

Research



Cite this article: Weckener M, Woodward LS, Clarke BR, Liu H, Ward PN, Le Bas A, Bhella D, Whitfield C, Naismith JH. 2023 The lipid linked oligosaccharide polymerase Wzy and its regulating co-polymerase, Wzz, from enterobacterial common antigen biosynthesis form a complex. *Open Biol.* **13**: 220373. <https://doi.org/10.1098/rsob.220373>

Received: 21 December 2022

Accepted: 27 February 2023

Subject Area:

structural biology

Keywords:

lipid, oligosaccharides, polymerase, Wzy, regulating

Authors for correspondence:

Chris Whitfield

e-mail: cwhitfie@uoguelph.ca

James H. Naismith

e-mail: James.naismith@strubi.ox.ac.uk

[†]These authors contributed equally to the study.

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.6461045>.

The lipid linked oligosaccharide polymerase Wzy and its regulating co-polymerase, Wzz, from enterobacterial common antigen biosynthesis form a complex

Miriam Weckener^{1,2,†}, Laura S. Woodward^{3,†}, Bradley R. Clarke⁴, Huanting Liu³, Philip N. Ward^{1,2}, Audrey Le Bas^{1,2}, David Bhella⁵, Chris Whitfield⁴ and James H. Naismith^{1,2}

¹Structural Biology, The Rosalind Franklin Institute, Harwell Campus, Didcot OX11 0QS, UK

²Division of Structural Biology, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

³Centre Biomedical Sciences, North Haugh, University of St Andrews, St Andrews KY16 9ST, UK

⁴Department of Molecular and Cellular Biology, The University of Guelph, Guelph, ON, Canada

⁵MRC—University of Glasgow Centre for Virus Research, University of Glasgow, Sir Michael Stoker Building, Garscube Campus, 464 Bearsden Road, Glasgow G61 1Q, UK

MW, 0000-0002-1419-9378; PNW, 0000-0003-2546-3426; ALB, 0000-0002-1100-6833; JHN, 0000-0001-6744-5061

The enterobacterial common antigen (ECA) is a carbohydrate polymer that is associated with the cell envelope in the *Enterobacteriaceae*. ECA contains a repeating trisaccharide which is polymerized by WzyE, a member of the Wzy membrane protein polymerase superfamily. WzyE activity is regulated by a membrane protein polysaccharide co-polymerase, WzzE. Förster resonance energy transfer experiments demonstrate that WzyE and WzzE from *Pectobacterium atrosepticum* form a complex *in vivo*, and immunoblotting and cryo-electron microscopy (cryo-EM) analysis confirm a defined stoichiometry of approximately eight WzzE to one WzyE. Low-resolution cryo-EM reconstructions of the complex, aided by an antibody recognizing the C-terminus of WzyE, reveals WzyE sits in the central membrane lumen formed by the octameric arrangement of the trans-membrane helices of WzzE. The pairing of Wzy and Wzz is found in polymerization systems for other bacterial polymers, including lipopolysaccharide O-antigens and capsular polysaccharides. The data provide new structural insight into a conserved mechanism for regulating polysaccharide chain length in bacteria.

1. Introduction

Gram-negative bacteria are defined by a complex cell envelope format, which includes an outer membrane that provides a permeability barrier, which excludes large molecules including some antibiotics [1], affords structural strength [2] and helps protect bacteria from components of host immune defenses [3]. The inner leaflet of the outer membrane is formed of glycerophospholipids, but the outer leaflet consists almost entirely of a characteristic glycolipid called lipopolysaccharide (LPS). LPS molecules are complex structures [4], containing a highly conserved lipid A molecule made up of a phosphorylated disaccharide modified with four to seven acyl chains [5]. Lipid A is itself glycosylated by the core oligosaccharide (core OS), which shows some structural variation between species. In many bacteria, the core OS is further decorated by an O-antigen polysaccharide

(OPS). The OPS polymer exhibits heterogeneous chain length distribution and repeat unit composition can vary considerably between species and between O-antigen serotypes within a species [3]; for example, there are more than 180 different O-antigens known for *Escherichia coli* [6]. Bacteria in which the OPS-biosynthesis genes have been deleted show reduced virulence due to impaired resistance to the immune response [3]. The diversity of OPS arises from the identities of the sugars that comprise the repeating unit, how they are connected, the extent of any branching to the linear arrangement, and chemical modification (e.g. acetylation) [3].

LPS is not the sole polysaccharide associated with the outer membrane or found on the cell surface of Gram-negative bacteria. Some bacteria are also protected from the immune system by a layer of high-molecular weight capsular polysaccharides (CPSs) [7]. Like OPS, CPSs have hypervariable repeat unit structures with more than 80 capsular serotypes in *E. coli*. Most members of the Enterobacteriaceae also produce a carbohydrate polymer called the enterobacterial common antigen (ECA) that is associated with the cell envelope. ECA can be linked to either diacylglycerolphosphate (ECA_{PG}) or LPS lipid A core (ECA_{LPS}), or it can exist in a lipid-free cyclic form in the periplasm (ECA_{CYC}), and its precise role (or roles) in bacterial physiology is enigmatic [8]. ECA contains a trisaccharide repeat unit (*N*-acetyl-D-glucosamine-(α 1-4)-*N*-acetyl-D-mannosaminuronic acid(β 1-4)-acetamido-4,6-dideoxy-D-galactose) [8].

ECA [8], most OPSs [3] and some CPSs [7] are synthesized by a conserved strategy, employing an assembly line of enzymes often referred to as the Wzx/Wzy pathway [9]. In ECA biosynthesis, the enzymatic cascade is initialized by a phosphoglycosyltransferase called WecA, and this enzyme participates in production of OPSs in many bacteria, including all OPS types in *E. coli* [3,8]. WecA transfers an acetamido sugar phosphate onto membrane-embedded undecaprenyl phosphate (Und-P) at the interface with the cytosol. In other OPSs and many CPSs, bacteria use a homologue of the WbaP phosphoglycosyltransferase to transfer a hexose phosphate to Und-P [3,7]. In each case, the reaction products are elaborated into an Und-PP-linked oligosaccharide repeat unit by membrane-associated glycosyltransferases. The lipid-linked repeat unit is then translocated across the inner membrane into the periplasm by a flippase Wzx (designated Wzx_E in ECA to distinguish it from homologues from other examples of the assembly strategy found in the same cell). Wzx proteins belong to the multidrug and toxic compound extrusion (MATE) exporter family [10]. Once flipped, ECA repeat units are polymerized by transfer of the growing chain from its Und-PP carrier to the non-reducing terminus of an incoming Und-PP-linked repeat unit. This reaction is catalysed by the integral membrane polymerase WzyE, which belongs to a larger shape, elongation, division and sporulation (SEDS) glycosyltransferase family [11]. After polymerization, the fate of the final polysaccharide polymer depends on the pathway. OPS (and presumably ECA_{LPS}) is transferred onto the lipid A core by the O-antigen ligase WaaL [3,12,13] and transferred to the outer leaflet of the outer membrane by the LPS transport (Lpt) machine [14]. Diacylglycerol phosphate is added to ECA_{PG} from a phosphatidylglycerol donor [15], but the enzyme responsible is uncertain, as is the export fate of this species. In Wzy-polymerized CPS, the polymer is transported to the surface through the dedicated outer membrane translocon Wza [16,17].

A characteristic feature of the Wzy-polymerization systems for OPS, ECA and CPS is the regulation of the polymerization phase by inner membrane proteins belonging to the polysaccharide co-polymerase (PCP) family [18]. This is best documented for OPS, where a PCP-1 protein (called Wzz) controls polymerization to create an abundant fraction of OPS species with a more homogeneous (modal) distribution of chain lengths [3]. A related PCP-1 protein (Wzz_E) operates in ECA biosynthesis [8,18]. Wzz proteins consist of two transmembrane helices flanking a large periplasmic domain; there is no cytoplasmic domain. The periplasmic domains of various Wzz homologues were found to crystallize in different oligomerization states [19–21] and the structures of full length, decameric Wzz_{SE} (from OPS biosynthesis in *Salmonella enterica* serovar Typhimurium) and octameric Wzz_{EC} (OPS biosynthesis in *E. coli* K-12) were recently solved by cryo-electron microscopy (cryo-EM) [22,23]. Wzz homologues have also been reported as (non-octameric) oligomers in a native-like lipid environment using cryo-EM [24]. All Wzz structures show a central cavity in the membrane region that is presumably filled by membrane lipids [22–24]. Wzy-dependent polymerization of CPS in Gram-negative bacteria exploits a more complicated PCP-2a protein (Wzc from *E. coli* is the prototype), distinguished from PCP-1 proteins by possession of an additional functionally essential cytoplasmic kinase domain [7]. Wzc is an octamer and the transmembrane helices are arranged in a non-close packed circular arrangement that creates a central cavity in the membrane similar to Wzz [25].

No direct catalytic activity has been described for any Wzz protein but mutants lacking *wzz* can no longer produce high-molecular weight modal OPS, suggesting it plays a regulatory and/or structural role [3]. The regulation of the OPS polymer length has been reconstituted *in vitro* by addition of Wzz to Wzy suggesting they may form a complex [26]. This concept is supported by *in vivo* cross-linking and co-immunoprecipitation studies in *Shigella flexneri* [27–30] and a yeast two hybrid experiment in *Rhizobium leguminosarum* [31]. The C-terminal transmembrane helix in Wzz appears to interact with Wzy and multiple residues in Wzy have been shown to be important for the interaction with Wzz [28–30]. Wzz has also been shown to co-immunoprecipitate with O-antigen, suggesting a direct interaction with the reaction product [28–30,32,33]. Mutations that affect OPS chain length are distributed throughout Wzz structures [34–36]. Notably variants of Wzz that are unable to oligomerize are unable to regulate chain length [36,37], but no simple relationship between polymer chain length and Wzz oligomerization state has been established [19]. This has led to several hypotheses for Wzz's mode of action [13,18,19,22,36,38–40], including: (i) the circumference of the periplasmic domain acting as some form of ruler for chain length [36]; (ii) Wzz as a scaffold around which multiple Wzy molecules assemble [19]; (iii) Wzz binding the growing OPS in such a way as to facilitate catalysis until polymer length breaks the interaction [40]; and (iv) OPS extending within the lumen of Wzz until it reaches the top of the periplasmic domain [13].

Here, we report a series of biochemical and structural studies on Wzy_{EPA} and Wzz_{EPA} from the ECA pathway of *Pectobacterium atrosepticum* (77% and 70% sequence identity to *E. coli* K12 WzyE and WzzE, respectively) that validate and characterize complex formation *in vitro* and *in vivo*.

