The molecular basis of regulation of bacterial capsule assembly by Wzc

Yun Yang1,2,3,8, Jiwei Liu1,2,8, Bradley R. Clarke4, Laura Seidel4, Jani R. Bolla5,6, Philip N. Ward1,2,3, Peijun Zhang2,7, Carol V. Robinson5,6, Chris Whitfield4✉& James H. Naismith1,2,3✉

Bacterial extracellular polysaccharides (EPSs) play critical roles in virulence. Many bacteria assemble EPSs via a multi-protein “Wzx-Wzy” system, involving glycan polymerization at the outer face of the cytoplasmic/inner membrane. Gram-negative species couple polymerization with translocation across the periplasm and outer membrane and the master regulator of the system is the tyrosine autokinase, Wzc. This near atomic cryo-EM structure of dephosphorylated Wzc from E. coli shows an octameric assembly with a large central cavity formed by transmembrane helices. The tyrosine autokinase domain forms the cytoplasm region, while the periplasmic region contains small folded motifs and helical bundles. The helical bundles are essential for function, most likely through interaction with the outer membrane translocon, Wza. Autophosphorylation of the tyrosine-rich C-terminus of Wzc results in disassembly of the octamer into multiply phosphorylated monomers. We propose that the cycling between phosphorylated monomer and dephosphorylated octamer regulates glycan polymerization and translocation.
High-molecular-weight extracellular polysaccharides (EPSs) play prominent roles in interactions between bacteria (pathogens, commensals, and symbionts) and their hosts (humans, livestock, and plants). Some EPSs are also important bioproducts in foods and other commercial applications. Although the structures of bacterial surface-associated and secreted polysaccharides are remarkably diverse, the majority are produced by a conserved and widely distributed assembly strategy (the Wzx-Wzy system) (reviewed in ref. 1). The process begins with the synthesis of polyprev了一遍 oligosaccharide repeat units at the cytoplasmic face of the inner membrane (IM). The lipid-linked repeat units are then flipped across the IM by the Wzx flippase, where they act as substrates for the Wzy polymerase, an integral membrane protein with a catalytic site located at the external face of the IM. The subsequent assembly stages differ, depending on the bacterial species and the glycoconjugate, but all involve a polysaccharide co-polymerase (PCP) protein that is important for establishing the size distribution of the polymeric product. This is vital for the biological properties and biological functions of the polymers. In the production of capsular polysaccharides (CAPS) or secreted EPSs in Gram-negative bacteria (Fig. 1a), Wzc (PCP-2a) proteins are thought to be the master regulator for both polymerization and translocation but the mechanism of regulation remains one of the most important questions in this field. Wzc possesses a C-terminal cytosolic protein tyrosine kinase (PTK) belonging to the BY-kinase family, which contains Walker A and B motifs, as well as a tyrosine-rich tail presenting several residues for phosphorylation (reviewed in ref. 2) (Fig. 1a). In contrast, the biosynthesis of lipopolysaccharide (LPS)-linked O-antigen polysaccharides by Wzx-Wzy pathways involves Wzz (PCP-1) homologs that lack a kinase domain and possess a shorter divergent periplasmic sequence compared to Wzc. This reflects some functional differences: Wzz is important for regulating O-antigen polymer chain length but plays no apparent role in the translocation of the final products. Structural data are available for Wzz but not for full-length Wzc.

In *Escherichia coli* EPS assembly prototypes, the Wzc autokinase is paired with a cognate soluble protein tyrosine phosphatase (PTP). The structure of one PTP, Wzb, has been determined. The corresponding genes are found in the EPS biosynthesis genetic locus but many isolates also possess unlinked genes encoding an additional Wzc and Wzb homolog pair (designated Etk and Etp) 

We advance a mechanistic model for the higher resolution apo-structure. Non-phosphorylated full-length Wzc is an octameric protein. Native tyrosine-phosphorylated Wzc was purified but the protein eluted as a broad peak in gel filtration and negatively-stained EM images showed heterogeneity in the samples (Supplementary Fig. 1a–c). Native protein mass spectrometry detected predominantly monomeric species and abundant phosphorylation adducts possessing 3, 4 or 5 phosphate groups (Fig. 1b). Further phosophopeptomic analysis suggested that the penta-phosphorylated species (5 P) was modified on tyrosines 708, 713, 715, 717 and 718. The strongest peak in tetra-phosphorylated (4 P) protein was modified on residues 708, 713, 715 and 717, but a species with phosphotyrosines at 708, 713, 717 and 718 were also detected. The triple-phosphorylated (3 P) was more heterogeneous (Supplementary Fig. 2a, b).

