenyl phosphate *N*,*N*′-diacetylbacillosamine-1-phosphoryl transferase involved in N-linked protein glycosylation in *Campylobacter concisus* [25–27]. The structure of PglC from *C. concisus* ($PglC_{Cc}$) shows a novel architecture consisting of a single re-entrant membrane helix on the cytoplasmic side of the membrane [22]. Bifunctional monoPGTs (biPGTs) possess a PglC-like domain within a larger protein with an additional functional domain; for example, PglB from *Neisseria gonorrhoeae* features an N-terminal PGT domain similar to PglC and a C-terminal amino-sugar acetyltransferase domain [27]. Large monoPGTs (LgPGTs) have four additional transmembrane helices (TM) at the N-terminus [28]; WbaP, a uridine diphosphate (UDP)-galactose::undecaprenol-phosphate galactose-1-phosphoryl transferase from O-antigen biosynthesis in *Salmonella enterica*, serves as the archetype (figure 1a) [28–31]. The 4.9 Å and 4.1 Å resolution single-particle cryo-electron microscopy (cryo-EM) structures of WbaP from *Escherichia coli* ($WbaP_{Ec}$) and *Salmonella enterica* ($WbaP_{Se}$) were recently reported [28,32]. The C-terminal catalytic domain of WbaP is structurally closely related to the SmPGT PglC [22,28,30,33,34].

Substantial progress has been made in describing the catalytic activity of monoPGTs by combining bioinformatics with activity assays, enabling wide-scale investigation of substrate specificities from bacteria in different genera [24]. While polyPGTs use a one-step mechanism via the formation of a ternary complex [16], it was proposed that SmPGTs use a two-step ping-pong mechanism involving the formation of a covalent phospho-sugar intermediate from the nucleotide diphosphosugar substrate [19,22]. The intermediate is utilized by the second substrate, undecaprenol-phosphate (und-P) to yield the undecaprenol pyrophosphate-linked monosaccharide product. Catalytically important residues have been identified in PGTs [28,30,33,35,36], including putative nucleophiles, and positively charged residues could coordinate the phosphates [37–39].

Studies of LgPGT proteins, difficult to express and purify, are limited by challenging *in vitro* assays and a lack of enzyme–substrate complex structures.

To further illuminate this important class of enzyme, we carried out structural and functional studies of the *E. coli* LgPGT WbaP from group 1 capsule biosynthesis. We report the 3.0 Å cryo-EM structure, describing in detail a distinctive domain swap feature. A second, 3.0 Å structure was solved in the presence of UDP-galactose (UDP-Gal) and the homology with SmPGTs was used to locate the active site residues, composed mostly of arginine and lysine residues, and guide site-directed mutagenesis. Using an *in vivo* assay, the significance of this positively charged cluster was assessed. These data give new insight into this important class of enzymes.

2. Material and methods

2.1. Cloning

A DNA fragment containing *wbaP* with an N-terminal decahistidine (His_{10}) tag was amplified from *E. coli* (strain O9a:K30 – E69 [40]) and ligated into the arabinose-inducible expression vector pBAD24 [41] (table 1). Different site-directed mutagenesis experiments modified this vector into different constructs, using Q5 polymerase and restriction enzyme DpnI (NEB) following a published method [41]: N-terminal TEV-cleavable His_{10} tag (pBAD His_{10}-TEV-*wbaP*), N-terminal TEV-cleavable His_6 tag (pBAD His_6-TEV-*wbaP*), and C-terminal TEV-cleavable His_{10} tag (pBAD *wbaP*-TEV-His_{10}). The C-terminal TEV-His_{10} tag removal used the pBAD-*wbaP*-TEV-His_{10} plasmid as a template with specific primers to make the untagged pBAD-*wbaP* vector (table 1). Mutations in the *wbaP* active site (4AS mutant) were engineered by site-directed mutagenesis of a pBAD His_{10}-TEV-*wbaP* and of a pBAD *wbaP*-TEV-His_{10} vector templates using specific primers (table 1). For the GS sequence insertion (5GS, 7GS and 9GS), specific primer pairs were used to remove the sequence encoding the β-hairpin between M246 and Q269 from the pBAD *wbaP*-TEV-His_{10} plasmid template and to insert different lengths of GS repeats (table 1). All the desired codon changes were confirmed by DNA sequencing.

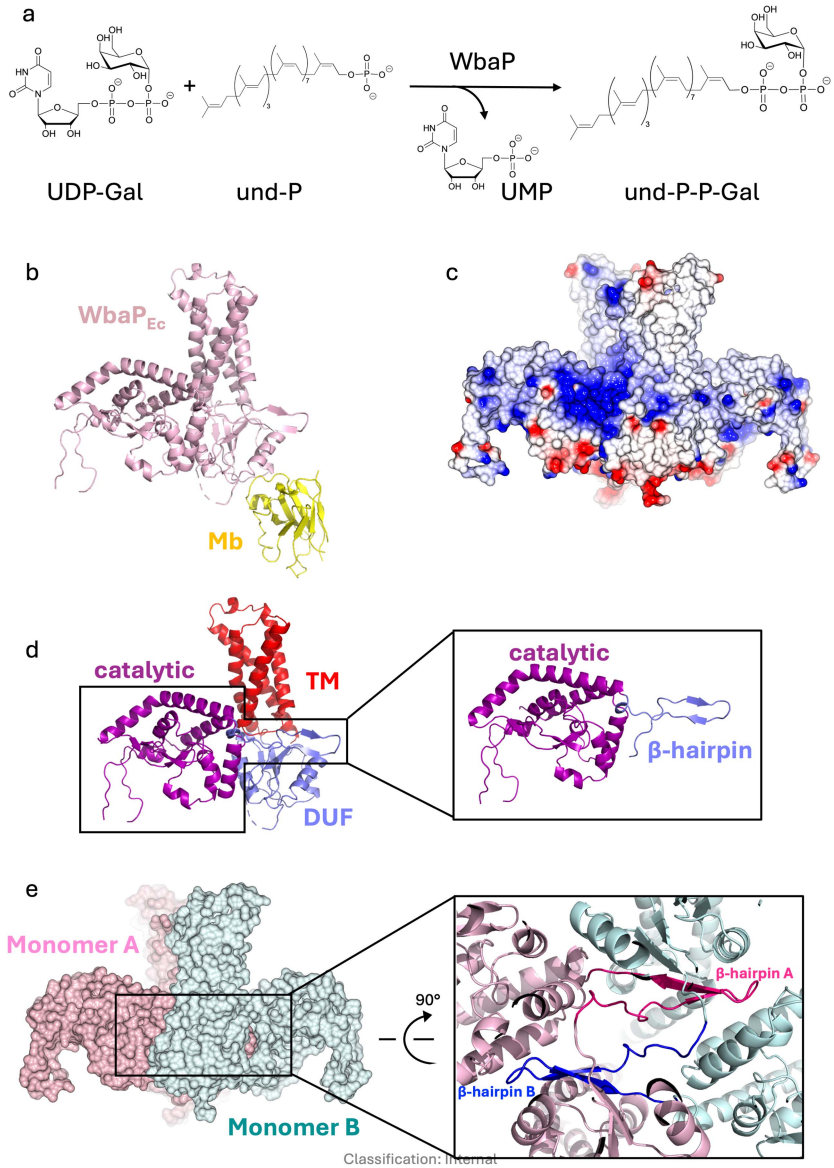


Figure 1. WbaP$_{Ec}$ reaction and cryo-EM structure. (a) Phosphoglycosyl transferase reaction by WbaP. WbaP catalyses the transfer of glucose-1-phosphate from UDP-Gal to the undecaprenol-phosphate acceptor, releasing UMP as a by-product. (b) Cryo-EM structure of *E. coli* apo-WbaP$_{Ec}$. The monomeric form (chain A) is shown in cartoon representation in pink, while the megabody Mb58 (chain C) is in yellow. (c) Electrostatic surface of dimeric apo-WbaP$_{Ec}$. (d) The different domains of one apo-WbaP$_{Ec}$ monomer are shown in cartoon representation, with the TM domain in red, the DUF domain in slate and the catalytic domain in magenta. A zoomed-in section of the catalytic domain (magenta) and the β-hairpin loop (slate) is highlighted. (e) Domain swapping between monomers in apo-WbaP$_{Ec}$ is represented in surface mode (chain A in pink and chain B in cyan). A zoomed-in section of the domain swapping is highlighted. The protein is shown in a cartoon representation, and the β-hairpin loops are represented in magenta (chain A) and blue (chain B).