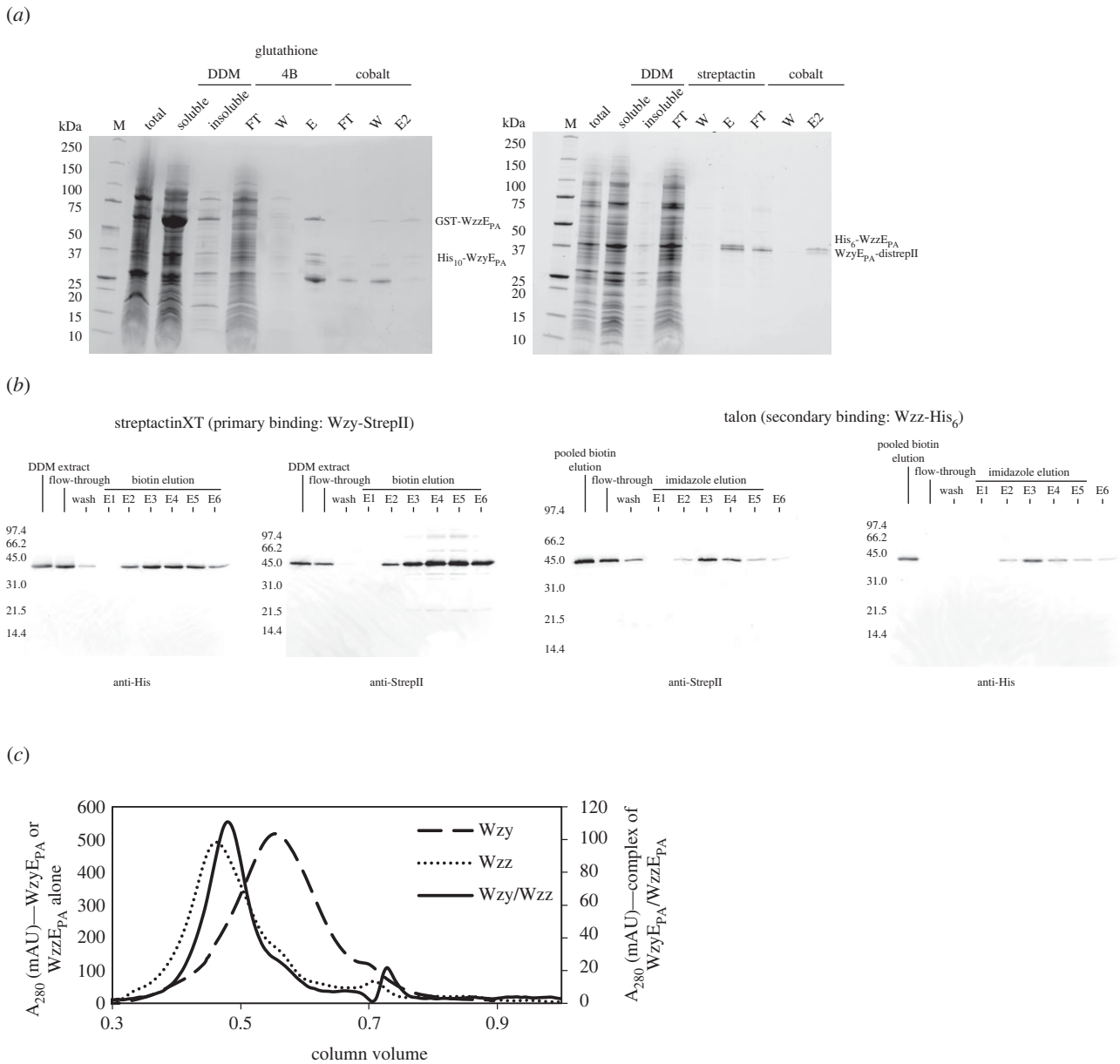


Figure 1. Isolation of Wzy:Wzz complexes. (a) SDS–PAGE of purifications of different combinations of affinity-tagged Wzy_{EPA}/Wzz_{EPA} complexes. The left panel shows GST-Wzz_{EPA} in complex with Wzy_{EPA}-His₁₀ and was purified by first applying the protein to glutathione 4B resin and then applying the eluate to cobalt resin. Note that the GST fusion substantially changes the size of the corresponding Wzz fusion. The predicted molecular weights of the fusion proteins are 66 kDa for GST-Wzz_{EPA}, 55.5 kDa for Wzy_{EPA}-His₁₀, 57 kDa for Wzy_{EPA}-distrepII and 42 kDa for His₆-Wzz_{EPA}. M, marker; FT, flow-through; W, wash and E, eluate. E2 denotes the eluate from the second affinity column. The right panel shows the SDS–PAGE gel of the purification of the His₆-Wzz_{EPA}/Wzy_{EPA}-distrepII complex by streptactin resin followed by cobalt resin. (b) Western blot images of samples taken during tandem affinity purification of His₆-Wzz_{EPA} and Wzy_{EPA}-distrepII by streptactin pulldown followed by TALON affinity chromatography. The western blots on the left are probing the samples taken after streptactin pulldown with the anti-His antibody recognizing His₆-Wzz_{EPA} and the anti-strepII antibody recognizing Wzy_{EPA}-distrepII. The western blots on the right show the samples taken during the cobalt affinity column, probed with the same antibodies as before. The predicted molecular weights of the fusion proteins are 57 kDa for Wzy_{EPA}-distrepII and 42 kDa for His₆-Wzz_{EPA}. E1–6 denote eluate fractions 1–6, collected during elution. (c) The SEC profiles of individually expressed and purified Wzy_{EPA}-distrepII and His₆-Wzz_{EPA} and the co-expressed and co-purified Wzy_{EPA}-distrepII: His₆-Wzz_{EPA} complex.

These findings support a general molecular model for how the polymerization of Wzy is regulated by Wzz.

2. Results

2.1. Isolation and validation of the Wzy_{EPA}/Wzz_{EPA} complex

The ECA biosynthesis proteins from *P. atrosepticum* were selected for a model system based on expression screening

of multiple homologues. Affinity-tagged variants of Wzz_{EPA} and Wzy_{EPA} were co-expressed in an *E. coli* host, with different combinations of tags used. The predicted complexes were then purified by sequential application to two different affinity columns, with each column specific for one tag. As a result, proteins in the final elution step (E2) must result from a complex possessing both Wzz_{EPA} and Wzy_{EPA} fusion proteins (figure 1*a,b*). To rule out any non-specific effects of the fusion partners, two different affinity tag combinations were used (GST-Wzz_{EPA} and Wzy_{EPA}-His₁₀; His₆-Wzz_{EPA} and Wzy_{EPA}-distrepII). For each tag

combination, WzzE_{PA} and WzyE_{PA} co-eluted from the second affinity column (electronic supplementary material, figures S1 and S2). The same outcome was observed with either affinity tag combination, and when the order of the affinity columns was reversed (electronic supplementary material, figures S1 and S2). Control experiments of applying the individual purified proteins to the non-cognate affinity resin showed no significant non-specific binding to the affinity resin (electronic supplementary material, figure S3), demonstrating that non-specific binding does not account for the co-purification.

Analysis of the affinity-purified WzyE_{PA}-distrepII:His₆-WzzE_{PA} complex by blue-native PAGE showed a smeared band and western immunoblots indicated this material contained both proteins (electronic supplementary material, figure S4A). When examined alone by BN-PAGE and western immunoblot, WzyE_{PA}-distrepII exhibits a broadly smeared band, with a considerably larger size distribution compared to the sample of the complex. To provide more insight, size exclusion chromatography was performed. This indicated a shift in size of the complex, relative to the individual component proteins (figure 1c). When examined alone, WzyE_{PA} and WzzE_{PA} both eluted as broad peaks, whereas the complex eluted as a sharper peak, suggesting conversion to a more homogeneous sample. The elution fraction of the complex was larger than WzyE_{PA} but surprisingly smaller than WzzE_{PA}. The latter finding may reflect the complex being more compact than the WzzE_{PA} homo-oligomer. However, an alternative (and perhaps more likely) interpretation is that the average oligomeric state of WzzE_{PA} is constrained upon interaction with WzyE_{PA}. This is consistent with the observation of various oligomeric states of isolated Wzz homocomplexes [22–24]. Quantitative immuno-dot blotting was used to estimate the molar ratio of His₆-WzzE_{PA}:WzyE_{PA}-distrepII (electronic supplementary material, figure S4B–D). Known quantities of each protein were spotted in a two-fold serial dilution and the chemiluminescence used to generate a standard curve for each. A serial dilution of the co-purified complex was analysed, and the standard curve used to determine the quantity of each component. This method yielded a stoichiometry of His₆-WzzE_{PA}:WzyE_{PA}-distrepII to be 9.6 (±1.8):1.

2.2. *In vivo* studies of WzyE/WzzE interaction

To support the isolation of the heterocomplex, the interaction of WzyE_{PA} and WzzE_{PA} was studied *in vivo* in *E. coli* cells. WzyE_{PA} and WzzE_{PA} were fused with fluorophores mClover3 and mCherry, respectively (figure 2a), to investigate interactions by Förster resonance energy transfer (FRET). To assess potential non-specific interactions due to the partners being membrane proteins, the PglL oligosaccharyltransferase from *Neisseria meningitidis* [41,42], was expressed as mClover3 and mCherry fusions for negative controls (figure 2a). PglL is a membrane protein with multi-spanning membrane topology and, like WzyE_{PA}, belongs to the SEDS superfamily but is otherwise unrelated [43]. Each construct was expressed, and western immunoblotting was used to confirm the presence of the desired fusion proteins (figure 2b). Fluorescence measurements were conducted on suspensions of *E. coli* cells expressing the different fluorophore-labelled constructs and their combinations. The proportion of baseline fluorescence (i.e. not due to FRET) at 620 nm was measured for

each construct by exciting the samples at the mClover3 excitation wavelength (488 nm). The authentic FRET signals were measured for the following pairs of proteins: WzyE_{PA} + WzzE_{PA}, WzyE_{PA} + PglL and PglL + WzzE_{PA} (figure 2c). A FRET signal was only observed for the WzyE_{PA}:WzzE_{PA} pair indicative that *in vivo* the two proteins form a specific complex *in situ* in the cell, consistent with the data described above for the isolated proteins (figure 2c).