To test the possibility that phosphorylation led to heterogeneity and aggregation of purified Wzc, a sample was dephosphorylated by treatment with phosphatase Wzb and then reapplied to the size exclusion column but it remained like the native protein (Supplementary Fig. 1d–f). Previous work showed that mutation of K540 in the Walker box abrogated Wzc phosphorylation. In contrast to phosphorylated and enzymatically-dephosphorylated protein, purified non-phosphorylated WzcK540M was homogenous, oligomeric (based on molecular mass in gel filtration) and yielded discrete octameric particles in negatively-stained EM images (Supplementary Fig. 1h–j). We re-examined the Wzb-treated sample and collected 1787 cryo-EM images which were analysed by Topaz to identify 356,511 particles (Supplementary Fig. 1g). Through rounds of reference-free 2D classifications, 421 particles converged into clear views of an octameric Wzc (Supplementary Fig. 1g). In the raw micrographs, these particles (Supplementary Fig. 1g) resembled octamers of WzcK540M. However, the small octameric population (~0.1%) in the Wzb-treated sample made structure determination impractical, so we proceeded with WzcK540M.

A cryo-EM structure of WzcK540M at 2.85 Å resolution was determined with C1 symmetry (Supplementary Table 1, Supplementary Fig. 2a–c). Although the data were processed in C1, much of the structure displayed eightfold symmetry and the application of C8 symmetry improved the resolution to 2.30 Å (Supplementary Fig. 2f, g). However, C8 symmetry breaks down for part of the periplasmic region, so local C1 refinement was used to complete the structure (Supplementary Fig. 2a, h, i). WzcK540M in complex with ADP was determined to 2.6 Å with C8 symmetry (Supplementary Fig. 2a–d) and is essentially identical to the apo-structure, with the exception of additional density for ADP and Mg2+ (Supplementary Fig. 2e). Below, we focus on the higher resolution apo-structure.
Fig. 1 Capsule assembly and Wzc. a Undecaprenyl diphosphate-linked oligosaccharide repeat units are synthesized by a phosphoglycosyltransferase (PGT; (1)) and serotype-specific glycosyltransferases (GT(s); (2)). These building blocks are flipped across the membrane by Wzx (3) and polymerized by Wzy (4) in a reaction regulated by the PCP-2a family autokinase, Wzc. Wzc cycles between phosphorylated and dephosphorylated (catalysed by Wzb) states for its function (6). The polymer is translocated across the outer membrane by Wza (7); the closed octameric structure of Wza19 is shown on the right. Wza is also regulated by Wzc. In the prototype and some other species, Wzi supports the organization of translocated polymer into the surface-associated capsule structure (8) but this protein is absent from systems that produce secreted EPSs. In Gram-positive, bacteria step (7) is absent and nascent polymer is instead attached to peptidoglycan in the cell wall. In the absence of Wzc or Wza, short oligosaccharides representing one to a few repeat units are incorporated into LPS molecules in an off-pathway reaction (hatched box). The process has recently been reviewed1,4. The figure has been adapted from1.

b Purified, native phosphorylated Wzc appears only as a monomer (peak series highlighted in blue) in native mass spectrometry. The 14+ charge state shows multiple phosphorylation (denoted P) states of Wzc with four sites being the most prevalent (inset). The theoretical and measured masses for all the observed species are shown in the table. Supplementary Fig. 2a identifies phosphopeptides. The peak series with a measured mass of 66,212 ± 2 Da was a contaminant protein.
For ease of discussion, each protomer can be decomposed into three portions: the periplasmic region (residues 51–427), the transmembrane region (31–50, 428–447) and the cytoplasmic region (17–30, 448–721) (Fig. 2a, b). The WzcK540M octamer is ~165 Å high (cytosol to periplasm) and 146 Å wide and possesses a large central cavity with side portals that open to the membrane bilayer (Fig. 2c, d). The high quality of the experimental map (Fig. 2e and Supplementary Fig. 3j–m) allows accurate assignment of the sequence to structural elements. The cytoplasmic region comprises a short N-terminal helix (α1, Ile 17-Arg 30), the
The structure of WzcK540M determined by single-particle cryo-EM. a The monomer of WzcK540M contains a cytoplasmic region (red), tyrosine-rich tail (yellow), two transmembrane helices (green) and the periplasmic region (blue, salmon and ochre). b A schematic representation of the structure of WzcK540M coloured as a The first residue of each structural block is shown. c WzcK540M forms an octamer. The transmembrane helices are not close-packed and create portals to a large central cavity, clearly visible in the space fill representation. The structure has both a periplasmic and a cytoplasmic ring. d The WzcK540M octamer viewed from the periplasm, reveals the periplasmic region also forms a ring-like arrangement. The central cavity is open to both the cytoplasm and lipid bilayer. Residues 65 to 84 could not be located experimentally and may occlude the entrance to the cavity from the periplasm.

c. 1

e The high quality of the EM map is illustrated by the holes for the aromatic residues. f The C-terminal tail from one monomer (coloured yellow and red) has residues labelled in normal text. Y717 is at the active site of the kinase domain from the other monomer (coloured cyan). Key kinase residues in the other monomer are labelled in bold italics. ADP and Mg2+ are shown and labelled. g With Y717 at the active site, Y715 sits in a pocket where it makes a hydrogen bond with E675 from the neighbouring subunit. A phosphorylated tyrosine at position i would be disfavoured by size and charge. Disruption of binding at position i could perturb binding at position i + 2, the active site. Thus phosphorylation of Y717 seems most likely to follow (not precede) phosphorylation of Y717. Residues from the kinase domain are labelled in italics.
also present in Wzz (R279-R98) but there are compensating salt bridges with D66 and E162 in Wzz (Fig. 3d).

