Structural studies of the Gram-negative O antigen ABC transporter WzmWzt

Christopher Alpert Caffalette
Woodbridge, CT

B.S. Biology, Psychology (Dual), Rensselaer Polytechnic Institute, 2006

A Dissertation Presented to the Graduate Faculty of the
University of Virginia in Candidacy for the
Degree of Doctor of Philosophy

Biophysics Degree-Granting Program
Department of Molecular Physiology and Biological Physics

University of Virginia
December, 2020

Abstract

Extracellular glycan biosynthesis is a widespread microbial protection mechanism. In Gram-negative bacteria, the O antigen polysaccharide represents the variable region of outer membrane lipopolysaccharides. O antigen structures are serotype specific and form extended cell surface barriers endowing many pathogens with survival benefits. In the ABC transporter-dependent biosynthesis pathway, O antigens are assembled on the cytosolic side of the inner membrane on a lipid anchor and reoriented to the periplasmic leaflet by the WzmWzt ABC transporter for ligation to the lipopolysaccharide core sugars. In agreement with its role as a polysaccharide transporter, WzmWzt forms a continuous transmembrane channel in a nucleotide-free state.

Here, I present the high-resolution structures of two distinct conformations of ATP-bound *Aquifex aeolicus* WzmWzt. In Chapter 2, I provide the crystal structure of ATP-bound WzmWzt and compare this conformation with the crystal structure of nucleotide-free WzmWzt. Large structural changes within the nucleotide-binding and transmembrane regions push conserved hydrophobic residues at the substrate entry site towards the periplasm and provide a model for polysaccharide translocation. With ATP bound within the nucleotide binding domains, the transporter forms a large transmembrane channel, with openings toward the membrane and periplasm, that is wide enough to accommodate an O antigen polysaccharide. The channel's periplasmic exit is sealed by detergent molecules that block solvent permeation. In a molecular dynamics (MD) simulation of water flux, an O antigen-occupied transporter is sealed, preserving the membrane potential. Further MD analyses suggest that, in a biological membrane, lipid molecules occupy the channel's periplasmic exit and prevent water flux in the transporter's resting state, leading to the proposal of a lipid-gating hypothesis.

In many cases, O antigen transport by WzmWzt depends on the chemical modification of the O antigen's non-reducing terminus, sensed by WzmWzt via a carbohydrate-binding domain (CBD) that extends its nucleotide-binding domain (NBD). In Chapter 3, I present the cryo-electron microscopy structure of the full-length WzmWzt transporter bound to ATP and in a lipid environment, revealing a highly asymmetric transporter organization. The CBDs dimerize and associate with only one NBD protomer. Conserved loops at the CBD dimer interface straddle a conserved peripheral NBD helix. The CBD dimer is oriented perpendicularly to the NBDs and its putative ligand-binding sites face the transporter to likely modulate ATPase activity upon O antigen binding. Further, our structure reveals a novel closed WzmWzt conformation in which an aromatic belt near the periplasmic channel exit seals the transporter in a resting, ATP-bound state. The sealed transmembrane channel is asymmetric, with one open and one closed cytosolic and periplasmic portal. The structure provides important insights into O antigen recruitment to and translocation by WzmWzt and related ABC transporters. Together, these analyses further our understanding of the WzmWzt transport mechanism culminating in a revised model of processive transport for the channel-forming O antigen transporter WzmWzt.

Dedication Page

I dedicate this work to my wife Sarah and my son Wil for all of their support, thank you both.

List of Abbreviations

β-ME, β-mercaptoethanol
AAA-ATPases, ATPases associated with diverse cellular activities-ATPases
ABC transporter, ATP-binding cassette transporter
ADP, adenosine diphosphate
ADP, atomic displacement parameters (Phenix)
Anap, 3-(6-acetylnaphthalen-2-ylamino)-2-aminopropanoic acid
APBS, Adaptive Poisson-Boltzmann Solver
ATP, adenosine triphosphate
C12E8, polyoxyethylene(8)-dodecyl ether
C15-PP, farnesyl-diphosphate
C5-PP, isopentenyl-diphosphate
C55-P, undecaprenyl-phosphate
CBD, carbohydrate-binding domain
CBM, carbohydrate-binding module
CDP, cytidine-diphosphate
CG, coarse-grained
CHAP, channel annotation package
CHARMM, Chemistry at HARvard Macromolecular Mechanics
CHARMM-GUI, CHARMM graphical user interface
CHL, coupling helix-loop
CMP, cytidine monophosphate
CPA, common polysaccharide antigen
CPS, capsular polysaccharides
Cryo-EM, cryo-electron microscopy
CTF, contrast transfer function
DAG, diacylglycerol
DMSO, dimethyl sulfoxide
Dol-P, dolichyl-phosphate
Dol-PP, dolichyl-diphosphate
Dol, dolichol or dolichyl
DolK, dolichol kinase
DTT, dithiothreitol
DW, dose-weighting or dose-weighted
EDTA, ethylenediaminetetraacetic acid
EH, external helix
EM, electron microscopy
EMDB, Electron Microscopy Data Bank
ETL, *E. coli* total lipid

FFT, fast Fourier transform
FRET, Förster resonance energy transfer or fluorescence resonance energy transfer
FSC, Fourier shell correlation
GA, glucuronic acid
Gal, galactose
Galf, galactofuranose
GalNAc, N-acetyl-galactosamine
Galp, galactopyranose
GF, gel filtration
GH, gate helix
GIF, Gatan imaging filter
GlcNAc, N-acetylglucosamine
GlcNAc-P, N-acetylglucosamine-1-phosphate
GPU, graphics processing unit
GT, glycosyltransferase
H-loop, histidine loop
HCl, hydrochloride
HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hGPT, human GlcNAc-1-P-transferase
His tag, histidine repeat motif tag
IF helix, interface helix
IM, inner membrane;
IMAC, immobilized metal affinity chromatography
IPTG, isopropyl-β-d-thiogalactoside
kDa, kilodalton
Kdo, 3-deoxy-d-manno-oct-2-ulosonic acid
l-Ara4N, 4-amino-4-deoxy-l-arabinose
LB, lysogeny broth
LDAO, lauryl-N,N-dimethylamine-N-oxide
LG-loop, leucine-glycine-loop
Lipid II, Und-PP-MurNAc-pentapeptide-GlcNAc
LoG, Laplacian-of-Gaussian
LLO, lipid-linked oligosaccharide
LPS, lipopolysaccharide
LTA, lipoteichoic acid
lyso-PG, lysophosphatidylglycerol
Man, mannose
ManNAc, N-acetylmannosamine
ManNAcA, N-acetylmannosaminuronic acid
MATE, multi-antimicrobial extrusion

MBOAT, membrane-bound O-acyltransferase
MD, Molecular dynamics
Mg^{2+}, magnesium
$MgCl_2$, magnesium chloride
MO, monoolein
MOP, multidrug/oligosaccharide-lipid/polysaccharide
MPI, message passing interface
MSP, membrane scaffold protein
MTF, modulation transfer function
MurNAc-pentapeptide, N-acetyl-muramoyl-pentapeptide
MurNAc, N-acetyl-muramic acid
MVF, mouse virulence factor
MWCO, molecular weight cutoff
NaCl, sodium chloride
NADH, nicotinamide adenine dinucleotide reduced
NBD, nucleotide-binding domain
NCS, non-crystallographic symmetry
Ni-NTA, nickel-charged nitrilotriacetic acid
NSLS-II, National Synchrotron Light Source II
OD, optical density
OM, outer membrane;
OSA, O-specific antigen
OST, oligosaccharyltransferase;
Pa-Wzx, Pseudomonas aeruginosa Wzx
PDB, Protein Data Bank
PDB, Protein Data Bank
PEG, polyethylene glycol
PG, phosphatidylglycerol
PG1, periplasmic gate helix 1
PG2, periplasmic gate helix 2
PGP, phosphatidylglycerol-phosphate
Pi, inorganic phosphate
PM, protein marker
PMSF, phenylmethylsulfonyl-fluoride
POPE, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POPG, palmitoyl-oleoyl-phosphatidylglycerol
PP, pyrophosphate
PPHPT, polyisoprenyl-phosphate hexose-1-phosphate transferases
PPNPT, polyisoprenyl-phosphate N-acetylhexosamine-1-phosphate transferases
PST, polysaccharide transport

RAM, random-access memory
RMSD, root-mean-square deviation
RSR, real-space refinement
RT, room temperature
S200, Superdex 200 size exclusion chromatography resin
SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC, size exclusion chromatography
smFRET, single molecule Förster or fluorescence resonance energy transfer
SN2, substitution nucleophilic bi-molecular
SSM, secondary structure matching
TAP, transporter associated with antigen processing
TCEP, tris(2-carboxyethyl)phosphine
TEV, tobacco etch virus
TLS, translation, libration, screw
TM, transmembrane
tmFRET, transition metal ion Förster or fluorescence resonance energy transfer
TMD, transmembrane domain
UDP, uridine diphosphate
Und-P, undecaprenyl-phosphate
Und-PP, undecaprenyl-diphosphate
Und, undecaprenol or undecaprenyl
UppS, undecaprenyl pyrophosphate synthase
Vi, inorganic vanadate
VMD, visual molecular dynamics
WT, wild type
WTA, wall teichoic acid

List of Figures

Appendices

List of Tables

Table of Contents

1 Chapter 1: Biosynthesis and export of bacterial glycolipids

Christopher A. Caffalette, Jeremi Kuklewicz, Nicholas Spellmon, and Jochen Zimmer

Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia 22908, USA

2020. *Annu Rev Biochem* 89: 741-68

The following introduction is adapted from the review article published in *Annual Review of Biochemistry*, volume publication date June 2020.

1.1 Overview

Complex carbohydrates are essential for many biological processes, from protein quality control to cell recognition, energy storage, cell wall formation. Many of these processes are performed in topologically extracellular compartments or on the cell surface; hence, diverse secretion systems evolved to transport hydrophilic molecules to their sites of action with polyprenyl lipids serving as ubiquitous anchors and facilitators of these transport processes.

Starting with recent insights into the generation, recognition, and recycling of polyprenyl lipids, this introductory chapter will summarize and compare bacterial biosynthesis pathways relying on the recognition and transport of lipid-linked complex carbohydrates. In particular, membrane transporters facilitating the essential process of *N*-linked glycan transport will be discussed and compared to transporters in the biosynthetic pathways for

O antigens, teichoic acids, and capsular polysaccharides, all implicated in the defense against host immune responses (1-3).

1.2 Polyprenyl lipids as facilitators of glycolipid transport

The phospholipid bilayers of cell and organelle membranes are hydrophobic barriers that are essentially impermeable to solutes, including organic or inorganic ions, carbohydrates, and nucleic acids. As such, diverse cellular uptake and secretion systems evolved to facilitate exchange with the environment, including facilitated diffusion, active transport, and passive flow along concentration gradients.

Polyprenyl-phosphate lipids are ubiquitous anchors facilitating transport of complex carbohydrates across biological membranes (**Fig. 1-1**). These lipids are essential for *N*-linked glycosylation in prokaryotes and eukaryotes, peptidoglycan and wall teichoic acid (WTA) biosyntheses, as well as capsular and lipopolysaccharide (LPS) membrane transport.

1.2.1 Polyprenyl lipid biosynthesis

In prokaryotes, the most abundant polyprenyl lipid is undecaprenyl-phosphate (Und-P or C55-P), a 55-carbon-long lipid built from 11 isoprene units. Generally, Und-P biosynthesis occurs in multiple steps by membrane-associated prenyltransferases (4), including the addition of 8 isoprenyl units from isopentenyl-diphosphate (C5-PP) onto farnesyl-diphosphate (C15-PP) and the dephosphorylation of the undecaprenyl-diphosphate (Und-

Fig. 1-1

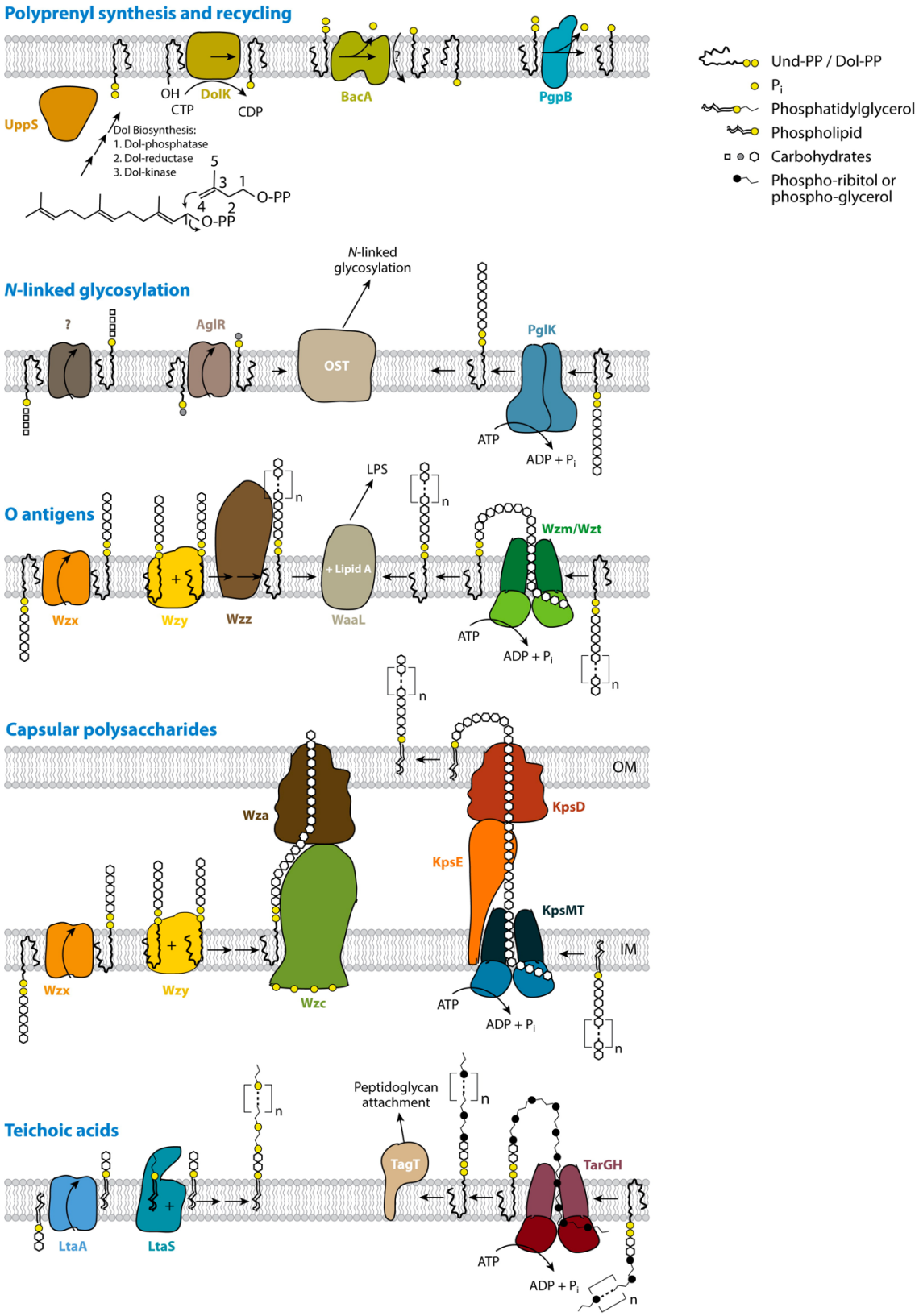

Polyprenyl synthesis and recycling
UppS
OH
CTP
DolK
CDP
BacA
PgpB
Dol Biosynthesis:
1. Dol-phosphatase
2. Dol-reductase
3. Dol-kinase
O-PP
O-PP
N-linked glycosylation
N-linked glycosylation
?
AglR
OST
PglK
ATP
ADP + Pi
O antigens
LPS
+ Lipid A
Wzm/Wzt
Wzx
Wzy
Wzz
WaaL
ATP
ADP + Pi
n
Capsular polysaccharides
OM
Wza
KpsD
KpsE
KpsMT
IM
Wzx
Wzy
Wzc
ATP
ADP + Pi
n
Teichoic acids
Peptidoglycan attachment
TagT
TarGH
LtaA
LtaS
ATP
ADP + Pi
n
Und-PP / Dol-PP
Pi
Phosphatidylglycerol
Phospholipid
Carbohydrates
Phospho-ribitol or phospho-glycerol

Fig. 1-1 Differences and commonalities of glycolipid transport pathways. The depicted pathways are described in the text. The transporter reorienting the glycolipid for *N*-linked glycosylation in archaea has not been identified yet; this reaction is catalyzed by the PglK ABC transporter in prokaryotes. Abbreviations: ABC, ATP-binding cassette; DolK, dolichol kinase; Dol-PP, dolichyl-diphosphate; IM, inner membrane; OM, outer membrane; OST, oligosaccharyltransferase; Und-PP, undecaprenyl-diphosphate; UppS, undecaprenyl pyrophosphate synthase.

PP) product to Und-P by undecaprenyl-pyrophosphate phosphatase (see next section). It should be noted, however, that the number of isoprene units, their degree of saturation, as well as the distribution of *cis* and *trans* double bonds within the molecule can vary slightly, as recently reviewed (5).

Eukaryotes and archaea primarily use dolichyl-phosphates (Dol-Ps) instead, which comprise 14 to 21 or more isoprene repeats and are saturated at the α-position, directly following the phosphate head group (6, 7).

1.2.2 Dephosphorylation of the lipid-linked diphosphate for *de novo* synthesis and recycling

Following biosynthesis, the pyrophosphate-linked lipids (e.g., Und-PP) are dephosphorylated to the monophosphates to serve as complex carbohydrate carriers. Further, as part of their transport function, the carriers are often attached to their carbohydrate cargoes via a diphosphate linker (**Fig. 1-1**). The linker remains attached to the lipid after the cargo has been transferred to the final acceptor, requiring the recycling of the pyrophosphate-linked pool in the extracellular membrane leaflet.

Dephosphorylation is catalyzed by phosphatases that fall into two main categories: the BacA and PAP2 superfamilies. Interestingly, all biochemically and structurally characterized polyprenyl phosphatases expose their active sites to the extracellular membrane surface where dephosphorylation for recycling occurs. No obvious candidate has been identified so far with a cytosolic active site that would be necessary for post-

biosynthesis dephosphorylation. However, recent structural analyses of *Escherichia coli* BacA support a model that could account for the elusive activity (8, 9).

The BacA structure revealed several surprises (**Fig. 1-2a**). First, the protein forms a homo-dimer with a large buried interface. Second, each protomer folds into six proper transmembrane (TM) helices and two pairs of short reentrant helices on either side of the membrane. Third, conserved catalytic residues are located on the periplasmic side. Fourth, each protomer comprises similar N- and C-terminal halves that are related by pseudo-two-fold symmetry similar to inverted domain repeats frequently observed in secondary transporters. Further, BacA forms a hydrophobic groove open to the periplasm and the extracellular membrane leaflet that ends in an electropositive putative catalytic pocket near the periplasmic water–lipid interface. In one structure, this groove is occupied by a monoolein molecule from the cubic phase crystallization medium (9).

BacA shares its topology with secondary transporters that couple solute transport with counter ion movement along its concentration gradient. Because these transporters usually adopt inward-and outward-facing conformations (10), a model was proposed in which BacA may also exhibit a transport function, perhaps facilitating the reorientation of the monophosphate reaction product (9). Further, because the BacA putative active site is formed within one repeat unit, it was speculated that the enzyme could exhibit two catalytic pockets, one on the cytosolic and one on the extracellular membrane side (8). Thereby, BacA could dephosphorylate *de novo* synthesized and enzymatically released Und-PP on

both sides of the membrane. However, these models remain to be tested and validated experimentally.

Fig. 1-2

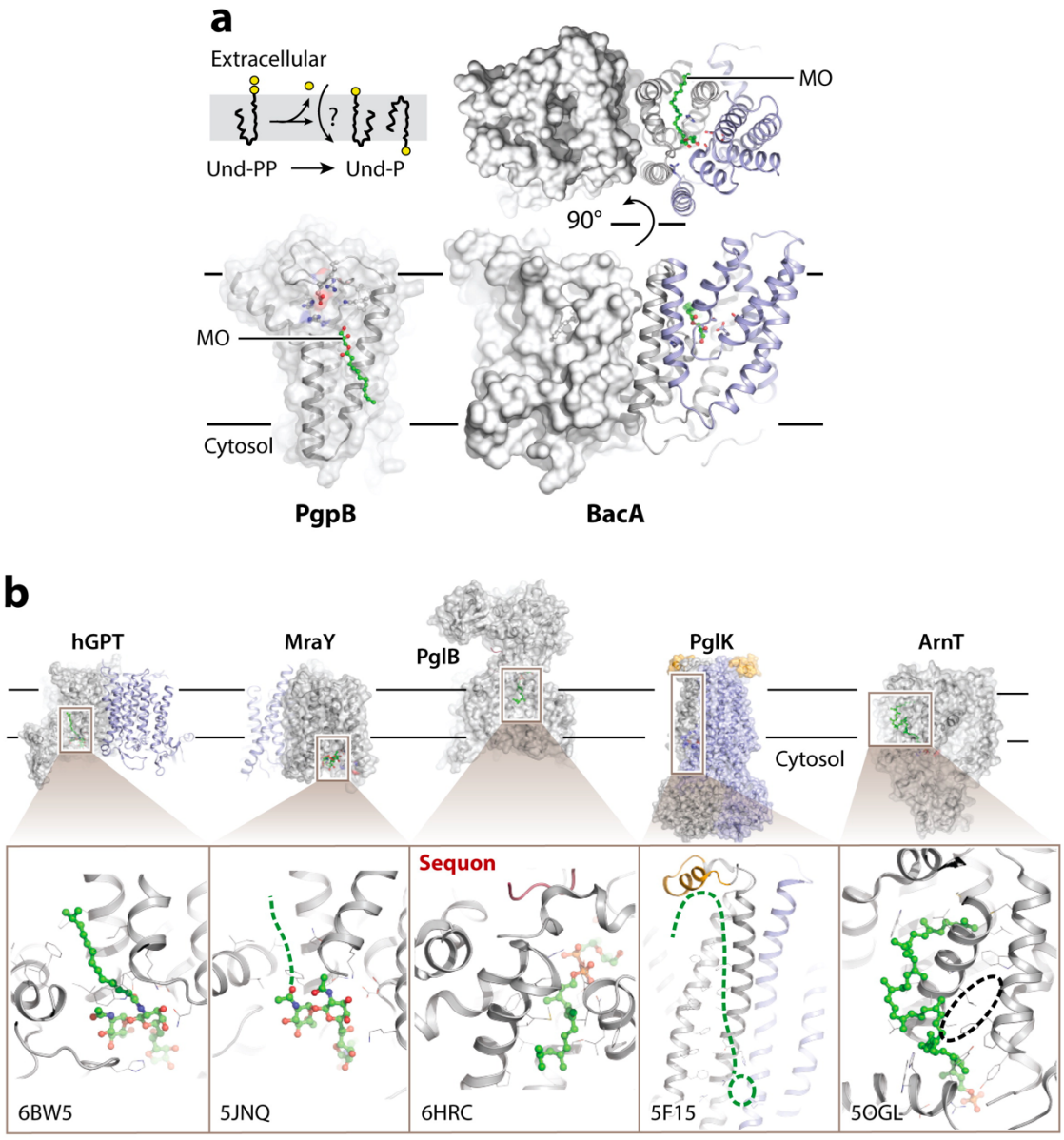


Fig. 1-2 Polyprenyl biosynthesis and recognition. **(a)** Crystal structure of *Escherichia coli* BacA and *Bacillus subtilis* PgpB (PDB identifiers 6CB2 and 5JKI). MO is shown as green sticks. **(b)** Comparison of different polyprenyl interaction motifs. Lipids or lipid-like ligands are shown as ball-and-stick models in green for carbon atoms. Proposed lipid-binding surfaces are indicated with green dashed lines. The putative lipid A–binding site on ArnT is indicated by a dashed oval. PDB identifiers are indicated in the subpanels. Abbreviations: MO, monoolein; PDB, Protein Data Bank; Und-P, undecaprenyl-phosphate; Und-PP, undecaprenyl-diphosphate.

BacA is not essential in *E. coli*. Knockout mutants revealed little *in vitro* growth defects, suggesting that alternative pathways exist to generate the essential Und-P pool (11). Indeed, although the total Und-PP phosphatase activity is substantially reduced in a *bacA* knockout strain, the approximately 25% residual activity was attributed to members of the PAP2 superfamily, in particular YbjG, YelU, and PgpB (12). Accordingly, triple knockouts of *bacA*, *pgpB*, and *ybjG* abolish *E. coli* growth, suggesting overlapping activities (12). Eukaryotic homologs of PgpB have been identified in yeast and mouse (termed CWH8 and DOLPP1, respectively) and localize to the endoplasmic reticulum (ER) membrane, where they are essential for N-linked glycosylation (13, 14).

The structures of *Bacillus subtilis* and *E. coli* PgpB have recently been determined (15, 16) (**Fig. 1-2a**). The enzymes form a helical bundle comprising six TM segments that form a V-shaped opening toward the periplasmic/extracellular membrane leaflet. On the extracellular side, the helical bundle is capped with amphipathic helices that run parallel to the membrane surface.

PAP2 enzymes contain three extracellular signature motifs (17) that create an electropositive pocket to accommodate the substrate's pyrophosphate group, which supports a function in dephosphorylating substrates in the extracellular membrane leaflet. *In vitro*, the *B. subtilis* PgpB enzyme dephosphorylates the substrates C55-PP, C5-PP, C15-PP, and phosphatidylglycerol-phosphate (PGP) with increasing specific activities, perhaps reflecting a primary function in phosphatidylglycerol (PG) biosynthesis (15).

1.2.3 Recognition of polyprenyls and their effects on membrane biophysics

An important, yet so far unanswered question is why polyprenyls serve as glycolipid anchors in essentially all kingdoms of life. The effects of Dol-P and Und-P on the membrane's fluidity, permeability, and phase transition behavior have recently been reviewed and support a role in facilitating lipid flip-flop (18-21). These observations are supported by recent structural analyses of membrane proteins that recognize polyprenyl-linked substrates (**Fig. 1-2b**).