2.3. High-resolution cryo-EM structure of WzzE_{PA}

A sample of the TEV protease cleaved (removal of affinity tag) and SEC-purified WzyE_{PA}:WzzE_{PA} complex was applied to cryo-EM grids, plunge frozen and reference-free 2D classes were obtained (electronic supplementary material, figure S5A). The classes reveal a detergent belt and the characteristic bell-shaped periplasmic domain of WzzE_{PA} oligomers that has been described previously for oligomers from other species [21–23]. However, no additional density attributable to WzyE_{PA} was evident on the outside or inside of the WzzE_{PA} transmembrane helices. This observation was interpreted as reflecting some combination of complex dissociation during cryo-EM grid preparation, sample heterogeneity and symmetry mismatch. The reconstructed 3D map shows only WzzE_{PA} and that WzzE_{PA} forms an octamer made up of four dimers arranged in C4 symmetry (figure 3a,b; table 1). By contrast to the recently published full-length structure of *E. coli* K-12 WzzB_{EC} (PDB 6RBG; C8 symmetry), the two transmembrane helices of each monomer (residues 39–54 and 326–337) in WzzE_{PA} were traced and surround a central membrane cavity with a diameter of 57 Å (measured between Ile335 residues, figure 3a,b). No contacts were formed between the transmembrane helices, although the termini of the transmembrane helices were disordered. The periplasmic domain is dominated by α -helices (figure 3c). A prominent conserved feature of the Wzz monomer is a long (approx. 100 Å) helix, which is approximately parallel to the long axis of the molecule [19–23,40]. This helix is connected by a helical turn-containing loop (residues 241–265) to two shorter α -helices, which are arranged approximately anti-parallel to the long helix. In the octamer, there are slight structural differences between the monomers localized at the loop that connects the long helix to the two shorter α -helices and this resulted in C4 symmetry. These helices connect to a region of structure identified as motif 1 [25], which is formed by residues 58–204 and 310–317. In WzzE_{PA}, motif 1 consists of a β -sheet with four anti-parallel β -strands and four α -helices (figure 3d,e). Two regions within motif 1, a loop (residues 153–163) (L-loop) and a loop–helix–loop (78–100), point into the large central periplasmic lumen (figure 3e). The L-loop extends further into the lumen and is closer to the membrane than the other region. The interactions between motif 1 of one monomer with its neighbour include salt bridges, hydrogen bonds and van der Waal's interactions, and are dominated by the large helix and the two helices to which it is connected. In the octamer, the periplasmic helices form a barrel that encloses the lumen (figure 3a,b). The transmembrane helices, which are not closely packed, enclose the disc-shaped cavity within the membrane resembling that seen for other Wzz structures [22,23] (figure 3a,b). However, comparison with other Wzz structures reveals some structural differences, most notably

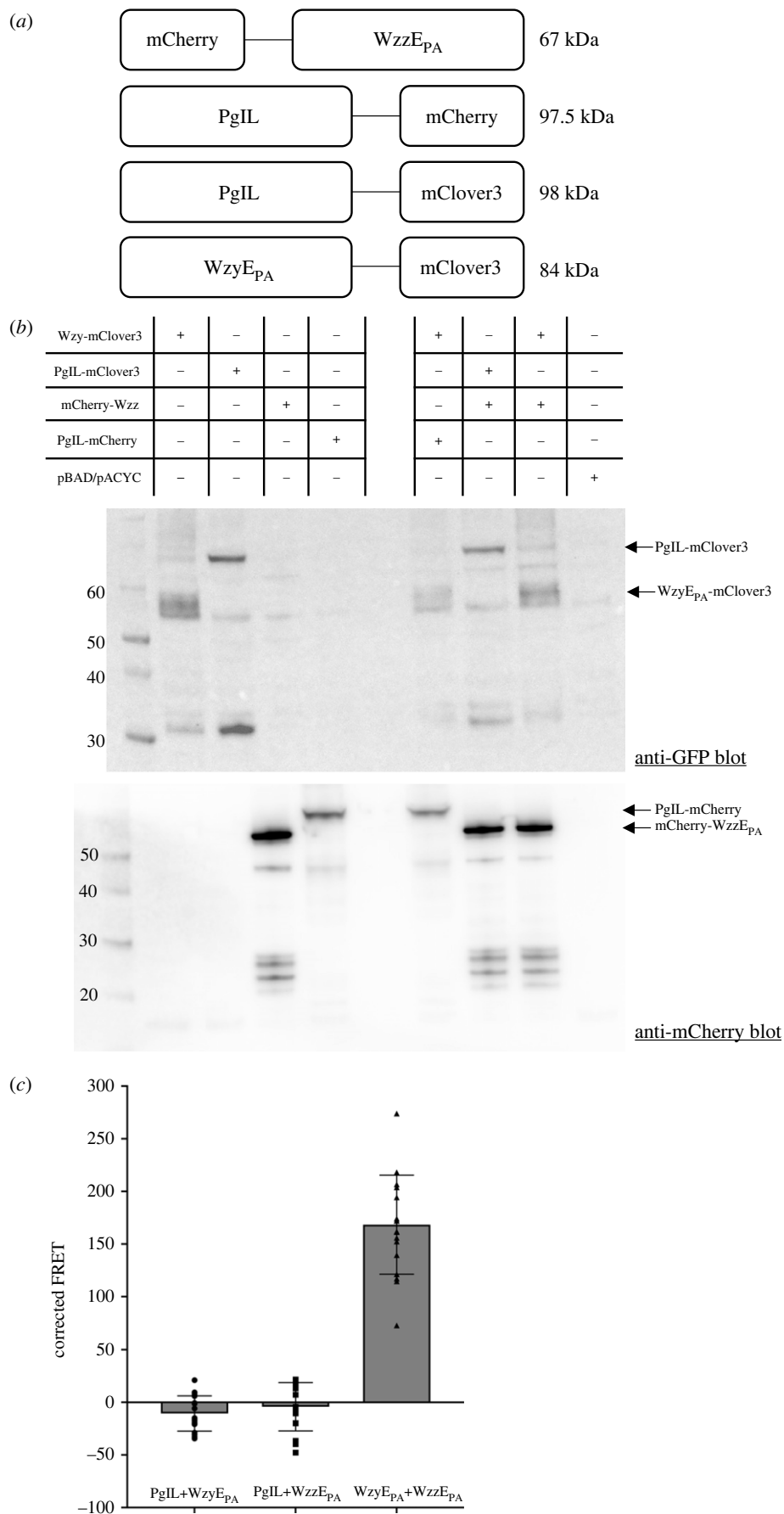


Figure 2. Identification of the *in vivo* WzyE_{PA} : WzzE_{PA} complex by FRET. (a) Overview of the domain architecture of the different fluorophore-labelled proteins. (b) Various combinations of fluorophore-labelled WzyE_{PA}, WzzE_{PA} and control proteins in *E. coli* lysates probed by western immunoblot analysis. (c). FRET was measured for the combinations of PgIL, WzyE_{PA} and WzzE_{PA} and corrected for background fluorescence and fluorophore bleed-through. $n = 18$; the data represent triplicates and error bars are standard deviation.

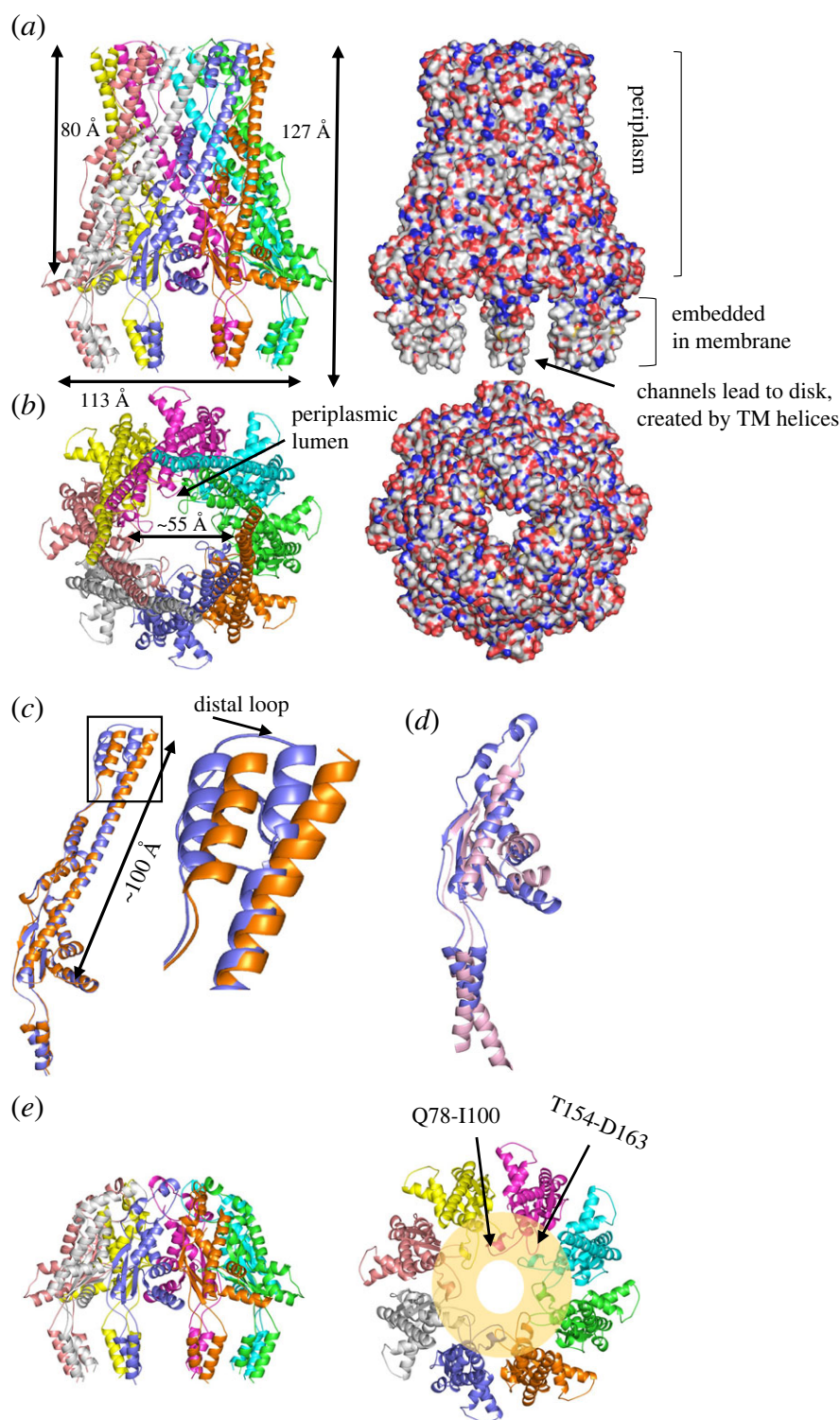


Figure 3. Structure of $WzzE_{PA}$. (a) $WzzE_{PA}$ (residue I39–V337) viewed from the side, the protein is represented in cartoon form (left) and in space fill (right). The periplasmic regions enclose a large central lumen. The transmembrane helices create a cavity in the membrane with channels. Residues 1–39 and 338–348 were not visualized. (b) $WzzE_{PA}$ viewed from the periplasm looking into the cytoplasm through the protein. The protein is represented in cartoon form (left) and in space fill (right). Space fill shows that the entrance to the lumen from the periplasm is almost occluded. (c) The monomers (coloured blue and orange) of $WzzE_{PA}$ are found in two conformations, distinguished by their structures at the periplasmic tip (A222–P280), highlighted in box. The monomers show an RMSD of 1.3 Å over 276 C α atoms. (d) Motif 1 (W58–A204 and R310–E317), the two transmembrane helices (TM1 38–53; TM2 326–337) and the connecting loops are structurally conserved in $WzzE_{PA}$ (blue) and in WzE_{EC} (pink). (e) The side and top view of motif 1 (defined as (d)). Motif 1 forms a ring at the periplasmic face of the membrane. Two regions of structure the L-loop (T153–D163) and a loop–helix–loop (Q78–I100) that reach into lumen are highlighted. The L-loop (T153–D163) lies closest to the membrane surface.

within motif 1. In the $WzzE_{PA}$ structures, the helix $\alpha 2$ (residues 103–113) connects to $\alpha 3$ (residues 115–124) by a kinked turn of approximately 90°. $\alpha 3$ connects via a loop on the outside of the barrel to $\alpha 4$ (residues 136–149), which connects to the strand $\beta 2$. This reflects a divergence from

the $WzzB$ structures (e.g. $WzzB_{EC}$, RCSB 6R8G [23]), where the $\alpha 2$ helix has no kink and is longer (residue 88–110). The kink in the helix in $WzzE_{PA}$ eliminates regions that participate in inter-subunit contacts in $WzzB_{EC}$. The helix is connected by a seven-residue loop to the strand $\beta 2$, where

Table 1. Data collection and processing statistics of WzzE_{PA} cryo-EM structure.