**The role of the tyrosine-rich tail.** In mass spectrometry, the native protein was shown to be phosphorylated and predominantly monomeric, echoing reports of the behaviour of isolated kinase domains. Systematic Tyr to Phe mutations established that the C-terminal tyrosine residues were essential for capsule production, but no single tyrosine was essential, nor was any single tyrosine sufficient. Reasoning the negatively charged Glu would mimic negatively charged pTyr (phosphotyrosine) and create a protein resembling a locked phosphorylated state, several Tyr-to-Glu mutants were constructed in Wzc<sup>K340M</sup>. With the exception of Y706E, each single Tyr to Glu replacement, as well as double (717 and 718) and triple
Fig. 3 The periplasmic and cytoplasmic rings of Wzc. a The class ii arrangement of motif 3 is shown. This is very different to the class i arrangement shown in Fig. 2a, suggesting motif 3 is dynamic. Motif 2 makes contact with α2 and thereby forging a structural connection between the periplasm and cytoplasm. The colour scheme is as in Fig. 2a. b The periplasmic ring of Wzc, using the colour scheme in Fig. 2a. The ring is predominantly held together by contacts between motif 1 of neighbouring monomers. The class ii arrangement is found as a pair where the helices from one monomer stack with the other monomer (circled). In the class iii arrangement, the helices are almost entirely disordered. c R410 is at the interface between the periplasmic domains of Wzc, where it makes interactions with Q62 and from the neighbouring subunit R262 and Q258 (bold italics). Interactions closer than 4 Å are shown as dashed lines. The interface in Wzz is shown in (d). d The interface in Wzz (PDB 6RBG)5, corresponding to that shown for Wzc in (e). In Wzz, R279 (equivalent to R410 in Wzc) is at the interface between the periplasmic domains, where it makes interactions with D66 and from the neighbouring subunit R98 and E616 (denoted bold italics). Interactions closer to 4 Å are shown as dashed lines. e The cytoplasmic ring of WzcK540M is held together by interactions between the neighbouring kinase domain and the tyrosine-rich C-terminus which is a target for phosphorylation (shown as a space fill sphere) that reach the active site of the neighbouring monomer. f Full-length Wzz (PDB GRGB2), shares the same arrangement of TM helices (pale green) and part of motif 1 (pale pink). The extended helical region (grey) in Wzz does not resemble any class of motif 3 in Wzc. Thus, the kinase domain, motif 2 and motif 3 of Wzz are all unique features. The octameric assembly of Wzz is shown in Supplementary Fig. 6b, c.

(715, 717 and 718) mutants formed octamers, as judged by EM (Supplementary Fig. 8a). We term this set of Y to E mutants “octamer capable”. A crude analysis of the percentage of octamers on the EM grid shows that as the number of Y to E substitutions increased, the portion of particles that were octameric reduced (Supplementary Fig. 8a). The capacity of these mutants to sustain autophosphorylation and polysaccharide production was assessed by constructing the same mutations in native (phosphorylation-proficient) Wzc protein. As expected, western blotting and mass spectrometry showed the amount of phosphorylation decreased with increasing numbers of replaced tyrosine residues (Fig. 4b, Supplementary Fig. 2c). Mass spectrometry showed these mutants are found as (with decreasing extent) phosphorylated monomers (Supplementary Fig. 2c), similar to the native protein. The “octamer capable” Y to E mutants (Supplementary Fig. 8a) all supported capsule production in vivo when introduced into the native wzc background (Supplementary Fig. 8b, c). Structural analysis suggests the Y706E mutant would perturb the interaction of α1 and α23, consistent with the observed functional impairment in single and double mutants involving Y706.

The introduction of four or more Glu residues resulted in severe defects in capsule production (Fig. 4a, Supplementary Fig. 8f). Phosphorylation of these proteins was not detected in whole-cell lysates (Fig. 4a) but this was attributed to reduced levels and limits of detection in lysates; phosphorylation was confirmed with the purified proteins (Fig. 4b, Supplementary Fig. 2c). A mutant with seven Glu residues did not support capsule production or autophosphorylation, as predicted by previous work24 (Supplementary Fig. 8f).

Analysis of cryo-EM images of WzcK540M with four Glu residues (designated WzcK540M4YE) revealed ~50% octamers, as well as other state(s) (Fig. 4c, Supplementary Fig. 9b–d). A 2.8 Å resolution structure of the WzcK540M4YE octamer was determined using the same approach as WzcK540M (Supplementary Fig. 9a–i). Overall, this structure has similar octameric (rmsd 3.1 Å over 4520 Ca atoms) and monomeric (rmsd of 2.5 Å over 651 Ca atoms) arrangements seen in WzcK540M (Supplementary Fig. 9e). The core kinase domain is unchanged between the structures (rmsd 0.4 Å for Ca atoms 452–697) but the tyrosine-rich tail has undergone a large change that now places Y708 at the active site (compared Y717 in WzcK540M) (Fig. 4d). Excluding the C-terminal tail and re-calculating the superpositions for monomer and octamer yields rmsd values of 1.8 Å and 2.3 Å respectively. The change in the tail results in a reduction of the buried surface area that holds the octamer together. In WzcK540M4YE, Y706 points into the same pocket as Y715 in WzcK540M (Fig. 2g) but this residue no longer makes the hydrogen bond with E675. In WzcK540M4YE, helix α3, is unwound, and the resulting loop no longer makes any contacts with any helix α1, whose position is altered as a result (Fig. 4d). The structure supports a model where the Tyr residues located C-terminal to Y708 have been phosphorylated, consistent with the observed multiple phosphorylation of the C-terminal peptide13–15.