2.2. Purification of WbaP

The pBAD-*wbaP*-TEV-His$_{10}$ vector was transformed into *E. coli* C43(DE3) cells (Sigma) [42]. *E. coli* C43(DE3)/pBAD-*wbaP*-TEV-His$_{10}$ cells were grown at 37°C with shaking in LB broth containing 100 µg ml^{-1} ampicillin until OD$_{600}$ 0.4–0.6 was reached. Expression of WbaP was then induced by the addition of 0.2% (w/v) L-arabinose, and the culture was grown for 3 h before harvesting cells by centrifugation at 4°C. The cell pellet was stored at −80°C. Cells were thawed and resuspended in 150 ml of cold PBS with 20 µg ml^{-1} DNase I before two passages through a cell disruptor at 30 kpsi. Unbroken cells and large debris were removed from the lysate by centrifugation at 4°C, 15 000 × *g* for 30 min. The bacterial cell membrane fraction was isolated by centrifugation at 4°C, 186 000 × *g* for 90 min. Membrane pellets were washed with cold PBS before storage at −80°C.

Membranes were solubilized overnight at 4°C in buffer A (50 mM sodium phosphate, pH 8.0; 250 mM NaCl; 20 mM imidazole; 2 mM β-mercaptoethanol; 0.1% (w/v) lauryl maltose neopentyl glycol (LMNG, Anatrace)). The insoluble fraction was removed by ultracentrifugation at 4°C, 186 000 × *g* for 90 min. The solubilized WbaP-TEV-His$_{10}$ was batch-bound to nickel NTA agarose resin (Agarose Bead Technologies) with stirring at 4°C for 1 h. Resin was washed with buffer A containing 45 mM imidazole and 0.003% (w/v) LMNG, and protein was eluted with the same buffer containing 300 mM imidazole. The buffer was exchanged to buffer B

Table 1. Primers used in this study.

primer	sequence	description
AJ14	GATCAAGCTTTTAATAAGCACCATCTTTCTTAAG	reverse primer for cloning *his*$_{10}$-*wbaP* into pBAD24 (HindIII)
AJ15	GATCGAATTCACCATGCATCATCATCATCATCATCATCATCATACGCA TTTGGCAAGAAACGTTA	forward primer for cloning *his*$_{10}$-*wbaP* into pBAD24 (EcoRI)
AJ18	ATGACGCATTTGGCAAGAAACGTTAGCTGCAGTATTACACTTGCGG TAGTGTGTAGGCTGGAGCTGCTTCG	lambda red primer for amplifying the kanamycin cassette from pKD4 to knock out *wbaP*
AJ19	ATAAGCACCATCTTTCTTAAGTACAACGCCAATTGTTTTAAATAAT ATAGCCATATGAATATCCTCCTTAG	lambda red primer for amplifying the kanamycin cassette from pKD4 to knock out *wbaP*
R319A_For	GCATGAAGCGGTAGGTAAAGGTGGGAAGCCATTCAAATGTTT	forward and reverse primer pairs used to engineer the four active sites mutant R319A, K331A, R377A and R401A into pBAD His$_{10}$-TEV-*wbaP*_WT plasmid template
R319A_Rev	TACCTACCGCTTCATGCCCATAAATCGCAGAGCCACC	
K331A_For	AATGTTTGGCATTTCGTTCAATGGTCGTTAATTCAAAAGAGG	
K331A_Rev	ACGAAATGCCAAACATTTGAATGGCTTCCCACCTTTACCT	
R377A_For	CAATTTCTAGCTCGAACAAGCCTTGACGAATTACCACAGT	
R377A_Rev	CTTGTTCGAGCTAGAAATTGTCCGATTTTAGTAATACGAGG	
R401A_For	GGGCCTGCTCCAATCATTACTGCTGAACTTGAACGC	
R401A_Rev	TGATTGGAGCAGGCCCCACAAGACTCATTTCTCCTTTC	
5GS_For	AGCGGCTCTGGTAGCAACCTAGCTAAATGGTCATCAC	forward and reverse primers used to engineer the insertion of a GS sequence of 5 residues (5GS) mutant into pBAD *wbaP*_WT-TEV-His$_{10}$ plasmid template
5GS_Rev	TTGCTACCAGAGCCGCTTCCTCGCAATGTTGGTATTACT	
7GS_For	TCTGGTAGCGGCTCTGGTAGCAACCTAGCTAAATGGTCATCAC	forward and reverse primers used to engineer the insertion of a GS sequence of 7 residues (7GS) mutant into pBAD *wbaP*_WT-TEV-His$_{10}$ plasmid template
7GS_Rev	CCAGAGCCGCTACCAGATCCTCGCAATGTTGGTATTACT	
9GS_For	AGCGGCTCCGGTTCTGGTAGCGGCTCTAACCTAGCTAAATGGTCATCAC	forward and reverse primers used to engineer the insertion of a GS sequence of 9 residues (9GS) mutant into pBAD *wbaP*_WT-TEV-His$_{10}$ plasmid template
9GS_Rev	CAGAACCGGAGCCGCTTCCTCGCAATGTTGGTATTACT	
wbaP-Cstop-For	GGTGCTTATTAACTCGAGGAAAACCTGTATTTTCAGGGCG	forward and reverse primers used to insert a stop codon after the C-ter tag using pBAD *wbaP*_WT-TEV-His$_{10}$ plasmid template
wbaP-Cstop-Rev	CCTCGAGTTAATAAGCACCATCTTTCTTAAGTACAACGC	

(20 mM Tris–HCl pH 8.0; 150 mM NaCl; 2 mM β-mercaptoethanol; 10% (v/v) glycerol; 0.003% (w/v) LMNG), before overnight incubation with 0.5 mg of TEV-protease at 4°C. A reverse IMAC step was used to separate WbaP from residual His-tagged WbaP and His$_7$-tagged TEV protease [43] before a size exclusion chromatography (SEC) step was run on a HiLoad 16/600 Superdex 200 column (Cytiva), eluted with buffer B containing 0.001% (w/v) LMNG at 4°C. Pooled fractions, identified by SDS-PAGE, were concentrated to approximately 5 mg ml^{-1}. Purified protein was used immediately or stored at 4°C for up to 48 h. Unless otherwise stated, protein expression and purification were assessed using Any-KDa Stain-Free SDS-PAGE gels (Bio-Rad) and imaged by the ChemiDoc system (Bio-Rad). The protein ladder used was Precision Plus Protein Unstained Standards (Bio-Rad).

2.3. Nanobody discovery, nanobody and megabody purification

To generate nanobodies, WbaP was purified following the purification protocol from above, except LMNG was replaced with 1% (w/v) DDM for extraction and 0.02% (w/v) DDM for purification. The final SEC was run in 20 mM Tris pH 8.5, 10% (v/v) glycerol, 1 mM TCEP, 0.02% (w/v) DDM. Nanobodies (Nb) were raised against purified WbaP and screened by the Steyaert laboratory as part of the Nanobody4Instruct facility, according to published methods [32,44]. Briefly, one llama was immunized over a period of 6 weeks with 900 µg WbaP$_{Ec}$ in total. Phage display libraries were generated, and specific binders were selected after two rounds of biopanning using WbaP immobilized on microtitre plate wells. This led to the discovery of 15 distinct Nb families. Nanobody expression and purification followed published protocols [39,44].

Selected nanobodies were reformatted to generate megabodies (Mbs) through molecular cloning techniques by the Steyaert laboratory through Nanobodies4Instruct using the 45 kDa HopQ scaffold protein (PDB-ID: 5LP2) [32]. Mb plasmids were transformed into Wk6 (su$^-$) *E. coli* cells and the cells were grown at 37°C in Terrific Broth to an OD$_{600}$ of 2.2–2.5. After induction with 0.4 mM isopropyl β-D-thiogalactopyranoside (IPTG), the cells were further grown (4 h) at 37°C, before being harvested

and resuspended in 50 mM Tris–HCl pH 8; 1 mM EDTA; 150 mM NaCl and 20% (w/v) sucrose (50 ml l^{-1}) with DNase I (20 µg ml^{-1}), before stirring for 2 h at 4°C. The cell suspension was centrifuged for 30 min at 18 000 × *g*, at 4°C and 5 mM $MgCl_2$ and 0.5 M NaCl were added to the supernatant before vacuum filtration with a 0.45 µm filter. Protein was batch-bound to Nickel NTA agarose resin (ABT) for 2 h at 4°C, washed with 100 mM Tris–HCl pH 7.4; 500 mM NaCl; 20 mM imidazole and eluted in the same buffer containing 500 mM imidazole. The eluate was concentrated before SEC using a HiLoad 16/600 Superdex S200 (Cytiva) with 10 mM Tris–HCl pH 7.4; 140 mM NaCl as the elution buffer. Protein fractions were pooled and concentrated to approximately 10 mg ml^{-1} before flash freezing and storage at −80°C.

2.4. High-resolution cryo-electron microscopy of WbaP

WbaP was incubated for 2 h at 4°C with nanobody Nb74 and megabody Mb58 derived from Nb73 (previously named Mb^{c7HopQ}_{Nb73}; renamed Mb58) at a WbaP:Nb74:Mb58 molar ratio of 1:1.2:1.2. The complex was loaded on a SEC column (HiLoad 16/600 Superdex 200; Cytiva) and eluted in buffer B. Fractions containing the WbaP-Nb74-Mb58 complex were pooled and concentrated to 3 mg ml^{-1} and centrifuged at 21 000 × *g* for 10 min at 4°C before vitrification on grids. Quantifoil 200 mesh gold grids with a R1.2/1.3 carbon foil were glow discharged at 30 mA twice for 30 s (GloQube, Quorum) and plunge frozen using a Vitrobot Mark IV (Thermo Fisher Scientific). Grids were clipped and data collected on a Titan Krios microscope (Thermo Fisher Scientific) equipped with a K3 camera and Bio-Quantum energy filter (Gatan, UK). Movies were collected in .tiff format in counted super-resolution mode using EPU (Thermo Fisher Scientific). Data collection parameters can be found in table 2.