<i>data collection and processing</i>	
frames	20
total dose (e ⁻ /Å ²)	85
pixel size (Å)	1.06
defocus (µm)	1.0–2.7
symmetry	C4
movies	2,347
particles	464,832
final number of particles	190,490
refined map resolution FSC = 0.143 (Å)	3.6
map sharpening B-factor (Å ²)	–151
<i>refinement</i>	
model resolution (Å)	2.1/3.7
FSC threshold	0.143/0.5
model resolution range (Å)	1.5–3.7
<i>model composition</i>	
non-hydrogen atoms	18,384
protein residues	2301
<i>B factors (Å²)</i>	
protein	50
<i>r.m.s. deviations</i>	
bond lengths (Å)	0.004
bond angles (°)	0.790
<i>validation</i>	
MolProbity score	1.8
clashscore	11.70
poor rotamers (%)	0.88
<i>Ramachandran plot</i>	
favoured (%)	96.3
allowed (%)	3.3
disallowed (%)	0.4

the structural elements of WzzE_{PA} and WzzB_{EC} once again superimpose (electronic supplementary material, figure S6). This region of the structure is on the outside of the octamer and some portions make contacts with the neighbouring monomers. Collectively, these changes in structure mean there is very little conservation in the motif 1 contact regions in WzzE_{PA} and WzzB_{EC}. Other smaller structural differences occur in the two regions of structures that protrude into the lumen (electronic supplementary material, figure S6).

2.4. Detection of WzyE_{PA} in WzyE_{PA}:WzzE_{PA} complexes using an antibody

To verify the presence and location of WzyE_{PA} in the complex, an antibody-labelling approach was adopted for cryo-EM sample preparation. A monoclonal antibody to distrep tag II (mAB) was added to a sample of affinity-purified His₆-WzzE_{PA}:WzyE_{PA}-GST-distrepII complex, which possesses

a WzyE_{PA} fusion protein containing a C-terminal GST and di-strep tag II. Addition of the antibody lead to a shift and broadening of the complex elution peak in SEC (figure 4a), whereas incubating the antibody with His₆-WzzE_{PA} alone did not (electronic supplementary material, figure S7A,B). Samples of the elution peak taken for SDS-PAGE and western blot confirmed the expected presence of all three proteins; WzyE_{PA}, WzzE_{PA} and mAB (figure 4b; electronic supplementary material, figure S5C–F). Images of negatively stained particles show two flask-shaped structures, characteristic of the WzzE_{PA} oligomeric periplasmic domains, are connected by extra density (figure 4c). Single-particle cryo-EM data collected for the WzyE_{PA}:WzzE_{PA}:mAB complexes showed dimerized WzzE_{PA} octamers (figure 4d,e). A model at approximately 20 Å resolution (electronic supplementary material, figure S7C) was created using these 2D classes, and shows the complex adopts a ‘dumbbell’ arrangement containing the two WzzE_{PA} octamers positioned end to end. The antigen-binding portion of the antibody appears in the centre of the membrane region of the WzzE_{PA} oligomers (figure 4e). The location of the antibody indicated that WzyE_{PA} is positioned in the centre of the transmembrane cavity formed by the transmembrane helices of octameric WzzE_{PA}. Each half of the dumbbell can accommodate one WzzE_{PA} octamer and the localization of the antibody indicates one WzyE_{PA} monomer is present in the lumen of WzzE_{PA}. This result is entirely consistent with the stoichiometry determined by dot blot above.

3. Discussion

The polymerization of polysaccharides from lipid-linked repeat units is central to the completion of key bacterial polymers, including ECA, OPS, CPS and peptidoglycan. With the exception of peptidoglycan, these systems exploit a conceptually conserved Wzx/Wzy-dependent pathway, where polymerization is catalysed by Wzy, an integral membrane protein. This biosynthesis strategy is arguably the most abundant format observed in known polysaccharide assembly systems in Gram-positive and Gram-negative bacteria. In addition to Wzy, a member of the PCP family is required to activate and regulate polymerization, resulting in the production of high-molecular weight polysaccharides that sometimes exhibit a modal length distribution, where the chain lengths of most products fall within a relatively narrow range [26,44]. In Gram-negative bacteria, most Wzy-associated PCPs belong to either the PCP-1 (Wzz) or the PCP-2a (Wzc) family [45]. Evidence from genetic data and co-purification methods has accumulated from OPS biosynthesis systems to support the possibility that PCPs form complexes with Wzy and here we extend those observations with ECA biosynthesis. Using FRET labelling and co-purification, we have confirmed that WzzE_{PA} and WzyE_{PA} form a complex *in vivo* (figure 2c). Furthermore, using multiple different tandem affinity tagging approaches, we isolated a WzzE_{PA}:WzyE_{PA} complex with a stoichiometry consistent with one WzyE_{PA} protein per WzzE_{PA} octamer (figure 1d). Although the proteins are stable and can be purified, we have not confirmed that they are functional in the native organism. The observed stoichiometry is consistent with the lower level of native expression of Wzy in comparison to that of Wzz seen in other bacteria [35,46]. We note that the

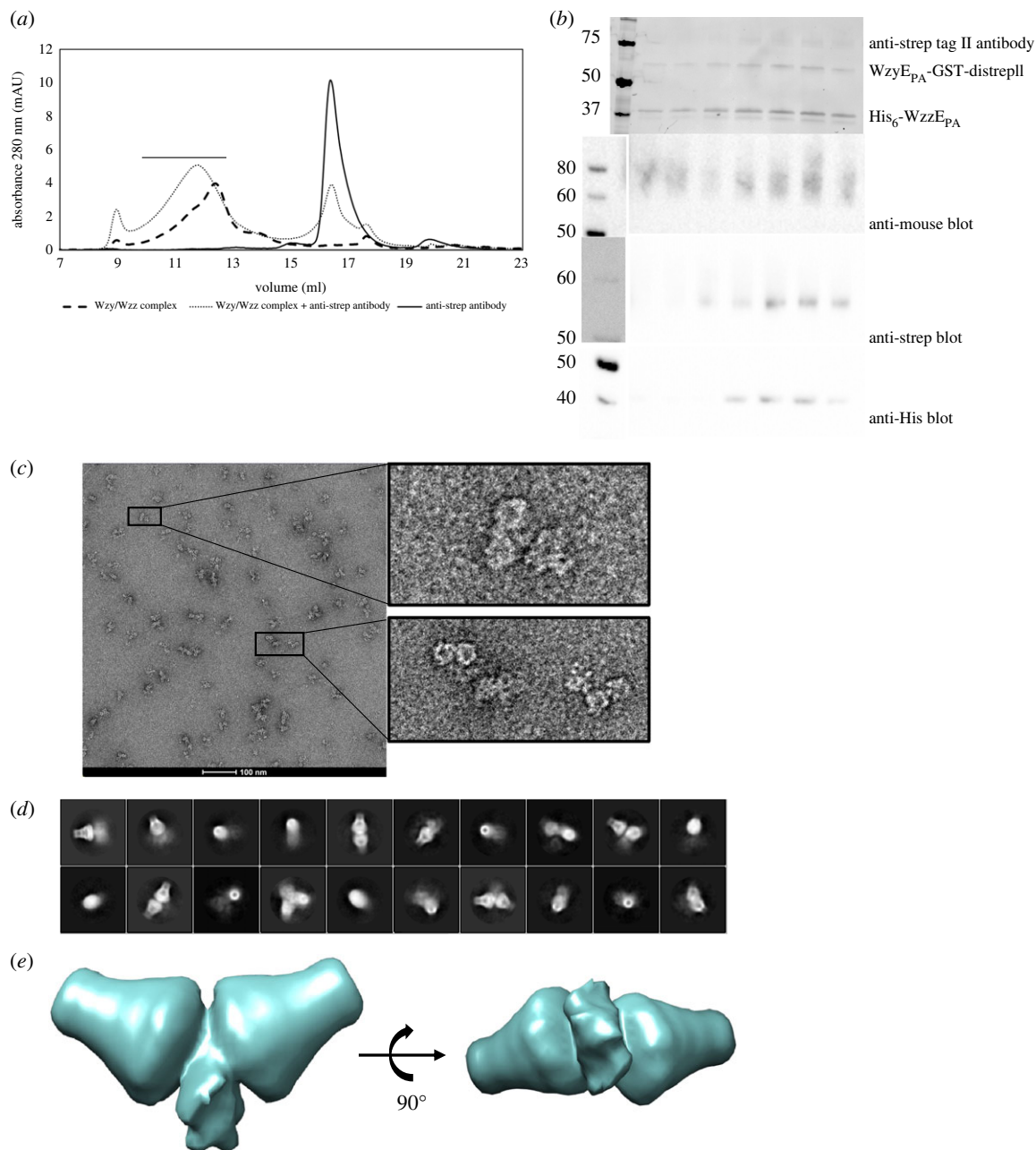


Figure 4. *Wzy*_{EPA} is located within *Wzz*_{EPA}. (a) Addition of the anti-strep tag II antibody leads to a shift in elution volume of the complex in SEC, as well as to a broadening of the elution peak. Control samples are shown. (b) The fractions under the broad peak (indicated by a line above the peak) were analysed by SDS-PAGE and western immunoblot. A polyclonal anti-mouse secondary antibody recognizes the mouse-derived anti-strep tag II antibody bound to *Wzy*_{EPA}-GST-distrepll; the HRP-coupled anti-strep tag II antibody was used to detect *Wzy*_{EPA}-GST-distrepll and the anti-His antibody was used to probe His₆-*Wzz*_{EPA}. For the western immunoblots, the lanes containing the molecular weight markers have been spliced from the same blot showing the protein lanes in order to align the lanes on the immunoblot containing sample from the same fractions. For the anti-strep immunoblot, a photograph had to be taken of the coloured marker as the antibody did not recognize it. The full-size images of the immunoblots can be found in the electronic supplementary material. (c) The concentrated peak fractions were applied to cryo-EM grids and negatively stained with uranyl formate. The inset shows dimeric *Wzz*_{EPA} particles, which were attributed to the antibody cross-linking of *Wzy*_{EPA}. (d) The reference-free 2D class averages from cryo-EM of the concentrated peak fractions. (e) The initial model was generated from 2D class averages obtained from cryo-EM experiments (d) showing two *Wzz*_{EPA} molecules adjacent to each other.