For example, biochemical and structural studies on hGPT and the bacterial paralog MraY revealed highly selective recognitions of Dol and Und lipids, respectively (22-25). The enzymes catalyze similar reactions using UDP-activated substrates: hGPT transfers 1-phospho-GlcNAc to Dol-P, and MraY transfers the peptidoglycan precursor phospho-MurNAc-pentapeptide to Und-P. *In vitro* transfer reactions using polyprenyl-linked substrates revealed that hGPT only accepts Dol-linked substrates, whereas MraY is only active with the Und-attached substrate (22). Because dolichol is saturated at the α-position, it was proposed that the enzymes specifically recognize the first isoprenyl unit, in combination with the attached phosphate group.

Crystal structures of both enzymes revealed hydrophobic grooves postulated to accommodate the polyprenyl tails. This groove is somewhat wider in MraY, perhaps explaining preferential binding of an unsaturated isoprenyl unit. Bound to hGPT, the lipid tail of the inhibitor tunicamycin fits neatly into the hydrophobic groove, further supporting its role in polyprenyl interaction (22).

Another mode of polyprenyl recognition has been proposed for the bacterial ATP-binding cassette (ABC) transporter PglK. PglK reorients Und-PP–linked heptasaccharides in the inner membrane (IM) for *N*-linked glycosylation (26). Homodimeric PglK adopts a classical bacterial exporter topology with an intracellular nucleotide-binding domain (NBD) fused to a TM permease domain (27). A unique architectural feature is a short periplasmic external helix (EH) that runs parallel to the membrane plane. *In vitro*, ATPase and translocation assays showed that mutagenesis of the EH, either by substituting selected residues or replacing it with a flexible linker, dramatically reduced the substrate-induced ATPase activity as well as the ability to complement a knockout *in vivo* (27). Accordingly, a model was proposed in which PglK recognizes the substrate's head group near the cytosolic water–lipid interface in a positively charged pocket (27, 28) and the lipid moiety on its membrane-spanning surface, in part via the EH. However, a direct involvement of the EH in lipid recognition has not been demonstrated so far.

ArnT is implicated in polymyxin resistance in Gram-negative bacteria (29). It is a family-C glycosyltransferase (GT) that transfers 4-amino-4-deoxy-l-arabinose (l-Ara4N) from Und-P to lipid A, thereby accounting for adaptive polymyxin resistance in many pathogens (30). The transfer reaction occurs on the periplasmic side of the IM and likely requires simultaneous binding of Und-P–l-Ara4N and lipid A.

ArnT binds Und-P via two cavities formed on its lipid-exposed surface (31). One cavity, near the water–lipid interface, clusters strictly conserved Lys, Arg, and Tyr residues to

coordinate the Und-P's phosphate group. Replacing any of these residues with Ala completely abolishes function *in vivo* (31). The polyprenyl tail of Und-P packs against a hydrophobic and lipid-exposed surface extending from the phosphate coordination site. In contrast to the putative active site, mutagenesis of the polyprenyl binding surface had only marginal effects on *in vivo* activity. Interestingly, ArnT forms a third cavity adjacent to the Und-P binding pockets that is speculated to accommodate lipid A, thereby allowing direct glycosyl transfer.

Similar results were obtained for the *Campylobacter lari* oligosaccharyltransferase PglB. The enzyme transfers an Und-PP–linked heptasaccharides to the acceptor Asn of the conserved sequon Asn-X-Ser/Thr. Although the native PglB substrate is the Und-PP–linked glycan, *in vitro* transfer reactions showed appreciable activity with synthetic substrates containing only two isoprenyl repeats (32, 33). This finding suggests that PglB primarily recognizes the glycan structure and the following diphosphate group. Indeed, the crystal structure of PglB in complex with a nonhydrolyzable nerylneryl-PP-GlcNAc substrate reveals extensive polar coordination of the diphosphate (32). The well-resolved nerylneryl tail, instead, is embedded in a hydrophobic groove spanning almost half the width of the TM segment.

As for ArnT, the substrate's diphosphate moiety packs into a conserved polar pocket near the sequon-binding site, which includes an Arg residue forming a salt bridge with the β-phosphate (32). The preceding α-phosphate is coordinated by multiple hydrogen bonds,

one to the hydroxyl group of a conserved Tyr residue, as observed in several other enzymes (34).

1.3 *N*-linked glycosylation

The discovery of an *N*-linked glycosylation system in *Campylobacter jejuni* was a landmark achievement (35) that enabled the identification of similar pathways in eukaryotes (36) and archaea (37). In contrast to eukaryotes, *N*-linked glycosylation is not essential in bacteria and is mediated by Pgl proteins [recently reviewed in (38)]. Here, a heptasaccharide is assembled intracellularly with Und-PP serving as membrane anchor (**Fig. 1-1**). Following synthesis, the glycolipid is reoriented to the periplasmic lipid leaflet by the ABC transporter PglK (discussed above) and released on the periplasmic side to serve as substrate for oligosaccharyltransferase PglB.

In archaea, however, synthesis and transport of the glycolipid resembles the eukaryotic pathway (36) (**Fig. 1-1**). Indeed, several of the archaeal enzymes were identified on the basis of similarities with the eukaryotic counterparts [exemplified by S-layer protein glycosylation in *Haloferax volcanii* and recently reviewed in (37)]. Archaea use Dol-Ps as lipid anchors [the degree of saturation and overall length show some differences compared with eukaryotic Dol-Ps (6)], and proteins implicated in biosynthesis and transport are abbreviated Agl. For example, AglU, AglG, AglL, and AglE synthesize a Dol-P–linked tetrasaccharide (37) that is reoriented to the extracellular lipid leaflet by an unidentified transporter. Following transfer of the tetrasaccharide to the sequon, the glycan is mannosylated by AglS, with Dol-P-Man serving as substrate. Dol-P-Man, in turn, is

transported to the extracellular lipid leaflet by AglR, a putative transporter belonging to the multi-antimicrobial extrusion (MATE)-like superfamily (37), similar to the Wzx and LtaA transporters discussed above (**Fig. 1-1**).

1.4 O antigen biosynthesis

1.4.1 Initiation of O antigen biosynthesis

The outer membrane (OM) of Gram-negative bacteria is rich in LPS, a complex glycolipid comprising lipid A connected to core oligosaccharides (lipid A-core), and a terminal O antigen polysaccharide implicated in innate immune response evasion (39). Biosynthesis of the lipid A-core and its transport across the IM by MsbA has been reviewed in detail (2, 40). In *Enterobacteriaceae*, O antigen biosynthetic genes are serotype specific and belong to one of three parallel pathways: the Wzx/Wzy-dependent pathway, the ABC transporter (WzmWzt)–dependent pathway, or the synthase-dependent pathway [reviewed by Raetz &Whitfield (41)]. Outside of *Enterobacteriaceae*, *Pseudomonas aeruginosa* is known to simultaneously express genes of both the Wzx/Wzy and WzmWzt pathways, yielding two LPS forms that terminate in either the strain-dependent O-specific antigen (OSA) or the common polysaccharide antigen (CPA, a d-rhamnose polymer), respectively (42).

O antigen biosynthesis starts with the transfer of an initial phosphorylated hexose or hexosamine sugar onto Und-P. This initiating GT reaction is catalyzed by two families of integral membrane enzymes, the polyisoprenyl-phosphate hexose-1-phosphate transferases (PPHPT) and polyisoprenyl-phosphate N-acetylhexosamine-1-phosphate transferases (PPNPT) (1). WbaP is a prototypical member of the PPHPT family and is the initiating

enzyme for the *Salmonella enterica* O antigen that transfers galactose-1-phosphate onto Und-P (43). The PPNPT family is represented by the enzymes WecA of *Enterobacteriaceae* and WbpL of *P. aeruginosa*. WecA transfers an N-acetylglucosamine-1-phosphate (GlcNAc-P) onto Und-P to initiate both O antigen and enterobacterial common antigen biosynthesis, though some observations of utilizing GlcNAc-P were reported in *E. coli* and *Yersinia enterocolitica* O8 (41, 44). WbpL is bifunctional and transfers GlcNAc-P to build CPA or N-acetylfucosamine-1-phosphate to initiate OSA biosynthesis (45).

1.4.2 The WecA glycosyltransferase

WecA is topologically similar to MraY (discussed above; see **Fig. 1-2b**), yet it contains 11 instead of 10 TM helices (44). Several biochemical studies on WecA revealed that the enzyme uses UDP-GlcNAc, magnesium, and Und-P as substrates and that polyprenyl tails either shorter or longer than C55, or containing dolichyl instead of C55, significantly decrease its catalytic activity (46). Interestingly, although *E. coli* WecA is catalytically inactive with Dol-P as carrier, the *Thermotoga maritima* homolog utilizes Dol-P, yet with approximately 40% less activity compared with Und-P (47). These observations support a role for the polyprenyl lipid in substrate binding as later proposed for MraY (discussed above); however, in the case of WecA, binding may be independent of lipid saturation at the α-position.

Detailed reaction mechanisms based on kinetic studies with purified WecA were proposed by Al-Dabbagh *et al*. (48). The authors favored a one-step reaction mechanism involving

a ternary complex comprising the enzyme Mg^{2+}/UDP-GlcNAc, and Und-P, similar to the S_N2-like mechanism proposed for MraY (48-50).

1.4.3 Wzx/Wzy-dependent pathway for O antigen biosynthesis

In Wzx/Wzy-dependent O antigen biosynthesis, serotype-dependent GTs build a single O antigen repeat unit onto the Und-PP–GlcNAc primer at the cytosolic leaflet of the IM [reviewed in (41) and (51)]. This glycolipid is then reoriented by Wzx to the membrane's periplasmic side, where the Wzy polymerase assembles the repeat units by transferring the nascent O antigen chain onto a new repeat unit. Wzz, a membrane-anchored periplasmic α-helical barrel, is involved in chain-length regulation of the O antigen, potentially by forming a complex with Wzy.

1.4.3.1 The Wzx translocase

Wzx is an integral membrane protein with 12 predicted TM helices (52). Despite a consistent TM topology and glycolipid translocase function, Wzx primary sequences are not well conserved (53, 54). Of note, although Wzx reorients lipid-linked O antigen repeat units, it can also translocate the Und-PP–GlcNAc primer produced by WecA (53). Further, complementation studies showed that Wzx homologs are interchangeable between serotypes and even species, suggesting a conserved recognition of an initial hexosamine unit, GlcNAc or GalNAc (55). Similarly, *P. aeruginosa* serotypes O2, O5, O16, O18, and O20 utilize a common Wzx translocase, yet the O antigen repeat units of these serotypes have only the initial N-acetylfucosamine in common (56). Therefore, a current model

postulates that recognition of the O antigen repeat unit boosts Wzx translocase activity over the basal level triggered by interaction with the Und-PP–linked initial hexosamine (57).

No structural data exist for Wzx or the other 11 members of the polysaccharide transport (PST) family. However, within the broader multidrug/oligosaccharide-lipid/polysaccharide (MOP) superfamily that includes PST proteins, structures of the MATE and mouse virulence factor (MVF) families have been solved, represented by NorM and MurJ (**Fig. 1-3a**), respectively. Hong *et al*. (51) recently reviewed the putative reaction mechanism of Wzx, drawing comparison with MOP family members. These secondary transporters expose their substrate-binding sites to either the *cis* or the *trans* side of the membrane, and substrate translocation is coupled to cotransport or counter transport of Na^+ or H^+ (51). In the absence of a Wzx structure, Islam *et al*. (58) constructed a homology model of Wzx on the basis of the *Vibrio cholerae* NorM X-ray structure showing how *P. aeruginosa* Wzx (Pa-Wzx) might adopt a periplasm-facing conformation (**Fig. 1-3b**). Further, reconstitution of purified Pa-Wzx into proteoliposomes showed that external protons, but not sodium ions, enhanced loss of reporter iodide from the proteoliposome lumen, suggesting that Pa-Wzx is a proton-coupled antiporter (59).

Recently determined conformational states of the Lipid II (Und-PP-MurNAc-pentapeptide-GlcNAc) flippase MurJ offer new insights into a potential Wzx translocation mechanism (60, 61). MurJ has 14 TM helices, with TM helices 1–12 forming the core topology shared among MOP superfamily members, including Wzx proteins (60, 61). Additional MurJ

Fig. 1-3

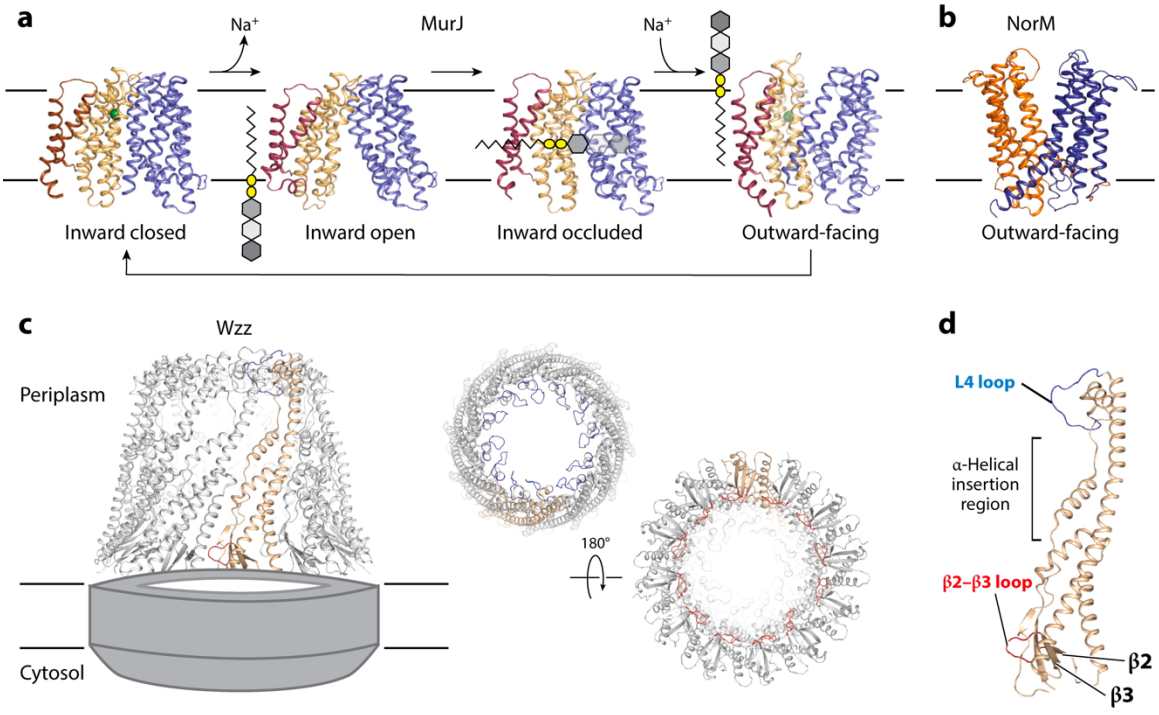


Fig. 1-3 Structural basis of translocation by Wzx flippases and the architecture of O antigen chain-length regulator Wzz. **(a)** Structures of the Wzx-like Lipid II flippase MurJ in each state of the putative translocation model: inward closed, inward open, inward occluded, and outward-facing (PDB identifiers 6NC6, 6NC7, 6NC8, and 6NC9, respectively). The two halves of the core MOP superfamily 12-TM bundle (TM1–6, violet; TM7–12, wheat) pivot along a central axis to expose the internal cavity to the cytosol or periplasm. Sodium (green sphere) binding and dissociation enables cation-dependent conformational changes that facilitate transport. The substrate is indicated as an Und-PP–linked O antigen fragment. **(b)** Outward-facing structure of the Wzx representative NorM (PDB identifier 3MKT). **(c)** Side, top, and bottom views of the Wzz periplasmic domain (PDB identifier 5NBZ). Wzz monomers (wheat) are arranged in a dodecameric α-helical barrel. The L4 loop (blue) and β2–β3 loop (red) sit at the apex and base of the Wzz barrel interior, respectively. TM and cytosolic domains are shown in gray. **(d)** Expanded view of the Wzz monomer highlighting the L4 loop, β2–β3 loop, and central α-helical region where insertions may extend the monomer's length in some Wzz homologs (PDB identifier 5NBZ). Abbreviations: MOP superfamily, multidrug/oligosaccharide-lipid/polysaccharide; PDB, Protein Data Bank; TM, transmembrane; Und-PP, undecaprenyl-diphosphate.

structures solved by Kuk *et al*. (61) include inward-facing closed, inward-facing occluded, and outward-facing states and show how the core 12-TM bundle translocates Lipid II across the IM for peptidoglycan biosynthesis (**Fig. 1-3a**). An identified sodium-binding site is important for MurJ folding; however, the role of the ion in Lipid II transport remains an open question (61). Two Arg residues at the center of the TM1 and TM8 interface form the pivot point between inward-facing and outward-facing states. These Arg residues in the *Thermosipho africanus* MurJ structures are proposed to stabilize the Lipid II pyrophosphate group during translocation, allowing the Und tail to remain in the membrane bilayer (61). Assuming similar interactions of the O antigen repeat unit's pyrophosphate linker or initial primer sugar with residues of the Wzx transport cavity, these interactions could support the basal level of flippase activity observed in truncation (53) and complementation (55) studies. The stabilization of species- or serotype-specific oligosaccharides within the cavity by the cognate Wzx may further enhance transport activity, as proposed (57).

1.4.3.2 The Wzy polymerase

Following translocation to the periplasm by Wzx, the Wzy polymerase extends the Und-PP–linked O antigen by one repeat unit per reaction. The elongation mechanism for Wzy originally proposed by Whitfield (62) involves growth of the O antigen chain at its reducing end, with new repeat units serving as acceptors. Although Wzy is an integral membrane protein, topology predictions range from 10 to 14 TM helices across species, which correlates with overall low primary sequence conservation (63-65). *In vitro* characterization by Woodward *et al*. (65) of purified *E. coli* O86 Wzy (WzyO86) showed

that WzyO86 alone can polymerize an O86 antigen up to 13 repeat units long from enzymatically synthesized Und-PP–linked O86 pentasaccharides, although shorter lengths were more abundant. This observation agrees with the existing model whereby Wzz, the O antigen chain-length regulator, is required for Wzy to polymerize longer-chain O antigens because, in its absence, primarily short-chain O antigens are produced (56). Strain-specific ranges of O antigen chain lengths, known as modal lengths, may then be controlled by a putative Wzy–Wzz complex or other interactions between the O antigen and Wzz (66). In support of this model, Woodward *et al*. (65) showed using their purified system that addition of *E. coli* O86 Wzz resulted in Wzy polymerizing specific O antigen modal lengths, shorter or longer, depending upon the strain-specific homolog of Wzz used. The authors also revealed that WzyO86 exhibits preference for Und lipids, as substrates containing Dol or lipids devoid of *cis* double bonds severely impact its function (65).

1.4.3.3 Wzz, the O antigen chain-length regulator

The Wzz chain-length regulator determines O antigen modal lengths in a strain-specific manner, though the mechanism is unknown. The proposed chain-length regulation model comprises two components: the stopwatch hypothesis involving the time of association between Wzy and Wzz in a putative complex and the ruler hypothesis pertaining to the interaction of the nascent O antigen with Wzz (67). Structural analyses of Wzz revealed a range of different oligomeric states from trimers to nonamers, and even dodecamers, depending on technique and sample preparation. Yet, overall, Wzz assembles into an α-helical bell-shaped ring of most likely 6–8 copies (67-71) (**Fig. 1-3c**). Although the

oligomeric state of biologically functional Wzz is an open question, existing evidence shows other factors are important for chain-length regulation as well (71-73).

Several regions within Wzz have been shown to affect O antigen modal length, including its periplasmic L4 loop, the α-helical insertion region, as well as the β2–β3 loop (**Fig. 1-3d**). Although mutagenesis of these regions impacts modal length, the molecular mechanisms responsible for the observed effects are unclear. Possible explanations include variations in Wzz barrel size (72), which in turn could affect the processivity of O antigen elongation; lack of interactions between Wzz and Wzy, resulting in altered catalytic activity (67, 73); and a putative carbohydrate-binding activity of Wzz's L4 loop, mediating interactions with the nascent O antigen (71).

1.4.4 The ABC transporter-dependent pathway for O antigen biosynthesis

1.4.4.1 The O antigen ABC transporter WzmWzt

An alternative route to synthesize and export O antigens is via the ABC transporter–dependent pathway (41). Here, the complete O polysaccharide is assembled intracellularly on an Und-PP anchor and exported across the IM by the WzmWzt ABC transporter (**Fig. 1-4**). Following O antigen translocation, the Wzx/Wzy-dependent and ABC transporter–dependent pathways converge at the WaaL ligase, which transfers the O antigen chain to the lipid A-core, thus forming mature LPS. This pathway has two distinctive modes to terminate and export the O polysaccharide. In one mode, the O antigen chain is capped at its nonreducing end with a chemical group to prevent further extension and signal biosynthesis completion (74); however, the other mode does not utilize end modification

Fig. 1-4

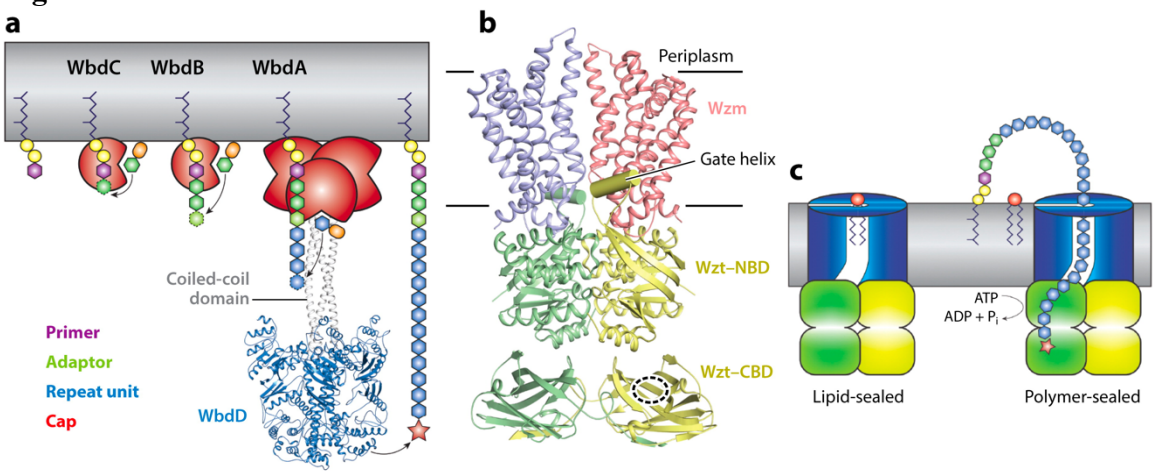


Fig. 1-4 ABC transporter–dependent O antigen biosynthesis. **(a)** Biosynthesis and length control of Und-PP–linked O antigen. Glycosyltransferases are shown in red and labeled. The O antigen is represented with hexagons colored to distinguish the primer, adaptor, and repeat units and a red star indicating the cap. In CBD-dependent export, the size of the O antigen is dictated by the size of the coiled-coil domain from WbdD, which terminates polymerization by end modification of the O antigen. The coiled-coil domain was modeled by CCBuilder and superimposed to the WbdD model (PDB identifier 4UW0). **(b)** Putative model of the full-length WzmWzt complex adopted from PDB identifiers 6M96 and 2R5O. The orientation of the CBD relative to the NBD is currently unknown. The likely cap-binding surface of the CBD is indicated by a dashed circle. **(c)** Lipid gating of the WzmWzt O antigen transporter. Abbreviations: ABC, ATP-binding cassette; CBD, carbohydrate-binding domain; NBD, nucleotide-binding domain; PDB, Protein Data Bank; Und-PP, undecaprenyl-diphosphate.

and instead appears to orchestrate the biosynthetic and export activities to control chain length (75). An important distinction for the export of capped O antigen chains is the presence of a carbohydrate-binding domain (CBD) at the C terminus of the ABC transporter's NBD. The CBD specifically interacts with the end modification, and this interaction is essential for O antigen secretion (76). This section highlights recent insights into the mechanism and control of these two processes, best demonstrated for *E. coli* O9a and *Klebsiella pneumoniae* O2a.

1.4.4.2 Adaptor attachment and O antigen processing

The Und-PP–linked O antigen polymer can be divided into four parts: a primer, an adaptor, the repeating units, and a terminal modification (if present; see **Fig. 1-4a**). As with the Wzx/Wzy-dependent pathway, the primer is a hexosamine-1-phosphate, generally GlcNAc-1-phosphate in *Enterobacteriaceae*, that is transferred to Und-P by WecA, thus forming Und-PP-GlcNAc. A set of GTs that localizes to the cytosolic membrane periphery in proximity to Und-PP-GlcNAc attaches the adapter and repeat unit sugars (41). The enzymes likely form a biosynthetic complex, as demonstrated for *E. coli* O9a and *K. pneumoniae* O2a (77, 78).

The O glycan adaptor is a short oligosaccharide, usually one to three sugars depending on the serotype, attached to the primer's nonreducing end by one or two GTs. For example, *E. coli* O9a expresses two monofunctional α(1→3) mannosyltransferases, WbdC and WbdB. WbdC transfers the first mannose to the primer, and WbdB builds two additional mannose units onto the former mannose (79, 80), resulting in Und-PP-GlcNAc-$(Man)_3$

tetrasaccharide. Similarly, *K. pneumoniae* O2a serotypes express two galactosyltransferases, WbbO and WbbN, possessing α(1→3) galactopyranosyl and β(1→3) galactofuranosyl activities, respectively (81, 82). WbbO transfers galactopyranose to the primer, and then WbbN links galactofuranose to the former galactopyranose, creating the Und-PP-GlcNAc-Gal*p*-Gal*f* trisaccharide (81, 82).