Initial data processing up to 2D classification was performed using the relion_it pipeline implemented at the UK National electron Bio-Imaging Centre (eBIC), Diamond Light Source [45,46]. For processing in C2 symmetry, movies were imported into CryoSPARC 5.1 [47] and motion corrected using Patch Motion Correction followed by Patch ctf Refinement. Particles were picked with Blob Picker before two rounds of 2D classification. Three Ab-Initio models were generated with C2 symmetry. The best model was refined in C2 symmetry using non-uniform refinement [48] and local refinement. Consecutive rounds of local ctf refinement and global ctf refinement were followed by local refinement in C2. Finally, reference-based motion correction followed by local refinement yielded the final C2 map.

We also processed in C1 symmetry with motion correction and aligned using MotionCor2 (RRID: SCR_016499) [49] in a 5 × 5 patch-based alignment in Relion 3.1 (RRID: SCR_016274) [50]. Ctf-estimation of full-frame non-dose-weighted micrographs was done using CTFFIND4.1 (RRID:SCR_016732) [51], and non-template-driven particle picking was performed in crYOLO (RRID: SCR_018392) [52]. DeepEMhancer (RRID: SCR_022573) [53] was used to perform post-processing and sharpening.

The model of monoPGT PglC from *C. concisus* (PDB-ID 8G1N) was manually placed into the map of WbaP in Chimera (RRID: SCR_004097) [54] and extended by Buccaneer (RRID: SCR_014221) [55,56]. A model of a nanobody (PDB-ID: 6ZBP) was manually placed into additional density observed adjacent to the DUF domain. The model was refined by jelly-body refinement in Refmac5 (RRID: SCR_014225) [57] within CCP-EM [58] and Phenix real-space refinement [59], supported by manual intervention and refinement in coot (RRID: SCR_014222) [60,61]. The map sharpened using DeepEMhancer was used for model building, whereas the unsharpened maps were used for all refinement steps. Maps sharpened within Relion and CryoSPARC were used for ligand detection.

2.5. UDP-Gal-WbaPEc structure

The WbaP-Nb74-Mb58 complex was produced and purified as described above. The protein complex at 2 mg ml^{-1} was incubated with a 500-fold molar excess of UDP-Gal for 80 s on ice before plunge freezing, using the same conditions as described above. Data were collected on a Titan Krios, as described above. Data collection parameters can be found in table 2.

The model of apo-$WbaP_{Ec}$ from refinement was manually placed into the refined map in Chimera and the placement was refined using Molrep in CCP-EM [58,62]. Manual refinement was done in Coot, and the model was refined using Phenix real space refinement (RRID: SCR_014224). The UDP-Gal ligand was manually placed in Coot, and the restraint file was generated using eLBOW [63].

2.6. Functional evaluation of WbaP variants by mutation complementation

To link the molecular structure of the active site to the functional mechanism, the catalytic activity of $WbaP_{Ec}$ was tested using the UDP-Glo assay [26,28]. However, we were unable to obtain enzyme activity *in vitro* that was distinguishable from background, a result we attributed to the complexities of our integral membrane protein purification. Switching to the N-terminal tagged WbaP still did not give activity reliably above controls in the *in vitro* assay [26,28].

As an alternative, the activity of WbaP constructs was assessed by restoration of capsular polysaccharide (CPS) production in a *wbaP* mutant from *E. coli* E69 (O9a:K30). The *wbaP* gene was replaced with a kanamycin-resistance cassette (from the template plasmid pKD4) using lambda-red mutagenesis to produce CWG861 [64] (table 1). *E. coli* CWG861 (Δ*wbaP*) was transformed with plasmids encoding individual WbaP derivatives, and whole-cell lysates were prepared [65]. Briefly, cultures were grown overnight in LB containing 100 µg ml^{-1} ampicillin with no induction. One OD_{600} unit of culture (approx. 5×10^8 cells) was collected by centrifugation at 12 000 × *g* in a microcentrifuge, resuspended in 100 µl of SDS-PAGE loading buffer, and heated in a boiling water bath for 10 min. Proteinase K was then added to 0.5 mg ml^{-1}, and the lysates were incubated at 55°C for 1 h. The lysates were separated on 8% SDS-PAGE gels in Tris–glycine buffer, and the glycans were transferred to nitrocellulose (Amersham Protran 0.45

Table 2. EM statistics for apo-WbaP$_{Ec}$ and UDP-Gal structures.

	apo-WbaP$_{Ec}$ (PDB ID 30US, EMD-58070)	apo-WbaP$_{Ec}$ (PDB ID 9S7S, EMD-54649)	UDP-Gal-WbaP$_{Ec}$ (PDB 30WC, EMD-58108)	UDP-Gal-WbaP$_{Ec}$ (PDB 9S8L, EMD-54664)
data collection and processing				
magnification	81 000	81 000	130 000	130 000
voltage (kV)	300	300	300	300
electron exposure (e^{-} Å^{-2})	40	40	50	50
defocus range (μm)	0.9–2.5	0.9–2.5	1.5–3.0	1.5–3.0
pixel size (Å px^{-1}) (super resolution)	1.072 (0.536)	1.072 (0.536)	0.658 (0.3254)	0.658 (0.3254)
symmetry imposed	C2	C1	C2	C1
initial particle images (picked)	7 899 145 (16 293 709)	1 587 399 (2 902 566)	1 714 443 (5 021 937)	1 455 541 (2 293 857)
final particle images	944 672	1 025 964	492 904	225 535
map resolution limit (Å)	3.0	3.3	3.0	3.3
at FSC threshold	0.143	0.143	0.143	0.143
map range (Å)	8.9–2.5	9.1–3.1	9.9–2.6	10.7–3.1
refinement				
initial model used	8G1N	8G1N	8G1N	8G1N
model resolution (Å)	3.0	3.2	3.0	2.8
FSC threshold	0.143	0.143	0.143	0.143
model resolution range (Å)	2.9–3.1	2.7–3.5	2.9–3.3	2.3–3.4
map sharpening *B* factor (Å^2)	−152	−200	−133	−88
model composition				
non-hydrogen atoms	18 199	18 199	18 043	18 059
protein residues	1121	1121	1106	1104
B factors (Å^2)				
protein	184	194	129	179
ligand			185	282
RMS deviations				
bond lengths (Å)	0.003	0.002	0.003	0.03
bond angles (°)	0.563	0.562	0.581	0.477
validation				
MolProbity score	1.5	1.5	2.1	1.3

(*Continued.*)

Table 2. (*Continued.*)

	apo-WbaP$_{Ec}$ (PDB ID 30US, EMD-58070)	apo-WbaP$_{Ec}$ (PDB ID 9S7S, EMD-54649)	UDP-Gal-WbaP$_{Ec}$ (PDB 30WC, EMD-58108)	UDP-Gal-WbaP$_{Ec}$ (PDB 9S8L, EMD-54664)
clashscore	4.95	5.3	7.7	3.7
poor rotamers (%)	0.41	0.1	1.9	0.1
Ramachandran plot				
favoured (%)	96.2	96.7	93.0	97.3
allowed (%)	3.7	3.2	6.84	2.6
disallowed (%)	0.09	0.1	0.18	0.1

µm NC, Cytiva) in 25 mM Tris, 150 mM glycine, 20% (v/v) methanol, pH 8.3. Blots were blocked in 5% (w/v) skimmed milk in TBS-T (10 mM Tris-Cl, pH 7.5, 150 mM NaCl, 0.05% (v/v) Tween-20). Restoration of CPS biosynthesis by the WbaP derivatives was assessed by reactivity with rabbit antiserum specific to K30 capsular antigen [66] (1:3000 dilution) as the primary antibody and goat anti-rabbit alkaline phosphatase (AP)-conjugated antibody (Cedar Lane, 1:3000 dilution) as the secondary. AP was detected using nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (Roche).

To detect WbaP derivatives, protein expression was induced in mid-exponential growth phase by the addition of 0.01% (w/v) L-arabinose for 1.5 h. Whole-cell lysates were prepared as described previously for the integral membrane protein, WbbF [67]. Protein SDS-PAGE was performed in 10% gels in Tris–glycine buffer. Proteins were transferred to nitrocellulose as described above, and detection was accomplished with mouse anti-Penta-His antibody (Qiagen, 1:3000 in TBST). The secondary antibody was goat-anti mouse horse radish peroxidase-conjugate (Cedar Lane, 1:3000) and chemiluminescent detection was performed using Immobilon Crescendo Western HRP substrate.