ECA operon is intact in the *E. coli* C43(DE3) strain used here and we cannot exclude the possibility that *Wzz*_{EC} monomers are incorporated into (predominantly) *Wzz*_{EPA} octamers to influence stoichiometry determination in the heterocomplex. However, we saw no evidence for significant sequence heterogeneity in the *Wzz*_{EPA} structure that would materially alter our estimate. Consistent with the results reported here, a complex between *Wzy*_{SF} (*S. flexneri*) and *Wzz*_{PHS-2} (or *Wzz*_{SF}) has also been observed using a comparable co-purification approach, but the stoichiometry was not established [28].

The 3.6 Å cryo-EM structure of full length *Wzz*_{EPA} that is found as an octamer with a large enclosed central periplasmic lumen connected to a membrane cavity defined by the transmembrane helices (figure 3*a,b*, electronic supplementary material, figure S5B). The oligomeric arrangement (figure 3*a*) has been seen in EM structures of *Wzz* from other species [19,21,23,24] but is different to hexamers/decamers seen in some structures of periplasmic domains [22]. Comparing the *Wzz*_{EPA} and *Wzc* structures shows that only the transmembrane helices and periplasmic motif 1 (figure 3*d*) are structurally conserved between PCP-1 and PCP-2a proteins

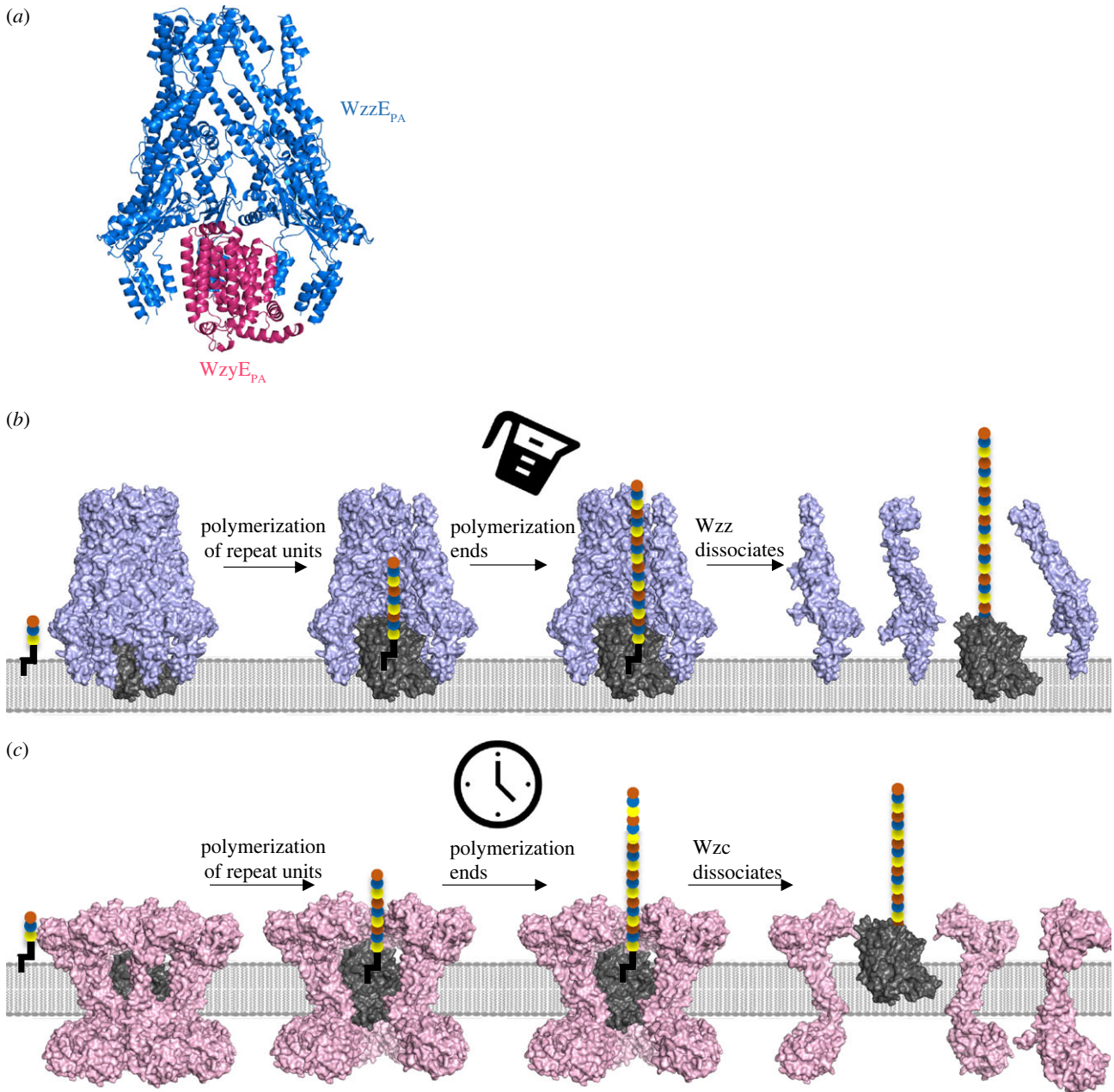


Figure 5. Model for the process of chain length regulation in the Wzx/Wzy pathway. (a) AlphaFold model of WzyE_{PA} docked into EM structure of WzzE_{PA}. The model of WzyE_{PA} was obtained from AlphaFold (ID: Q6CZF1) and placed by hand into the transmembrane cavity of WzzE. (b) The molecular measuring cylinder model of chain regulation of Wzy by Wzz (PCP-1) in ECA and OPS biosynthesis. (c) The molecular timer mechanism of chain regulation of Wzy by Wzc (PCP-2a) in CPS biosynthesis.

(figure 3d) [25]. In octamers of both Wzc and WzzE_{PA}, the transmembrane helices are arranged in a similar circular array that delimit the cylindrical cavity (approx. 57 Å in diameter), which is embedded within the membrane (WzzE_{PA} shown in figure 3a,b). The helices are not closely packed and thus create connecting channels between the bulk membrane and the central cavity (figure 3b). Motif 1 forms a ring-like structure in the periplasm with two regions extending into the space above the membrane cavity (figure 3e). There are several interactions between neighbouring monomers in the octamer that involve residues within motif 1 (figure 3e).

Using an antibody to a C-terminally tagged WzyE_{PA}-distrepII, we visualized the WzzE_{PA}:WzyE_{PA}:mAB complex in both negative stained and cryo-EM (figure 4c–e). Both methods show that the Fab fragment is centred on the cytoplasmic face of the characteristic Wzz cavity (figure 4c,e)

consistent with the 8:1 WzzE_{PA}:WzyE_{PA} *in vitro* stoichiometry (figure 1d). The diameter of the membrane embedded portion of the AlphaFold model [47,48] of WzyE_{PA}, obtained from the AlphaFold Protein Structure Database, is around 50 Å (figure 5a), thus the transmembrane portion of WzyE_{PA} can potentially be placed in the WzzE_{PA} (and Wzc) membrane cavity without van der Waal clashes. The surfaces of the transmembrane helices of WzzE_{PA}, facing WzyE_{PA}, and the membrane-embedded surface of WzyE_{PA} are mainly hydrophobic, consistent with the arrangement deduced from EM (figure 5a). AlphaFold-Multimer [49] predictions of a heterodimer consisting of a WzzE chain and WzyE places WzyE in the membrane cavity, in agreement with the EM result.

With WzyE_{PA} placed in the middle of the WzzE_{PA} octamer, a working model for the polymerization process emerges. The gaps between the transmembrane helices of

WzzE_{PA} would potentially allow diffusion of the lipid-linked oligosaccharide substrates to access WzyE_{PA}, and allow the spent lipid diphosphate carriers to diffuse out (figure 3a). In such an arrangement, the periplasmic face of Wzy would interact with the structurally conserved motif 1 (figure 3e) that extends into the lumen. Since OPS chain length increases in the presence of Wzz *in vivo* and during *in vitro* synthesis [26], it would suggest that this complex promotes processivity of Wzy and shields the growing polymer from the acceptor (or the action of subsequent enzymes like OPS ligases [3]) until the modal length is achieved. Insertions in motif 1 are known to disrupt the modal chain length regulation of O-antigen [37] consistent with our structural inference that the arrangement of Wzz around Wzy is crucial for function. Wzz mutants with reduced modal chain length are also less able to form stable oligomers [36,37], while Wzz mutants showing increased chain length are all found as oligomers [37]. Further support for this model comes from chimeric proteins composed of fused portions of *S. flexneri* WzzB_{SF} and WzzE_{SF}. In the bacterium, these Wzz homologues regulate different Wzy proteins to produce different polymers with different chain length distributions. Exchanging the TM1 of WzzB_{SF} and WzzE_{SF} had little effect. However, creation of a WzzE_{SF} chimera containing TM2 from WzzB_{SF} disrupted ECA modal length regulation, while the reciprocal chimera exhibited a reduced modal length control of the O-antigen [29]. TM1 faces outwards from and TM2 faces inward to the central membrane cavity, and this arrangement aligns with our structural model of WzzE_{PA}. Additional studies with *S. flexneri* WzyB_{SF} have identified a region of the protein that is important for its regulation by Wzz_{pHS-2} [28]. This region of WzyB_{SF} was not required for its regulation by Wzz_{SF} [28]. Unfortunately, the primary sequences of WzyE_{PA} and Wzy_{SF} are not well conserved and the AlphaFold models (Wzy_{SF} published in [28]) are sufficiently different that simple comparisons provide limited insight.

The data presented here invalidates one former proposal for Wzz mode of action, in which Wzz serves as a scaffold for multiple Wzy molecules to assemble and where the outer surface of Wzz acts as a ruler [19,22]. The data suggest the nascent polymer forms inside the lumen, consistent with the proposal that the interior surface of the periplasmic domain of WzzE (roughly 80 Å high (figure 3a)) may provide some form of a molecular ruler [36]. The interior periplasmic lumen being the key to controlling polymer length was originally suggested in studies with FepE from *E. coli* and WzzB from *S. flexneri*, where mutations within the lumen affected chain length regulation [50,51]. The current data supports a model where the polymer grows until it fills the periplasmic lumen in WzzE. Filling the lumen with polymer may then destabilize the Wzz octamer, which has been shown to exist in different oligomeric states [22–24]. We propose that the putative disruption of the Wzz oligomer serves two purposes. First, since oligomeric Wzz has been shown to be essential for extended polymerization [26,37], disruption of this oligomeric complex would substantially decrease processive polymerization. Second, partial disassembly of Wzz would expose the lipid-linked polymer to the enzymes necessary for transfer to its final carrier or translocation machinery. This explains why mutants lacking PCP function have an ‘unregulated’ phenotype, where product abundance decreases as chain length increases. Once the polymer is

transferred into the downstream steps in the pathway, Wzz would be able to re-assemble around Wzy and restart the polymerization cycle (figure 5b). This model implicitly introduces a third factor in chain length regulation. In addition to the size of the polymer and volume of the Wzz lumen, the stability of the Wzz oligomer is important. This third factor rationalizes the variation in modal length reported to arise from changes in motif 1 [37]; these would perturb oligomer stability. Although changes to individual residues in motif 1 (e.g. K321T in *Pseudomonas aeruginosa* Wzz2 [36]) have resulted in shifted modal length, mechanistic interpretation has not been possible. A comprehensive study in which a pentapeptide was inserted into different regions of Wzz_{SF} showed that changes in most regions resulted in proteins conferring decreased modal length, and some completely eliminated regulation [37]. Like the site-directed mutants, many of the phenotypes of insertions can be rationalized by invoking a decrease in oligomer stability. Interestingly, the L-loop of Wzz (electronic supplementary material, figure S6) is predicted to engage Wzy and insertions in this loop led to an increase in modal length [37].