The O antigen is then built on the adapter's nonreducing end by serotype-specific GTs. These GTs may come in the form of a multidomain enzyme or a multi-subunit complex (77, 83). One example of a multidomain GT includes WbdA from *E. coli* O9a. The enzyme is a family-4 bifunctional mannosyltransferase containing two GT-B domains in which the N- and C-terminal domains possess α(1→2) and α(1→3) transferase activity, respectively (83-85). Previous studies showed that WbdA utilizes both GT domains to synthesize a mannotetraose repeat by cycling turns, adding two mannose units linked α(1→2) followed by two mannose units linked α(1→3) (84, 85). An example of a multiprotein complex includes *K. pneumoniae* O2a, in which the O polysaccharide is synthesized by three GTs (WbbM, WbbN, and WbbO) that form a membrane-associated complex (77). As mentioned above, WbbO and WbbN synthesize the adapter, and WbbM synthesizes the O antigen repeat unit.

1.4.4.3 O antigen chain-length control and termination

The modal distribution of the O glycan is also serotype-specific and controlled differently by CBD-independent or CBD-dependent export modes as highlighted for the *K. pneumoniae* O2a and *E. coli* O9a prototypes, respectively. For the CBD-independent

pathway, O antigen secretion only occurs if the biosynthetic and export components are co-expressed (75). For the CBD-dependent pathway, however, O antigen biosynthesis can precede export, suggesting that the biosynthesis and export machineries function separately (74).

If synthesis and export are coupled, O antigen synthesis may be terminated during or after ABC transporter–mediated export. In support of this hypothesis, overexpressing the ABC transporter in *K. pneumoniae* O2a decreases the O antigen size distribution (75), perhaps owing to premature export of nascent O antigen chains. Most likely, the *K. pneumoniae* O2a ABC transporter recognizes the conserved Und-PP-GlcNAc primer to initiate transport and has less specificity for the repeat unit structure. As such, the *K. pneumoniae* O2a transporter also translocates other Und-PP–linked O polysaccharides, for example, the *E. coli* O9a mannan (75).

In CBD-dependent export, the length of the O antigen polymer is controlled by a modifying cap (i.e., β-Kdo, phosphate, or methylphosphate) that is attached to the polymer's nonreducing end (74). The modifying cap is transferred by specific transferases containing membrane-anchored coiled-coil domains and is required for O antigen export *in vivo* (86). The coiled-coil domain serves as a molecular ruler that measures the length of the nascent polysaccharide and separates the capping transferase activity from the membrane-associated GTs (87).

Chain-length control and termination have been well studied for *E. coli* O9a (78). This serotype utilizes WbdD, a multidomain, bifunctional kinase and methyltransferase that caps the O antigen with a phosphomethyl group (**Fig. 1-4a**). The catalytic domains are attached to a coiled-coil region that interacts with the membrane as well as the O antigen–synthesizing WbdA (78).

A recent WbdD structure including a short segment of the coiled-coil domain revealed a homotrimeric arrangement, which was confirmed in solution by small-angle X-ray scattering analysis (87). Thus, WbdD adopts a mushroom-like fold, with the coiled-coil domain forming the stem and the methyltransferase and kinase domains forming the dome. The coiled-coil domain dictates the length of the synthesized O antigen polymer by separating the catalytic domains from the membrane, thereby coupling the likelihood of the capping reaction to O antigen length.

1.4.4.4 Translocation of the O antigen polymer

The completed Und-PP–linked O antigen is reoriented to the periplasmic side of the IM by the WzmWzt ABC transporter, where it is accessible to the WaaL ligase [reviewed in (88)]. Wzm is the integral membrane component, and Wzt forms the NBD as well as the CBD in some species. Although the NBD contains the conserved sequence motifs necessary for ATP hydrolysis, the CBD reveals limited sequence conservation yet is required for proper exporter function by binding the O antigen's nonreducing end cap (76). Interestingly, the two domains do not have to be covalently linked. Co-expression of the separated transporter and CBD entities is sufficient for *in vivo* O antigen export, whereas *in vitro*,

interaction of the isolated CBD with the transporter increases its ATPase activity, even in the absence of the O antigen (76, 89).

Structures of isolated WztCBD were solved from *E. coli* O9a and *Raoultella terrigena* (76, 90). In both cases, two WztCBD protomers form a stable homodimer by swapping the C-terminal β-strands. Each WztCBD subunit forms a β-sandwich with an immunoglobulin-like fold, a topology similar to those identified in other CBD structures (91). The domain was shown to interact with the O antigen's terminal cap on the surface of a concave β-sheet, using moderately conserved residues (76, 90). The CBD and NBD are connected via a short ∼15-amino-acid linker, yet their precise arrangement is unknown (**Fig. 1-4b**).

Crystal structures of the CBD-truncated *Aquifex aeolicus* WzmWzt complex, nucleotide-free and ATP-bound, were recently solved (34, 89), providing the first insights into the O antigen translocation mechanism (**Fig. 1-4b**). One distinguishing structural feature is a small gate helix (GH) between the first two β-strands of the NBD that is positioned close to the Wzm protomer interface near the membrane boundary. This GH creates an electropositive pocket that may selectively bind the lipid head group of the Und-PP–linked glycan. Of note, a similar GH is found in WTA ABC transporters, which translocate Und-PP–linked polyglycerol or polyribitol in Gram-positive bacteria (3).

O antigens and teichoic acids are long polymers. Studies on the bacterial cellulose synthase revealed that cellulose spans the width of the membrane with approximately 10 glucosyl units (92). O antigens may exceed this length more than 10-fold; hence, the ABC

transporter has to accommodate the polymer during the translocation reaction. Supporting this hypothesis, the WzmWzt transporter forms a continuous solvent-accessible TM channel in nucleotide-free and ATP-bound states that is sufficiently large to accommodate a polysaccharide chain (34, 89). In the ATP-bound conformation, the channel displays a lateral opening toward the periplasmic lipid leaflet, which likely allows exit of the Und-PP head group into the bilayer phase following reorientation.

Molecular dynamics simulations revealed an intriguing WzmWzt gating mechanism (34). It was proposed that the transporter may always adopt a channel-forming conformation, yet in the absence of a translocating polymer, membrane lipids occupy the channel's periplasmic half, thereby forming a hydrophobic plug that seals the pore (**Fig. 1-4c**).

These analyses led to a model in which the Und-linked O antigen first interacts with the transporter's GH to initiate transport and then its charged head-group inserts into the channel and reorients to the periplasmic side with the polyprenyl chain remaining in the bilayer phase, followed by several cycles of ATP hydrolysis to translocate the polysaccharide (34). In this model, the initial interaction of the O antigen's cap with the WzmWzt CBD may increase its local concentration, thereby facilitating recognition and transport.

1.4.5 WaaL, the O antigen ligase

In the periplasm, the O antigen is transferred to the lipid A-core by the family-8 GT WaaL, thereby generating a mature LPS molecule (93). Progress toward dissecting the function of

WaaL is hindered by poor primary sequence conservation among WaaL homologs and the lack of a 3-dimensional structure. Nevertheless, several groups report progress in illustrating the key functional and structural features of WaaL from *V. cholerae*, *S. enterica*, *E. coli*, and *P. aeruginosa* (94-97). These homologs share 12 predicted TM helices together with an extended, yet variably sized, periplasmic loop between helices 9 and 10. This loop was proposed to form the active site. Two conserved loop residues, Arg and His, were shown to be critical for enzymatic activity (96), perhaps by coordinating the pyrophosphate moiety of the Und-PP O antigen donor, as observed in other related GTs (98).

1.5 Teichoic acids

1.5.1 Overview

Teichoic acids (from Greek *teīkhos*, meaning "defense wall") are among the most predominant glycopolymers in Gram-positive bacteria, in which they are important components of the cell surface architecture, performing essential functions in cell division and shape regulation. Teichoic acids occur in at least two different forms: one that is anchored to the peptidoglycan layer and known as WTA and another that is anchored to a diacylglycerol (DAG) glycolipid in the extracellular leaflet of the membrane [lipoteichoic acid (LTA)]. The distinct mechanisms that synthesize these polymers are briefly summarized and have recently been reviewed in detail (3, 99).

Teichoic acids are copolymers of polyol repeat units linked via phosphodiester bonds. Among the most common WTA polyols are ribitol (r) and glycerol (g) phosphates, and enzymes implicated in their biosyntheses and secretions are referred to as Tar- or Tag-,

respectively. LTAs commonly comprise glycerol-phosphate repeat units originating from PG lipids and linked to DAG via a di-saccharide unit, exemplified by the glucosyl-β1-6-glucosyl-β1-3-DAG glycolipid of *Staphylococcus aureus* and *B. subtilis* (99).

LTA and WTA are assembled by fundamentally different mechanisms that resemble, respectively, the Wzx/Wzy-dependent and ABC transporter–dependent pathways of O antigen biosynthesis discussed above. Although LTA is assembled from PG head groups on the membrane's extracellular side, WTA, in contrast, is generated intracellularly on an Und-PP anchor and transported to the cell surface by the Tar(g)GH ABC transporter (3, 99).

1.5.2 Wall teichoic acid biosynthesis

WTA is assembled by TagF (**Fig. 1-5a**). In 2010, Lovering *et al*. (100) reported the TagF structure from *Staphylococcus epidermidis* that catalyzes the transfer of a phospho-glycerol moiety from cytidine-diphosphate (CDP)–glycerol to the growing WTA polymer. TagF contains a membrane-targeting domain and a catalytic region formed by two GT-B domains (98). The membrane-targeting domain contains two helices, one hydrophobic and one electropositive, that could interact with the lipid's acyl chains and the negatively charged phospholipid headgroups, respectively. Accordingly, association with the membrane has been proposed to influence the length of the formed WTA polymer (101).

The TagF C-terminal GT-B domain binds the nucleotide moiety of a co-crystallized CDP-glycerol substrate via a conserved DYSSVXXD/E motif. Interestingly, CDP-glycerol

Fig. 1-5

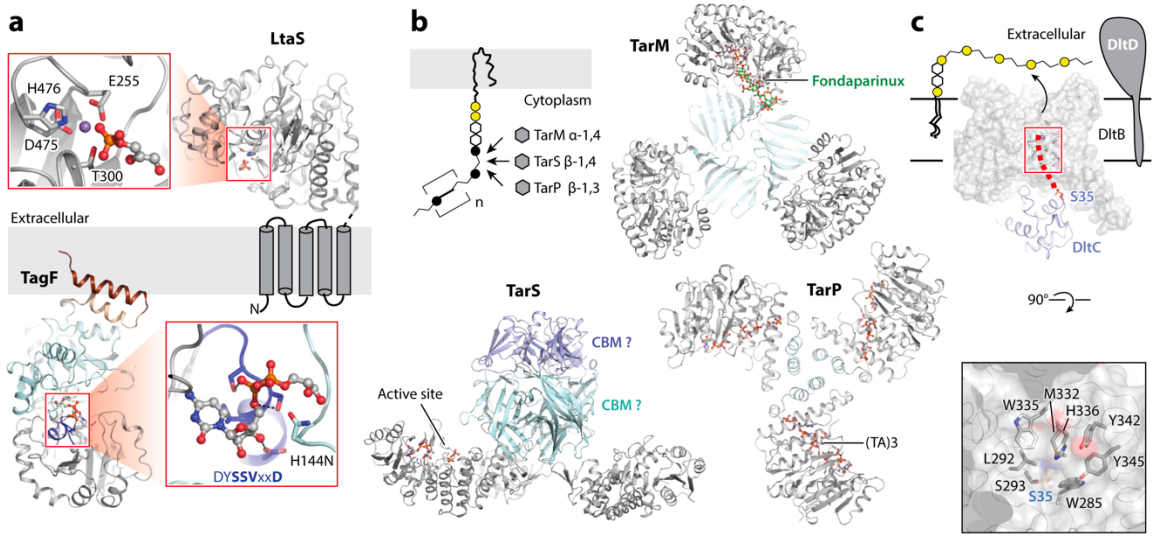


Fig. 1-5 Teichoic acid synthesis and modification. **(a)** LtaS and TagF synthesize wall and lipoteichoic acids on opposing membrane sides (PDB identifiers 2W5S, 3L7K, respectively). The membrane-spanning segment of LtaS is indicated by gray cylinders. **(b)** TarM, TarS, and TarP glycosylate WTA with different specificities. Staphylococcus aureus TarM was crystallized as a trimer (PDB identifier 4X7P) and as a monomeric mutant bound to fondaparinux, α-GlcNAc-glycerol, and UDP (PDB identifier 4X7R). An overlay of both structures is shown. Similarly, trimeric TarS (PDB identifier 5TZ8) was superimposed with a monomeric construct co-crystallized with UDP-GlcNAc (PDB identifier 5TZE). TarP is shown in its trimeric form (PDB identifier 6H4M) bound to UDP-GlcNAc and a ribitol-phosphate trimer (TA)3. **(c)** DltB-mediated d-alanylation of WTA. DltB and DltC are shown as surface and ribbon representations, respectively. DltD is illustrated as a diagram. The putative path of the phosphopantetheine group is indicated by a red dashed line (PDB identifier 6BUH). Abbreviations: CBM, carbohydrate-binding module; PDB, Protein Data Bank; WTA, wall teichoic acid.

binds in an extended conformation with the coordination of the CDP moiety in agreement with other nucleotide-bound GT-B enzymes. The glycerol moiety, however, is thought to be in a nonphysiological conformation where it occupies the acceptor-binding site. In a productive complex, the terminal glycerol moiety of the WTA polymer would occupy this position (i.e., serving as acceptor), which is in close proximity to the predicted base catalyst that deprotonates the acceptor hydroxyl during the S_N2-like transfer reaction (98).

1.5.3 Lipoteichoic acid biosynthesis

LTA polymers are assembled on the extracellular side of the cell membrane by LtaS (**Fig. 1-5a**). The DAG anchor is synthesized intracellularly and reoriented to the extracellular side by LtaA, a MATE-like transporter with similarity to Wzx (99).

LtaS is a membrane-integrated enzyme with a C-terminal extracellular catalytic domain. Recent crystal structures of the LtaS catalytic domains from *B. subtilis* (102) and *S. aureus* (103) provided important insights into substrate recognition and polymer biosynthesis. LtaS adopts a sulfatase-like structure containing an N-terminal α/β core shared with various members of the alkaline phosphatase superfamily (103).

Comparisons of the *S. aureus* enzyme with the structure of phosphoglycerate mutase revealed a conserved Mn^{2+} binding pocket opposing the membrane and formed by T300, D475, H476, and E255. Further, in a glycerol-phosphate–bound structure, the phosphate group is positioned near the Mn^{2+} ion and, importantly, in close proximity to T300. A glycerol-phosphate–bound structure is considered to represent a prehydrolysis state in

which glycerol-phosphate mimics the PG head group (103). Lu *et al*. (103) proposed a reaction mechanism in which the hydroxyl of T300 performs a nucleophilic attack on the PG phosphate group, thereby forming a covalent glycerol-phospho-threonine reaction intermediate. Biochemical data suggest that LTA is extended on its terminal repeat unit farthest from the lipid anchor (104). Thus, the 3′ hydroxyl of the terminal LTA glycerol group would then mediate a similar nucleophilic replacement reaction, thereby transferring the glycerol-phosphate from LtaS to LTA. This model was corroborated by the *B. subtilis* LtaS structure, which revealed additional density extending from the catalytic threonine, consistent with phosphorylation (102).

1.5.4 Export of wall teichoic acid

WTA export follows steps similar to those for O antigen biosynthesis by the ABC transporter–dependent pathway [discussed above and reviewed in (3)]. The WTA ABC transporter TarGH resembles WzmWzt, including a similar TM topology and the presence of a cytosolic GH. Complementation studies in *B. subtilis* showed that the *S. aureus* WTA transporter functionally replaces the native transporter, despite chemical dissimilarity of the WTAs (105). Hence, TarGH likely does not discriminate between different teichoic acid structures yet may recognize the conserved Und-PP-GlcNAc anchor.

1.5.5 Modification of teichoic acids

Following biosynthesis, WTA can be modified to alter its physicochemical properties. Prior to translocation, the polymer is frequently glycosylated with glucose and GlcNAc

moieties [recently reviewed in (3)]. Modifications occurring after membrane translocation include d-alanylation and peptidoglycan attachment.

S. aureus WTA is substantially glycosylated with GlcNAc at the ribitol C4 position, and the TarM and TarS enzymes catalyze the α- and β-transfer to this acceptor, respectively (3) (**Fig. 1-5b**). In 2018, Gerlach *et al*. (106) identified a third prophage-encoded enzyme, termed TarP, which transfers GlcNAc to the ribitol C3 position. Strikingly, TarP-modified WTA elicits a 7- to 40-fold lower IgG response in mice compared with strains containing TarS-modified WTA, and the enzymatic activity suffices to establish methicillin resistance in *S. aureus* (106).

TarM, TarS, and TarP all form trimers as biologically functional units (106-109), yet mutations that render the enzymes monomeric have little effect on their catalytic activities. Therefore, oligomerization may primarily support *in vivo* processivity. Several crystal structures in UDP- and substrate-bound states have been reported for the enzymes, revealing different architectures of their catalytic and trimerization domains. Of note, the C-terminal trimerization domain of TarS contains a putative CBD also found in *Anoxybacillus pullulanase*, suggesting it has a role in WTA interaction (109). TarP has been co-crystallized with a synthetic teichoic acid polymer (106), and TarM with fondaparinux (a heparin pentasaccharide), and both ligands bind to a surface-exposed basic groove (109).

1.5.6 D-alanylation of teichoic acids

Teichoic acid can be further modified by d-alanylation on the cell surface, which, for example, increases resistance to cell wall stress. The attachment of d-alanine units is catalyzed by a membrane-bound O-acyltransferase (MBOAT) (110), an enzyme family found in all kingdoms of life. Similar systems are believed to O-succinylate the exopolysaccharide poly-β-1,6-GlcNAc in Gram-positive bacteria [recently reviewed by Howell and colleagues (111)].

D-alanylation of teichoic acid is mediated by the dltABCD operon (112), of which DltB is an integral membrane protein and DltD is anchored to the cell surface via a single N-terminal TM helix. A d-alanine residue is first attached to the phosphopantetheine group of the carrier protein DltC (catalyzed by DltA) in the cytosol before it is transferred to teichoic acid by DltB. The periplasmic DltD is required for LTA modification *in vivo*, yet its precise function remains to be determined.

Crystal structures of DltB alone and in complex with phosphopantetheine-DltC provided the first mechanistic insights into the transfer reaction (113) (**Fig. 1-5c**). DltB adopts a funnel-shaped ring structure with a central tunnel that is closed by conserved Trp residues. DltC, a soluble cytosolic protein, binds to the funnel and positions the phosphopantetheine close to the tunnel entrance. Although only its phosphate group was resolved in the electron density map, the phosphopantetheine position near the channel entrance suggests that it could traverse the membrane through the DltB pore. Thus, upon association with DltB,

DltC likely exposes its d-alanyl-phosphopantetheine group to the cell surface, where transfer to teichoic acid takes place (113).

1.6 Biosynthesis of capsular polysaccharides

1.6.1 Overview

Capsular polysaccharides (CPS) are major virulence factors produced by many Gram-negative and Gram-positive pathogens. Virulence correlates with the extent and composition of the capsule (114), which aids in evading host innate immune responses (1, 115, 116). For example, CPS reduce complement-mediated killing and phagocytosis by masking cell-surface epitopes and/or mimicking host polysaccharides, such as polysialic acids, heparosan, or hyaluronan (117, 118).

In Gram-negative bacteria, CPS are frequently (but not always) lipid linked at their reducing end. Similar to O antigen and teichoic acid biosyntheses, two mechanistically distinct routes exist for CPS biosynthesis: the Wzx/Wzy-dependent and the ABC transporter–dependent pathways [reviewed in (1)]. Group 1 and 4 CPS are synthesized by the Wzx/Wzy-dependent pathway, whereas group 2 and 3 CPS require the ABC transporter export machinery (1).

1.6.2 Group 1 and 4 capsular polysaccharide biosynthesis

The biosynthesis of group 1 and 4 CPS shares similarities with the Wzx/Wzy-dependent O antigen biosynthesis pathway. The CPS repeat units are synthesized on an Und-PP anchor, reoriented to the periplasmic side, and polymerized by Wzy concomitant to export across

the OM through Wza. If, instead, the repeat units are transferred in the periplasm to the lipid A-core by WaaL (see O antigen discussion above), the resulting glycoconjugate is referred to as K_{LPS}, containing the lipid A-core extended by a K antigen instead of an O antigen (1).

The CPS length produced by Wzy is not random—in fact, most CPS exhibit fairly narrow size distributions (119). Wzy partners with the periplasmic Wzc polysaccharide copolymerase, which contains two TM segments, a large periplasmic region, as well as a cytosolic autokinase domain (**Fig. 1-1**). Autophosphorylation has been detected on several C-terminal Tyr residues, referred to as the C-terminal Y-cluster (120). Wzc is believed to function as an oligomer, similar to Wzz of O antigen biosynthesis (**Fig. 1-3c**).

Negative-stain electron microscopy analyses showed that Wzc forms a four-fold symmetric oligomer that interacts with the OM Wza subunit, thereby forming a large ∼180-Å-long multimeric complex (121). Recent crystallographic studies on a catalytically inactive mutant (K540M) of the Wzc cytosolic autophosphorylation domain reveal a ring-like, octameric assembly. Each subunit interacts extensively with its neighbors, and Tyr715 of the Y-cluster of one protomer points directly into the nucleotide-binding pocket of a neighboring subunit, suitably positioned for phosphorylation (122). Studies on the structurally related Tyr kinase CapB from *S. aureus* in phosphorylated and unphosphorylated states suggest that the ring-shaped octamers disassemble upon autophosphorylation (122). Accordingly, a model was proposed in which the Wzc complex dissociates upon autophosphorylation and reassociates following dephosphorylation by the

Wzb phosphatase. In contrast, Sachdeva *et al*. (116) proposed that Wzc phosphorylation and dephosphorylation primarily affect the assembly of the cytosolic autokinase domains, with perhaps only minor perturbations to the TM and periplasmic domains, aiding in CPS translocation.

During CPS biosynthesis, Wzc interacts with Wza in the OM, thereby coupling chain growth with OM transit (121). Wza is an OM porin (**Fig. 1-6a**), which assembles into a detergent-resistant multimer (123). This barrel comprises a stack of four eight-fold symmetric rings and traverses the OM with amphipathic C-terminal α-helices (124). *In vivo* cross-linking studies in *E. coli* K30 using a site-specifically incorporated, photo-inducible cross-linker demonstrated that the CPS indeed migrates through the Wza pore (125).

1.6.3 Group 2 and 3 capsular polysaccharide biosynthesis

Group 2 and 3 CPS are completely synthesized intracellularly before export. The polymers are attached to a phospholipid anchor and secreted to the cell surface in a single step by an ABC transporter–containing secretion system. Apart from differences in transcriptional regulation, groups 2 and 3 are synthesized in a similar fashion, and we focus on recent developments regarding group 2 K antigens [see (1, 115, 126) for additional information].

1.6.4 Lipid-linked primer

CPS biosynthesis starts with the attachment of 3-deoxy-d-manno-oct-2-ulosonic acid (Kdo) to lyso-PG (127) by a family-99 GT, termed KpsS in *E. coli* (128). This lipid adapter is then extended by 5–9 Kdo units by the bifunctional KpsC, with CMP-β-Kdo serving as

Fig. 1-6

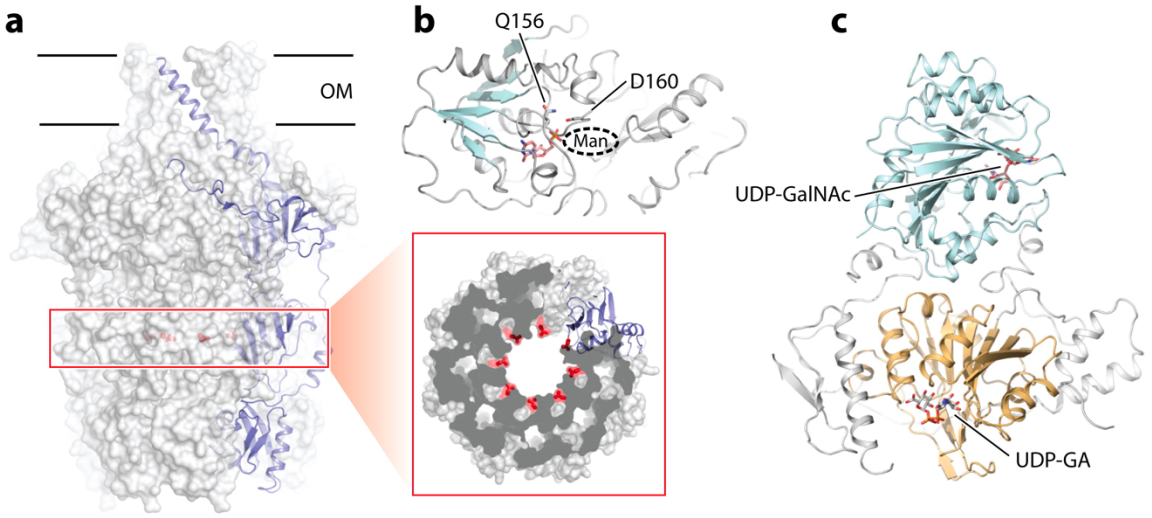


Fig. 1-6 Biosynthesis of capsular polysaccharides. **(a)** Structure of the OM-integrated Wza (PDB identifier 2J58). One subunit of the octameric complex is shown as a blue diagram. The inset shows a slice through the periplasmic Wza barrel viewed from the periplasm, with residues contacting the translocating CPS shown in red. **(b)** Crystal structure of the *Thermosulfurimonas dismutans* KpsC β-Kdo transferase (PDB identifier 6MGB) bound to CMP. The putative binding pocket for the mannosyl (Man) unit is indicated by a dashed oval. **(c)** Chondroitin synthase binds UDP-activated GalNAc and UDP-activated GA via different GT domains (blue and yellow, respectively) (PDB identifiers 2Z87 and 2Z86). Abbreviations: CPS, capsular polysaccharides; GA, glucuronic acid; GT, glycosyltransferase; OM, outer membrane; PDB, Protein Data Bank.

sugar donor (128, 129). Homologous activities are provided by HcsB/HcsA and LipB/LipA in *Haemophilus influenzae* (130) and *Neisseria meningitidis* (131), respectively. In *E. coli*, CMP-β-Kdo is synthesized by KpsU/KdsB (132) and is notoriously unstable *in vitro*, hence biochemical assays usually rely on its *in situ* formation.