The secondary structures of motif 1 and motif 2 in the periplasmic region are preserved in WzcK540M4YE but the structure comparison reveals some changes (rmsd 0.8 Å for 207 Ca atoms). Motif 3 still possesses the same three structural arrangements seen in WzcK540M but only two (not three) monomers have the fully ordered class i arrangement (rmsd of 0.4 Å for 133 Ca atoms). The larger rmsd value for the monomer and octamer, compare to the individual regions, reflects shifts in the relative positions of the regions. This can be visualised by superimposing the core kinase domains of monomer for WzcK540M and WzcK540M4YE, which shows a rigid body rotation of 11° at the other end of the molecule (Fig. 4e). Superposition using all the Ca atoms in octamer reveals shifts in the periplasmic helices, particularly those in class ii (Fig. 4f).

pY718 was observed in some peptides of the native protein (Supplementary Fig. 2a) but it is not a conserved residue (Supplementary Fig. 5a). An additional set of multiple Tyr to Glu mutants was constructed in order to test whether Y718 itself played a role and revealed that Y718 has no functional significance (Supplementary Fig. 8a, d, e).

The role of motif 3. The structures of the helical bundles appear to be dynamic within the octamer (resulting in three possible arrangements). Comparing WzcK540M vs WzcK540M4YE revealed the distribution of arrangements changed, suggesting the helical bundles are particularly sensitive to changes elsewhere in the structure. However, we are unable to exclude the possibility that this difference is due to a relatively smaller dataset of WzcK540M4YE compared with WzcK540M. To further probe the functional significance of the helical bundles, a construct lacking motif 3 (ΔMotif3) was engineered in either Wzc or WzcK540M backgrounds. As expected, purified WzcK540MΔMotif3 formed octamers, same as WzcK540M (Fig. 5a). However, the WzcΔMotif3 protein was unable to support capsule production establishing these helices are key to function, despite the preservation of its autokinase activity (Fig. 5b).

Discussion
Wzc participates in the production of critical CPS virulence factors in prominent multidrug-resistant ESKAPE pathogens Klebsiella pneumoniae25 and Acinetobacter baumannii26. This makes Wzc a candidate for antibacterial development to counter these and other pathogens. In the E. coli K30 prototype, Wzc (PCP-2a) proteins regulate polymerization, as deletion of wzc results in no accumulation or synthesis of any high-molecular-
An explicit link between Wzc and polymer chain length was also reported for *E. coli* colanic acid. A gain-of-function mutation in *A. venetianus* RAG1 *wzc* (equivalent to R410L in Wzc) leads to a hyper mucoid phenotype and increased apparent molecular weight (and potential value) of the commercial biopolymer known as emulsan.

Since Wzc and its O-antigen biosynthesis counterpart, Wzz, both regulate the polymerase Wzy, this aspect of their function(s) would be expected to reflect some structural conservation. Both proteins share most of motif 1 which forms an octameric ring and both have a large chamber in the membrane formed by the transmembrane helices (Supplementary Fig. 6b, c). The chamber offers a plausible location for Wzy and the portals would allow free exchange of lipid-linked oligosaccharides substrates. This location would imply contact between Wzy and α12 of Wzc and some experimental support for this proposal comes from mutants of the equivalent helix of Wzz that are known to alter Wzy behaviour.

Wzc possesses a cytoplasmic autokinase domain and two periplasmic sub-domains (motifs 2 and 3) that are absent in Wzz. Non-phosphorylated Wzc is an octamer held together by the interaction between tyrosine-rich C-terminal peptide of one monomer and the active site of the kinase domain of neighbouring monomer (Figs. 2c, 3e). The structure of WzcK540M4YE, chosen to mimic a phosphorylated state, showed a significant conformational change in the tyrosine tail, which would be
**Fig. 4** The effect of phosphorylation. 

*a* Western immunoblot of whole cell lysates probed with anti-K30 antiserum to detect cell surface polysaccharides (top panel), the anti-His antibody for expression (middle panel) and anti-pTyr antibody for pTyr (bottom panel). 

*b* Western immunoblot of purified Wzc proteins probed with anti-pTyr antibody and anti-His antibody is labelled as above. The phosphorylation of 3YE,4YE,5YE not visible in (a) is attributed to sensitivity. 

**Fig. 5** Function of Wzc periplasm motif 3. 

*a* Representative 2D class averages of Wzc<sup>K540M</sup> ΔMotif3 show the protein forms an octamer. Cryo-EM data were collected on Glacios (Thermo Fisher). Box size, 248 Å. 