Conceptually the model has parallels with the CPS system involving Wzc (PCP-2a), where polymerization of CPS requires both assembly and disassembly of the Wzc octamer; in the case of Wzc, the disassembly process is driven by phosphorylation [25]. AlphaFold models of the relevant Wzy show that the majority of the protein would be embedded within the membrane. This would suggest that the two TM helices and motif 1 of Wzz (and by analogy Wzc), which are located in the membrane and immediately above the location the Wzy polymerase would occupy, are central to the interaction between polymerase and co-polymerase. Consistent with this concept, we note that is only this region that is structurally conserved in Wzz and Wzc. We propose these regions engage and modulate Wzy. Modulation of Wzy would be achieved by filling the lumen in PCP-1 systems (akin to a molecular measuring cylinder) (figure 5b), while PCP-2a systems depend on the kinetics of the competing phosphorylation/dephosphorylation to exercise control (analogous to a molecular timer) (figure 5c).

Mutant complementation experiments with *wzx*, *wzy* and *wzz* knockout strains have led to the proposal that Wzx (which flips the Wzy substrate across the membrane), Wzy and Wzz form a multi-protein complex [52,53]. A super-complex of this type is enticing because it forms a central coordinating point to regulate OPS biosynthesis by interacting with other proteins and can be extended to the biosynthesis of other types of polysaccharides assembled by the same strategy. A wider network of regulation would be consistent with the observation that deletions in the proposed periplasmic loop between TMs 4 and 5 in WbaP (the phosphoglycosyltransferase initiating enzyme for OPS synthesis in *Salmonella*) affect OPS chain length while retaining catalytic activity [53]. It has been hypothesized that this deletion affects the interaction with the chain length regulator. Experimental evidence for this complex is currently lacking.

In conclusion, we have reported data consistent with new mechanistic model for control of Wzy-mediated polymerization by the PCP-1 proteins. A fuller molecular understanding of this and other PCP-regulated systems will require an atomic level resolution of the interactions between Wzy, its regulators, and other participants in a super-complex. This is the long-term goal of our research.

4. Methods

4.1. Cloning

The genes encoding WzzE_{PA} (Uniprot ID: Q6CZE3; gene locus ECA4209) and WzyE_{PA} (UniProt ID: Q6CZF1; gene locus ECA4201) were PCR amplified from genomic DNA of *P. atrosepticum* SCRI 1043 (ATCC BAA-672). The DNA fragment encoding WzzE_{PA} was cloned into pACYCGSTDuet (Novagen) and pEHisTEV [54] adding an N-terminal GST-tag or a TEV protease-cleavable His₆-tag, respectively, to the gene product. The DNA fragment encoding WzyE_{PA} was cloned into a proprietary vector constructed by Huaning Liu, or into pBAD10HisTEV, adding a C-terminal TEV protease-cleavable distrepII tag or His₁₀-tag, respectively. All constructs were cloned using NcoI and XhoI restriction sites introduced by PCR primers. For the WzyE_{PA}-GST-distrepII construct, *E. coli* GST was inserted between Wzy and the C-terminal distrep tag II. The GST-encoding fragment was amplified from plasmid pGEX-KT by PCR using KOD DNA polymerase. pGEX-KT was obtained from the American Type Culture Collection (ATCC77331). The vector containing the wzyE_{PA}-distrepII fragment was linearized with XhoI and the GST-encoding fragment was inserted at this site by Gibson Assembly (New England Biolabs). For the Wzy-distrepII-mClover3 construct, the mClover3 gene was cloned by PCR amplification from a plasmid purchased from Addgene (ID: 105802); EcoRI and HindIII restriction sites were in the primer sequences, and the amplification product was cloned into the pBAD10HisTEV vector containing Wzy-distrepII at the 3' end of the fragment encoding the distrepII tag. The vector was modified by site-directed mutagenesis to incorporate the EcoRI and HindIII restriction sites [55]. The same vector was used to express PglL-distrepII-mClover3. The mCherry fragment was PCR amplified, introducing BamHI and EcoRI restriction sites, and the amplification product was cloned into the pACYCDuet-1 vector. Subsequently, wzzE_{PA} was cloned into the same vector at the 3' end of mCherry by PCR amplification and restriction digest using EcoRI and XhoI. PglL was cloned into the mCherry containing pACYCDuet-1 vector by PCR amplification and restriction digestion using NcoI and BamHI, introducing pglL at the 5' end of mCherry. The primers used are listed in table 2.

4.2. Expression and purification

Chemically competent *E. coli* C43(DE3) F⁻ ompT gal dcm hsdSB(r_B⁻ m_B⁻) cells [56] were transformed sequentially by the two plasmids [57]. The culture volume for individual protein expression was 6 l, whereas the complexes of different affinity-tagged WzyE_{PA} and WzzE_{PA} were cultured in 12 l Luria Broth. WzyE_{PA}-His₁₀ was co-expressed with GST-WzzE_{PA} at 25°C for 4 h following induction of expression with 0.1 mM IPTG and 0.2% (w/v) arabinose at OD₆₀₀ 1.0. WzyE_{PA}-distrepII was co-expressed with His₆-WzzE_{PA} at 15°C overnight, following induction with 0.5 mM IPTG at OD₆₀₀ 0.4. Cells were harvested by centrifugation at 5000 g, 4°C and cell membranes were prepared by lysing the cells in PBS, pH 7.2 using a cell disrupter, removing cell debris at 15 000 × g, and ultracentrifugation at 150 000 × g. Membranes were solubilized in PBS containing 1% (w/v) DDM

Table 2. List of primers used in this study.

Wzz_fw	AAATTCATGGTAAATCAGAGAAGCTTG
Wzz_rv	AAATTTCTCGAGACGACGACGGCGAGCCAG
Wzy_fw	AAATTTATGGCGCTTGGGCAATTTGG
Wzy_rv	AAATTTCTCGAGTTATCTGCTCTTTATC
gstF1	GGTGATAAAGGAGCAGATACTCGAGGGCGGTGGCGTT CGTCCCTATACTAGGTTATTGGAAAATAAGG
gstR1	CCTGAAAATACAGGTTTCTCGAGCGATTTGGAGGATGGTCGCC
SDM_fw	TTCCCGTCGACAAGCTTGATCCGGCTGTAACAAAGC
SDM_rv	AAGCTTGTGACGGGAATTCTTTTTCGAATTGCGGATGACTCC
mClover3_fw	CCGAA TTCGTGAGCAAGGG
mClover3_rv	CGCAAGCTT CTACTGTACAGCTCGTC

for 2 h at 4°C. The supernatant containing solubilized proteins was obtained by centrifugation at 150 000 × g.

All affinity chromatography steps were done as batch purifications at 4°C. For the purification with the GST-tag, proteins were bound to glutathione sepharose 4B resin (Cytiva), washed in PBS containing DDM, and eluted in 50 mM Tris pH 8.0, 200 mM NaCl, 10 mM glutathione, 0.026% (w/v) DDM. The eluate was desalted into PBS, pH 8, 20 mM imidazole, 0.026% (w/v) DDM and bound onto HIS-Select cobalt resin (Sigma Aldrich). After washing with PBS, pH 8, 30 mM imidazole, 0.026% (w/v) DDM, the protein was eluted in PBS, pH 8, 400 mM imidazole, 0.026% (w/v) DDM. For the reverse purification, protein was bound to nickel resin (GE Healthcare), washed with PBS containing 20 and then 30 mM imidazole before being eluted using 400 mM imidazole. The eluate was dialyzed into PBS, pH 8.0, 0.026% (w/v) DDM and bound to glutathione sepharose 4B. The resin was washed with PBS, pH 8.0, 0.026% (w/v) DDM and bound protein was eluted using 50 mM Tris pH 8.0, 200 mM NaCl, 10 mM glutathione, 0.026% (w/v) DDM.

For the purification of His₆-WzzE_{PA} and WzyE_{PA}-distrepII, the supernatant was incubated with either HIS-Select cobalt resin (Sigma Aldrich) or Talon resin (Cytiva). The resin was washed with PBS, pH 7.3, 20 mM imidazole, 0.026% (w/v) DDM and bound proteins eluted with 400 mM imidazole. The eluate was dialyzed in 50 mM phosphate buffer, pH 8, 200 mM NaCl, 0.026% (w/v) DDM overnight and then incubated with streptactin or streptactinXT resin (IBA Lifesciences). The resin was washed with dialysis buffer and bound protein eluted in buffer containing 2.5 mM desthiobiotin.

The elution fraction was concentrated in a spin concentrator and either directly injected onto a superose 6 10/300, or incubated with TEV protease for 3 h at 4°C, prior to size exclusion chromatography.

4.3. SDS–PAGE analysis

Samples for SDS–PAGE analysis of the purification progress were prepared using 2× Laemmli sample buffer (Bio-Rad). Samples for analysis via SDS–PAGE followed by western immunoblotting were loaded onto a 4–12% Bis–Tris NuPAGE gel (Novex), along with Sharp Pre-stained protein standard (Novex) and run at 200V in 1× MES SDS running

buffer (NuPAGE). Samples for SDS–PAGE analysis were loaded onto an any-kD Mini Protean TGX Stain-free protein gel (Bio-Rad), along with Precision plus unstained marker (Bio-Rad) and run at 300 V in 1× Tris/Glycine/SDS buffer (Bio-Rad). Any-kD Mini Protean TGX Stain-free protein gels were visualized using the Chemidoc MP system (Bio-Rad). The loaded sample volume ranged between 15 and 20 μ l.

4.4. Blue-native PAGE and western immunoblot

For blue-native PAGE, samples were prepared by mixing with NativePage sample buffer and NativePage 5% G-250 Sample Additive (both Thermo Fisher Scientific) and separated by electrophoresis on a NativePAGE Novex Bis–Tris 4–16% precast gel (Thermo Fisher Scientific). NativeMark Unstained Protein Standard was used for reference (Thermo Fisher Scientific). For western immunoblotting, the PVDF membrane was air-dried after transfer, rinsed in methanol and water, blocked with 5% milk in PBS–T (PBS with 0.05% (v/v) Tween-20). For detection, the immunoblot was probed with HRP-conjugated anti-His tag (Sigma Aldrich) and anti-strep tag II (IBA Lifesciences) antibodies.