KpsC is a family-107–retaining GT with N- and C-terminal catalytic domains. Recently, biochemical and structural analyses on KpsC delineated the order of the enzyme's activities (133). The authors used a fluorescently labeled PG acceptor that was extended by a single β-Kdo unit in a reaction catalyzed by KpsS. This reaction product was then used for β-Kdo transfer reactions using the isolated N- and C-terminal domains of *E. coli* KpsC. The results clearly demonstrated that the KpsC C-terminal domain attaches the second Kdo unit to the C7 position of the Kdo-lipid primer formed by KpsS. Subsequently, the KpsC N-terminal domain attaches multiple Kdo units to the C4 position of the disaccharide, thereby producing a β-Kdo-(2→4)-β-Kdo-(2→7)-β-Kdo lipid anchor (128). The precise number of 2→4 linked Kdo units *in vivo* is strain specific.

The KpsC N-terminal GT domain was co-crystallized with CMP, which identified the substrate-binding pocket (133) (**Fig. 1-6b**). This pocket extends into a deep, narrow tunnel believed to accommodate the Kdo moiety. The nucleotide's phosphate group localizes adjacent to an invariant QXXXD motif and may aid in the GT reaction, as proposed for the WbbB GT-99 β-Kdo transferase (133).

Most characterized group 2 and 3 CPS biosynthesis systems contain similar KpsS and KpsC functions with some modifications (133). A significant exception, however, is found in *S. enterica*, which produces the Vi antigen. KpsS and KpsC homologs have not been found in this species; instead, the Vi antigen contains an N-acetylhexosamine residue at the reducing end that is modified with two β-hydroxyl acyl chains, thereby resembling half of a lipid A molecule (134). Biochemical analyses suggest export of the Vi antigen by an ABC transporter system, which has recently been reviewed (135).

1.6.5 Biosynthesis of the capsular polysaccharide repeat unit structure

Following biosynthesis of the complete Kdo-lipid anchor, nonprocessive, often bifunctional GTs extend this primer with the CPS chain. Biochemical analyses of *Pasteurella multocida* heparosan and hyaluronan synthases (PmHS and PmHAS) or of *E. coli* K4 chondroitin polymerase revealed modest enzymatic activity in the absence of and significantly increased catalysis in the presence of lipid-linked or soluble primers (136-138).

Crystal structures of the *E. coli* K4 chondroitin synthase provided the first insights into the architecture and polymerization mechanism of a bifunctional CPS synthase (138) (**Fig. 1-6c**). Chondroitin comprises alternating β-1,4-linked glucuronic acid (GA) and β-1,3-linked N-acetylgalactosamine units (139), which the enzyme adds nonprocessively to the nonreducing end via two GT-A domains with different substrate specificities (140). Chondroitin synthase most likely releases the nascent polymer product after each

elongation reaction to reposition the new terminal sugar unit for the next transfer reaction, and biochemical analyses of PmHAS and PmHS support this model (136, 137, 141-143).

Although bifunctional GTs are commonly found in CPS operons, in some cases the enzymatic functions are encoded in separate genes. This is the case for the *E. coli* K5 heparosan, synthesized by KfiA (attaching GA) and KfiC (attaching GlcNAc) (144). Additionally, some CPS are glycosylated after biosynthesis, exemplified by the K4 chondroitin CPS, which is decorated with fructose at GA's C3 position by KfoE (145).

1.6.6 Capsular polysaccharide export

After CPS biosynthesis, the polymer must be transported across the IM, the periplasm, and the OM to reach the cell surface. The envelope-spanning CPS secretion machinery may resemble tripartite efflux pump systems, in which the OM TolC porin associates with membrane fusion proteins and IM transporters [recently reviewed in (146)].

CPS ABC transporters are usually heterodimeric transporters and referred to as KpsT and KpsM in *E. coli* for the nucleotide-binding and TM domains, respectively (**Fig. 1-1**). Deletion of either KpsM or KpsT leads to intracellular accumulation of CPS (147, 148).

In 2013, Willis *et al*. (127) identified lyso-PG as the lipid anchor for the polysialic acid CPS of the *E. coli* K1 and *N. meningitidis* group B as well as the heparosan-like capsule of *E. coli* K5. In all cases, the lipid is attached via a β-linkage to Kdo, which could serve as a recognition motif initiating transport. This concept is partially supported by cross-species

complementation assays; for example, the *Pasteurella haemolytica* A1 ABC transporter complements an *E. coli* K1 kpsM/kpsT knockout (149). *P. haemolytica* produces CPS comprising N-acetylmannosaminuronic acid (ManNAcA) β-1,4-linked with N-acetylmannosamine (ManNAc), whereas *E. coli* K1 produces a polysialic acid polymer (149).

CPS ABC transporters share several conserved features with O antigen and teichoic acid transporters, including a similar TM topology and an N-terminal amphipathic helix. However, although O antigen and teichoic acid transporters contain the cytosolic GH that likely recruits the Und-PP–linked polymer, this extension is missing in CPS ABC transporters, perhaps reflecting the utilization of a different glycolipid anchor.

Past the IM, the CPS transporter likely couples to multimeric periplasmic and OM subunits, referred to as KpsE and KpsD in *E. coli*, with KpsD predicted to adopt a Wza-like fold, as recently discussed (66).

An interesting observation from *S. enterica* raises questions as to the single-step CPS secretion model. Liston *et al*. (150) identified a pectate lyase homolog in *Burkholderiales* (VexL) that specifically degrades the O-acetylated Vi antigen. This enzyme, when recombinantly expressed in *S. enterica*, localizes to the periplasm and degrades the Vi antigen, suggesting that the polymer is accessible to periplasmic proteins during transit. Whether this periplasmic exposure is unique to the Vi secretion machinery or is more generally applicable to CPS secretion systems remains to be determined.

1.7 Scope of dissertation research

Complex carbohydrates are employed by prokaryotes and eukaryotes alike in a beautifully complex array of roles throughout the cell. Cell surface polysaccharides such as teichoic acids, capsular polysaccharides and O antigens enable microbial pathogens to evade innate immune responses, resist desiccation, adhere to surfaces, and deter antibiotics (2, 151). In my graduate work, I have focused on how the O antigen helps Gram-negative bacteria to protect themselves from their surrounding environment. Toward this end, I have taken a structure-function approach to study the biosynthesis and transport of the Gram-negative bacterial O antigen. How the ABC transporter WzmWzt translocates the undecaprenyl-linked O antigen polysaccharide across the inner membrane during the initial steps of lipopolysaccharide biosynthesis remains an open question, and my work has sought to further understanding of the WzmWzt transport mechanism. Chapters 2 and 3 will focus on the two aims of my thesis proposal. Chapter 2 presents progress on Aim 1: to determine the crystal structure of ATP-bound WzmWzt for comparison with the nucleotide-free WzmWzt crystal structure previously solved by Yunchen Bi (89). Chapter 3 presents progress on Aim 2: to determine the cryo-EM structure of ATP-bound WzmWzt as a full-length construct to visualize the position of the Wzt CBD relative to the known structure of the WzmWzt transporter.

2 Chapter 2: A lipid gating mechanism for the channel-forming O antigen ABC transporter

Christopher A. Caffalette[1], Robin A. Corey[2], Mark S.P. Sansom[2], Phillip J. Stansfeld[2] and Jochen Zimmer[1]

[1] Molecular Physiology and Biological Physics, University of Virginia School of Medicine, Charlottesville, VA 22908, USA

[2] Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK

Nat Commun 10, 824, doi:10.1038/s41467-019-08646-8 (2019)

In this chapter, I will present an article published in *Nature Communications* presenting the ATP-bound crystal structure of WzmWzt alongside molecular dynamics simulations data to support the hypothesis that phospholipids from the periplasmic leaflet of the bacterial inner membrane may gate (close) the WzmWzt channel, preventing water flux across the inner membrane. This work was initiated in August 2017, and the article was published in February 2019.

2.1 Introduction

An abundant defense mechanism of Gram-negative pathogens against the host innate immune response is the exposure of lipid-linked complex carbohydrates on their cell surfaces (152-154). These sugary coats establish extended barriers around the cell,

reducing the risk of complement-mediated killing, regulating host-pathogen interactions, and protecting against desiccation (155). O antigens represent one class of glycolipid conjugates and form the variable region of the lipopolysaccharide (LPS) molecules in the outer membrane (OM) of Gram-negative bacteria (41). The O antigen polymers extend the conserved LPS core, which comprises lipid-A and the inner and outer core oligosaccharides. O antigens consist of serotype-defining oligosaccharide units of 3–5 sugars that are repeated multiple times to form hypervariable polysaccharides up to ~100 sugars long (41, 156).

O antigens are assembled by two fundamentally different but equally abundant mechanisms. One pathway assembles the polymer from individual undecaprenyl pyrophosphate (UndPP)-linked repeat units on the periplasmic side of the inner membrane (IM) (41). In the ABC transporter-dependent pathway, however, the full-length and also UndPP-linked O antigen is first assembled on the cytosolic side of the IM, and then translocated to the periplasmic side by an ABC transporter for ligation to the LPS core (41, 88). For both pathways, complete LPS molecules are shuttled to the OM via the Lpt pathway, while the UndPP is recycled via BacA (8, 157).

ABC transporter-mediated translocation of UndPP-linked O antigens requires the recognition of the substrate at the IM's cytosolic side; the reorientation of the UndPP moiety and its lateral release into the periplasmic membrane leaflet; and the translocation of the polysaccharide chain across the membrane. In an extended conformation, about 8–10 sugars of a linear polysaccharide suffice to span the average width of a lipid bilayer

(92). Because O antigens can exceed this number many times (87, 158), their membrane translocation requires the formation of a continuous polysaccharide channel and likely a processive translocation mechanism.

The recent structure of the *Aquifex aeolicus* (Aa) WzmWzt O antigen ABC transporter in a nucleotide-free conformation provided the first insights into its translocation mechanism (89). Firstly, in accordance with its function in biopolymer translocation, the transporter forms a continuous transmembrane (TM) channel sufficiently wide to accommodate a polysaccharide chain. Secondly, the transporter's Wzt nucleotide-binding domain (NBD) contains a unique α-helical extension (referred to as the 'gate helix') near the cytosolic water-lipid interface. Because an extended loop is also found in this region of the ABC transporters of Gram-positive bacteria that translocate UndPP-linked teichoic acids (159), it has been speculated that it is required for substrate recognition. Thirdly, the TM channel formed by Wzm is lined with aromatic residues that likely interact with the translocating polysaccharide via CH-π stacking interactions (160).

While the O antigen polymer may prevent conductance of water and small molecules during translocation, how the ABC transporter pushes the polysaccharide into its channel and how this channel closes in a resting state to maintain the membrane's permeability barrier remain important unresolved questions. To address these points, we determined the high-resolution crystal structure of AaWzmWzt in an ATP-bound conformation. The structure reveals the close association of the transporter's NBDs that coordinate two ATP molecules. Compared to the nucleotide-free state, the transporter undergoes large rigid-

body movements of its nucleotide-binding and TM domains. Furthermore, TM helix 1 and a conserved loop at the putative substrate entry site bend towards the channel entrance and likely account for polysaccharide translocation. These conformational rearrangements in the ATP-bound state result in a continuous TM channel with two lateral exits towards the periplasmic membrane leaflet. The channel's periplasmic opening is accessible to detergent molecules that form a hydrophobic plug and prevent solvent molecules from spanning the entire channel. Molecular dynamics (MD) simulations of the membrane-embedded transporter confirm that, in a biological membrane, lipid molecules readily occupy the channel's periplasmic opening after substrate release and suggest their crucial role in transporter gating.

2.2 Results

2.2.1 Structure of the ATP-bound O antigen transporter

To gain mechanistic insights into O antigen polysaccharide translocation, we stabilized and crystallized a catalytically inactive AaWzmWzt ABC transporter in an ATP-bound conformation. ATP hydrolysis was prevented by introducing a single point mutation in the transporter's Walker B motif, replacing Glu167 with glutamine (referred to as AaWzmWztEQ) (161). In addition, the transporter's carbohydrate-binding domain (CBD) at the C-terminus of its NBD was removed to facilitate crystallization, as described previously (89). The CBD extensions are found in a subset of O antigen ABC transporters and are likely involved in recruiting the substrate to the transporter and initiating translocation (162). CBD-truncated ABC transporters are fully functional *in vivo* if the CBD is co-expressed as a separate domain (90).

In its ATP-bound state, the AaWzmWztEQ transporter formed well-diffracting crystals, allowing structure determination at one of the highest reported resolutions for an ABC transporter (**SI, Table 1**). The ATP-bound AaWzmWztEQ structure was determined by molecular replacement at a resolution of 2.05 Å and includes residues 1–256 of Wzm and 2–235 of Wzt. The transporter contains one ATP molecule per active site. Our high-resolution structure confirms the previously determined lower resolution nucleotide-free architecture of AaWzmWzt (89), with the exception that Wzm's N-terminal interface helix (IF helix, amino acids 3–19) was out of register by one residue.

ATP binding stabilizes the closed conformation of the transporter's NBDs (**Fig. 2-1**) (163). Although crystallized in the presence of ATP and magnesium chloride, we observe that the nucleotide's phosphate groups are only coordinated by water molecules and protein. The transporter binds ATP at its NBD using conserved Walker A and B, H-loop, and signature motifs that have been previously described (**Fig. 2-1 and SI, Fig. S2-1**) (163, 164). The signature motif in O antigen and related ABC transporters consists of the sequence Y_{142}-S-S/T-G-M-X-X-R/K-L-A/G-F_{152}, of which Ser143 contacts ATP's γ-phosphate and Met146 packs into a hydrophobic pocket formed by Tyr142 and Phe134 (**SI, Fig. S2-1**).

Within the membrane-spanning segment, the Wzm subunits contact each other only via TM helices 1 and 5, similar to the nucleotide-free state (89). The N-terminal IF helix rests at the interface of the nucleotide-binding and TM domains and is connected to TM helix 1

Fig. 2-1

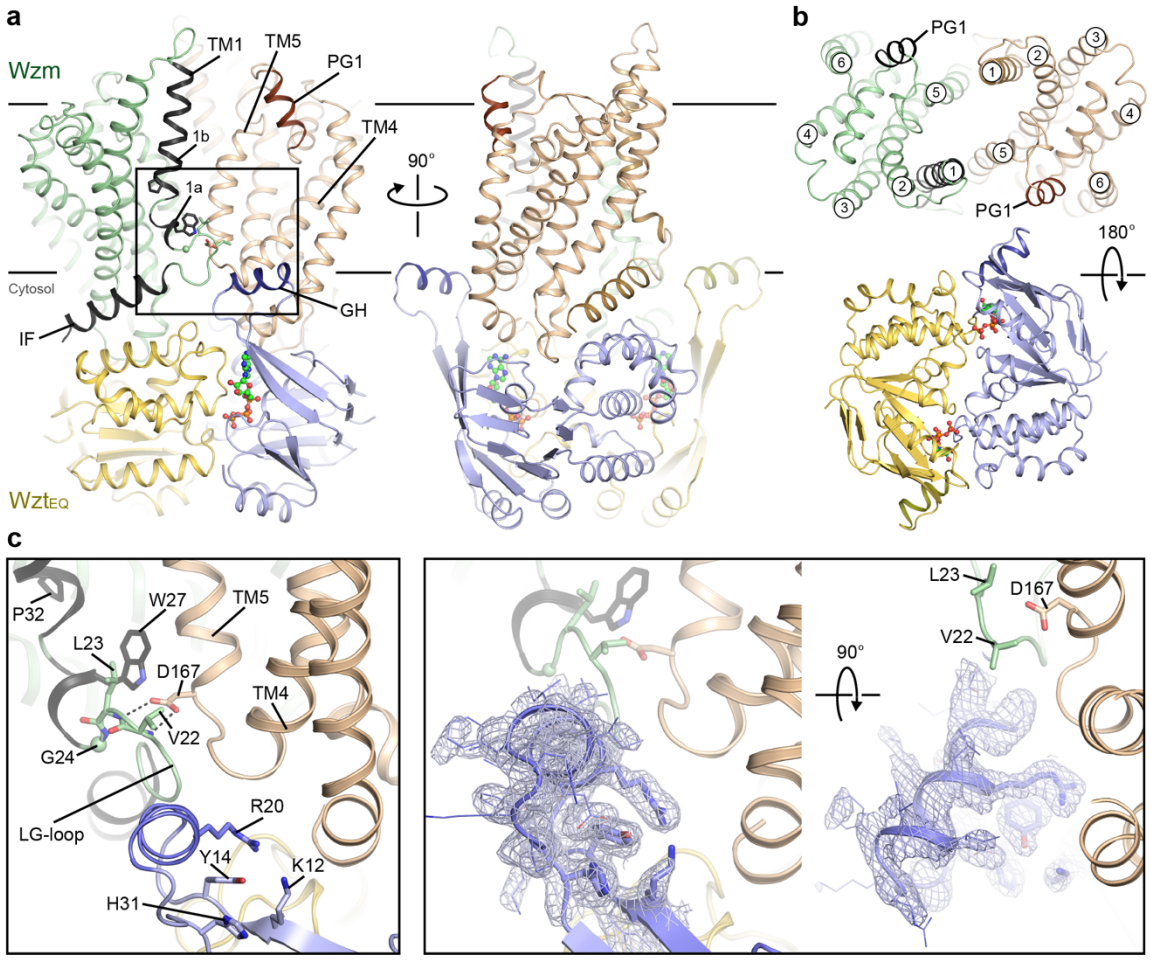


Fig. 2-1 ATP-bound structure of the O antigen ABC transporter. **(a)** Two orthogonal views of AaWzmWztEQ with ATP bound. The Wzm protomers are shown in green and wheat with the N-terminal interface helix (IF) and TM helix 1 highlighted in black. The Wzt nucleotide-binding domains are shown in yellow and blue with bound ATP portrayed as ball-and-sticks. An extended loop of Wzt contains the cytoplasmic gate helix (GH, purple) that sits at the Wzm protomer interface. IF and TM1 are connected by the LG-loop. Lines show approximate lipid bilayer boundaries. **(b)** Periplasmic and cytosolic views of AaWzmWzt. Each Wzm protomer contains six transmembrane helices, and the loop between Wzm TM5/6 forms a short periplasmic gate helix (PG1). **(c)** The gate helix forms an electropositive pocket near the Wzm interface. Left: Interaction of the GH and LG-loop with the Wzm protomers. Right: Two orthogonal views of the GH with a 2Fo-Fc composite omit electron density map contoured at 1σ shown as a gray mesh.

via an extended loop that packs against the connection of TM helices 4 and 5 of the opposing subunit (**Fig. 2-1**). The extended loop (hereafter referred to as 'LG-loop') ends with a conserved L_{23}-G motif and is stabilized through several conserved interactions. These include side chain contacts between Leu23 and Trp27 within the same subunit and backbone contacts of Val22 and Leu23 with the carboxylate of the conserved Asp167 in the opposing subunit (**Fig. 2-1c**). Past the LG-loop and only observed in the ATP-bound conformation, TM helix 1 is divided by the conserved Pro32 into a short 3_{10}-helix (TM1a, residues 25–28) and TM1b (residue 30–47) (**Fig. 2-1**).

The transporter's cytosolic gate helix, formed by an extended connection of β-strands 1 and 2 of the NBD, packs with its N-terminus against the TM4–5 loop of the same half-transporter as well as the LG-loop of the opposing AaWzmWztEQ heterodimer (**Fig. 2-1c**). At this position, the helix forms a positively charged pocket at its interface with the transmembrane domain (TMD) containing the hydroxyl group of the conserved Tyr14 at its center (**Fig. 2-1c**). An electropositive nature of this pocket is conserved among O antigen and teichoic acid transporters and established in AaWzt by the side chains of Lys12, Arg20, and His31 that point straight into it (**Fig. 2-1c**).

It has previously been speculated that the cytosolic gate helix forms the substrate-binding site for recruitment of UndPP-linked O antigens or teichoic acids (89). Indeed, tyrosine residues surrounded by positively charged side chains are frequently found in undecaprenyl-phosphate and UndPP binding pockets, as exemplified by the enzymes aminoarabinose (31), farnesyl (165), and oligosaccharyltransferase (32), among others. In

these cases, the tyrosine's hydroxyl together with arginine and lysine side chains contact the substrate's phosphate group, while the lipid moiety resides in a hydrophobic pocket: a similar coordination is likely in AaWzmWzt (**Fig. 2-1**).

2.2.2 ATP-bound WzmWzt forms a continuous TM channel

In a nucleotide-free conformation, AaWzmWzt forms a TM channel wide enough to accommodate a translocating polysaccharide chain (89). Strikingly, this TM channel is maintained in the ATP-bound conformation, albeit with a slightly smaller diameter (**Fig. 2-2**). Using a 2.5 Å radius probe, a continuous solvent-accessible pore is identified starting at the cytosolic water-lipid interface near the gate helix and ending on the periplasmic side with a large funnel-shaped opening (**Fig. 2-2a**). The periplasmic opening contains lateral exits towards the periplasmic bilayer leaflet formed at the Wzm protomer interface between TM helix 1 and the periplasmic gate helix (PG1) (**Fig. 2-2a**). Of note, a rescue mutation that restores ATPase activity of the related *Staphylococcus aureus* wall teichoic acid transporter in the presence of a small molecule inhibitor localizes to this periplasmic lateral exit (159). This suggests that conformational changes of the TMDs and lateral exit opening are coupled to ATP hydrolysis.

The channel radius is widest within the cytosolic membrane leaflet, constricts to about 2.5 Å at its midpoint, and widens significantly towards the periplasm and laterally towards the periplasmic lipid leaflet. Atomistic MD simulations in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) lipid bilayer confirm that the ATP-bound transporter architecture is very stable with a constant pore profile over the course of a

Fig. 2-2

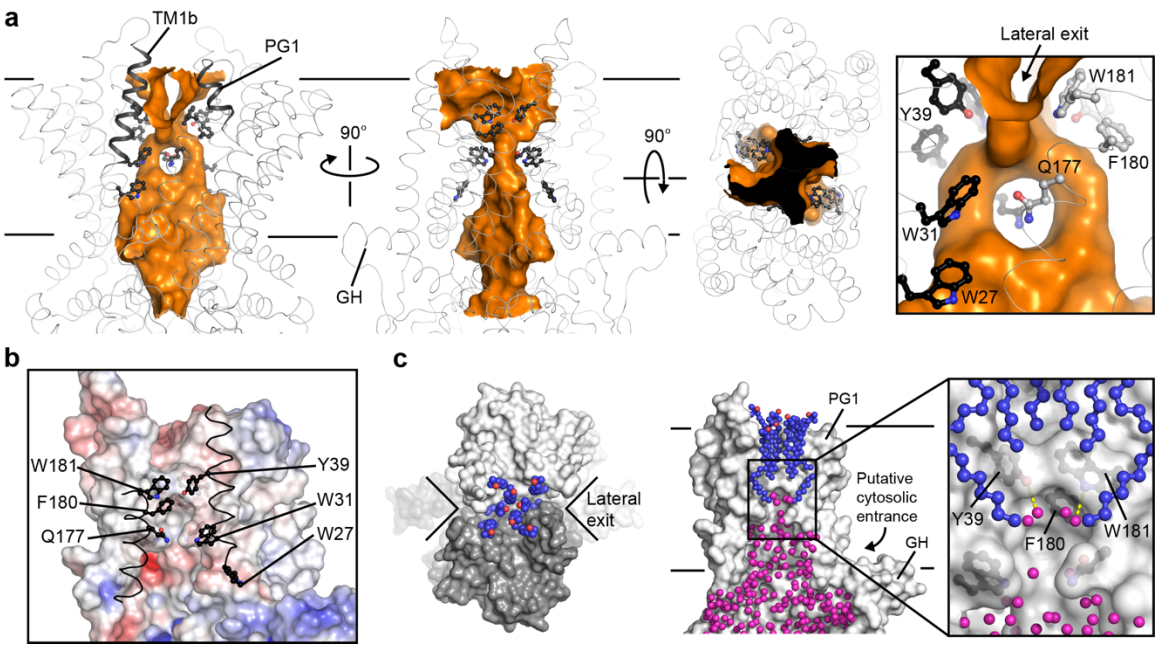


Fig. 2-2 WzmWzt forms a large transmembrane channel in the ATP-bound conformation. **(a)** Surface representation of the AaWzmWztEQ channel calculated with a 2.5 Å-radius probe shown as an orange surface. Right panel, zoomed in view of the channel near the periplasmic lateral exit. Aromatic residues lining the channel are shown as black and gray ball-and-sticks for different Wzm protomers. The transporter backbone is shown as a gray ribbon. **(b)** Electrostatic surface potential of the channel calculated with the APBS plugin in PyMOL (scale −5.0 (red) to +5.0 (blue) kT/e). TM1 and 5 of the opposing Wzm protomer are shown as a black ribbon with selected aromatic residues in ball-and-sticks. **(c)** LDAO molecules occupy the periplasmic opening of the channel. Left: Periplasmic view of AaWzmWztEQ. Right: The AaWzmWztEQ channel is shown with one WzmWzt half transporter removed. Crystallographically resolved water molecules inside the Wzm channel are shown as magenta spheres. AaWzmWztEQ is shown as a gray surface and LDAO molecules as blue and red spheres.

100 ns simulation (**SI, Fig. S2-2**). The constriction halfway across the membrane is formed by Gln177 of TM helix 5, which points towards the channel lumen and interacts with its symmetry mate in the opposing half transporter (**Fig. 2-2a**). A different side chain orientation of Gln177 would eliminate this constriction.

As observed in the nucleotide-free conformation, the Wzm channel has mainly hydrophobic character with Trp27, Trp31, Tyr39, Gln177, Phe180, and Trp181 forming putative contact points with the translocating O antigen. In particular, Tyr39, Phe180, and Trp181 encircle the channel in an 'aromatic belt' forming a hydrophobic ring that may regulate polymer and solvent permeation (see below) (**Fig. 2-2a, b**).