2.7. Generation of figures

For Wbap$_{Ec}$, WbaP$_{Se}$ (PDB-ID: 8T53) and PglC$_{Cc}$ structures (PDB-ID: 8E37 and 8G1N), figures were made using Chimera and ChimeraX (RRID: SCR_015872) [54], PyMOL (RRID: SCR_000305), QtMG [68], LigPlot+ (RRID: SCR_018249) [69] and AlphaFill [70]. SEC traces were drawn using GraphPad Prism (SCR_002798). Sequence alignment was made using Clustal Omega (RRID: SCR_001591) [71] and Espript 3.0 [72]. The RMSDs for structure alignments were calculated using the superposition/alignment plugin of Pymol with the super method over 10 cycles of outlier rejection.

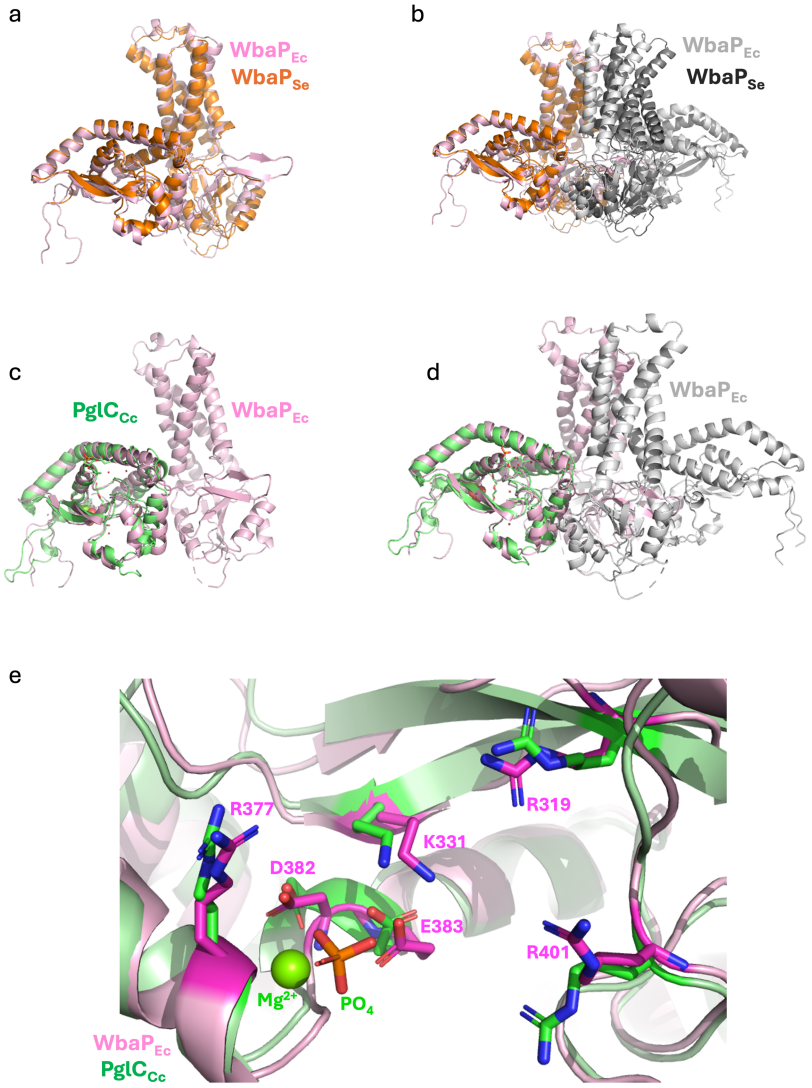
3. Results and discussion

Monotopic PGTs initiate the synthesis of CPSs formed by Wzx/Wzy-dependent pathways, where Wzx is the MATE-family exporter, and Wzy is the polymerase [7,8,73]. The condensation with the lipid is catalysed by either WbaP (which uses a UDP-Gal donor; figure 1a) or WcaJ (UDP-Glc donor), depending on the serotype. WbaP from *E. coli* O9a:K30 was selected as a prototype as part of our efforts to describe the biochemistry of the CPS biosynthetic pathway. The WbaP$_{Ec}$ protein shares 57% sequence identity with the published WbaP$_{Se}$ from O-antigen synthesis in *Salmonella enterica* [28] (supplementary material, figure S1). The catalytic domains of WbaP$_{Ec}$ (residues 274–476) and SmPGT PglC$_{Cc}$ (residues 1–187) share 35% sequence identity [22].

3.1. Structure of apo-WbaP^{Ec}

We set out to determine the structure of apo-WbaP$_{Ec}$. After failing to obtain crystals, we screened a library of 21 nanobodies to stabilize the WbaP$_{Ec}$ protein using SEC [39]. Although crystals were obtained, the resolution was limited to 5–6 Å. Megabodies (engineered nanobodies grafted onto a selected protein scaffold) were used to increase the complex size and aid in particle alignment for single-particle cryo-EM [32]. A cryo-EM apo-WbaP$_{Ec}$ structure at 4.9 Å was obtained with a megabody named Mb58 here (previously named Mb$^{c7HopQ}_{Nb73}$ [32] as Mb58 was itself derived from Nb73) in amphipol A8-35 [72]; however, its resolution was limited due to preferred orientation. Both the addition of a second nanobody (Nb74) and the replacement of amphipols with the detergent LMNG were required to overcome these limitations. This led to a 3.0 Å cryo-EM map in C2 symmetry (figure 1b–e, table 2; supplementary material, figures S2a,b and S3–S6) and an almost complete structure which was built using the PGT PglC$_{Cc}$ structure [21,22].

The apo-WbaP$_{Ec}$ structure aligns well with the published WbaP$_{Se}$ structure (RMSD 1.651 Å over 327 Cα atoms, using PDB-ID: 8T53 chain A and our apo-structure chain A; figure 2a,b). Briefly, in WbaP$_{Ec}$ the N-terminal first four transmembrane helices (residues 1–142) directly connect to a globular α/β domain (residues 143–242) which is followed by two strands (residues 245–269)



Classification: Internal

Figure 2. Structural comparison of apo-WbaP$_{Ec}$ with WbaP$_{Se}$ and PglC$_{Cc}$. Structural alignment of WbaP$_{Ec}$ (pink for chain A and light grey for chain B) with (a,b) WbaP$_{Se}$ (PDB-ID: 8T53, orange for chain A and dark grey for chain B) and (c–e) PglC$_{Cc}$ (PDB-ID: 8G1N, green). (a,c) The monomeric structure (chain A); (b,d) the dimeric structure of WbaP (chains A and B). (e) A close-up of the active site of WbaP$_{Ec}$ (chain A) and PglC$_{Cc}$. The conserved catalytic residues (arginine and lysine residues and the catalytic dyad aspartate–glutamate) are shown as sticks. The numbering corresponds to the WbaP$_{Ec}$ amino acids. The magnesium ion and the phosphate molecule are only present in the PglC$_{Cc}$ crystal structure.

that form a β-hairpin (figure 1d,e). The β-hairpin connects to the C-terminal catalytic domain. An additional region of density did not belong to the apo-WbaP$_{Ec}$ and was modelled as a nanobody domain. The 2D classes show the density extends beyond a simple nanobody and was thus assigned as the megabody Mb58 (figure 1b; supplementary material, figures S2a,b and S5). However, the remaining scaffold of the megabody could not be resolved, potentially due to flexibility. The megabody binds predominantly to the α/β domain of one monomer but also makes contact with the catalytic domain of the other monomer. This bridging of the two monomers, we speculate, rigidifies the dimer. We did not observe any density for nanobody Nb74: either it was disordered, or the complex had dissociated. Similar to the WbaP$_{Se}$ structure, we were unable to fully resolve residues 182–189 of the α/β domain of unknown function (DUF) and residues 340–362 of the PGT domain loop (close to the presumed active site) in the experimental map.

The β-hairpin feature inserts between the α/β and catalytic domains of the other monomer; thus, it can be considered a domain swap and would be predicted to be important for dimer stability (figure 1d,e). PglC is active as a monomer, and the C-terminal domain of WbaP (catalytic core containing the last transmembrane helix) superposes well with PglC and was shown to retain the catalytic activity, ruling out an essential role for the dimer in catalysis (figure 2c,d) [22,35,36]. To probe the role of the β-hairpin in dimer stability, residues M246 to Q269 were replaced with a 5-, 7- or 9-residue serine-glycine (GS) repeat sequence. Only 7GS and 9GS constructs were detectably expressed by stain-free SDS-PAGE gels, but neither could be purified due to a low protein yield, thus preventing definitive assessment (supplementary material, figure S7a).