*b* Western immunoblot of whole cell lysates probed with anti-K30 antiserum to detect cell surface polysaccharides (top panel), the anti-pTyr antibody for pTyr (middle panel) and anti-His<sub>α1</sub> antibody for expression (bottom panel). Bacteria were grown for 16 h without L-arabinose induction to achieve amounts of Wzc similar to those from chromosomal copies. Uncropped gels with molecular weight markers are shown in Supplementary Fig. 10a. Experiments were performed in technical triplicate. 

**Fig. 6** Function of Wzc periplasm motif 3. 

*a* Capping effect of octamer on the Wzc periplasmic domain. 

*b* Representative 2D class averages of Wzc<sup>K540M</sup> ΔΔMotif3 (ΔMotif2ΔMotif3) show the protein forms a tetramer. Cryo-EM data were collected on Glacios (Thermo Fisher). Box size, 248 Å. 

**Fig. 7** The effect of phosphorylation on motility. 

*a* Western immunoblot of whole cell lysates probed with anti-K30 antiserum to detect cell surface polysaccharides (top panel), the anti-His<sub>α1</sub> antibody for expression (middle panel) and anti-pTyr antibody for pTyr (bottom panel). 

*b* Western immunoblot of purified Wzc proteins probed with anti-pTyr antibody and anti-His<sub>α1</sub> antibody is labelled as above. The phosphorylation of 3YE,4YE,5YE not visible in (a) is attributed to sensitivity. Uncropped gels with molecular weight markers are shown in Supplementary Fig. 10a. Experiments were performed in technical triplicate. 

**Fig. 8** The effect of phosphorylation on motility. 

*a* Western immunoblot of whole cell lysates probed with anti-K30 antiserum to detect cell surface polysaccharides (top panel), the anti-His<sub>α1</sub> antibody for expression (middle panel) and anti-pTyr antibody for pTyr (bottom panel). 

*b* Western immunoblot of purified Wzc proteins probed with anti-pTyr antibody and anti-His<sub>α1</sub> antibody is labelled as above. The phosphorylation of 3YE,4YE,5YE not visible in (a) is attributed to sensitivity. Uncropped gels with molecular weight markers are shown in Supplementary Fig. 10a. Experiments were performed in technical triplicate. 

**Fig. 9** The effect of phosphorylation on motility. 

*a* Western immunoblot of whole cell lysates probed with anti-K30 antiserum to detect cell surface polysaccharides (top panel), the anti-His<sub>α1</sub> antibody for expression (middle panel) and anti-pTyr antibody for pTyr (bottom panel). 

*b* Western immunoblot of purified Wzc proteins probed with anti-pTyr antibody and anti-His<sub>α1</sub> antibody is labelled as above. The phosphorylation of 3YE,4YE,5YE not visible in (a) is attributed to sensitivity. Uncropped gels with molecular weight markers are shown in Supplementary Fig. 10a. Experiments were performed in technical triplicate.
Wzc motif 3 (Fig. 3f). Deletion of the periplasmic helices in motif 3 resulted in a Wzc variant protein that retained the ability to form octamers and possessed normal autokinase activity (Fig. 5a, b). However, the mutant abolished the production of polysaccharides uncoupling these activities from phosphorylation (Fig. 5b). This phenotypic outcome resembles deletion of Wza in vivo. The data is consistent with these helices being key to the control of Wza and the interaction of Wzc with Wza being required for activation of Wzy.

We propose that the machine that drives synthesis and export of EPS is a ternary complex minimally comprising octameric Wzc, octameric open-formed Wza and Wzy. Autophosphorylation of Wzc results in octamer dissociation, leading to the dissolution of the machine which halts polymerisation and closes Wza, of Wzc results in octamer dissociation, leading to the dissolution of EPS is a ternary complex minimally comprising octameric Wzc, octameric open-formed Wza and Wzy. Autophosphorylation of Wzc results in octamer dissociation, leading to the dissolution of Wzc and rationalises the essential requirement for activation of Wzy.

The molecular insights into this system provide the essential foundation to investigate these predictions and offer opportunities for efforts to target polysaccharide biosynthesis for both therapeutic and industrial benefit.

**Methods**

**Cloning, expression and purification of Wzc.** The wzc gene from *E. coli* (ET10) was cloned into vector pBAD24 to generate pBR901, in order to express Wzc with a C-terminal hexa-histidine tag. *E. coli* TOP 10 cells transformed with pBR901 were grown at 37 °C until OD600 reached approximately 0.8, then Wzc expression was induced with 0.002% arabinose at 20 °C overnight. Cells were harvested by centrifugation and stored at –80 °C. Cell pellets were resuspended with lysis buffer (20 mM Na phosphate, pH7.0, 500 mM NaCl) and lysed by passage through a Constant Systems cell disruptor. Unbroken cells were removed by centrifugation at 20,000×g for 1 h at 4 °C. Membranes were collected by ultracentrifugation at 186,000×g for 1 h at 4 °C and solubilized with 20 mM Na phosphate, pH7.0, 500 mM NaCl, 1% DDM at 4 °C. Residual membrane debris was removed by ultracentrifugation and the supernatant containing Wzc was purified under gravity using AB7 nickel resin (Cat. No. 68618-TAMAN-100). The protein was eluted with 20 mM Na phosphate, pH 7.0, 500 mM NaCl, 0.003 % Lauryl Maltose Neopentyl Glycol (LMNG) containing 300 mM imidazole after successive column washes with 20 mM imidazole and 50 mM imidazole. Purified protein was buffer exchanged to buffer (20 mM Na phosphate, pH 7.0, 500 mM NaCl, 0.003 % Lauryl Maltose Neopentyl Glycol (LMNG)) using a CentriPure P100 column and concentrated with a 100 kDa concentrator. Concentrated protein was applied to a Superose 6 10/30 increase column equilibrated with 20 mM HEPES, 150 mM NaCl, 0.001 % LMNG, 2 mM tris(2-carboxyethyl)phosphine (TECP), pH 7.3. Wzc mutant derivatives were made through site-directed mutagenesis or by Gibson assembly following the manufacturer’s protocols (New England Biolabs). Expression and purification of Wzc mutants were performed as described for the native protein.