Interestingly, the experimental electron density reveals six complete and four partially ordered detergent molecules, assigned as lauryl-N,N-dimethylamine-N-oxide (LDAO) molecules, within the periplasmic channel opening and lateral exit, respectively (**Fig. 2-2c and SI, Fig. S2-3**). Most of the interactions with the transporter occur through TM helices 1 and 5, which form the seam between the Wzm protomers. In this position, the detergent molecules occupy the entire channel volume, thereby excluding solvent. Indeed, crystallographically resolved water molecules inside the TM channel are only observed within the first (cytosolic) section of the channel up to the aforementioned aromatic belt and are blocked from reaching the periplasm by the detergent molecules (**Fig. 2-2c**).

The transporter's TM channel is sufficiently wide to accommodate a polysaccharide chain. Indeed, MD simulations with a modeled UndPP-linked *E. coli* O9a repeat unit (166) inside

the channel require no significant changes in pore dimensions (**SI, Fig. S2-4**). In this model, the oligosaccharide resides within the channel's periplasmic half, and the undecaprenyl moiety extends into the bilayer through the lateral exit. Channel lining residues, including Tyr39, Phe180, and Trp181 of the aromatic belt pack against the O antigen and/or hydrogen bond with its hydroxyl groups (**SI, Fig. S2-4c**). Combined, the observed channel properties suggest that our structure represents WzmWzt's ATP-bound conformation during polysaccharide translocation, requiring no further channel widening to accommodate a polysaccharide (**SI, Figs. S2-2 and S2-4**).

2.2.3 Lipids and substrate control water flux through the TM channel

Like all membrane transporters, WzmWzt needs to carefully gate its TM channel; the pore should be sufficiently wide to permit substrate passage only, with no space for additional water flux. To probe the effects of substrate binding on the permeation of water through the Wzm channel, we performed a series of atomistic MD simulations on the *E. coli* O9a antigen-bound transporter and analyzed water flux through its pore (**Fig. 2-3**). Both the conformation of the docked O antigen and that of the transporter were stable over 100 ns, with extensive contacts made between them as outlined above (**SI, Fig. S2-4**).

Analysis of water occupancies using VMD and the CHAP package (167-169) reveals a clear solvent-excluded volume in the channel near its aromatic belt in the presence of the substrate, while in its absence, this volume is filled with water molecules (**Fig. 2-3**). Therefore, the translocating substrate appears to prohibit water permeation through the channel.

Fig. 2-3

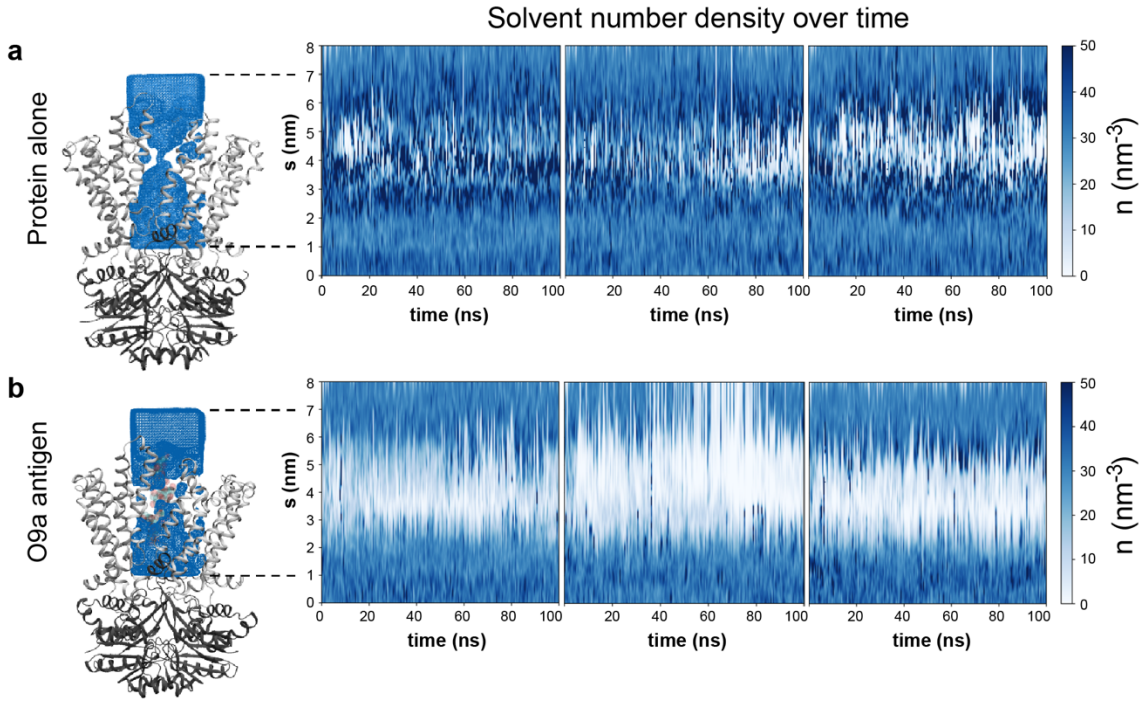


Fig. 2-3 Water flux through WzmWzt is prevented by substrate. **(a)** Snapshot of AaWzmWzt following 100 ns atomistic simulation, showing only the protein (Wzm in light gray, Wzt in dark gray) and a small region of water density around Wzm (blue). A column of water can be seen through the center of Wzm. Right: Quantification of the water solvent density, n, across the 100 ns simulation. The Y-axis, s, represents the position along the channel axis. Data from three repeats are shown. **(b)** Substrate-engaged transporter, with the sugars of the substrate shown as green and red transparent spheres. Right: Quantification of the water solvent density plotted the same as in (a).

This raises the question of how the transporter maintains the membrane's permeability barrier in the absence of a polysaccharide chain. To test this, we ran simulations of the channel in a coarse-grained (CG) representation (170, 171), bound to a hepta-mannose oligosaccharide and embedded in a POPE bilayer (**Fig. 2-4a**). The oligosaccharide was then slowly pulled laterally from the channel over a period of 4 μs. In each of these simulations, before the oligosaccharide leaves the channel, a lipid molecule tightly associates with and enters the transporter through the lateral exit (**Fig. 2-4a and SI, Supplementary Movie 2-1, online content**). As the oligosaccharide leaves, the lipid molecule replaces it inside the channel (**Fig. 2-4a**) such that the pore is constantly occluded. Indeed, MD simulations of the lipid-bound state reveal that water permeation is blocked, comparable to the O antigen inserted state (**Fig. 2-4b**). This suggests the presence of a 'lipid-plug' to seal the channel in the absence of substrate.

2.2.4 ATP-induced movements suggest a stepwise transport mechanism

The introduced Wzt Walker B mutation allowed trapping of the transporter in an ATP-bound state, revealing a closed conformation of its NBDs, as usually observed in nucleotide-bound ABC transporters (**Figs. 2-1 and 2-5 and SI, Fig. S2-1**). Upon ATP binding, each NBD rotates by about 10° and moves by about 9 Å toward the other, thereby closing the ATP binding sites at their interface (**SI, Supplementary Movie 2-2, online content**). Within each NBD, the ATP-induced rotation of the RecA-like subdomain toward the helical region moves the cytoplasmic gate helix by about 8 Å towards the channel entrance (**Fig. 2-5 and SI, Fig. S2-5**) (172, 173). This closure is stabilized by aromatic

stacking interactions of the conserved Tyr11 with ATP's adenine group (**SI, Figs. S2-1 and S2-5a**).

Fig. 2-4

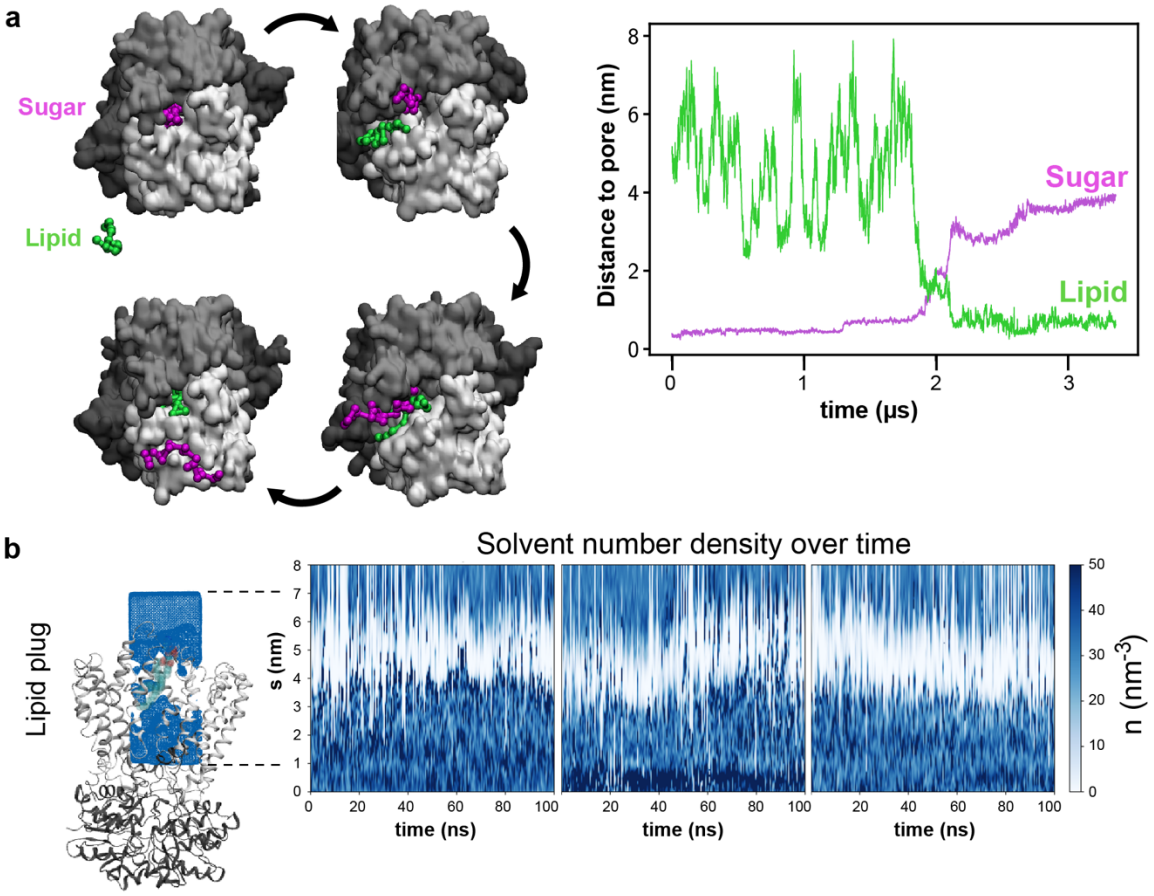


Fig. 2-4 Lipids gate the WzmWzt channel. **(a)** Details of the CG pulling simulations, with snapshots of a representative system viewed from the periplasm. Wzt is shown as black surface, and the two Wzm protomers are shown as light gray and white surfaces. The sugar molecule, which is pulled from the channel, is shown in pink and the replacing POPE molecule in green. The rest of the membrane is omitted for clarity. Right: Quantification of sugar–lipid swap for a representative simulation. **(b)** Post-100ns atomistic simulation snapshot of the lipid-engaged AaWzmWzt transporter and quantification of the water solvent density over the simulation, as per Fig. 2-3. The lipid molecule is shown as green and red transparent spheres.

Fig. 2-5

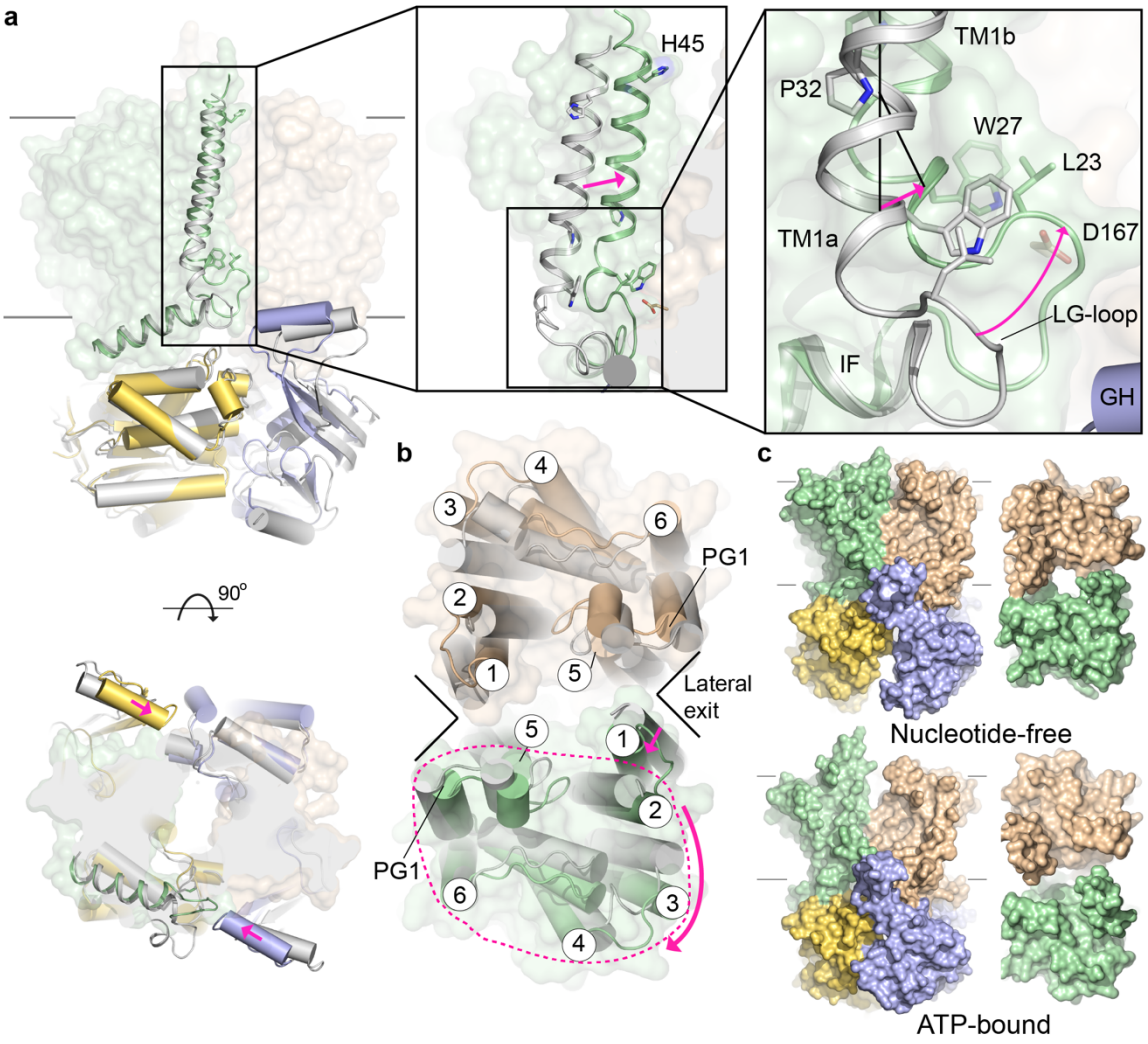


Fig. 2-5 ATP-induced conformational changes of WzmWzt. **(a)** The Wzm and Wzt subunits of the nucleotide-free state (shown in gray) were individually structurally aligned with the corresponding subunits in the ATP-bound conformation. For Wzt, alignment was based on the helical subdomain (residues 82–192) and for Wzm, on TM helices 2–6 by secondary structure matching in Coot. **(b)** Global alignment based on the NBD-dimer of the nucleotide-free (gray) and ATP-bound states. Shown is a periplasmic view of the Wzm protomers. **(c)** Surface representations of the nucleotide-free and ATP-bound WzmWzt conformations. In all panels, Wzt and Wzm are shown in yellow/blue and green/wheat, respectively.

The inward movement of the gate helix is possible due to a similar displacement of the TMDs. The Wzm protomers slide against each other in the plane of the bilayer, while maintaining their interactions between TM helices 1 and 5. Overall, TM helices 2 to 6 move as a rigid body with only minor structural changes compared to the transporter's nucleotide-free state (**Fig. 2-5b and SI, Fig. S2-5b**). TM helix 1, however, shifts significantly by ~10 Å towards the channel axis primarily by sliding along TM helices 2 and 3 (**SI, Fig. S2-5b**). This transition could be induced by the inward movement of the NBD's gate helix and facilitated by the bending of TM helix 1 around Pro32 (see above). During these transitions, the LG-loop remains associated with the invariant Asp167 of TM helix 5 in the opposing half transporter (**Fig. 2-5a**). Because this interaction persists, the LG-loop flips from a position parallel to the water-lipid interface in the absence of nucleotide to perpendicular to it in the ATP-bound conformation (**Fig. 2-5a**).

Combined, the ATP-induced conformational changes push the gate helix, LG-loop, and TM helix 1 towards the periplasm. Relative to a reference point at the channel center near the periplasmic exit, the N terminus of the cytosolic gate helix and the conserved Leu23 of the LG-loop move ~4 Å and 9 Å towards this point, respectively, while His45 at the periplasmic end of TM helix 1 translocates ~2 Å away from it (**Fig. 2-5a and SI, Supplementary Movie 2-2, online content**). Assuming a direct interaction of these motifs with the translocating polysaccharide chain, the observed conformational changes likely account for polymer translocation.

In the full-length AaWzmWzt transporter, the NBD's C terminus is attached to a CBD (**SI, Fig. S2-6a**). CBDs are found in ABC transporters that translocate 'capped' O antigens, which are modified at their non-reducing termini after completion of O antigen biosynthesis (76, 88, 166). The CBDs specifically recognize these modifications. Crystallographic analysis of isolated CBDs from *E. coli* and *R. terrigena* Wzts confirm their jelly-roll like fold and dimeric organization (90, 162), and the isolated AaWzt-CBD dimer interacts with the truncated transporter (89). The interaction surface with the transporter likely also contains the binding site for the O antigen cap (90).

Because the distance between the NBDs' C termini remains between 40–48 Å in the nucleotide-free and ATP bound states (**SI, Fig. S2-6b**), we speculate that the CBDs do not restrict the movements of the NBDs during catalysis, yet conformational changes and functional properties could be modulated upon binding of the O antigen cap.

2.3 Discussion

O antigen biosynthesis is a multistep process that results in the exposure of glycoconjugates on the bacterial cell surface. Biosynthesis relies on either the polymerization of lipid-linked O antigen repeat units in the periplasm or the orchestrated activities of membrane-localized cytosolic glycosyltransferases (41). When fully assembled intracellularly, a high molecular weight polymer has to be transported to the IM's periplasmic side where both biosynthesis pathways converge. Here, the WaaL ligase covalently links the polysaccharide to the LPS core (41).

At a minimum, membrane transport of biological polymers requires substrate recognition on the membrane's cytosolic side, initiation of transport and insertion of the polymer into its transporter, as well as polymer translocation through a TM channel. For lipid conjugated polymers, including O antigens and teichoic acids, the lipid moiety also has to reorient in the membrane during this process, such that the lipid anchor moves from the cytosolic to the periplasmic membrane leaflet.

For O antigens and teichoic acids, the lipid anchor is undecaprenyl-phosphate, which is modified with an acetylated amino-sugar (in most cases GlcNAc), thereby generating an UndPP-GlcNAc anchor (156). Despite the absence of a substrate molecule, the AaWzmWzt architecture and, in particular, the location and physical properties of its cytosolic gate helix strongly argue that the substrate-binding site is located at the interface of the TMD with the gate helix (**Figs. 2-1 and 2-6**). At this location, the substrate would directly interact with the LG-loop at the TMD protomer interface, which in turn is likely to facilitate the insertion of the lipid head group into the transporter.

As substrate insertion occurs at the interface of the TMDs, with TM helices 1 and 5 forming the contacts (**Fig. 2-1**), the lipid anchor could reorient with its undecaprenyl moiety remaining in the bilayer and its charged head group in the transporter proper, as proposed for 'flipping' of UndPP-linked oligosaccharides by PglK (27). Flexibility of the side chain packing at this interface likely forms a 'soft seam' to allow the passage of the undecaprenyl chain. Further, the distance over which the polyprenyl group has to be 'dragged' through the channel wall is significantly shortened by the channel's lateral exit within the

periplasmic bilayer leaflet (**Fig. 2-6**). The negative membrane potential likely contributes to the reorientation of the negatively charged UndPP head-group to the periplasmic side, while also preventing its backsliding, from the periplasmic to the cytosolic side.

Fig. 2-6

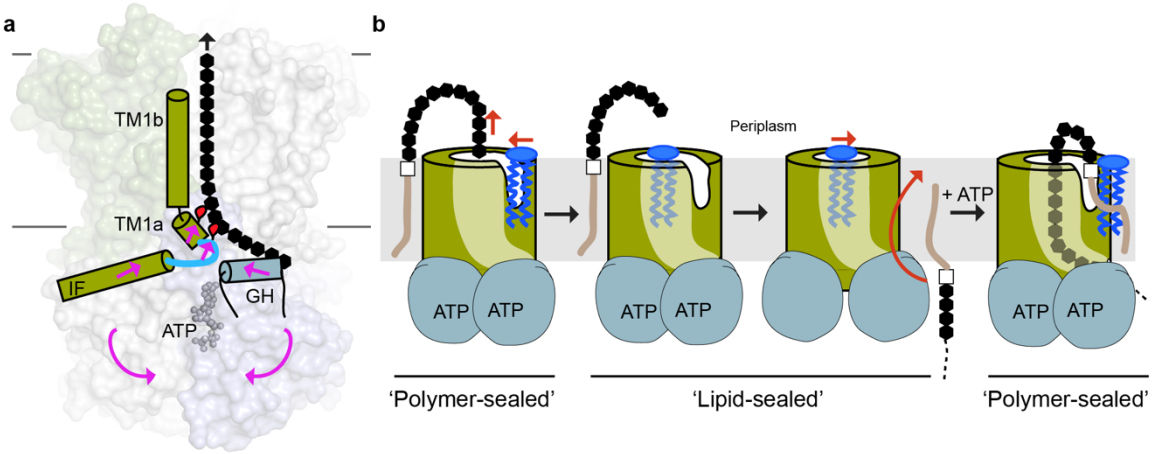


Fig. 2-6 Model of O antigen membrane translocation. **(a)** Upon ATP binding and NBD closure, the cytosolic gate helix (GH) pushes against the LG loop connecting Wzm's IF and TM1. This movement correlates with the bending of TM1 and upward movement of TM1a. Conserved hydrophobic residues (Leu23 and Trp27, red ovals) interact with and push the polysaccharide chain into the transmembrane channel. **(b)** Membrane lipids gate the WzmWzt channel. Concomitant with the final translocation step that releases the polysaccharide into the periplasm, lipid molecules diffuse into the channel to form a hydrophobic plug. This plug is replaced by the substrate upon insertion of a new lipid-linked O antigen into the channel. The substrate's lipid moiety is released into the periplasmic membrane leaflet through the lateral exit, which also allows lipid diffusion as a gating mechanism. The substrate is shown as a brown line connected to a white square for UndPP and black hexagons for the O antigen.

The insertion of the substrate into the channel near the cytosolic gate helix would place the polysaccharide chain in direct contact with the LG-loop and the following TM helix 1a. The ATP-induced flipping motion of this loop and bending of TM1a are likely important transitions that mediate polysaccharide translocation. Both, the conserved Leu23 and Trp27 within these motifs could translocate the polysaccharide based on steric interactions.

Similarly, polypeptide translocation by hexameric AAA-ATPases, such as ClpX, Vps4, and Hsp104, also relies on steric interferences between bulky (usually aromatic) residues of a moveable loop and the translocation substrate (174-176). In comparison, translocation through cellulose synthase is mediated by the movement of a short α-helix that interacts with the polysaccharide, thereby transporting the polymer one sugar unit at a time (177). Our data suggest that O antigen ABC transporters use a similar translocation mechanism, combining a loop movement and helical bending motion (**Fig. 2-6a**). The distance over which the LG-loop and TM1a move are consistent with a processive translocation of 1–2 sugar units per ATP-induced conformational transition; similar step-sizes have been observed for cellulose synthase and ClpX (174, 177).

While our data provide a plausible mechanism for the stepwise forward translocation of the polysaccharide, how the transporter resets between translocation steps without retro-translocating the substrate remains unknown. Possible mechanisms could rely on kinetic differences of active translocation versus Brownian diffusion and/or the presence of a unidirectional ratchet based on steric interactions. Yet, these detailed mechanistic analyzes must await structural information on substrate-bound WzmWzt states.

ATP-energized processive polymer translocation requires the opening and closing of the transporter's NBDs while the polymer resides in its TM channel, i.e. NBD closure should not collapse the TM channel. Our MD simulation data demonstrate that the translocating polysaccharide prevents formation of a proton wire, crucial for maintaining the membrane's proton motive force during translocation. Yet, this permeability barrier has to be maintained in the absence of a translocating substrate. Instead of adopting a channel-free closed conformation, WzmWzt likely accomplishes this task by allowing lipid molecules to occupy its periplasmic channel opening in the absence of a substrate, such that membrane lipids, instead of protein domains, gate the TM channel (**Fig. 2-6b**).

Our MD simulation data suggest that a single POPE molecule suffices to gate the Wzm channel, although the periplasmic funnel is wide enough to accommodate multiple LDAO molecules. It is likely that the extended lipid acyl chains together with Wzm's aromatic belt form a sufficiently hydrophobic seal to prevent water permeation. This is supported by the observation that, in our crystal structure, solvent does not permeate past the most deeply inserted two detergent molecules near the channel's aromatic belt.

A lipid gating mechanism is possible because their diffusion likely exceeds the O antigen translocation rate by several orders of magnitude, ensuring that the transporter's channel is always occupied by either lipid molecules or the translocation substrate, but not solvent. Similarly, upon initiation of O antigen translocation, the UndPP head group could displace the lipid plug. Furthermore, as the lateral exit only opens in the ATP-bound state, lipid

release into the bilayer can only occur during or after the ATP-induced power stroke. Thus, on substrate binding, the UndPP moiety likely reorients while displacing the lipid plug, and the O antigen inserts into the channel for processive translocation and prevention of water permeation.