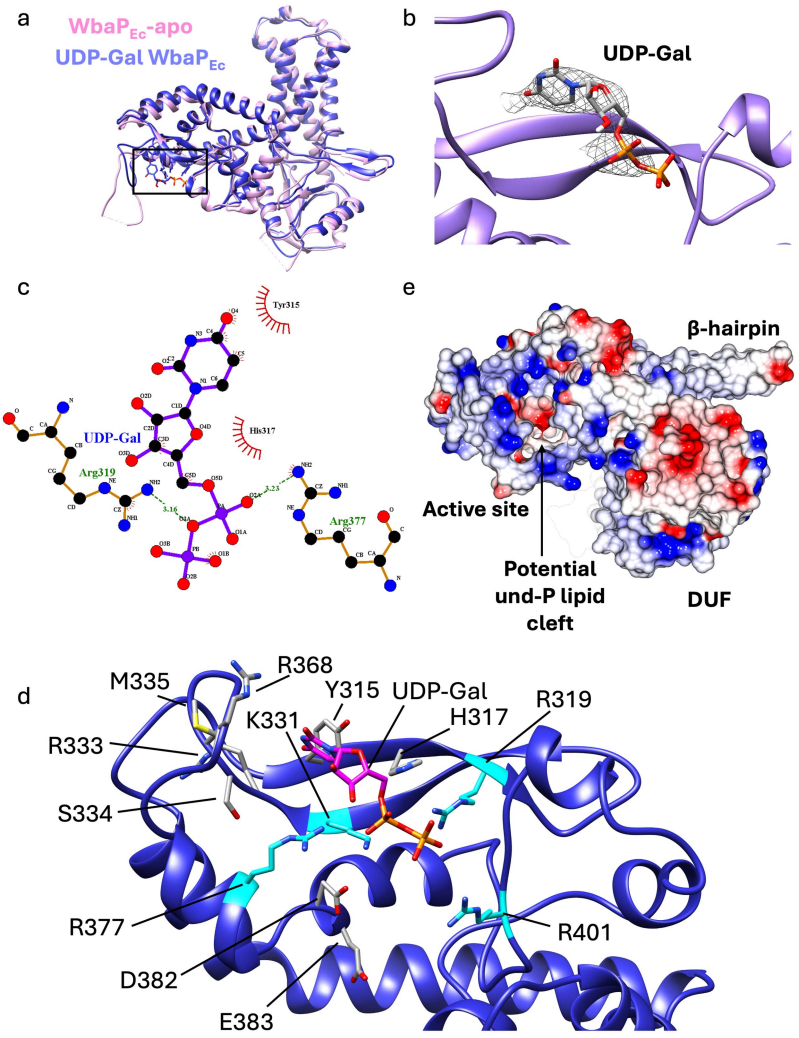


Figure 3. UDP-Gal-WbaP$_{Ec}$ complex. (a) Alignment of the monomeric cryo-EM structures of apo-WbaP$_{Ec}$ (pink) and WbaP$_{Ec}$ in complex with UDP-Gal (slate). One molecule of UDP-Gal is present in each active site of the 'PglC-like' domains (chain A shown here). (b) Close-up of the UDP-Gal substrate binding site of WbaP$_{Ec}$. WbaP$_{Ec}$ chain A is shown in a cartoon representation from the periplasmic side, and UDP-Gal as a stick. The coulomb density for UDP-Gal is represented in grey mesh at an σ level of 23 in Pymol. (c) Interaction of WbaP chain A residues with UDP-Gal, shown as a Ligplot+ plot. Hydrophobic residues are represented in red; hydrogen bonds in green. (d) Cartoon representation of the WbaP$_{Ec}$ active site with residues (slate and cyan) interacting with UDP (as sticks) highlighted. The cyan residues represent the arginine and lysine residues between WbaP and PglC chosen for mutational analysis. (e) Electrostatic surface of the UDP-Gal WbaP$_{Ec}$ monomer (chain A), highlighting the active site with a potential lipid cleft binding site according to [74], the TM and the β-hairpin of the DUF domain.

3.2. WbaPEc structure in complex with UDP-Gal

We determined a 3.0 Å structure of WbaP$_{Ec}$ in the presence of the substrate UDP-Gal, Nb74 and Mb58 (figure 3a, table 2; supplementary material, figures S2c and S8). The overall structure of WbaP$_{Ec}$ is essentially unchanged (RMSD 0.877 Å, over 375 Cα atoms) within the resolutions of the structures (figure 3a; supplementary material, figure S2c). At the local level we observed changes in the kinked membrane helix (residues 292–313), two short helices (residues 367–385 and 400–422) and a loop (residues 440–455), all within the catalytic domain (supplementary material, figure S9). There was additional density located in the catalytic domain, which we modelled as the UDP moiety of UDP-Gal (figure 3b–d; supplementary material, figures S9 and S10). Positioning of the UDP-Gal has some uncertainty given the quality of the map, and we were unable to resolve density for the galactose in our C2 map (figure 3b). In the C1 DeepEMhancer [53], there was some additional density consistent with galactose (supplementary material, figure S8f) in a pocket bounded by three helices and two loops, but the low quality of the map precludes analysis. The uracil moiety π-stacks Y315 and forms hydrophobic interactions with Y315 and H317 close by (figure 3c,d), with M335 and R333 close by. The pyrophosphate makes contacts with the positively charged R319 and R377 residues. K331 is close to the pyrophosphate (approx. 5 Å), but is slightly too far to make a polar contact in our model.

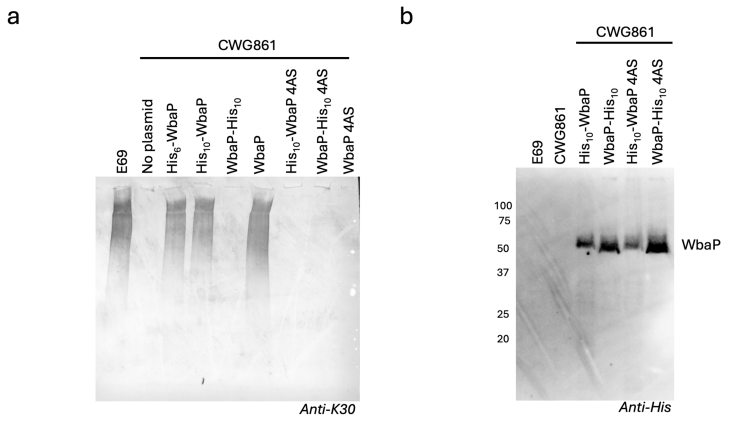


Figure 4. *In vivo* activity assay of $WbaP_{Ec}$. *In vivo* activity assay of $WbaP_{Ec}$ wild-type and catalytic mutants. Western blot of *E. coli* whole-cell extracts using an anti-K30 antibody (a) or anti-His tag antibody (b). E69 is the wild-type *E. coli* strain (O9:K30), and CWG861 is the *wbaP* deletion strain of E69 (Δ*wbaP*). The wild-type $WbaP_{Ec}$ protein with an N-terminal or C-terminal His-tag was tested, along with the active site mutant 4AS, containing alanine mutations of four potentially catalytic residues.

A recent molecular simulation of $PglC_{Cc}$ shows that residues 61–81, which form a loop, close over bound UDP-diNAcBac and und-P [74]. This loop corresponds to residues 333–370 in $WbaP_{Ec}$; consistent with the simulation, part of the loop (339–366) is not defined in our structure, presumably because it is dynamic. The superposition of both structures shows that the ordered portion of the loop is in a similar conformation for both apo-$WbaP_{Ec}$ and UDP-Gal-$WbaP_{Ec}$ structures (supplementary material, figure S9). This corresponds to the 'open' conformation seen in $PglC_{Cc}$ observed in the absence of substrate or in the presence of only one of the substrates. This result is consistent with the molecular simulations that show that the $PglC_{Cc}$ loop only reorganizes upon ligand binding [74].

3.3. Active site of WbaPEc

The C-terminus of WbaP is adjacent to the active site, which could potentially impair the protein's function. This prompted us to compare N-terminally tagged (both His_6 and His_{10}) and C-terminally tagged (His_{10}) WbaP. The activities of those differently tagged $WbaP_{Ec}$ were probed using a complementation assay where the *wbaP* gene was deleted in *E. coli* CWG861, thereby eliminating K30 capsule production (figure 4a,b). Capsule production was reinstated by transforming with a plasmid-encoded $WbaP_{Ec}$. Whilst both N-terminally tagged variants restored activity, the C-terminal one did not (figure 4a; supplementary material, figures S3a and S7b) [26,28]. It is important to note that our structures were obtained using a C-terminally tagged WbaP (due to protein expression yields), which was inactive under our assay conditions. However, given that our apo-structure aligns well with the published $WbaP_{Se}$ structure (RMSD 1.651 Å over 327 Cα atoms), our structures still retain relevance for structural interpretation.