For the dephosphorylation, Wzc was incubated with Wzb at room temperature in a molar ratio of 1:5. Samples were taken after 0, 1, 2 and 3 h intervals and the presence of phosphorylation on tyrosine were monitored as described below.

**Western immunoblot detection of phosphorylated Wzc proteins.** In total, 8 μl of 0.2 mg/ml purified proteins were analyzed by western immunoblotting using mouse monoclonal anti-phospho-tyr antibody (Sigma, Cat. No. P4110, dilution 1:5000) as primary and an HRP conjugated anti-mouse IgG antibody (Promega, Cat. No. W402R, dilution 1:5000) as secondary antibody. An anti-polyHistidine-peroxidase-antiperoxidase antibody (Sigma, Cat. No. A7058, dilution 1:5000) was used to detect hexa-histidine tagged Wzc protein, for showing the loading amount.

Wzc variants were also examined in whole-cell lysates prepared from *E. coli* CWG285 transformants grown in LB medium (containing 100 μg/ml ampicillin where required) at 37 °C. CWG285 is derived from *E. coli* ET10 and contains one insertion in wzc, and another insertion that exerts a polar inactivating effect on the wzc homolog, etb. Lysates were prepared by harvesting 1 OD600 equivalent of bacteria by centrifugation at 12,000 × g for 2 min. The pellets were lysed in 100 μl of Laemmli buffer (125 mM Tris-Cl, pH 6.8, 20 % (w/v) glycerol, 4% (w/v) SDS, 0.004 % (w/v) bromophenol blue) and heated at 100 °C for 10 min. The whole-cell lysates were analysed by SDS-PAGE and transferred to Protran 0.45 μm nitrocellulose membrane (Amersham). Hexa-histidine-tagged Wzc was identified using mouse anti-endoHis antibody (Qiagen, 1:2000) and phosphorylated Wzc was detected with mouse PY20 antibody (Sigma, Cat. No. P4110, 1:2000). Goat anti-mouse horseradish peroxidase (HRP) conjugated antibody (Cedar Lane, 1:3000) was used as the secondary antibody. HRP was detected by chemiluminescence using the Crescendo Western HRP substrate (Millipore). The uncropped images of gels are available in Supplementary Fig. 10.

**Western immunoblot analysis of cell surface polysaccharides.** Wzc variants were tested for function by transforming the corresponding plasmids into *E. coli* CWG285. Bacteria were grown and whole-cell lysates prepared as described above, with the exception that samples of CPS analysis were treated with proteinase K (0.5 mg/ml) at 37 °C for 1 h. Whole-cell lysates were separated and transferred to nitrocellulose as described above. K30 CPS was detected using rabbit anti-K30 antiserum as the primary antibody at a 1:3000 dilution and goat anti-rabbit alkaline phosphatase (AP) conjugated antibody (Cedar Lane, 1:3000) as the secondary antibody. AP was detected with nitro blue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (Roche).
EM analysis. For negative stain, 3.5 μl of purified protein (around 0.05 mg/ml) was applied to glow-discharged 400-mesh carbon-coated copper grids, which were then washed three times with ddH₂O and stained with 0.75% uranyl formate. Images were acquired using an FEI Tecnai T12 microscope at a magnification of ×42,000, with a calibrated pixel size of 2.63 Å.

For cryo-EM of WzcK540M and WzcK540M4YE, 3.5 μl purified sample at ~2 mg/ml was applied to glow-discharged Quantifoil gold r1/1.3/100 mesh grids and the grids were blotted for about 3 s in 1% solutions of 100% humidity and 4 °C before vitrification in liquid ethane using Vitrobot (FEI). All datasets were collected on a Titan Krios equipped with a K3 direct electron detector at eBIC, Diamond Light Source, UK. The apo- and ADP-complex datasets were collected by SerialEM at a magnification of ×100,000 with the physical pixel size of 0.829 Å/pixel. The total dose was 55 e Å⁻² and 53.6 e Å⁻² for apo- and ADP-complex datasets, respectively, and the defocus range was ~0.5 μm to ~2.5 μm. The 4YE dataset was collected by EPU software at a magnification of ×105,000, with the physical pixel size of 0.831 Å/pixel, a calibrated pixel size of 2.63 Å.