Other microbial glycolipids implicated in pathogen survival are capsular polysaccharides, of which groups 2 and 3 are also secreted by ABC transporters (1). Here, the transporter partners with periplasmic and OM components to form a continuous conduit across the cell envelope, similar to Type-1 secretion and multi-drug extrusion systems (178-180). How these transporters select their phospholipid-linked substrates (127) and maintain the membrane's permeability barrier in a resting state are important unresolved questions.

Type-2 ABC exporters frequently reorient hydrophobic molecules within lipid bilayers (181-183). The observed ATP-induced conformational changes of WzmWzt, in particular of its TMD, may thus represent a general mechanism by which these ubiquitous transporters operate.

2.4 Methods

2.4.1 Cloning, protein expression, and purification

The AaWzmWztEQ mutant used for crystallization was generated using the QuikChange method on an existing pETDuet wild type AaWzmWzt expression plasmid. Expression and purification of the mutant construct was carried out as previously described with certain exceptions as noted below (89). AaWzmWztEQ was overexpressed in *E. coli*

C43(DE3) cells grown at 37 °C in lysogeny broth (LB) medium containing 100 μg mL−1 ampicillin. Upon reaching an optical density at 600 nm (OD600) of 0.6, expression was induced by the addition of 0.5 mM isopropyl-β-d-thiogalactoside (IPTG) and cells were incubated at 37 °C for 4 h until harvest by centrifugation.

Cell pellets were resuspended in RB buffer containing 20 mM Tris HCl pH 7.5, 100 mM NaCl, and 5 mM β-mercaptoethanol (β-ME) and lysed in a microfluidizer. The membrane fraction was pelleted by ultracentrifugation at 200,000 × g for 60 min in a Beckman Coulter Type 45 Ti rotor. Membrane pellets were then solubilized for 60 min at 4 °C in SB buffer containing 50 mM sodium phosphate pH 7.2, 100 mM NaCl, 20 mM imidazole, 5 mM β-ME, and 2% (w/v) polyoxyethylene(8)-dodecyl ether (C12E8, Anatrace). Remaining insoluble material was cleared by ultracentrifugation at 200,000 × g for 30 min in a Beckman Coulter Type 45 Ti rotor, and the supernatant was batch incubated with Ni-NTA agarose resin (ThermoFisher Scientific) for 60 min at 4 °C. Next, the Ni-NTA resin was packed into a gravity flow chromatography column and washed successively with three buffers: 50 mL WB1 buffer (RB buffer containing 22 mM imidazole and 5 mM LDAO; 50 mL WB2 buffer (RB buffer containing 40 mM imidazole and 5 mM LDAO); and 50 mL WB3 buffer (RB buffer containing 1.5 M NaCl, 22 mM imidazole and 5 mM LDAO). After washing, AaWzmWztEQ was eluted in 35 mL EB buffer (RB containing 300 mM imidazole and 5 mM LDAO).

The eluate was concentrated to 10 mL using 50 kDa cut-off centrifugal filters (Amicon) and dialyzed overnight against a buffer containing 20 mM Tris pH 7.5, 100 mM NaCl,

5 mM magnesium chloride, 5 mM β-ME, and 5 mM LDAO. The following day, the dialysed protein was purified over a S200 gel filtration column (GE Healthcare) equilibrated with GF buffer containing 20 mM Tris HCl pH 7.5, 100 mM NaCl, 5 mM magnesium chloride, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP), and 5 mM LDAO. The main peak fraction was concentrated to a 15 mg mL−1 final concentration using a 50-kDa cut-off centrifugal filter (Amicon).

2.4.2 Crystallization

Prior to crystallization, the AaWzmWztEQ transporter was combined 1:1 with GF buffer containing 4 mM ATP to achieve final concentrations of 7.5 mg mL−1, 2 mM and 5 mM for AaWzmWztEQ, ATP and magnesium, respectively. Protein samples containing ATP and magnesium were incubated on ice for a minimum of 1 h before setting up crystal trays. Crystals were grown by sitting-drop vapor diffusion at 17 °C after combining 1 µL of well solution (21–25% v/v PEG 200, 0.02 M sodium acetate pH 4.0, 0.05–0.2 M ammonium sulfate, and 0.02 M sodium chloride) and 1 µL of protein solution. Crystals grew to their final size within 4 days and were harvested at 7 days. The crystals were cryoprotected by incremental addition of glycerol to the drop to a final concentration of 20%. Crystals were then harvested and flash cooled in liquid nitrogen prior to data collection.

2.4.3 Structure determination

Diffraction data for crystals belonging to space group $P6_522$ were collected at the AMX beamline at Brookhaven National Laboratories (NSLS-II). Data were integrated in XDS (184) and reduced in Aimless using the CCP4 program suite (185) yielding a 2.05 Å dataset

with one half-transporter in the asymmetric unit (**SI, Table 1**). The $CC_{1/2}$ metric reported in **SI, Table 1** is defined as the correlation between intensities from random half-datasets (186).

The AaWzmWztEQ structure was solved by molecular replacement using the transmembrane (Wzm) and nucleotide-binding domains (Wzt) of the AaWzmWzt nucleotide-free structure (PDB 6AN7) as search model ensembles with the program Phaser (187). An initial round of rigid-body refinement was performed in Refmac5 as part of the CCP4 program suite. All subsequent refinement was performed with Phenix.refine (188) using TLS parameters (189), and the model was manually completed using Coot (190). Although strong densities for ATP and LDAO molecules were observed early in refinement, these ligands were modeled at late stages of refinement. The structure was refined to R_{work}/R_{free} values of 16.4 and 18.2% with 98.0 and 0.2% of residues in the favored and disallowed regions of the Ramachandran diagram, respectively. Coordinates and structure factors for AaWzmWztEQ have been deposited in the Protein Data Bank (PDB). Figures were generated using PyMOL [The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC] and Wzm channel dimensions were analyzed using HOLLOW (191).

2.4.4 MD simulations

2.4.4.1 Docking substrate

A model substrate was created comprising a single repeat of the O9a antigen attached to UndPP (molecule was Me-P-3-αMan-(1–2)-αMan-(1–2)-αMan-(1–3)-αMan-(1–3)-αMan-

(1–3)-αMan-(1–3)-αMan-(1–3)-βGlcNAc-UndPP). This model substrate was docked into the ATP-bound AaWzmWztEQ structure using AutoDock Vina (192), with the best scoring pose favoring a periplasmic positioning of the UndPP pyrophosphate chosen (**SI, Fig. S2-4**).

2.4.4.2 Atomistic MD simulation

All simulations were run using GROMACS 5.1.2 (193). Simulations were built using the AaWzmWztEQ ATP-bound coordinates, and described using the Charmm36 force field (194). Where used, all atom O9a antigen parameters were generated using CHARMM-GUI (195). The protein-membrane structures were built into POPE membranes with explicit TIP3P water and sodium and chloride ions to a neutral charge and concentration of 0.15 M. The systems were energy minimized using the steepest descents method over 5000 steps, then equilibrated with positional restraints on heavy atoms for 1 ns in the NPT ensemble at 310 K with the V-rescale thermostat and semi-isotropic Parrinello-Rahman pressure coupling (196, 197). Production simulations were run without positional restraints with 2 fs time steps over 100 ns. Systems were analyzed for pore size and water flux using the CHAP package (168, 198).

2.4.4.3 Coarse-grained simulation

Using the three independent post-100 ns substrate-bound simulations as a starting point, coarse-grained (CG) simulations were run using Martini 2.2 (170, 171) of the AaWzmWztEQ complex in a POPE membrane bound to a hepta-mannose sugar (with the methyl phosphate, β-GlcNAc and UndPP removed). These systems were built using the

MemProtMD protocol (199). Briefly, the protein-sugar complex was placed in a simulation box with POPE lipids orientated randomly around it. In addition to the bonds implicit in the Martini force field, elastic bonds of 1000 kJ mol^{-1} nm^{-2} were applied between protein beads within 1 nm. The systems were then solvated with Martini water and ions to a neutral charge and 0.15 M, before 100 ns of simulation with 20 fs time steps, to allow the POPE bilayer to form around the protein. The substrate was then pulled from the pore, where the sugar-protein distance was increased in the xy directions at a rate of 1 nm per μs. Lipid binding was observed in all three repeats, and binding was quantified by following the distance between the specific lipid and the center of geometry of the residues labelled in **SI, Fig. S2-4c**. Post 4 μs snapshots were converted back to an atomistic description (200) and simulated as described above.

2.5 Data availability

Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under accession code 6M96.

2.6 Acknowledgements

We thank Yunchen Bi and Finn Maloney for comments on the manuscript and the staff at AMX beamline of the National Synchrotron Light Source II, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under Contract No. DE-SC0012704. C.C. was supported by the Cell and Molecular Biology NIH training grant NIH-5T32GM008136 and J.Z. is supported by NIH grant 1R01GM129666. R.C. M.S. and P.S. were funded by Wellcome

(208361/Z/17/Z) and the BBSRC (BB/P01948X/1, BB/R002517/1, and BB/S003339/1). Atomistic simulations were carried out on the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk), provided by HECBioSim, the UK High End Computing Consortium for Biomolecular Simulation (hecbiosim.ac.uk), which is supported by the EPSRC (EP/L000253/1).

2.7 Supplementary Information

Table 1 Data collection and refinement statistics (molecular replacement)

	AaWzmWzt$_{EQ}$
Data collection	
Space group	$P6_522$
Cell dimensions	
a, *b*, *c* (Å)	144.75, 144.75, 201.47
α, β, γ (°)	90, 90, 120
Resolution (Å)[a]	25-2.05 (2.09-2.05)[b]
R_{pim}	0.030 (0.495)
$CC_{1/2}$[c]	0.998 (0.582)
$I / \sigma I$	14.6 (1.7)
Completeness (%)	99.9 (99.9)
Redundancy	12.1 (9.8)
Refinement	
Resolution (Å)	25-2.05
No. reflections	78248
R_{work} / R_{free} (%)	16.4/18.2
No. atoms	
Protein	4230
ATP	31
LDAO	256
All other non-protein/solvent	94
Water	371
B-factors (Å^2)	
Protein	49.1
ATP	29.2
LDAO	81.1
All other non-protein/solvent	77.6
Water	64.2
R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	0.970

[a]Data collected on a single crystal

[b]Values in parentheses are for highest-resolution shell

[c]Correlation between intensities from random half-data sets (186)

Fig. S2-1

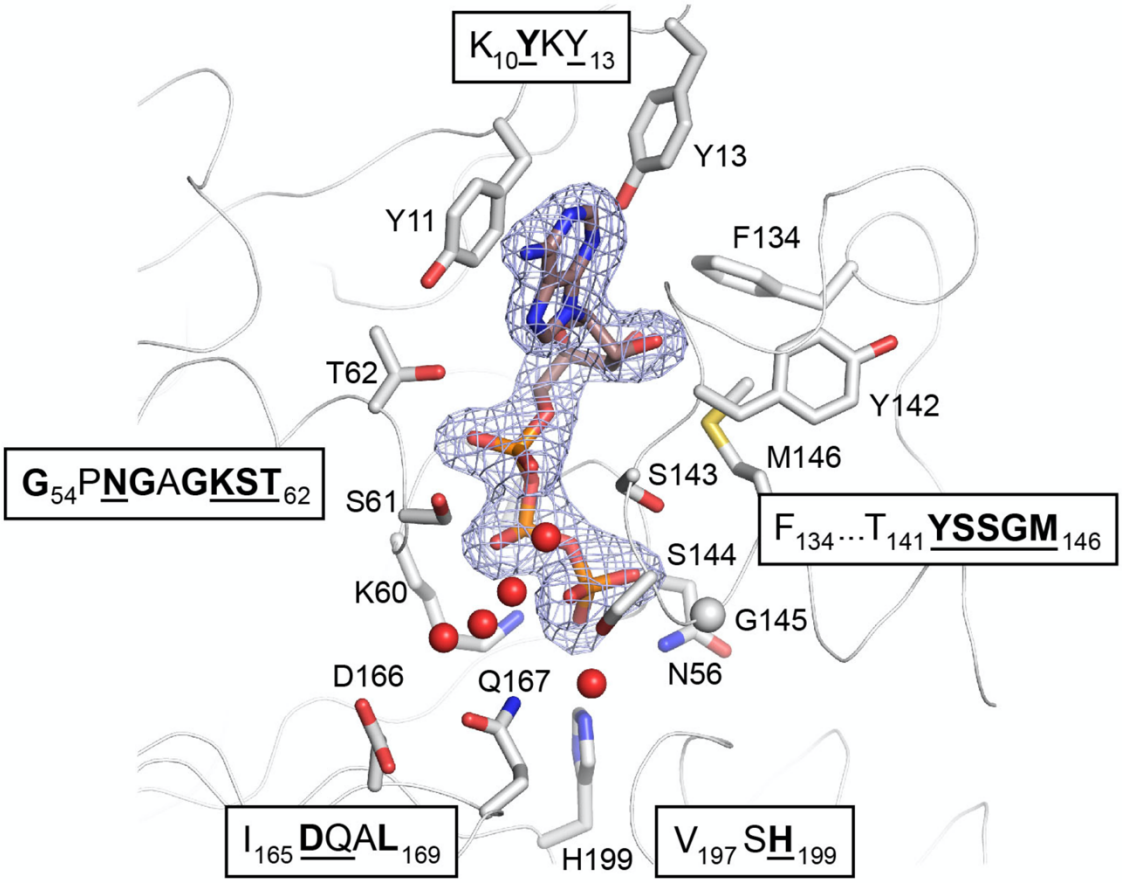


Fig. S2-1 Coordination of ATP at the Wzt dimer interface. The Wzt NBDs are shown as a ribbon with side chains and water molecules contacting ATP shown as sticks and red spheres, respectively. ATP-coordinating sequence motifs are shown as a text box with conserved residues in bold, residues shown in the image are underlined. ATP is shown as sticks together with an unbiased SigmaA-weighted Fo-Fc difference electron density contoured at 4.5σ (blue mesh).

Fig. S2-2

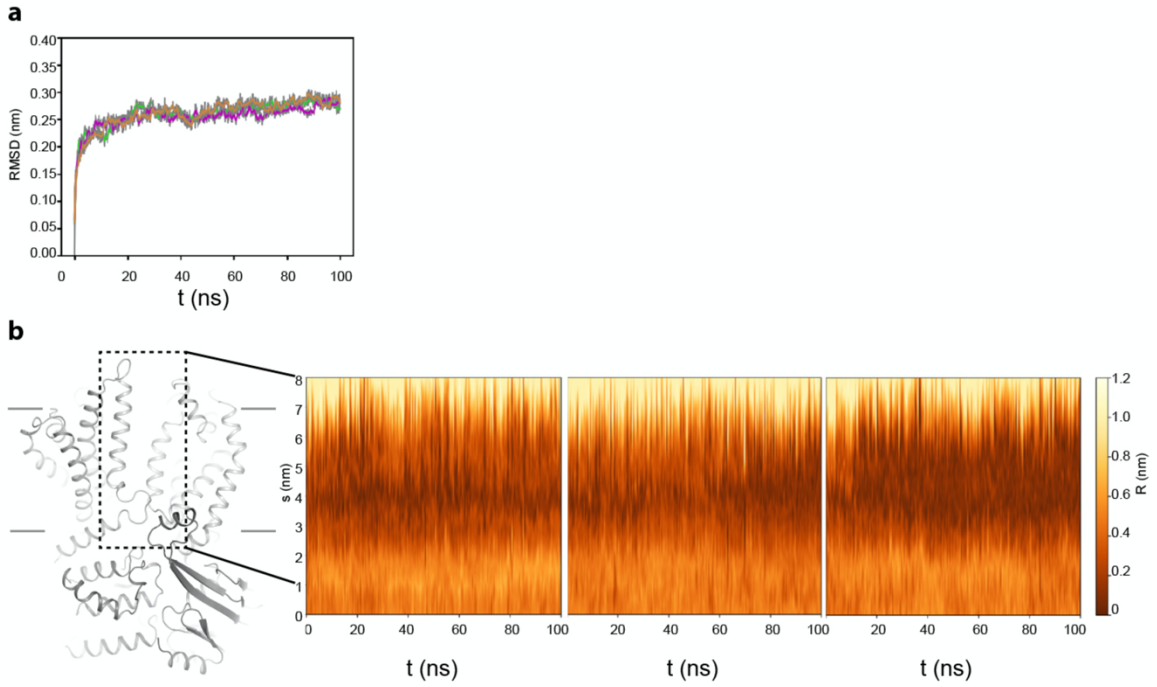


Fig. S2-2 Channel dimensions of the ATP-bound WzmWzt transporter. **(a)** RMSD of protein C-alpha carbons from three atomistic 100 ns molecular dynamics simulations of ATP-bound AaWzmWztEQ in a POPE lipid bilayer. **(b)** Channel dimensions over each of the 100 ns simulations, as measured using the CHAP package. The Y-axis, s, represents the position along the channel axis as indicated by the dashed box. R: Channel radius at position s.

Fig. S2-3

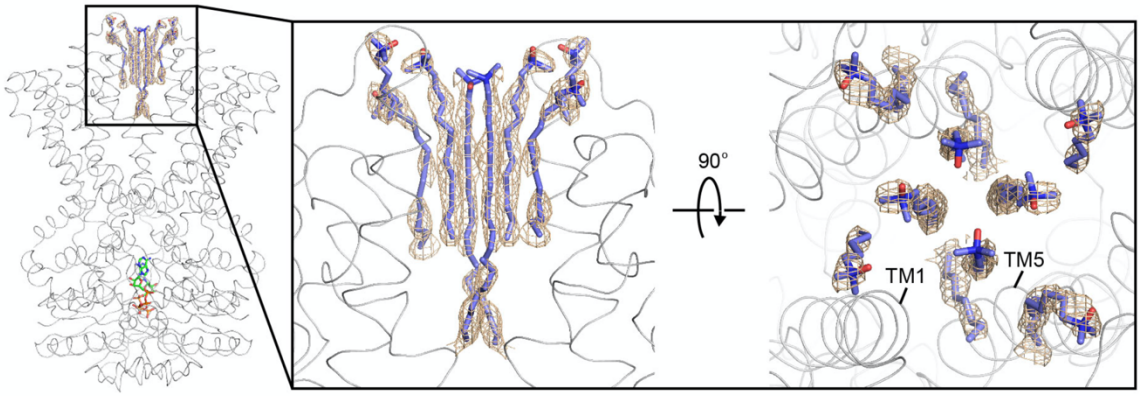


Fig. S2-3 Detergent molecules plug the periplasmic channel exit. Unbiased SigmaA-weighted Fo-Fc difference electron density contoured at 2.5σ for six full-length and two partial LDAO molecules within the periplasmic channel opening (wheat mesh). LDAO molecules are shown as red and blue sticks, ATP is shown as colored sticks, and WzmWzt is shown as a ribbon.

Fig. S2-4

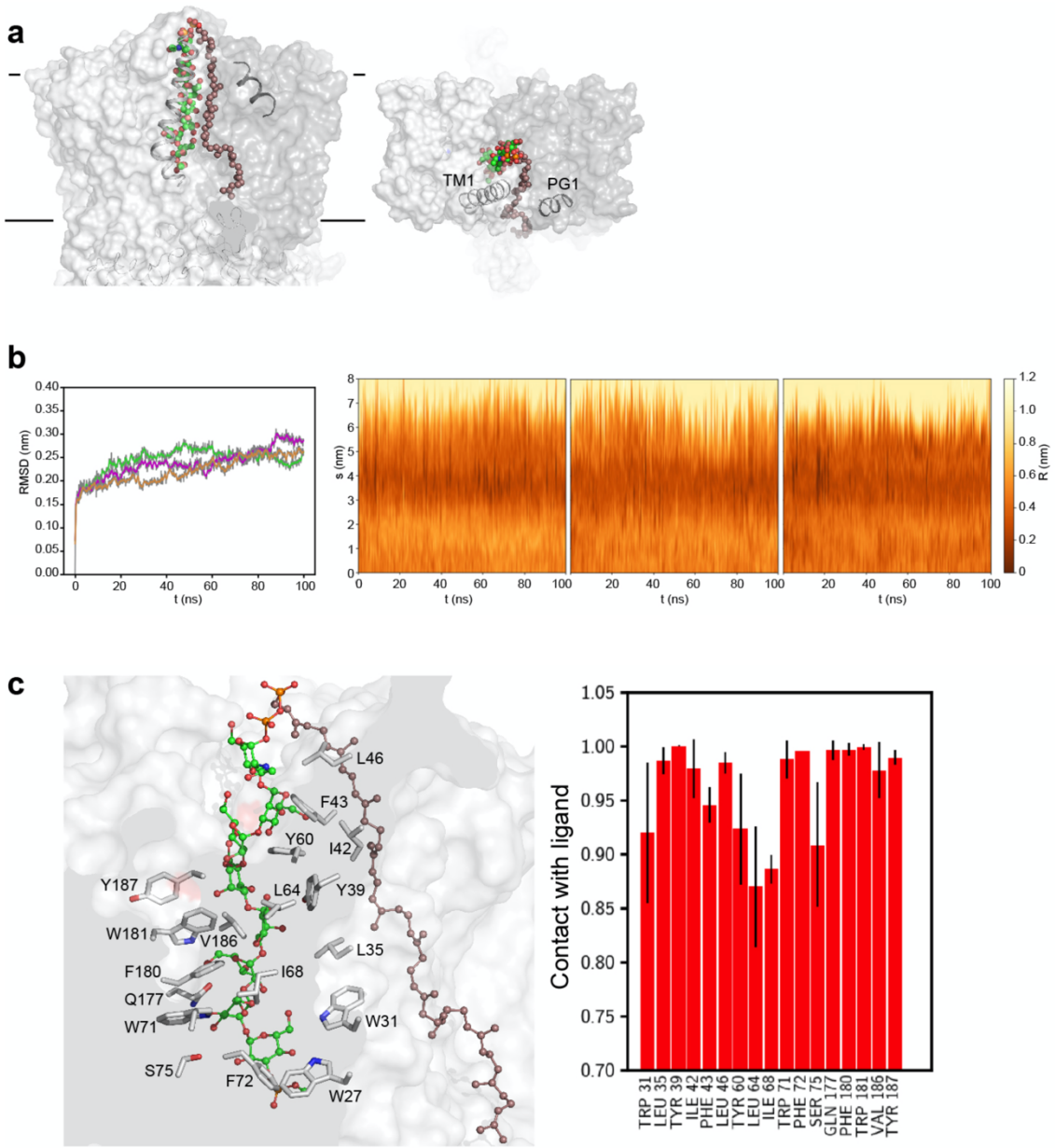

a
TM1
PG1
b
RMSD (nm)
t (ns)
s (nm)
R (nm)
c
L46
F43
Y60
I42
Y187
L64
Y39
W181
V186
L35
F180
I68
Q177
W71
W31
S75
F72
W27
Contact with ligand
TRP 31
LEU 35
TYR 39
ILE 42
PHE 43
LEU 46
TYR 60
LEU 64
ILE 68
TRP 71
PHE 72
SER 75
GLN 177
PHE 180
TRP 181
VAL 186
TYR 187

Fig. S2-4 MD analysis of substrate-bound WzmWzt. **(a)** Side and periplasmic views of the modeled AaWzmWztEQ transporter bound to an *E. coli* O9a antigen repeat unit with a methylphosphate derivatized non-reducing end after a 100 ns all atom MD simulation. The oligosaccharide is shown as spheres in green for carbon atoms. The undecaprenyl moiety is shown as violet beads. TM helix 1 and the periplasmic gate helix (PG1) are shown as cartoons. **(b)** RMSD and pore dimensions of the O9a-bound Wzm channel in three simulations, as per SI, Fig. S2-2. **(c)** Left: Coordination of the O9a antigen inside the Wzm channel. Wzm is shown as a clipped surface with selected oligosaccharide- coordinating residues shown as sticks. Residues are labeled for one protomer only. Right: Contact analysis representing interactions between channel-lining residues and the O9a oligosaccharide during the 100 ns simulations for three simulations. Contact is defined as inter-residue distances of less than 4.5 Å and is expressed as the mean normalized % interaction with error bars showing one standard deviation.

Fig. S2-5

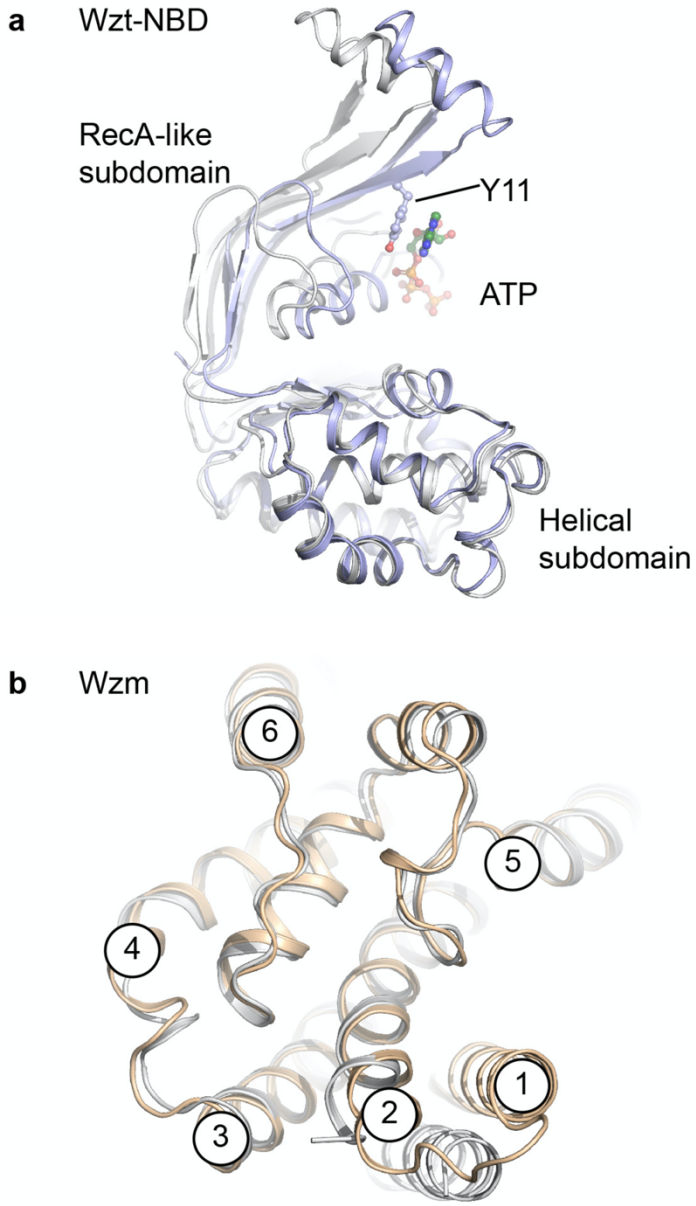


Fig. S2-5 Individual comparison of the Wzt and Wzm structures in the ATP-bound and nucleotide-free states. The domains were individually aligned based on the helical subdomain (Wzt) **(a)** or TM helices 2-6 (Wzm) **(b)** by Secondary Structure Matching (SSM) in Coot and visualized as cartoon representations. The nucleotide-free state is colored gray and ATP is shown as 'ball-and-sticks'.