The conserved residues R319, K331, R377 and R401 interact with or are close to the UDP-Gal molecule (Cα within 11 Å) and are located on the dynamic loop (figures 2e and 3c,d; supplementary material, figure S10). To determine the activity of those residues, mutation of these residues *en bloc* in mutant 4AS (R319A-K331A-R377A-R401A) was probed using the same *in vivo* complementation assay. The 4AS mutant did not restore *in vivo* activity (figure 4a,b), suggesting that some of those four positively charged residues are essential for protein activity, independent of the His-tag position. This result is consistent with a single mutation in $WbaP_{Se}$ where incorporation of radioactivity from ^{14}C-labelled donor was abolished [28,35]. The equivalent residues in $PglC_{Cc}$ are R47, K59, R88 and R112 [74] (figure 2e). Previous studies of $PglC_{Cc}$ have proposed that the pyrophosphate of the UDP-diNAcBac substrate is coordinated by K59 (K331 $WbaP_{Ec}$) whilst R112 (R401 $WbaP_{Ec}$) interacts with the nucleotide base [22,27]. Both residues were identified to be involved in binding UDP-Gal in the complex structure of $WbaP_{Ec}$ and UDP-Gal.

$PglC_{Cc}$ residues M63, F43 and Q45 were identified as interacting with both substrates (residues M335, Y315 and H317, respectively, in $WbaP_{Ec}$), with Q45 proposed to align the und-P phosphate to attack UDP-diNAcBac [27,72]. An adventitious phosphate molecule in the structure of $PglC_{Cc}$ was suggested as the location for und-P phosphate (figure 2e; supplementary material, figure S10) [22]. However, we note the different relative positioning of UDP-Gal in $WbaP_{Ec}$ compared to this phosphate molecule (supplementary material, figure S10). Furthermore, the lack of molecular description and coordinates availability for the simulations of $PglC_{Cc}$ does not allow us to compare the binding site of UDP-Gal in our $WbaP_{Ec}$ structure with the binding site of UDP-diNAcBac in $PglC_{Cc}$ [74].

MonoPGTs follow a ping-pong mechanism involving a covalent protein aspartate phospho-sugar intermediate (D93 in $PglC_{Cc}$, demonstrated by trapping the analogous thio-intermediate) [19,22]. D93 is part of a dyad with E94, which is thought to coordinate a Mg^{2+} ion, and both residues are essential for catalysis (figure 2e; supplementary material, figure S10) [19,22,75]. PolyPGT proteins, such as MraY and DPAGT1, also coordinate a magnesium ion using a single aspartic acid residue [17,76]. The corresponding D382 and E383 dyad in $WbaP_{Ec}$ is located at the active site and was shown to be essential for catalysis in $WbaP_{Se}$ (figures 2e and 3d; supplementary material, figure S10) [35]. In our complex, the carboxylate of D382 points towards the UDP-Gal phosphate, consistent with the formation of a covalent mechanism. However, the 8 Å distance means that either the protein

and/or ligand would have to adjust its position for a catalytic role. In our structure, the resolution precluded identification of Mg^{2+}. AlphaFill [70], an algorithm that uses sequence and structure similarity to 'transplant' missing ligands and ions, positions a magnesium ion in $WbaP_{Ec}$ close to the Asp-Glu dyad, similar to the $PglC_{Cc}$ structure, but which only interacts with D382 (supplementary material, figure S11a–c) [35]. Altering the position of the adjacent dynamic loop and concomitant changes at the active site could alter the AlphaFill prediction.

AlphaFill also positions a di-palmitoyl-3-sn-phosphatidylethanolamine molecule (chosen as a simpler mimic of undecaprenyl phosphate) in the active site of $WbaP_{Ec}$, similar to the $PglC_{Cc}$ structure but on the periplasmic face of the membrane [21] (supplementary material, figure S11d,e). Recent simulation experiments with $PglC_{Cc}$ suggest that the phosphate of the und-P molecule interacts with R88 (R377 $WbaP_{Ec}$), R145 (R434 $WbaP_{Ec}$) and K179 (no equivalent in WbaP) and the magnesium ion [22,27,74,75]. In our model of $WbaP_{Ec}$, R377 is close to the pyrophosphate, whereas R434 is distant but is located on a partially disordered loop, thus may require binding of und-P before it becomes ordered (figure 3c,d). The residue R333 (the conserved K59 in $PglC_{Cc}$) points away from the substrate and seems unlikely to interact directly with UDP-Gal. It could instead interact with und-P to stabilize it during catalysis, as previously suggested [75]. While no und-P complex could be obtained with $WbaP_{Ec}$, the recent simulations of $PglC_{Cc}$ suggest the presence of a hydrophobic lipid cleft where the C55 lipid tail could potentially bind on the cytoplasmic face of the membrane (figure 3e) [74].

The experimental cryo-EM structure of $WbaP_{Ec}$ reveals a dimeric arrangement formed by the transmembrane helical bundles with a domain-swapping β-hairpin in the cytoplasmic domain. The role of the highly conserved α/β domain (the DUF) remains unexplained. Computationally annotated as a Co-A binding domain, Dodge *et al.* [28] have suggested a nucleotide sensing regulatory function. We queried the DALI server [77] against the α/β domain, which confirmed a nucleotide-binding capacity. However, our structure does not provide any further insight into the role of the α/β domain, nor do we see any density for UDP-Gal at the α/β domain in our holo-structure. The presence of UDP-Gal now provides experimental evidence for the location of the active site and helps rationalize a number of site-directed mutagenesis studies. The structure suggests that an active site loop is dynamic, consistent with modelling studies which show conformational changes upon substrate binding. The synthesis and transport of lipid-linked oligosaccharides is a potential therapeutic target in bacteria, and WbaP (and its paralogues) is the first committed (and so potentially flux-limiting) step in the biosynthesis of many of these molecules. Given that potent intracellular inhibitors of the polyPGTs are known, such as the natural product tunicamycin, it may well be that corresponding targeting of bacterial monoPGTs, enabled by the structures now revealed here, could prove a fruitful avenue.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. The WbaP EM maps and models were deposited to EMBD and the Protein Data Bank (wwPDB OneDep) under accession codes PDB-ID: 9S7S and EMD-54649 for $WbaP_{Ec}$-apo refined in C1 symmetry; and PDB-ID: 9S8L and EMD-54664 for UDP-Gal-$WbaP_{Ec}$ in C1 symmetry. The C2 symmetry refinement maps and model for the $WbaP_{Ec}$-apo were deposited under PDB-ID: 30US and EMD-58070 for $WbaP_{Ec}$-apo; and PDB-ID: 30WC and EMD-58108 for UDP-Gal-$WbaP_{Ec}$. The raw micrographs are deposited in EMPIAR under accession number EMPIAR-12930 for $WbaP_{Ec}$-apo and EMPIAR-12929 for UDP-Gal-$WbaP_{Ec}$. The WbaP plasmids used in this study are deposited with ADDGENE: pBAD-wbaP-TEV-His_{10} (242398), pBAD-His_6-TEV-wbaP (246461), pBAD-His_{10}-TEV-wbaP (246462), pBAD-wbaP (246463), pBAD-His_{10}-TEV-wbaP_4AS (246464), pBAD-wbaP_4AS-TEV-His_{10} (246465), pBAD-wbaP_4AS (246466), pBAD-wbaP_5GS-TEV-His_{10} (246467), pBAD-wbaP_7GS-TEV-His_{10} (246469), pBAD-wbaP_9GS-TEV-His_{10} (246470). The nanobody and megabody sequences have been deposited with NanoSaurus (Nb74: SD-PV3V and Mb58: SD-8GJ9) and plasmids are available from the VUB, Brussels, upon request.

Supplementary material is available online [78].

Declaration of AI use. AI was not used in the preparation of the manuscript.

Authors' contributions. M.W.: conceptualization, formal analysis, investigation, methodology, writing—original draft, writing—review and editing; A.L.B.: conceptualization, formal analysis, investigation, methodology, writing—original draft, writing—review and editing; P.N.W.: conceptualization, formal analysis, methodology, writing—original draft, writing—review and editing; B.R.C.: conceptualization, formal analysis, investigation, methodology, writing—original draft, writing—review and editing; P.J.H.: investigation, methodology, writing—review and editing; S.A.M.: investigation, methodology, writing—review and editing; H.L.: investigation, methodology, writing—review and editing; A.E.: investigation, methodology, writing—review and editing; A.M.G.: investigation, methodology, writing—review and editing; E.P.: resources, writing—review and editing; A.L.: investigation, methodology, writing—review and editing; J.S.: resources, supervision, writing—review and editing; B.G.D.: supervision, writing—review and editing; C.W.: conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, supervision, writing—original draft, writing—review and editing; J.H.N.: conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. J.H.N., A.L.B., P.N.W, M.W. and P.J.H. were supported by the Wellcome Trust (grant nos. 220526/Z/20/Z and 100209/Z/12/Z). The Rosalind Franklin Institute is funded by UK Research and Innovation through the Engineering and Physical Sciences Research Council (UKRI-EPSRC, EP/V011359/1, EP/T012021/1, EP/X527245/1). Work in C.W.'s laboratory is funded by a Foundation grant (FDN-2016-148364, to C.W.) from the Canadian Institutes of Health Research and by the Canada Research Chairs programme. The Membrane Protein Laboratory at Diamond Light Source, an Instruct-ERIC centre, is funded by grants from the Wellcome Trust (202892/Z/16/Z and 223727/Z/21/Z), with additional support provided by Diamond Light Source and the Research Complex at Harwell, both Instruct-ERIC centres. EM screening was performed using the Research Complex at Harwell Glacios at Diamond Light Source (proposal NT29493). We thank Diamond Light Source for access and support of the cryo-EM facilities at the UK National electron Bio-Imaging Centre (eBIC), proposals NR27436 and NT29493. Additional support was provided by Diamond Light Source and the Research Complex at Harwell, both Instruct-ERIC centres. We thank INSTRUCT-ERIC, part of the European Strategy Forum on Research Infrastructures (ESFRI), and the Research Foundation—Flanders (FWO) for their support of the Nanobody discovery. The nanobodies and megabodies were obtained via Nanobodies4INSTRUCT, Brussels, project PID2899-VID4581 and PID8707-VID15847 (Nb74 (reference CA13974), megabody Mb58 (reference CA15858) derived from Nb73 (reference CA13973)).