Contrast transfer function values were estimated by Gctf or CTFFIND. Particles were picked using Laplacian-of-Gaussian (LoG), based auto-picking in Relion3.0. After extraction and normalization in Relion3.0, particles were imported into cryoSPARC for further processing. Multiple rounds of 2D classification were carried out. The initial model was generated from a selection of 2D classes in cryoSPARC. Other rounds of 3D classifications were carried out without applying symmetry. C8 symmetry was observed for the resulting maps, except motif 3. For the ADP-complex and 4YE mutant maps, the best 3D classes were reconstructed using non-uniform refinement in cryoSPARC applying C1 or C8 symmetries. For the apo-structure, after cleaning through multiple rounds of 2D and 3D classification, 673,112 particles were used for further 3D classification with C1 symmetry. One representative class with the most complete motif 3 was chosen for non-uniform refinement to generate a 2.85 Å map at C1 symmetry. A mask covering only the periplasmic region was generated and local refinement was carried out long with the C8 symmetry. The resulting 2.77 Å map showed good density for the side chains of the periplasmic motif 3. For the C8 reconstruction, further 3D classifications were carried out and 546,088 particles were subjected to one final round of non-uniform refinement with C8 symmetry to yield a 2.3 Å map (Supplementary Fig. 3a).

For WzcK540M, 3.5 μl of purified sample at ~1.8 mg/ml was used to make the grids in the same way described above. In total, 1787 micrographs were collected on Glacios (ThermoFisher) at a magnification of ×92,000 with the physical pixel size 1.55 Å/pixel. Defocus range was ~1.0 μm to ~2.5 μm. Total dose was 60 e Å⁻². Using convolutional neural networks based algorithm Topaz23 we were able to pick 356,511 particles. Multiple rounds of 2D classifications were carried out, and the class averages showing clear Wzc occtamer end views were selected and analysed.

Grids of WzcK540M-Detox13 at ~1.9 mg/ml were made in the same way described above. Dataset was collected on Glacios at a magnification of ×92,000 with the physical pixel size 1.55 Å/pixel. The total dose was 60.38 e Å⁻². Data was processed similar to Wzc24-26 to get 2D class averages.

Model building and refinement. For the apo-structure, the cytoplasmic kinase domain was refined from a homology model (PDB code 3LA6). The other regions were built manually in Coot27. The 2.3 Å C8-symmetry map was used to build the structure, with the exception of motif 3. The locally refined 2.77 Å C1 symmetry map was used to build the three classes of motif 3. The entire structure was assembled in Coot and refined against the 2.85 Å C1 map. The structure without motif 3 was refined against the C8 map. For the ADP-bound complex structure, ligands were manually added in Coot and the structure was refined in PHENIX28. For the 4YE mutant structure, the mutation points were manually built in Coot, based on the apo-structure, and refinement was carried out in PHENIX. Figures of cryo-EM maps and structures were generated with PyMOL (The PyMOL Molecular Graphics System, Version 2.1 Schrödinger, LLC) and UCSF Chimera31.

Antibodies. Monoclonal anti-poly-Histidine-peroxidase antibody produced in mouse was purchased from Sigma (Cat. No. A7058). Monoclonal anti-phosphotyrosine antibody produced in mouse was purchased from Sigma (Cat. No. P4110). HRP conjugated anti-mouse IgG (H+L) antibody was purchased from Promega and used as the secondary antibody for detecting phosphotyrosine (Cat. No.W4028); AP-conjugated goat-anti-rabbit IgG was purchased from Cedarlane (Cat. No. CLC43008). Peroxidase-conjugated goat-antimouse IgG was purchased from Cedarlane (Cat. No. 115-036-003). Penta-his antibody was purchased from Qiagen (Cat. No. 34660).

Data availability

The data that supports this study are available from the corresponding authors upon reasonable request. EM maps and models are deposited in the EMDB and wwpdb under accession codes EMD-12338 and PDB 7NH (C1 WzcK540M); EMD-12339 and PDB 7NH (C8 WzcK540M); EMD-12340 (WzcK540M periplasmic localized map); EMD-12360 and PDB 7NI (C1 WzcK540M ADP complex); EMD-12359 and PDB 7NH (C8 WzcK540M ADP complex); EMD-12349 and PDB 7NE (WzcK540M4YE); EMD-12349 and PDB 7NE2 (WzcK540M4YE). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD025820. All constructs are available from the authors upon request until their deposition and release by ADDGENE. Source data are provided with this paper.