Fig. S2-6

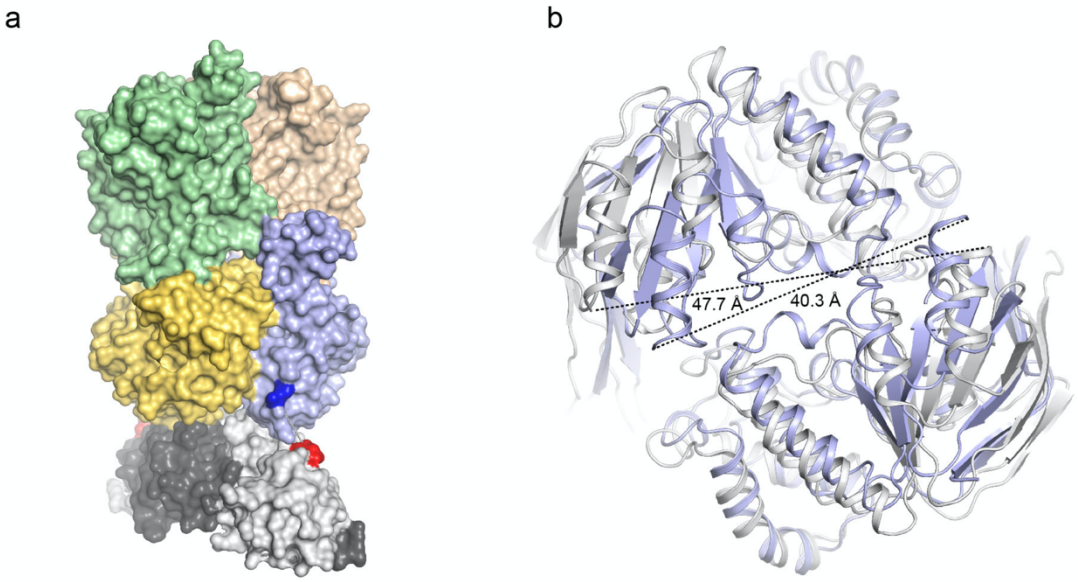


Fig. S2-6 Nucleotide-induced movement of the Wzt-NBD C termini. **(a)** Localization of the C-terminal carbohydrate-binding domains (CBDs) relative to the truncated AaWzmWzt transporter. The CBD was manually docked to form the WzmWzt-CBD complex using CBD coordinates PDB:5HNO. The CBD dimer is shown in light and dark gray. The Wzt C-terminal residue is highlighted in blue, and the CBD dimer N-terminal residues are highlighted in red. **(b)** Cytosolic view of the ATP-bound and nucleotide-free AaWzmWzt transporter structures shown in blue and gray, respectively.

Supplementary Movie 2-1

https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-019-08646-8/MediaObjects/41467_2019_8646_MOESM4_ESM.mpg

Movie showing the replacement of the bound oligosaccharide with the lipid plug.
Shown are frames between 1.8 and 2.4 μs, where each frame is 50 ns. Colors and representations as per Fig. 4a. The remaining lipids and solvent molecules are removed for clarity.

Supplementary Movie 2-2

https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-019-08646-8/MediaObjects/41467_2019_8646_MOESM5_ESM.mov

Nucleotide induced conformational changes of AaWzmWztEQ. Shown is a morph from the nucleotide-free to ATP-bound conformation of AaWzmWztEQ in different orientations. Each view of the transporter shows two open to ATP-bound cycles. The transporter is shown as a cartoon representation and semi-transparent surface.

3 Chapter 3: Cryo-EM structure of the full-length WzmWzt ABC transporter required for lipid-linked O antigen transport

Christopher A. Caffalette and Jochen Zimmer

Molecular Physiology and Biological Physics, University of Virginia School of Medicine, Charlottesville, VA 22908, USA

In this chapter, I will present my work on the ATP- and Mg^{2+}-bound cryo-electron microscopy (cryo-EM) structure of the full-length O antigen ABC transporter WzmWzt.

3.1 Introduction

Cell surface polysaccharides are ubiquitous features of microbial pathogens that enable them to evade innate immune responses, resist desiccation, adhere to surfaces, and deter antibiotics (1, 2, 151). A vast array of surface-anchored and secreted polysaccharides have evolved to extend the cell envelope surrounding and protecting the plasma membrane (201). To achieve this defense, Gram-negative bacteria employ capsular and lipo-polysaccharides, as well as colanic acid; Gram-positives utilize wall- and lipo-teichoic acids; and the mycobacterial outer membrane is fortified by numerous capsular polysaccharides (1-3, 202, 203).

Lipopolysaccharide (LPS) is a glycolipid comprised of lipid A, core oligosaccharides, and a terminal polysaccharide known as the O antigen (41). The O antigen extends the cell envelope and has been shown to confer resistance to complement-mediated lysis (204-209). Its defensive properties are enhanced by its hypervariability among Gram-negative species, forming the basis for serotyping bacterial strains based on the O antigen repeat unit structure (210). Additionally, the length of the O antigen chain can vary from only a single repeat unit 3-4 sugars long to ~100 sugar units, thereby creating an efficient extracellular barrier (211).

The *E. coli* O9a serotype exemplifies the O antigen biosynthesis pathway that utilizes an ABC transporter, known as WzmWzt (162, 211, 212). In this pathway, the O antigen is first synthesized in its entirety at the cytosolic leaflet of the inner membrane, anchored to an undecaprenyl-phosphate (Und-P) lipid. Following completion of biosynthesis, WzmWzt reorients the polymer to the periplasmic side of the inner membrane where it is transferred *en-bloc* to the lipid A-core oligosaccharide by the ligase WaaL. The mature LPS molecule is then integrated into the outer leaflet of the outer membrane via the Lpt pathway (213).

ABC transporter-dependent O antigen biosynthesis is similar to wall teichoic acid (WTA) production in Gram-positives where the Tar/gGH ABC transporter secretes WTA across the cell membrane (105). Further, an analogous process was proposed for mycobacterial species where the Rv3781-Rv3783 ABC transporter ortholog is predicted to translocate an isoprenyl-linked galactan (214, 215).

WzmWzt is a tetrameric exporter that has been crystallized in nucleotide-free (apo) and ATP-bound states (34, 89). In both conformations, the transporter forms a continuous transmembrane (TM) channel sufficiently wide to accommodate a polysaccharide chain. Inside the channel, the polymer is likely coordinated by aromatic and hydrophilic residues, as also observed for other polysaccharide-interacting proteins (92, 177). Strikingly, in the ATP-bound conformation, the periplasmic channel exit forms profound lateral gates toward the periplasmic membrane leaflet, proposed to be important for releasing the lipid-linked O antigen into the lipid bilayer. It has been suggested that WzmWzt uses a processive step-by-step translocation mechanism in which a cytosolic LG-loop near the channel entrance interacts with and moves the polymer towards the channel upon ATP binding (34, 89).

Some bacteria signal completion of O antigen biosynthesis by chemically modifying the non-reducing end of the growing polysaccharide with, for example, methyl-, phosphomethyl-, or β-linked 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) groups (74). This capping reaction terminates polymer elongation, and thus controls O antigen chain length, and primes it for translocation by WzmWzt (87).

A hallmark of ABC transporters secreting capped O antigens is a carbohydrate-binding domain (CBD) extending the C-terminus of the Wzt nucleotide-binding domain (NBD). The CBD recognizes the terminal cap in the context of the O antigen repeat unit it is attached to (212). This interaction is necessary for efficient transport of the lipid-linked O

antigen across the inner membrane *in vivo*. Structural analyses of CBDs from *E. coli*, *Raoultella terrigena* and *Aquifex aeolicus* revealed that the CBD adopts a jelly-roll conformation and dimerizes by exchanging the C-terminal β-strand between protomers (76, 90, 216). Mutagenesis and *in vitro* binding studies with the *E. coli* O9a and *R. terrigena* CBDs suggest that a concave jelly-roll surface forms the binding site for the O antigen cap (76, 212).

We currently have separately determined structures of the CBD-truncated *A. aeolicus* (Aa) WzmWzt in apo and ATP-bound states, as well as the structure of its isolated Wzt CBD dimer. Separately, however, these structures do not inform on whether O antigen cap binding influences catalytic activity and thus transport, which are important mechanistic questions. Here, we present the cryo-EM structure of the full-length AaWzmWzt transporter in an ATP-bound state in a lipid nanodisc. The structure reveals an unexpected position of the CBD dimer and, more broadly, provides new insights into how biopolymer-translocating ABC transporters gate their central channel and couple processive transport to ATP hydrolysis.

3.2 Results

3.2.1 Overview

To gain insights into how WzmWzt recognizes and transports the O antigen, we determined the cryo-EM structure of the full-length, catalytically inactive AaWzmWzt transporter reconstituted into palmitoyl-oleoyl-phosphatidylglycerol (POPG) lipid nanodiscs in a Mg^{2+}/ATP-bound state (**SI, Figs. S3-1A and S3-2**). ATP hydrolytic activity of the wild type transporter after reconstitution into nanodiscs was demonstrated for both E. coli total lipid extract and POPG nanodiscs prior to structure determination (**SI, Fig. S3-1B**). The resultant cryo-EM map has an overall estimated resolution of 3.6 Å and is of excellent quality showing density for the vast majority of side chains in the Wzm transmembrane (TM) and Wzt NBD regions (**SI, Fig. S3-2 and Table 2**). The initial map showed poor density in some regions of the Wzt CBD due to flexibility. However, signal subtraction and local 3D classification generated a second map of the Wzt dimer with an overall estimated resolution of 3.6 Å, showing improved density for the Wzt CBD sub-domain (**SI, Fig. S3-1 C-E**). This map was used to generate a separate model of the Wzt dimer.

Fitting the CBD of this separate Wzt model into the initial full-length map allowed building an entire WzmWzt model. This model includes residues 2-256 of each Wzm, excluding residues 48-54 of one Wzm protomer due to poor density of the TM1-TM2 loop, and residues 1-395 of each Wzt, excluding residues 242-252 and 241-253 of both Wzt protomers, respectively, due to disordered linker regions with missing densities in the EM map.

3.2.2 Overall structure of the full-length WzmWzt transporter

Similar to the eukaryotic ABC exporters ABCA1, ABCG2, and ABCG5/8, WzmWzt is a prokaryotic example of the Type-5 ABC exporter topology comprised of two Wzm protomers, each containing six TM helices, that are coupled to the cytosolic ATP-binding Wzt dimer (217). In addition to six proper TM helices, two re-entrant helices, PG1 and PG2, connect TM helices 5 and 6 and an amphipathic N-terminal interface helix (IF) lies perpendicular to the TM helices at the Wzm-Wzt interface (**Fig. 3-1**). Further, an extended loop between TM helices 2 and 3 inserts into the NBD forming a coupling helix-loop (CHL) in a similar fashion to other ABC transporters (218). Wzt adopts a classical NBD fold featuring an additional gate helix (GH) extension that packs against the interface of the Wzm protomers (**Fig. 3-1A**) (217).

As also observed for some *E. coli* and *K. pneumoniae* serotypes, Aa Wzt contains a C-terminal CBD. Published structures of *E. coli*, *R. terrigena*, and most recently Aa CBDs show that the isolated CBD itself dimerizes via β-strand exchange, which is preserved in the full-length WzmWzt structure (76, 90, 216).

Our cryo-EM structure reveals the position of the CBD dimer relative to the known core WzmWzt structure (**Fig. 3-1A**). Strikingly, the dimeric CBD interacts with only one NBD. Conserved loops from each CBD protomer at the dimer interface straddle the C-terminal helix of only one NBD (hereafter referred to as ‘hinge helix’). This arrangement centers the CBD dimer under one NBD at an approximately 15 degree angle relative to the plane of the inner membrane. In this position, the putative O antigen cap binding sites of both

Fig. 3-1

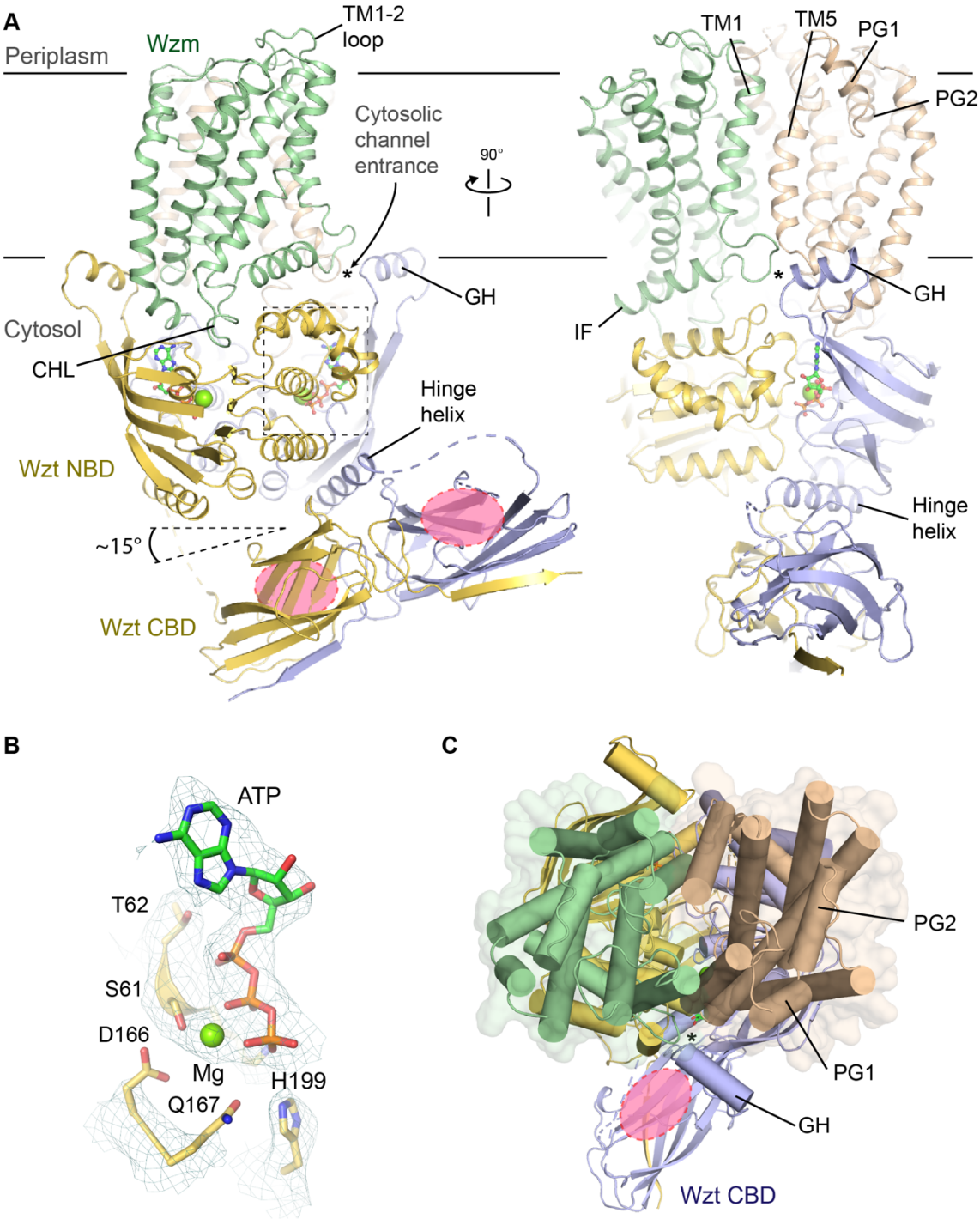

A
Periplasm
Wzm
TM1-2
loop
Cytosolic
channel
entrance
90°
GH
Cytosol
CHL
Hinge
helix
Wzt NBD
~15°
Wzt CBD
TM1
TM5
PG1
PG2
IF
GH
Hinge
helix
B
ATP
T62
S61
D166
Mg
Q167
H199
C
PG2
PG1
GH
Wzt CBD

Fig. 3-1 Overall architecture of full-length WzmWzt. **(A)** Atomic model of ATP-bound full-length Aa WzmWzt. Wzm protomers are shown in green and wheat, Wzt protomers are colored in yellow and purple, respectively. Wzt contains a nucleotide-binding domain (NBD) and a carbohydrate-binding domain (CBD). ATP is shown as sticks and magnesium is shown as a green sphere. The cytosolic channel entrance is indicated with a black asterisk. The putative O antigen terminal cap binding sites at the CBD dimer are indicated as red ellipses. IF: Interface helix, PG: Periplasmic gate helix, CHL: Coupling helix-loop, GH: Gate helix. **(B)** ATP density at the NBD. ATP is shown in sticks and the cryo-EM map as a teal mesh (contoured at 8σ). **(C)** Periplasmic view of WzmWzt with the atomic model of the Wzm dimer shown as cylinders and semi-transparent surfaces. One O antigen terminal cap-binding site of the CBD dimer (red ellipse) sits below the cytosolic channel entrance (black asterisk).

CBDs face the NBD dimer and lie within the same vertical plane as the cytosolic channel entrances (34). However, due to the special location and tilting of the CBD dimer relative to the NBDs, the distances between the proposed cap binding sites on the CBDs and the GHs near the WzmWzt cytosolic channel entrances are significantly different (discussed below, **Fig. 3-1**). The overall structure of the CBD dimer is consistent with its crystal structure in isolation, yet subtle conformational changes near the ligand binding sites could have functional implications, as discussed in detail below.

Although the structures of the TM and NBD regions are consistent with our previously determined ATP-bound WzmWzt crystal structure (PDB 6M96), we note several functionally important differences. First, while the crystallized transporter contained only ATP and water molecules at the active sites, the cryo-EM map indicates the presence of a Mg^{2+}/ATP complex for each NBD, allowing these to be built into the final model (**Fig. 3-1B**). Accordingly, the side chain density of the conserved Ser61 points toward the magnesium density to coordinate the ion, while it points away from it in the crystal structure (**SI, Fig. S3-3**). Placement of three water molecules into the final, refined model supports octahedral magnesium coordination (**SI, Fig. S3-3 and Table 3**). Second, the short Wzm loops connecting TM1 to TM2 at the periplasmic side of the transporter adopt different conformations in the crystal and cryo-EM structures. These loops point away from the transporter in the crystal structure, due to crystal contacts, and pack against the surface formed by the remaining TM3-TM6 helices in our cryo-EM map (**Fig. 3-1A**). Third, the continuously open TM channel described in the crystal structure is narrowed and closed in

Fig. 3-2

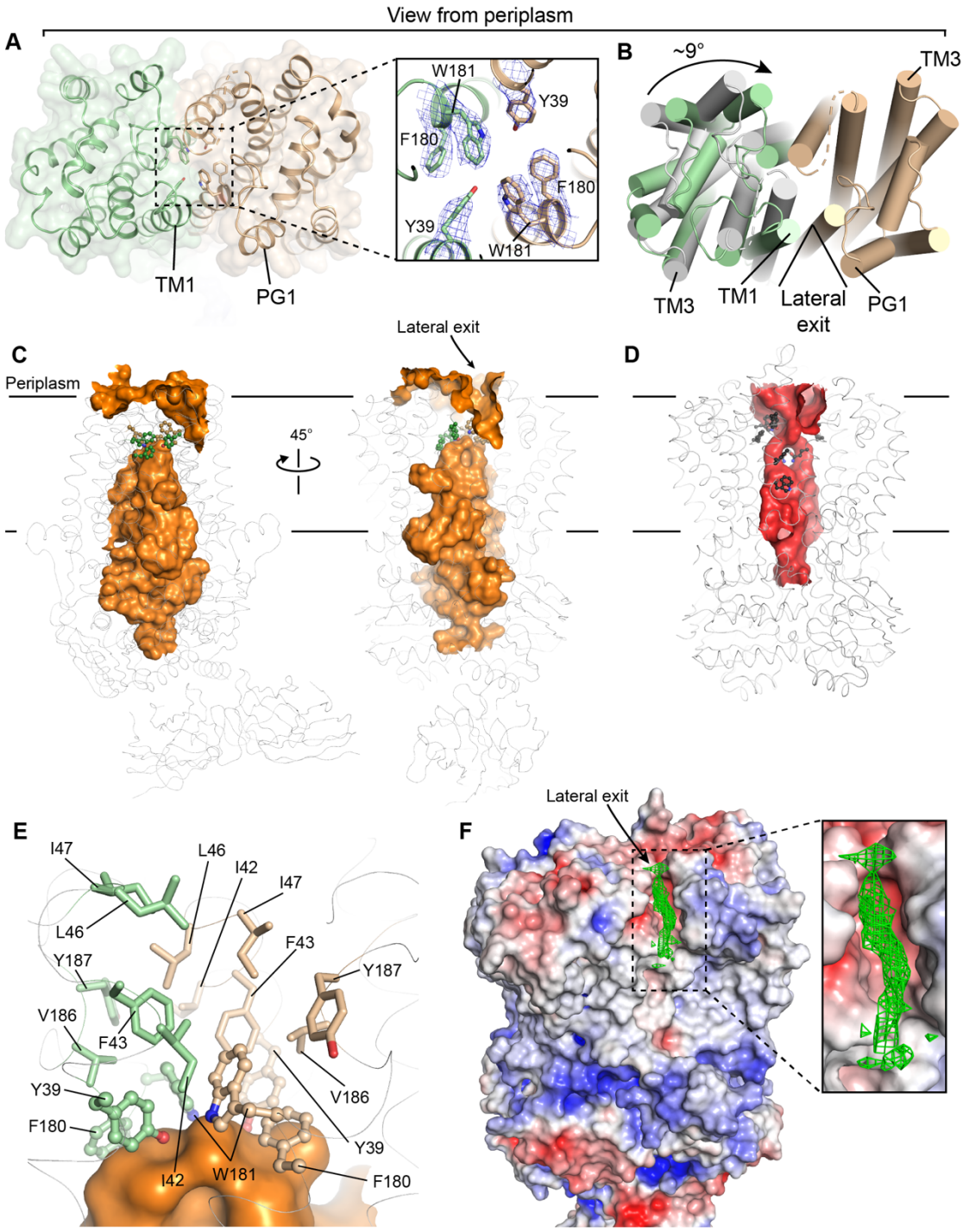

View from periplasm
A
W181
Y39
F180
F180
Y39
W181
TM1
PG1
B
~9°
TM3
TM3
TM1
Lateral exit
PG1
C
Periplasm
45°
Lateral exit
D
E
I47
L46
I42
I47
L46
F43
Y187
Y187
V186
F43
V186
Y39
F180
Y39
I42
W181
F180
F
Lateral exit

Fig. 3-2 The closed conformation of the WzmWzt transmembrane channel. **(A)** View of WzmWzt from the periplasm. Inset: Residues of the Wzm aromatic belt are shown as sticks with cryo-EM map density contoured at 8σ shown as a blue mesh. **(B)** WzmWzt viewed from periplasm as in A, showing Wzm dimer asymmetry. One half-transporter (containing Wzm in wheat) is superimposed on the other half-transporter (containing Wzm in green) by aligning their Wzt NBDs and is displayed in gray. The lateral exit is formed by TM helix 1 and PG1 from different protomers. **(C)** Surface representation of the WzmWzt channel displayed as an orange surface. **(D)** Surface representation of the WzmWzt channel calculated for the ATP-bound crystal structure (PDB 6M96) displayed in red. Aromatic residues forming the hydrophobic belt are shown as sticks. **(E)** Zoomed-in view of the aromatic belt for the cryo-EM WzmWzt structure. Residues directly implicated in channel closure are shown in green and wheat as ball-and-sticks. Additional hydrophobic residues in the periplasmic exit of the channel are shown as sticks. The protein backbone is shown as gray ribbon in C, D and E. (F) Electrostatic surface potential of WzmWzt prepared in PyMOL using the APBS plugin (scale -5.0 (red) to +5.0 (blue) kT/e). Non-protein cryo-EM map density within the lateral exit contoured at 6σ is shown as a green mesh.

the new structure (**Figs. 3-1C and 3-2A**). Thus, the full-length WzmWzt cryo-EM structure is representative of an ATP-bound state in the absence of the O antigen substrate, i.e. a resting state.

3.2.3 An asymmetric closed TM channel

Previous WzmWzt crystal structures in apo and ATP-bound states revealed transporter conformations with continuous TM channels. These channel-forming conformations are expected to occur during polymer translocation (34, 89). In a nanodisc-reconstituted and ATP-bound state, however, Wzm protomers are arranged closer together toward the channel axis, allowing contacts between the previously identified aromatic belt residues (Tyr39, Phe180, Trp181) near the periplasmic channel exit (34), which close the channel (**Fig. 3-2A, B**).