4.5. Dot blot to determine stoichiometry in protein complexes

Samples were applied directly onto a nitrocellulose membrane using the Minifold 96-well dot-blot system (GE Healthcare). To establish a standard curve, three different amounts of co-purified complex were applied, along with a serial dilution of His₆-WzzE_{PA} and WzyE_{PA}-distrepII. After membrane blocking with 5% milk in PBS–T, the membrane was probed with HRP-coupled anti-His tag (Sigma) or anti-strep tag II (IBA Lifesciences) antibodies. The signals were quantified using the BioRad ChemiDoc MP system with background correction. The circles on the dot blot indicate the areas used for densitometry. U1 indicates the area used for background correction. Wzz exhibited a double band on SDS–PAGE, only the top band could be detected by the anti-His antibody during western blotting, whereas the second lower band did not. To account for this untagged Wzz in the sample, the ratio between tagged/untagged protein was determined from an SDS–PAGE gel using the BioRad ChemiDoc MP system.

4.6. Sample preparation for electron microscopy

For the high-resolution data collection for WzzE_{PA}, tandem affinity-purified and TEV-cleaved WzzE_{PA}:WzyE_{PA} was subjected to SEC on a superose 6 column and the peak fractions containing purified complex were pooled and concentrated to 5 mg ml⁻¹. To ensure that the complex was present in ice suspended over holes, the sample was applied to grids that had been glow-discharged in the presence of pentyl amine and then plunge frozen on a Vitrobot mark IV (Thermo Fisher Scientific). Grids were imaged on a Thermo Fisher Scientific Titan Krios equipped with a Gatan BioQuantum energy filter and K2 camera. Data were recorded at a nominal magnification of 130 k×, corresponding to 1.06 Å pixel⁻¹ at the specimen scale. The total electron dose was 85 e⁻/Å², partitioned over 20 movie frames, with defocus ranging between -1 and -2.7 μ m; 2347 micrographs were recorded and

processed to correct for specimen movement in Relion 3.1 [58]. Defocus estimation was performed using CtfFind4 [59]. From these data, a total of 464 832 putative particle images were picked for further processing in Relion 3.1 [58]. 2D and 3D classification defined a final dataset of 190 490 particle images that were refined to calculate a final reconstruction with a resolution estimate of 3.6 Å. To aid the interpretation of density in less well resolved areas of the map, the reconstruction was subject to deep-learning based density modification using DeepEMhancer [60]. Model building was performed using Coot [61] and Phenix [62–64].

For negative staining of His₆-WzzE_{PA}:WzyE_{PA}-GST-distrepII complexes with anti-strep tag II antibody (mAB), the elution from the streptactin resin (IBA Lifesciences) (or in case of His₆-WzzE_{PA}, the elution from the cobalt resin), was concentrated in a spin concentrator and incubated with the anti-strep tag II antibody overnight at 4°C. The mixture was injected onto a superose 6 10/300 (Cytiva) column. A sample of the peak fraction was applied on CU400 grids and stained with 2% uranyl acetate. The grids were imaged on a FEI Tecnai T12 microscope.

For single-particle cryo-EM of WzyE_{PA}-distrepII:His₆-WzzE_{PA} in complex with the IgG-class anti-strep tag II antibody, fractions containing the proteins of interest were pooled and concentrated. Sample was applied to glow-discharged Quantifoil 200 Au mesh R1.2/1.3 grids and plunge frozen using a Vitrobot Mark IV. 1621 micrographs were collected on a FEI Glacios microscope equipped with a Falcon III camera in linear integrating mode. Processing was conducted in Relion 3.1 [57,58] and motion correction was performed using MotionCor2 [64,65]. Ctf estimation was done using Gctf [65] and particles were manually picked to create a template. The template was used for auto-picking in Relion 3.1, and 15 442 particles were picked and used to generate an initial model without symmetry application. Based on the observed symmetry of the obtained model, the same particles were used for initial model generation with C2 symmetry applied followed by a round of 3D refinement.

4.7. *In vivo* Förster resonance energy transfer

The individual vectors as well as plasmids expressing the various combinations of WzyE_{PA}-mClover3, mCherry-WzzE_{PA} and the PglL fusion proteins, PglL-mClover3 and PglL-mCherry were transformed into *E. coli* C43(DE3) cells. The combination of the empty pBAD and pACYCDuet-1 vectors were also transformed. Single colonies of each transformant were selected and grown in LB media in the presence of antibiotics and 1% (w/v) D-glucose overnight at 37°C. The culture was diluted 25-fold into LB with appropriate antibiotics and grown at 37°C to OD 1.2–1.3. Cloned gene expression was induced with 0.002% (w/v) L-arabinose, 0.4 mM IPTG or both and the cells were grown overnight at 15°C. Cells were harvested by centrifugation, washed in PBS, diluted 2-fold in PBS and cell density was measured at 450 nm and used for OD correction to 1. Fluorescence was measured in a Clariostar plate reader (BMG Labtech) at 488–14 nm/535–30 nm for mClover3 and at 570–15 nm/620–20 nm for mCherry. FRET measurements were conducted at 488–14/620–20 nm. Cell density was measured at 450 nm and used for OD correction to 1. To obtain the corrected FRET value, measurements were corrected for background using the empty vector control and for baseline fluorescence by measuring the single vector

control samples [65]. Measurements were conducted in triplicate. Data were plotted in GraphPad Prism 9 as a scatter plot with bar diagram.

To confirm expression of the fusion proteins, samples of the cells from the FRET experiments were taken for western immunoblot analysis using the HRP-conjugated anti-GFP-tag monoclonal antibody and the polyclonal anti-mCherry antibody (Proteintech).

Data accessibility. EM maps and models are deposited in the EMDB and wwPDB under accession codes EMD-16364 and PDB 8C0E. Raw data are deposited in EMPIAR under accession code EMPIAR-11355. All plasmids are being deposited with ADDGENE, until their release they are available as *E. coli* agar stabs upon request to J.H.N.

The data are provided in electronic supplementary material [66].

Authors' contributions. M.W.: formal analysis, investigation, methodology, validation, writing—original draft, writing—review and editing; L.S.W.: conceptualization, formal analysis, investigation, methodology, writing—original draft, writing—review and editing; B.R.C.: formal analysis, methodology, resources, writing—review and editing; H.L.: formal analysis, methodology, resources, writing—review and editing; P.N.W.: resources, writing—review and editing; A.L.B.: resources, writing—review and editing; D.B.: formal analysis,

methodology, resources, writing—review and editing; C.W.: conceptualization, formal analysis, funding acquisition, supervision, writing—original draft, writing—review and editing; J.H.N.: conceptualization, formal analysis, funding acquisition, supervision, 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., M.W., A.L.B. and P.N.W. were supported by Wellcome Trust (grant nos. 20289/Z/16/Z and 100209/Z/12/Z). C.W. received a Canada Research Chair. The research of C.W. and B.R.C. was supported by a Canadian Institutes of Health Research Foundation grant (FDN-2016-148364 to C.W.). D.B. acknowledges funding MC_UU_12014/7 from the Medical Research Council. We thank Diamond Light Source for access to the Cryo-EM facilities at the UK national electron bio-imaging centre (eBIC), proposal EM16637-1, funded by the Wellcome Trust, MRC and BBSRC. Computation used the Oxford Biomedical Research Computing facility, a joint development between the Wellcome Centre for Human Genetics and the Big Data Institute supported by Health Data Research UK and the National Institute for Health Research. Support was provided by the Wellcome (grant no. 203141/Z/16/Z). The OPIC electron microscopy facility was founded and supported by Wellcome (grant nos 060208/Z/00/Z and 093305/Z/10/Z).

References

- 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)
- Rojas ER *et al.* 2018 The outer membrane is an essential load-bearing element in Gram-negative bacteria. *Nature* **559**, 617–621. (doi:10.1038/s41586-018-0344-3)
- Whitfield C, Williams DM, Kelly SD. 2020 Lipopolysaccharide O-antigens-bacterial glycans made to measure. *J. Biol. Chem.* **295**, 10 593–10 609. (doi:10.1074/jbc.REV120.009402)
- 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**, 15 767–15 821. (doi:10.1021/acs.chemrev.0c01321)
- Whitfield C, Trent MS. 2014 Biosynthesis and export of bacterial lipopolysaccharides. *Annu. Rev. Biochem.* **83**, 99–128. (doi:10.1146/annurev-biochem-060713-035600)
- Liu B *et al.* 2020 Structure and genetics of *Escherichia coli* O antigens. *FEMS Microbiol. Rev.* **44**, 655–683. (doi:10.1093/femsre/fuz028)
- 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)
- Rai AK, Mitchell AM. 2020 Enterobacterial common antigen: synthesis and function of an enigmatic molecule. *MBio* **11**, e01914–20. (doi:10.1128/mBio.01914-20)
- Woodward L, Naismith JH. 2016 Bacterial polysaccharide synthesis and export. *Curr. Opin. Struct. Biol.* **40**, 81–88. (doi:10.1016/j.sbi.2016.07.016)
- Hong Y, Liu MA, Reeves PR. 2018 Progress in our understanding of Wzx Flippase for translocation of bacterial membrane lipid-linked oligosaccharide. *J. Bacteriol.* **200**, e00154–17. (doi:10.1128/JB.00154-17)
- Maczuga NT, Tran ENH, Morona R. 2022 Topology of the *Shigella flexneri* enterobacterial common antigen polymerase WzyE. *Microbiology (Reading)* **168**, 001183. (doi:10.1099/mic.0.001183)
- Ashraf KU *et al.* 2022 Structural basis of lipopolysaccharide maturation by the O-antigen ligase. *Nature* **604**, 371–376. (doi:10.1038/s41586-022-04555-x)
- Islam ST, Lam JS. 2014 Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway. *Can. J. Microbiol.* **60**, 697–716. (doi:10.1139/cjm-2014-0595)
- Lundstedt E, Kahne D, Ruiz N. 2021 Assembly and maintenance of lipids at the bacterial outer membrane. *Chem. Rev.* **121**, 5098–5123. (doi:10.1021/acs.chemrev.0c00587)
- Morris KN, Mitchell AM. 2023 Phosphatidylglycerol is the lipid donor for synthesis of phospholipid-linked enterobacterial common antigen. *J. Bacteriol.* **205**, e00403–22. (doi:10.1128/jb.00403-22)
- Dong C, Beis K, Nesper J, Brunkan-LaMontagne AL, Clarke BR, Whitfield C, Naismith JH. 2006 Wza the translocator for *E. coli* capsular polysaccharides defines a new class of membrane protein. *Nature* **444**, 226–229. (doi:10.1038/nature05267)
- Nickerson NN, Mainprize IL, Hampton L, Jones ML, Naismith JH, Whitfield C. 2014 Trapped translocation intermediates establish the route for export of capsular polysaccharides across *Escherichia coli* outer membranes. *Proc. Natl Acad. Sci. USA* **111**, 8203–8208. (doi:10.1073/pnas.1400341111)
- 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)
- Tocilj A *et al.* 2008 Bacterial polysaccharide co-polymerases share a common framework for control of polymer length. *Nat. Struct. Mol. Biol.* **15**, 130–138. (doi:10.1038/nsmb.1374)
- Chang CW, Tran ENH, Ericsson DJ, Casey LW, Lonhienne T, Benning F, Morona R, Kobe B. 2015 Structural and biochemical analysis of a single amino-acid mutant of WzzBSF that alters lipopolysaccharide O-antigen chain length in *Shigella flexneri*. *PLoS ONE* **10**, 1–21. (doi:10.1371/journal.pone.0138266)
- Kalynych S, Cherney M, Bostina M, Rouiller I, Cygler M. 2015 Quaternary structure of WzzB and WzzE polysaccharide copolymerases. *Protein Sci.* **24**, 58–69. (doi:10.1002/pro.2586)
- Collins RF, Kargas V, Clarke BR, Siebert CA, Clare DK, Bond PJ, Whitfield C, Ford RC. 2017 Full-length, oligomeric structure of Wzz determined by cryoelectron microscopy reveals insights into membrane-bound states. *Structure* **25**, 806–815.e3. (doi:10.1016/j.str.2017.03.017)
- Wiseman B, Nitharwal RG, Widmalm G, Högbom M. 2021 Structure of a full-length bacterial polysaccharide co-polymerase. *Nat. Commun.* **12**, 369. (doi:10.1038/s41467-020-20579-1)
- Larue K, Kimber MS, Ford R, Whitfield C. 2009 Biochemical and structural analysis of bacterial O-antigen chain length regulator proteins reveals a conserved quaternary structure. *J. Biol. Chem.* **284**, 7395–7403. (doi:10.1074/jbc.M809068200)
- Yang Y *et al.* 2021 The molecular basis of regulation of bacterial capsule assembly by Wzc. *Nat. Commun.* **12**, 4349. (doi:10.1038/s41467-021-24652-1)