Received: 12 March 2021; Accepted: 29 June 2021; Published online: 16 July 2021

References

1. Whitfield, C., Wear, S. S. & Sande, C. Assembly of bacterial capsular polysaccharides and exopolysaccharides. Annu. Rev. Microbiol. 74, 521–543 (2020).
2. Morona, R., Purins, L., Tocilj, A., Matte, A. & Cygler, M. Sequence-structure relationships in polysaccharide co-polymerase (PCP) proteins. Trends Biochem. Sci. 34, 78–84 (2009).
3. Jutard, F. et al. BYKdb: the bacterial protein tYrosine kinase database. Nucleic Acids Res. 40, D321–D324 (2012).
4. Whitfield, C., Williams, D. M. & Kelly, S. D. Lipopolysaccharide O-antigens: bacterial glycans made to measure. J. Biol. Chem. 295, 10593–10609 (2020).
5. Wiseman, B., Nitharwal, R. G., Widmalm, G. & Högbom, M. Structure of a membrane-bound state of Escherichia coli Wzb reveals a novel glycophosphorylase domain. J. Biol. Chem. 281, 19570–19577 (2006).
6. Hageleuken, G., Huang, H., Mainprize, I. L., Whitfield, C. & Naismith, J. H. Crystal structures of Wzb of Escherichia coli and CspB of Streptococcus pneumoniae, representatives of two families of tyrosine phosphatases that regulate capsule assembly. J. Mol. Biol. 392, 678–688 (2009).

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-021-24652-1 | www.nature.com/naturecommunications
10. Ilan, O. et al. Protein tyrosine kinases in bacterial pathogens are associated with virulence and production of exopolysaccharide. *EMBO J.* **18**, 3241–3248 (1999).
11. Vincent, C. et al. Relationship between exopolysaccharide production and protein-tyrosine phosphorylation in gram-negative bacteria. *J. Mol. Biol.* **304**, 311–321 (2000).
12. Doublet, P., Grangeasse, C., Obadia, B., Vaganay, E. & Cozzone, A. J. Structural organization of the protein-tyrosine autokinase Wzx within *Escherichia coli*. *J. Biol. Chem.* **277**, 37339–37348 (2002).
13. Olivares-Illana, V. et al. Structural basis for the regulation mechanism of the tyrosine-kinase Wzc activity on colanic acid synthesis in *Escherichia coli* K12. *J. Mol. Biol.* **367**, 42–53 (2007).
14. Nadler, C. et al. Cycling of Etk and Etp phosphorylation states is involved in formation of group 4 capsule by *Escherichia coli*. *PLoS ONE* **7**, e37984 (2012).
15. Bechet, E. et al. Identification of structural and molecular determinants of the tyrosine-kinase Wzx and implications in capsular polysaccharide export. *Mol. Microbiol.* **77**, 1315–1335 (2010).
16. Temel, D. B., Dutta, K. & Ghose, R. Sequence-specific backbone H3, 4C and 15N assignments of the catalytic domain of the Escherichia coli protein tyrosine kinase, Wzx. *Biol. Mass Spectrom. NMR Assign.* **8**, 37–41 (2014).
17. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).
18. Paiment, A., Hocking, J. & Whitfield, C. Impact of phosphorylation of specific residues in the tyrosine autokinase, Wzx, on its activity in assembly of group 1 capsules in *Escherichia coli*. *J. Bacteriol.* **184**, 6437–6447 (2002).
19. Ernst, C. M. et al. Adaptive evolution of virulence and persistence in *carbapenem-resistant Klebsiella pneumoniae*. *Nat. Med.* **26**, 705–711 (2020).
20. Talansky, Y. et al. Capsule carbohydrate structure determines virulence in *Acinetobacter baumannii*. *PLoS Pathog.* **17**, e1009291 (2021).
21. Adams, P. D. et al. PHENIX: a comprehensive python-based system for structure determination. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **50**, 50–57 (1994).
22. MRC and BBSRC. All grids pre-screened and selected using the cryo-EM facility (OPIC) acknowledge Diamond for access and support of the CryoEM facilities at the UK national infrastructure program. J. W. F. Gill and E. M. S. C. thank the regulator of lipopolysaccharide O-chain length in *Escherichia coli* for access and support of the CryoEM facilities at the UK national infrastructure program.
23. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
24. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
25. Adams, P. D. et al. PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 (2020).
26. Metterson, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

### Acknowledgements

J.H.N., Y.Y., and P.W. are supported by Wellcome Trust (100209/Z/12/Z). J.R.B. and C.V.R. are supported by MRC Programme grants MR/N020413/1 and MR/V028839/1. J.W.F. and C.V.R. are supported by the Canadian Institutes of Health Research Foundation grant (FDN-2016-148364 to C.W.). J. L. and P.Z. are supported by Wellcome Trust Investigator Award (204622/Z/17/Z). We acknowledge Diamond for access and support of the CryoEM facilities at the UK national infrastructure program (EMBL-EBI proposal EM202231), funded by the Wellcome Trust, MRC and BBSRC. All grids pre-screened and selected using the cryo-EM facility (OPIC) in the Division of Structural Biology, University of Oxford, part of the UK Centre of Instruct-ERIC. The Franklin is a core funded research institute of the EPSRC.

### Author contributions

Y.Y. purified proteins with contributions from P.N.W., Y.Y., J.L., P.Z., and J.H.N. carried out the E.M. structural studies and analysis. B.C., L.S., and C.W. carried out the in vivo assay and analysis of Wzx mutant function. J.B. and C.V.R. carried out the mass spectrometry. J.H.N. and C.W. led the study. All authors contributed to the analysis of data and writing of the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-24652-1.

**Correspondence** Please replace [contact information] with the appropriate email addresses or institutions.

**Peer review information** *Nature Communications* thanks [list of reviewers] for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