Acknowledgements. We thank Katleen Willibal for the technical assistance during nanobody discovery and Eva Beke for her contributions in reformatting nanobodies into megabodies.

References

1. Di Lorenzo F, Duda KA, Lanzetta R, Silipo A, De Castro C, Molinaro A. 2022 A journey from structure to function of bacterial lipopolysaccharides. *Chem. Rev.* **122**, 15767–15821. (doi:10.1021/acs.chemrev.0c01321)
2. Sun J, Rutherford ST, Silhavy TJ, Huang KC. 2022 Physical properties of the bacterial outer membrane. *Nat. Rev. Microbiol.* **20**, 236–248. (doi:10.1038/s41579-021-00638-0)
3. Tytgat HLP, Lebeer S. 2014 The sweet tooth of bacteria: common themes in bacterial glycoconjugates. *Microbiol. Mol. Biol. Rev.* **78**, 372–417. (doi:10.1128/mmbr.00007-14)
4. Liston SD, Willis LM. 2021 Racing to build a wall: glycoconjugate assembly in Gram-positive and Gram-negative bacteria. *Curr. Opin. Struct. Biol.* **68**, 55–65. (doi:10.1016/j.sbi.2020.11.013)
5. Whitfield C, Williams DM, Kelly SD. 2020 Lipopolysaccharide O-antigens—bacterial glycans made to measure. *J. Biol. Chem.* **295**, 10593–10609. (doi:10.1074/jbc.REV120.009402)
6. Raetz CRH, Whitfield C. 2002 Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **71**, 635–700. (doi:10.1146/annurev.biochem.71.110601.135414)
7. Whitfield C. 2006 Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem.* **75**, 39–68. (doi:10.1146/annurev.biochem.75.103004.142545)
8. Whitfield C, Wear SS, Sande C. 2020 Assembly of bacterial capsular polysaccharides and exopolysaccharides. *Annu. Rev. Microbiol.* **74**, 521–543. (doi:10.1146/annurev-micro-011420-075607)
9. Morona R, Purins L, Tocilj A, Matte A, Cygler M. 2009 Sequence-structure relationships in polysaccharide co-polymerase (PCP) proteins. *Trends Biochem. Sci.* **34**, 78–84. (doi:10.1016/j.tibs.2008.11.001)
10. Brown S, Santa Maria JP, Walker S. 2013 Wall teichoic acids of Gram-positive bacteria. *Annu. Rev. Microbiol.* **67**, 313–336. (doi:10.1146/annurev-micro-092412-155620)
11. Jankute M, Cox JAG, Harrison J, Besra GS. 2015 Assembly of the mycobacterial cell wall. *Annu. Rev. Microbiol.* **69**, 405–423. (doi:10.1146/annurev-micro-091014-104121)
12. Whitfield C, Williams DM, Kelly SD. 2020 Lipopolysaccharide O-antigens—bacterial glycans made to measure. *J. Biol. Chem.* **295**, 10593–10609. (doi:10.1074/jbc.REV120.009402)
13. Ascari A, Morona R. 2025 Recent insights into Wzy polymerases and lipopolysaccharide O-antigen biosynthesis. *J. Bacteriol.* **207**, e0041724. (doi:10.1128/jb.00417-24)
14. Ramírez AS, Locher KP. 2023 Structural and mechanistic studies of the N-glycosylation machinery: from lipid-linked oligosaccharide biosynthesis to glycan transfer. *Glycobiology* **33**, 861–872. (doi:10.1093/glycob/cwad053)
15. O'Toole KH, Bernstein HM, Allen KN, Imperiali B. 2021 The surprising structural and mechanistic dichotomy of membrane-associated phosphoglycosyl transferases. *Biochem. Soc. Trans.* **49**, 1189–1203. (doi:10.1042/BST20200762)
16. Al-Dabbagh B, Olatunji S, Crouvoisier M, El Ghachi M, Blanot D, Mengin-Lecreulx D, Bouhss A. 2016 Catalytic mechanism of MraY and WecA, two paralogues of the polyprenyl-phosphate N-acetylhexosamine 1-phosphate transferase superfamily. *Biochimie* **127**, 249–257. (doi:10.1016/j.biochi.2016.06.005)
17. Liu Y *et al.* 2016 New insight into the catalytic mechanism of bacterial MraY from enzyme kinetics and docking studies. *J. Biol. Chem.* **291**, 15057–15068. (doi:10.1074/jbc.M116.717884)
18. Al-Dabbagh B, Mengin-Lecreulx D, Bouhss A. 2008 Purification and characterization of the bacterial UDP-GlcNAc:undecaprenyl-phosphate GlcNAc-1-phosphate transferase WecA. *J. Bacteriol.* **190**, 7141–7146. (doi:10.1128/JB.00676-08)
19. Das D, Kuzmic P, Imperiali B. 2017 Analysis of a dual domain phosphoglycosyl transferase reveals a ping-pong mechanism with a covalent enzyme intermediate. *Proc. Natl Acad. Sci. USA* **114**, 7019–7024. (doi:10.1073/pnas.1703397114)
20. O'Toole KH, Imperiali B, Allen KN. 2021 Glycoconjugate pathway connections revealed by sequence similarity network analysis of the monotopic phosphoglycosyl transferases. *Proc. Natl Acad. Sci. USA* **118**. e2018289118. (doi:10.1073/pnas.2018289118)
21. Anderson AJ, Dodge GJ, Allen KN, Imperiali B. 2023 Co-conserved sequence motifs are predictive of substrate specificity in a family of monotopic phosphoglycosyl transferases. *Protein Sci.* **32**, e4646. (doi:10.1002/pro.4646)
22. Ray LC, Das D, Entova S, Lukose V, Lynch AJ, Imperiali B, Allen KN. 2018 Membrane association of monotopic phosphoglycosyl transferase underpins function. *Nat. Chem. Biol.* **14**, 538–541. (doi:10.1038/s41589-018-0054-z)
23. Allen KN, Entova S, Ray LC, Imperiali B. 2019 Monotopic membrane proteins join the fold. *Trends Biochem. Sci.* **44**, 7–20. (doi:10.1016/j.tibs.2018.09.013)
24. Durand T, Dodge GJ, Siuda RP, Higinbotham HR, Arbour CA, Ghosh S, Allen KN, Imperiali B. 2024 Proteome-wide bioinformatic annotation and functional validation of the monotopic phosphoglycosyl transferase superfamily. *Proc. Natl Acad. Sci. USA* **121**, e2417572121. (doi:10.1073/pnas.2417572121)
25. Lukose V, Walvoort MT, Imperiali B. 2017 Bacterial phosphoglycosyl transferases: initiators of glycan biosynthesis at the membrane interface. *Glycobiology* **27**, 820–833. (doi:10.1093/glycob/cwx064)
26. Das D, Walvoort MTC, Lukose V, Imperiali B. 2016 A rapid and efficient luminescence-based method for assaying phosphoglycosyltransferase enzymes. *Sci. Rep.* **6**, 33412. (doi:10.1038/srep33412)
27. Lukose V, Luo L, Kozakov D, Vajda S, Allen KN, Imperiali B. 2015 Conservation and covariance in small bacterial phosphoglycosyl transferases identify the functional catalytic core. *Biochemistry* **54**, 7326–7334. (doi:10.1021/acs.biochem.5b01086)
28. Dodge GJ, Anderson AJ, He Y, Liu W, Viner R, Imperiali B. 2024 Mapping the architecture of the initiating phosphoglycosyl transferase from *S. enterica* O-antigen biosynthesis in a liponanoparticle. *eLife* **12**, RP91125. (doi:10.7554/eLife.91125)
29. Jiang X-M, Neal B, Santiago F, Lee SJ, Romana LK, Reeves PR. 1991 Structure and sequence of the rfb (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). *Mol. Microbiol.* **5**, 695–713. (doi:10.1111/j.1365-2958.1991.tb00741.x)
30. Wang L, Liu D, Reeves PR. 1996 C-terminal half of *Salmonella enterica* WbaP (RfbP) is the galactosyl-1-phosphate transferase domain catalyzing the first step of O-antigen synthesis. *J. Bacteriol.* **178**, 2598–2604. (doi:10.1128/jb.178.9.2598-2604.1996)
31. Wang L, Reeves PR. 1994 Involvement of the galactosyl-1-phosphate transferase encoded by the *Salmonella enterica* rfbP gene in O-antigen subunit processing. *J. Bacteriol.* **176**, 4348–4356. (doi:10.1128/jb.176.14.4348-4356.1994)