26. Woodward R *et al.* 2010 In vitro bacterial polysaccharide biosynthesis: defining the functions of Wzy and Wzz. *Nat. Chem. Biol.* **6**, 418–423. (doi:10.1038/nchembio.351)
27. Nath P, Morona R. 2015 Detection of Wzy/Wzz interaction in *Shigella flexneri*. *Microbiology* **161**, 1797–1805. (doi:10.1099/mic.0.000132)
28. Ascari A, Tran ENH, Eijkelkamp BA, Morona R. 2022 Identification of the *Shigella flexneri* Wzy domain modulating Wzz_{PHS-2} interaction and detection of the Wzy/Wzz/Oag complex. *J. Bacteriol.* **204**, e0022422. (doi:10.1128/jb.00224-22)
29. Leo V, Tran E, Morona R. 2020 Polysaccharide copolymerase WzzB/WzzE chimeras reveal that the transmembrane 2 region of WzzB is important for interaction with WzyB. *J. Bacteriol.* **203**, e00598-20. (doi:10.1128/JB.00598-20)
30. Leo V, Teh MY, Tran ENH, Morona R. 2021 Identification of a region in *Shigella flexneri* WzyB disrupting the interaction with WzzPHS2. *J. Bacteriol.* **203**, e00413-21. (doi:10.1128/JB.00413-21)
31. Marczak M, Dźwierzyska M, Skorupska A. 2013 Homo- and heterotypic interactions between Pss proteins involved in the exopolysaccharide transport system in *Rhizobium leguminosarum* *bv. trifolii*. *Biol. Chem.* **394**, 541–559. (doi:10.1515/hsz-2012-0161)
32. Daniels C, Griffiths C, Cowles B, Lam JS. 2002 *Pseudomonas aeruginosa* O-antigen chain length is determined before ligation to lipid A core. *Environ. Microbiol.* **4**, 883–897. (doi:10.1046/j.1462-2920.2002.00288.x)
33. Tang KH, Guo H, Yi W, Tsai MD, Wang PG. 2007 Investigation of the conformational states of Wzz and the Wzz-O-antigen complex under near-physiological conditions. *Biochemistry* **46**, 11 744–11 752. (doi:10.1021/bi701181r)
34. Franco AV, Liu D, Reeves PR. 1998 The Wzz (cld) protein in *Escherichia coli*: amino acid sequence variation determines O-antigen chain length specificity. *J. Bacteriol.* **180**, 2670–2675.
35. Daniels C, Morona R. 1999 Analysis of *Shigella flexneri* Wzz (Rol) function by mutagenesis and cross-linking: Wzz is able to oligomerize. *Mol. Microbiol.* **34**, 181–194. (doi:10.1046/j.1365-2958.1999.01591.x)
36. Kintz EN, Goldberg JB. 2011 Site-directed mutagenesis reveals key residue for O antigen chain length regulation and protein stability in *Pseudomonas aeruginosa* Wzz. *J. Biol. Chem.* **286**, 44 277–44 284. (doi:10.1074/jbc.M111.273979)
37. Papadopoulos M, Morona R. 2010 Mutagenesis and chemical cross-linking suggest that Wzz dimer stability and oligomerization affect lipopolysaccharide O-antigen modal chain length control. *J. Bacteriol.* **192**, 3385–3393. (doi:10.1128/JB.01134-09)
38. Morona R, van den Bosch L, Manning PA. 1995 Molecular, genetic, and topological characterization of O-antigen chain length regulation in *Shigella flexneri*. *J. Bacteriol.* **177**, 1059–1068.
39. Bastin DA, Stevenson G, Brown PK, Haase A, Reeves PR. 1993 Repeat unit polysaccharides of bacteria: a model for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. *Mol. Microbiol.* **7**, 725–734. (doi:10.1111/j.1365-2958.1993.tb01163.x)
40. Kalynych S, Valvano MA, Cygler M. 2012 Polysaccharide co-polymerases: the enigmatic conductors of the O-antigen assembly orchestra. *Protein Eng. Des. Sel.* **25**, 797–802. (doi:10.1093/protein/gz057)
41. Power PM, Seib KL, Jennings MP. 2006 Pilin glycosylation in *Neisseria meningitidis* occurs by a similar pathway to wzy-dependent O-antigen biosynthesis in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **347**, 904–908. (doi:10.1016/j.bbrc.2006.06.182)
42. Warren MJ, Roddam LF, Power PM, Terry TD, Jennings MP. 2004 Analysis of the role of pglI in pilin glycosylation of *Neisseria meningitidis*. *FEMS Immunol. Med. Microbiol.* **41**, 43–50. (doi:10.1016/j.femsim.2004.01.002)
43. Meeske AJ *et al.* 2016 SEDS proteins are a widespread family of bacterial cell wall polymerases. *Nature* **537**, 634–638. (doi:10.1038/nature19331)
44. Kenyon JJ, Reeves PR. 2013 The Wzy O-antigen polymerase of *Yersinia pseudotuberculosis* O:2a has a dependence on the Wzz chain-length determinant for efficient polymerization. *FEMS Microbiol. Lett.* **349**, 163–170. (doi:10.1111/1574-6968.12311)
45. Morona R, van den Bosch L, Daniels C. 2000 Evaluation of Wzz/MPA1/MPA2 proteins based on the presence of coiled-coil regions. *Microbiology (NY)* **146**, 1–4. (doi:10.1099/00221287-146-1-1)
46. Guo H, Lokko K, Zhang Y, Yi W, Wu Z, Wang PG. 2006 Overexpression and characterization of Wzz of *Escherichia coli* O86:H2. *Protein Expr. Purif.* **48**, 49–55. (doi:10.1016/j.pep.2006.01.015)
47. Varadi M *et al.* 2022 AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444. (doi:10.1093/nar/gkab1061)
48. Jumper J *et al.* 2021 Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589. (doi:10.1038/s41586-021-03819-2)
49. Evans R *et al.* 2021 Protein complex prediction with AlphaFold-Multimer. *BioRxiv* 2022:2021.10.04.463034. (doi:10.1101/2021.10.04.463034)
50. Tran ENH, Morona R. 2013 Residues located inside the *Escherichia coli* FepE protein oligomer are essential for lipopolysaccharide O-antigen modal chain length regulation. *Microbiology (NY)* **159**, 701–714. (doi:10.1099/mic.0.065631-0)
51. Tran ENH, Papadopoulos M, Morona R. 2014 Relationship between O-antigen chain length and resistance to colicin E2 in *Shigella flexneri*. *Microbiology (NY)* **160**, 589–601. (doi:10.1099/mic.0.074955-0)
52. Marolda CL, Tatar LD, Alaimo C, Aebi M, Valvano MA. 2006 Interplay of the Wzx translocase and the corresponding polymerase and chain length regulator proteins in the translocation and periplasmic assembly of lipopolysaccharide O-antigen. *J. Bacteriol.* **188**, 5124–5135. (doi:10.1128/JB.00461-06)
53. 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 (NY)* **154**, 440–453. (doi:10.1099/mic.0.2007/013136-0)
54. Liu H, Naismith JH. 2009 A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expr. Purif.* **63**, 102–111. (doi:10.1016/j.pep.2008.09.008)
55. 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)
56. 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)
57. Hanahan D. 1983 Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557–580. (doi:10.1016/S0022-2836(83)80284-8)
58. Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJ, Lindahl E, Scheres SHW. 2018 New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166. (doi:10.7554/eLife.42166)
59. Rohou A, Grigorieff N. 2015 CTFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221. (doi:10.1016/j.jsb.2015.08.008)
60. 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)
61. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010 Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501. (doi:10.1107/S0907444910007493)
62. Afonine PV *et al.* 2012 Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367. (doi:10.1107/S0907444912001308)
63. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung LW, Read RJ, Adams PD. 2007 Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 61–69. (doi:10.1107/S090744490705024X)
64. Terwilliger TC, Adams PD, Read RJ, McCoy AJ, Moriarty NW, Grosse-Kunstleve RW, Afonine PV, Zwart PH, Hung L-W. 2009 Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta*

Crystallogr. D Biol. Crystallogr. **65**, 582–601. (doi:10.1107/S09074444909012098)

65. Skeldal S, Kjaergaard MM, Alwassel S, Nyengaard JR. 2015 Establishing a cellular FRET-based fluorescence plate reader assay to monitor proNGF-induced

cross-linking of sortilin and the neurotrophin receptor p75NTR. *Int. J. Biochem. Mol. Biol.* **6**, 17–25.

66. Weckener M, Woodward LS, Clarke BR, Liu H, Ward PN, Le Bas A, Bhella D, Whitfield C, Naismith JH.

2023 The lipid linked oligosaccharide polymerase Wzy and its regulating co-polymerase, Wzz, from enterobacterial common antigen biosynthesis form a complex. Figshare. (doi:10.6084/m9.figshare.c.6461045)