32. Uchański T *et al.* 2021 Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. *Nat. Methods* **18**, 60–68. (doi:10.1038/s41592-020-01001-6)
33. Saldías MS, Patel K, Marolda CL, Bittner M, Contreras I, Valvano MA. 2008 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. (doi:10.1099/mic.0.2007/013136-0)
34. Schäffer C, Wugeditsch T, Messner P, Whitfield C. 2002 Functional expression of enterobacterial o-polysaccharide biosynthesis enzymes in *Bacillus subtilis*. *Appl. Environ. Microbiol.* **68**, 4722–4730. (doi:10.1128/AEM.68.10.4722-4730.2002)
35. Patel KB, Furlong SE, Valvano MA. 2010 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. (doi:10.1093/glycob/cwq104)
36. Patel KB, Ciepichal E, Swiezewska E, Valvano MA. 2012 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. (doi:10.1093/glycob/cwr114)
37. Perez C, Köhler M, Janser D, Pardon E, Steyaert J, Zenobi R, Locher KP. 2017 Structural basis of inhibition of lipid-linked oligosaccharide flippase PglK by a conformational nanobody. *Sci. Rep.* **7**, 46641. (doi:10.1038/srep46641)
38. Perez C, Gerber S, Boilevin J, Bucher M, Darbre T, Aebi M, Reymond JL, Locher KP. 2015 Structure and mechanism of an active lipid-linked oligosaccharide flippase. *Nature* **524**, 433–438. (doi:10.1038/nature14953)
39. Le Bas A *et al.* 2025 Structure of WzxE the lipid III flippase for enterobacterial common antigen polysaccharide. *Open Biol.* **15**, 240310. (doi:10.1098/rsob.240310)
40. Homonylo MK, Wilmot SJ, Lam JS, MacDonald LA, Whitfield C. 1988 Monoclonal antibodies against the capsular K antigen of *Escherichia coli* (09: K30(A): H12): characterisation and use in analysis of K antigen organisation on the cell surface. *Can. J. Microbiol.* **34**, 1159–1165. (doi:10.1139/m88-204)
41. Liu H, Naismith JH. 2008 An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91. (doi:10.1186/1472-6750-8-91)
42. Miroux B, Walker JE. 1996 Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298. (doi:10.1006/jmbi.1996.0399)
43. Kapust RB, Tözsér J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS. 2001 Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng.* **14**, 993–1000. (doi:10.1093/protein/14.12.993)
44. Pardon E *et al.* 2014 A general protocol for the generation of nanobodies for structural biology. *Nat. Protoc.* **9**, 674–693. (doi:10.1038/nprot.2014.039)
45. Fernandez-Leiro R, Scheres SHW. 2017 A pipeline approach to single-particle processing in *RELION*. *Acta Crystallogr. D* **73**, 496–502. (doi:10.1107/S2059798316019276)
46. Horstmann A, Riggs S, Chaban Y, Clare DK, de Freitas G, Farmer D, Howe A, Morris KL, Hatton D. 2024 A service-based approach to cryoEM facility processing pipelines at eBIC. *Acta Crystallogr. D* **80**, 174–180. (doi:10.1107/S2059798324000986)
47. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. 2017 cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296. (doi:10.1038/nmeth.4169)
48. Punjani A, Zhang H, Fleet DJ. 2020 Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221. (doi:10.1038/s41592-020-00990-8)
49. Zheng SQ, Palovcak E, Armache J-P, Verba KA, Cheng Y, Agard DA. 2017 MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332. (doi:10.1038/nmeth.4193)
50. Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, Scheres SH. 2018 New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166. (doi:10.7554/eLife.42166)
51. Rohou A, Grigorieff N. 2015 CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221. (doi:10.1016/j.jsb.2015.08.008)
52. Wagner T *et al.* 2019 SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218. (doi:10.1038/s42003-019-0437-z)
53. Sanchez-Garcia R, Gomez-Blanco J, Cuervo A, Carazo JM, Sorzano COS, Vargas J. 2021 DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* **4**, 874. (doi:10.1038/s42003-021-02399-1)
54. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004 UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612. (doi:10.1002/jcc.20084)
55. Cowtan K. 2006 The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002–1011. (doi:10.1107/S0907444906022116)
56. Cowtan K. 2008 Fitting molecular fragments into electron density. *Acta Crystallogr. D* **64**, 83–89. (doi:10.1107/S0907444907033938)
57. Brown A, Long F, Nicholls RA, Toots J, Emsley P, Murshudov G. 2015 Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D* **71**, 136–153. (doi:10.1107/S1399004714021683)
58. Burnley T, Palmer CM, Winn M. 2017 Recent developments in the CCP-EM software suite. *Acta Crystallogr. D* **73**, 469–477. (doi:10.1107/S2059798317007859)
59. Liebschner D *et al.* 2019 Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877. (doi:10.1107/S2059798319011471)
60. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010 Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501. (doi:10.1107/S0907444910007493)
61. Emsley P, Cowtan K. 2004 Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132. (doi:10.1107/S0907444904019158)
62. Vagin A, Teplyakov A. 2010 Molecular replacement with MOLREP. *Acta Crystallogr. D* **66**, 22–25. (doi:10.1107/S0907444909042589)
63. Moriarty NW, Grosse-Kunstleve RW, Adams PD. 2009 Electronic ligand builder and optimization workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D* **65**, 1074–1080. (doi:10.1107/S0907444909029436)
64. Datsenko KA, Wanner BL. 2000 One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645. (doi:10.1073/pnas.120163297)
65. Hitchcock PJ, Brown TM. 1983 Morphological heterogeneity among *Salmonella* lipopolysaccharide chemotypes in silver-stained polyacrylamide gels. *J. Bacteriol.* **154**, 269–277. (doi:10.1128/jb.154.1.269-277.1983)
66. Dodgson C, Amor P, Whitfield C. 1996 Distribution of the rol gene encoding the regulator of lipopolysaccharide O-chain length in *Escherichia coli* and its influence on the expression of group I capsular K antigens. *J. Bacteriol.* **178**, 1895–1902. (doi:10.1128/jb.178.7.1895-1902.1996)

67. Wear SS, Hunt BA, Clarke BR, Whitfield C. 2020 Analysis of the topology and active-site residues of WbbF, a putative O-polysaccharide synthase from *Salmonella enterica* serovar borreze. *J. Bacteriol.* **202**, e00625-19. (doi:10.1128/JB.00625-19)
68. Winn MD *et al.* 2011 Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242. (doi:10.1107/S0907444910045749)
69. Laskowski RA, Swindells MB. 2011 LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51**, 2778–2786. (doi:10.1021/ci200227u)
70. Hekkelman ML, de Vries I, Joosten RP, Perrakis A. 2023 AlphaFill: enriching alphafold models with ligands and cofactors. *Nat. Methods* **20**, 205–213. (doi:10.1038/s41592-022-01685-y)
71. Madeira F, Madhusoodanan N, Lee J, Eusebi A, Niewielska A, Tivey ARN, Lopez R, Butcher S. 2024 The EMBL-EBI job dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res.* **52**, W521–W525. (doi:10.1093/nar/gkae241)
72. Robert X, Gouet P. 2014 Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324. (doi:10.1093/nar/gku316)
73. Sande C, Whitfield C. 2021 Capsules and extracellular polysaccharides in *Escherichia coli* and *Salmonella*. *EcoSal Plus* **9**, eESP00332020. (doi:10.1128/ecosalplus.ESP-0033-2020)
74. Higinbotham HR, Arbour CA, Imperiali B. 2025 Correlating membrane-protein dynamics with function: integrating bioinformatics, molecular dynamics, and single-molecule FRET. *Protein Sci.* **34**, e70352. (doi:10.1002/pro.70352)
75. Majumder A, Vuksanovic N, Ray LC, Bernstein HM, Allen KN, Imperiali B, Straub JE. 2023 Synergistic computational and experimental studies of a phosphoglycosyl transferase membrane/ligand ensemble. *J. Biol. Chem.* **299**, 105194. (doi:10.1016/j.jbc.2023.105194)
76. Dong YY *et al.* 2018 Structures of DPAGT1 explain glycosylation disease mechanisms and advance TB antibiotic design. *Cell* **175**, 1045–1058.e16. (doi:10.1016/j.cell.2018.10.037)
77. Holm L. 2022 Dali server: structural unification of protein families. *Nucleic Acids Res.* **50**, W210-W215. (doi:10.1093/nar/gkac387)
78. Weckener M *et al.* 2026 Supplementary material from: Structural insights into the polytopic *Escherichia coli* phosphoglycosyl transferase WbaP. Figshare. (doi:10.6084/m9.figshare.c.8588522)