The side chains of Tyr39, Phe180, and Trp181 from each Wzm protomer are well resolved in the EM density map and establish that these residues are indeed in close proximity, the narrowest point being a 3.4 Å distance between the Trp181 side chains of each Wzm protomer (**Fig. 3-2A**). A similar distance between phenylalanine residues was observed in the human ABC transporter ABCG2 serving as an intracellular gate that seals the outward open conformation (219). Further, the recent structure of the WTA TarGH ABC transporter, homologous to WzmWzt, also shows a closed state in which Tyr and Leu residues form a similar, aromatic belt-like seal (**SI, Fig. S3-4**) (220).

A water accessibility analysis of the WzmWzt channel using the software HOLLOW (191) with a 1.4 Å radius probe further underscores the closed channel conformation (**Fig. 3-2 C and D**). Past the aromatic belt, the channel is sealed by a hydrophobic water-excluding region, indicating that our new structure is indeed a closed resting state of the transporter in the absence of an O antigen (**Fig. 3-2E**).

Emphasized by the channel shape, the cryo-EM structure represents an asymmetric closed transporter conformation with only one open lateral exit (**Fig. 3-2C**). Superimposing the WzmWzt transporter halves on their NBDs highlights the different conformations of the Wzm subunits in each half transporter. One Wzm protomer is rotated as a rigid body around TM helix 3 relative to the other subunit by approximately 9 degrees, thereby breaking symmetry in the Wzm dimer (**Fig. 3-2B**). This asymmetry results in closure of one of the periplasmic lateral channel exits previously observed between TM1, TM5 and PG1 in the symmetric ATP-bound crystal structure (34) (**Fig. 3-2 C and D**).

In addition to asymmetry at the periplasmic lateral exits, we also observe asymmetry at the cytosolic channel entrances adjacent to each GH, a proposed substrate entry site (34). Within one Wzm-Wzm dimer interface, the open periplasmic lateral exit coincides with a closed cytosolic entry site, and the opposite configuration (closed lateral gate, open cytosolic entry site) is observed at the other Wzm dimer interface (discussed below and **SI, Fig. S3-5**).

Of note, the asymmetric location of the CBD dimer does not bias the conformation of the Wzm dimer. Our initial 3D particle classification identified a subset in which the openings at the cytosolic and periplasmic sites are reversed, relative to the location of the CBD dimer (**SI, Fig. S3-5)**.

Within the open lateral exit, we observe non-protein density, perhaps a partially ordered lipid molecule, between TM1 and PG1 in the cryo-EM map, suggesting that the lateral exit can accommodate lipid acyl chains (**Fig. 3-2F**). The bound molecule appears to stabilize the open lateral exit. Other, lower resolution, maps obtained during data processing lacking the compound showed different lateral exit conformations with either wide (open), intermediate or narrow (closed) distances between TM1 and PG1. This demonstrates that the Wzm dimer samples different conformations in a resting, polymer-free state.

While the thereby generated asymmetry of the Wzm protomers is remarkably subtle, other examples of ATP-uncoupled conformational changes of ABC transporters within their TM segments have been correlated with large structural transitions (221).

3.2.4 WzmWzt modulates TM channel width through rigid body movements

Despite the overall conformational differences observed across all three available WzmWzt structures (apo, ATP-bound open, ATP-bound closed), the Wzm protomers behave as rigid bodies. With the exception of TM1, aligning the Wzm protomers from all three structures shows that they are essentially identical for TM2-TM6 and the IF helix (**Fig. 3-3A**).

Fig. 3-3

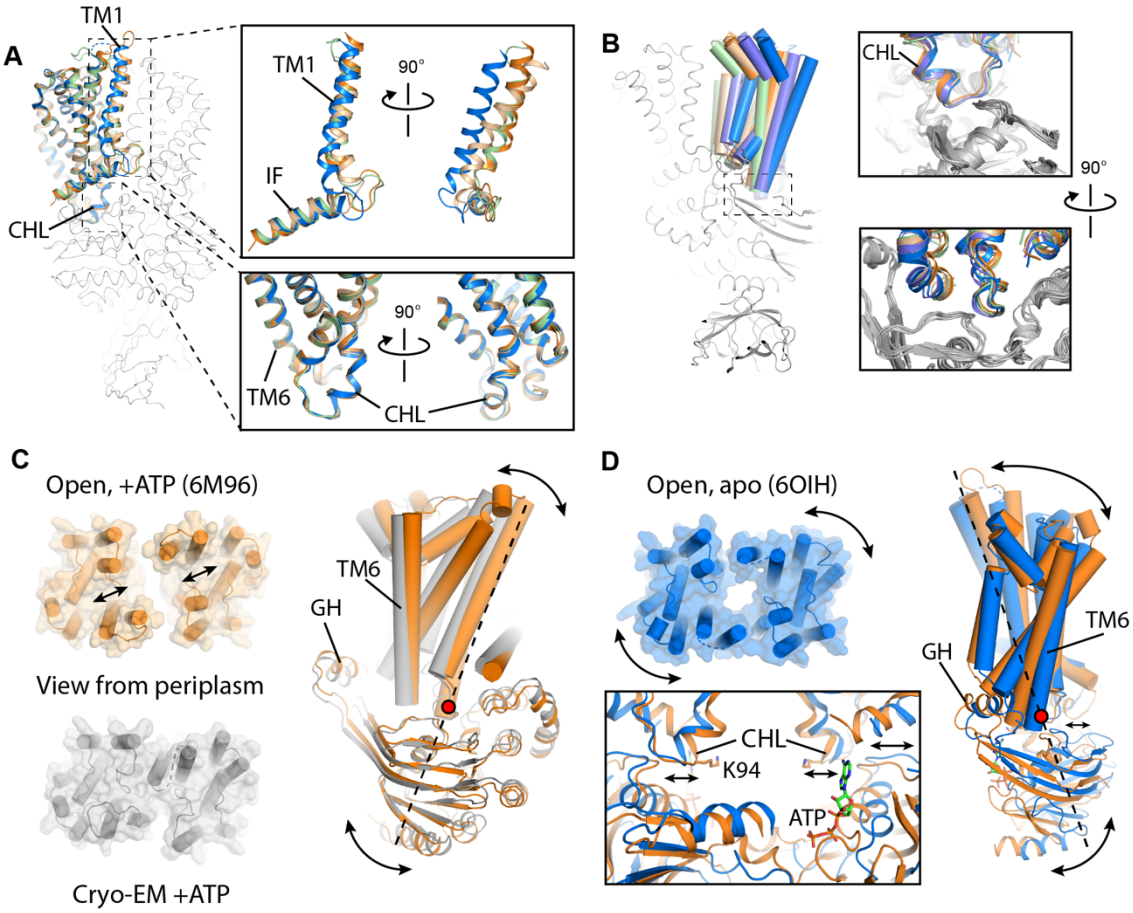

A
TM1
CHL
TM1
90°
IF
90°
TM6
CHL
B
CHL
90°
C
Open, +ATP (6M96)
View from periplasm
Cryo-EM +ATP
TM6
GH
D
Open, apo (6OIH)
CHL
K94
ATP
GH
TM6

Fig. 3-3 WzmWzt modulates channel width by coupling NBD and TM conformational changes. **(A)** Wzm alignment (based on TM2-6) and superposition of Wzm protomers from the apo-state crystal structure (blue, PDB 6OIH), ATP-bound crystal structure (orange, PDB 6M96), and cryo-EM structure (protomer 1 green, protomer 2 wheat). **(B)** Superposition of all Aa WzmWzt half-transporter structures based on Wzt NBD. Only the Wzm protomers are displayed as cylindrical cartoons and colored as in A. Right, insets show two views of the aligned half-transporters with Wzm protomers in color and Wzt NBD protomers in gray. This alignment includes the second half transporter of the apo-state (PDB 6OIH) excluded in panel (A) and shown in purple. **(C)** Left, view from the periplasm of the ATP-bound crystal structure (orange, PDB 6M96) and cryo-EM structure (gray) both shown as cylindrical helices and semi-transparent surface. Arrows indicate rigid body movements. Right, ATP-bound crystal structure (orange) and cryo-EM structure (gray) aligned on the Wzt NBD in the background. The orange conformation results from a tilt about the red dot along the axis shown as a dotted line. **(D)** Left top, view from the periplasm of the apo-state crystal structure (PDB 6OIH) shown as cylindrical helices and semi-transparent surface. Arrows depict Wzm rotations from the apo to ATP-bound state depicted on the right. Left bottom, view parallel to the plane of the membrane showing the Wzm-Wzt interface at the CHL. Arrows depict Wzm separation when transitioning from ATP-bound to the apo state of the NBDs. Right, comparison of the ATP-bound (orange) and apo (blue) crystal structures aligned on the CHL.

Aligning WzmWzt half-transporters based on their Wzt NBDs shows that while Wzm adopts a range of angles relative to the channel axis, the CHL is found in a common position across all structures, suggesting a path to couple ATP hydrolysis with Wzm rigid body movement (**Fig. 3-3B**). Indeed, transitioning from a closed to an open TM channel in the ATP-bound state (representing resting and O antigen inserted states, respectively), involves the rotation of a WzmWzt half-transporter around an axis running along TM3 and through the CHL, a central fulcrum (**Fig. 3-3C**).

Extending this analysis to the transition from ATP-bound to nucleotide-free states with an open TM channel, as required during O antigen transport, reveals a similar coupling effect. Here, transitioning from ATP-bound to the apo state results in further opening of the channel and requires an opposing rotation of the Wzt NBD and Wzm around the CHL axis (**Fig. 3-3D and SI, Fig. S3-6**).

Further, the increased diameter of the TM channel in the nucleotide-free state, compared to the open channel in the ATP-bound conformation, results from the increased distance between the NBDs, which moves the pivot points located at the CHLs farther apart (**Fig. 3-3D**).

3.2.5 Asymmetry of the full-length WzmWzt ABC transporter

The interaction of the Wzt CBD dimer with the hinge helix of only one NBD generates profound asymmetry of WzmWzt. The interactions of the pseudo 2-fold symmetric central CBD-dimer loops with the hinge helix are made possible by various hydrophilic and

Fig. 3-4

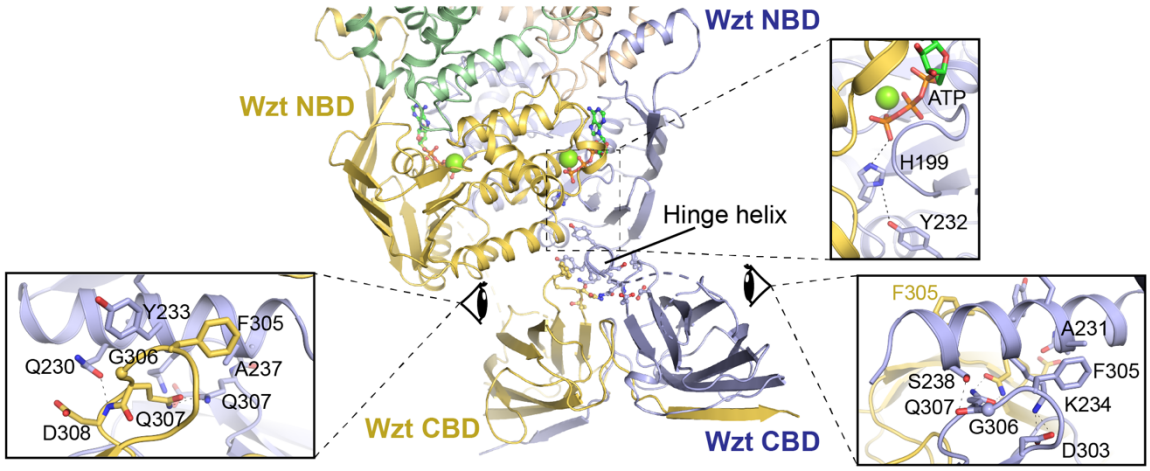


Fig. 3-4 The Wzt CBD dimer interacts with one Wzt NBD. View of WzmWzt from within the cytosol down the barrel of the NBD hinge helix. Top right inset shows the interaction of hinge helix Y232 with ATP via the H-loop H199. ATP and side chains are shown as sticks and magnesium is shown as a green sphere. Bottom left and right insets show interactions of the CBD central loops with the hinge helix. Side chains are shown as sticks and glycines are shown as spheres for the C-alpha position. Wzt protomers are colored purple and yellow, respectively.

hydrophobic interactions (**Fig. 3-4**). Specifically, Phe305 of one CBD loop faces Ala231 in the hinge helix, and the backbone carbonyl of the loop Gly306 hydrogen bonds with Ser238 of the hinge helix. This knob-in-hole type interaction is further stabilized by a salt bridge between Asp303 (loop) and Lys234 (hinge helix) (**Fig. 3-4**).

On the opposing side of the hinge helix, similar interactions occur with the other CBD loop. Two knob-in-hole interactions are observed, one formed by Phe305 (loop) and Ala237 (hinge helix) and a second formed by Gly306 (loop) and Tyr233 (hinge helix). Further, the side chain of Gln230 (hinge helix) is in hydrogen-bonding distance to the backbone amide of loop residue Asp308. The CBD loops are stabilized in an upright position by an intermolecular glutamine bridge between Gln307 residues of each CBD (**Fig. 3-4**). This intricate network of interactions explains how the isolated CBD-dimer supplied *in-trans* can complement a CBD-truncated WzmWzt transporter by stimulating ATPase or translocation activity *in vitro* and *in vivo*, respectively (76, 89).

3.2.6 Wzt exposes the CBD O antigen-binding site to the NBD and TM domains

In the full-length WzmWzt structure, the putative CBD O antigen-binding surfaces are oriented toward the NBDs, rather than towards the cytosol (**Figs. 3-1, 3-4 and 3-5**). This transporter-facing orientation creates a direct interface with the NBD and provides a means for coupling O antigen cap binding at the CBD to effects at other transporter domains implicated in ATP hydrolysis or O antigen channel insertion. Indeed, the NBD hinge helix, serving as the attachment point for the CBD, contains the invariant Tyr232. In both the ATP-bound cryo-EM and crystal structures, its side chain points directly at H-loop His199,

Fig. 3-5

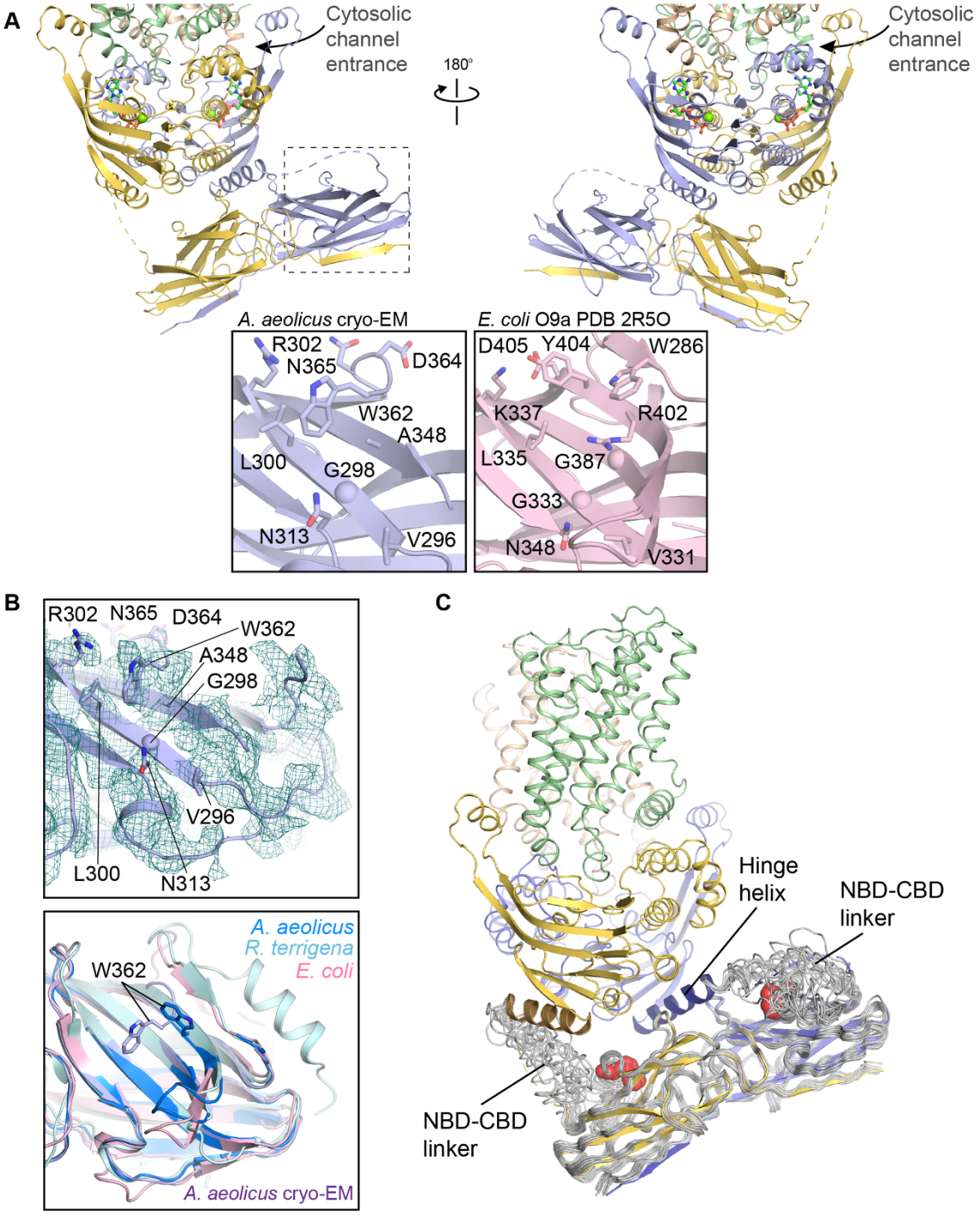

A
Cytosolic channel entrance
180°
Cytosolic channel entrance
A. aeolicus cryo-EM
R302
N365
D364
W362
A348
L300
G298
N313
V296
E. coli O9a PDB 2R5O
D405
Y404
W286
K337
R402
L335
G387
G333
N348
V331
B
R302
N365
D364
W362
A348
G298
V296
L300
N313
A. aeolicus
R. terrigena
E. coli
W362
A. aeolicus cryo-EM
C
Hinge helix
NBD-CBD linker
NBD-CBD linker

Fig. 3-5 Wzt CBD presents the O antigen cap-binding site to the transporter. **(A)** Views of the Wzt dimer showing the alignment of the Wzt CBD with the cytosolic channel entrance. The disordered NBD-CBD linker regions are depicted as dashed lines. Inset left, zoomed-in view of the cryo-EM Wzt CBD concave jelly-roll surface highlighted in the dashed-line box. Inset right, corresponding surface in the *E. coli* O9a Wzt CBD crystal structure (salmon, PDB 2R5O). Residues lining the putative O antigen cap-binding pocket are shown as sticks or spheres for C-alpha positions. **(B)** Upper panel, Wzt CBD cryo-EM map density (teal mesh) shown at a map threshold of 10.25σ and superimposed with the Aa Wzt CBD dimer cryo-EM structure (light blue). Lower panel, alignment and superposition of Wzt CBDs from the cryo-EM full length WzmWzt structure (light blue) and the isolated crystal structures of Aa (royal blue), *R. terrigena* (pale cyan, PDB 5HNO), and *E. coli* (salmon, PDB 2R5O). **(C)** An ensemble of 15 Robetta structure predictions of Wzt (gray ribbon) aligned with the cryo-EM WzmWzt structure (colored as in Fig. 3-1). The hinge helices are highlighted in dark yellow and dark purple. Residue W362 is shown as red spheres.

which is critical for ATP hydrolysis (222, 223). Accordingly, O antigen binding to the CBD could affect ATP hydrolysis (**Fig. 3-4 and SI, Fig. S3-7**).

Mapping important *E. coli* O9a CBD residues implicated in O antigen cap binding (Trp286, Asn348, Arg402, Tyr404, Asp405, Gly333 and Gly387) to the corresponding residues of the Aa CBD (Trp362, Asn313, Asp364, Asn365, Gly298, Ala348) suggests functional similarities (**Fig. 3-5A**). Of these residues, Trp362 is particularly important as its position is dramatically different in the isolated CBD crystal (PDB code 6O14) and full-length WzmWzt structures (**Fig. 3-5B**).

In all currently available isolated CBD crystal structures residues Cys359-Ile363 begin the final, NBD-proximal β-strand of the putative O antigen-binding site. However, in the full-length transporter, Trp362 at the center of this β-strand is shifted towards the ligand binding pocket and inserts into a cavity formed by the conserved residues Gly298, Leu300 and Ala348. Further, the β-strand immediately upstream of Trp362 is disordered and not resolved in our cryo-EM map, suggesting that the entire Gly352 to His361 region is dynamic in full-length WzmWzt (**Fig. 3-5B**).

The linkers connecting the NBDs and CBDs are also disordered in the cryo-EM map and were not modeled. At low thresholds, however, weakly defined density extending from the CBDs indicates the locations of the linker regions in multiple conformations (**SI, Fig. S3-8A**). For the CBD closest to the membrane, a 25 Å distance between the NBD and CBD termini accommodates 11 residues, whereas for the other CBD protomer, the 22 Å NBD-

CBD termini distance would be bridged by 13 residues (due to the different number of unmodeled residues). Robetta (224, 225) structure predictions suggest that the linker lengths suffice to span these distances in ATP-bound and nucleotide-free states (**Fig. 3-5C and SI, Fig. S3-8**).

The 15 degree angle of the CBD relative to the plane of the inner membrane places the two CBD ligand binding sites at ~54 Å (membrane-proximal) and ~67 Å (membrane-distal) from their corresponding channel entry sites. Additionally, the interaction of the CBD dimer with only one NBD offsets the CBD dimer by ~21 Å from the central axis of the transporter, such that one CBD sits directly underneath the center of the transporter and the other is outside its projected footprint (**Fig. 3-5C and SI, Fig. S3-8**).

3.3 Discussion

The proposed model of WzmWzt-dependent transport of the Und-PP-linked O antigen starts with an initial tethering event, in which the O antigen cap binds to the Wzt CBD, thereby increasing its local concentration near the transporter. Reorienting the lipid anchor then loads the O antigen polysaccharide into the central channel, beginning ATP-driven processive transport (34, 89).

Spanning the distances from the cap-binding CBD surfaces to the proposed channel entrances would require a polymer at least 10-15 sugar units long (assuming a length of 5 Å per sugar unit). However, owing to the profound conformational flexibility of polysaccharides, even significantly larger polymers could interact with the CBD and

channel entrance simultaneously, in accordance with estimated O antigen chain lengths (211).

Using the conserved Tyr232 within the hinge helix as a marker for sequence alignments, modeling of the NBD/CBD interactions in the *E. coli* O9a and *R. terrigena* WzmWzt transporters suggests a similar mode of interactions (**SI, Fig. S3-9**). In *R. terrigena*, one knob-in-hole interaction could involve hinge helix residues Ala244, Ala248, and Gly252 accommodating Gln355 and a second could include hinge helix residue Leu255 "knob" fitting into a "hole" formed by Gly356 and Gly357 of one CBD loop. Within the opposing CBD loop, this same double Gly motif is well-suited to accommodate Leu247 and Glu250 from the hinge helix (**SI, Fig. S3-9 C and E**). Similarly, in *E. coli* O9a, one side of the hinge helix likely has a knob-in-hole interaction with L236 positioned adjacent to Gly341 of the CBD loop and, on its other side, hinge helix Gly241 would face Gly341 of the opposing CBD protomer (**SI, Fig. S3-9 D and F**). Thus, the off-center position of the CBD dimer relative to the NBDs is likely a conserved feature of O antigen ABC transporters.

The CBDs are connected with the NBDs via disordered linkers 11 and 13 residues long for the CBD closest to and farthest away from the membrane, respectively. An additional residue was modeled at the C-terminus of the hinge helix (Asn241) and the N-terminus of the membrane proximal CBD (Gly253), resulting in the shorter 11-residue linker. Due to the tilt and off-center position of the CBD dimer and supported by weak map density, the linker leading into the membrane proximal CBD appears to sit right above the CBD ligand binding site, thereby potentially occluding it (**Fig. 3-5C and SI, Fig. S3-8 A and B**). Cap

binding to this CBD would require significant conformational changes of the linker region and/or the CBD dimer tilt. On the opposite side, however, a combination of weak map density and Robetta models suggest that the linker connecting with the membrane distal CBD is in a stretched conformation extending away from the CBD ligand binding site. Here, cap binding is probably minimally affected by the linker, raising the possibility that only this CBD interacts with the O antigen cap (**Fig. 3-5C**). Further studies with the native, capped O antigen are necessary to determine its precise interactions with the CBD and assess binding site preferences, if any.

O antigen binding to the CBD has the potential to modulate the hydrolytic activity of the NBDs. One such coupling mechanism may exist through disorder-to-order transitions of one β-strand of the CBD. While residues 359 to 363 adopt a β-strand conformation in the isolated CBD crystal structures, the region from Gly352 to His361 is disordered in the context of the full-length transporter, so that Trp362 can occupy the CBD's putative cap-binding pocket on its concave jelly-roll surface (**Fig. 3-5A**). Displacement of Trp362 (or similar aromatic residues in the *E. coli* and *R. terrigena* CBDs) by the O antigen cap could then induce conformational changes within the Gly352-His361 region as well as the NBD/CBD linker that are propagated to the NBD via the hinge helix to affect transporter function.

One candidate residue for modulating ATP hydrolysis is the highly conserved Tyr232 located at the center of the hinge helix (**SI, Fig. S3-9A**) (89). This residue points its hydroxyl group at the H-loop histidine His199, a crucial motif in the hydrolysis of ATP

(**Fig. 3-4 and SI, Fig. S3-7A**) (222, 223). Thus, structural rearrangements of the CBD and linker region upon cap binding could exert a direct effect on ATP hydrolysis by manipulating the interaction of Tyr232 with His199. This is supported by the observation that in the 2.05 Å ATP-bound crystal structure of CBD-truncated WzmWzt, these residues are within hydrogen-bonding distance whereas they are more separated in our cryo-EM structure (**SI, Fig. S3-7B**). Subtle coordination differences of the H-loop and coupled networks could have profound effects on transporter function.

Further, Tyr232 also forms a knob-in-hole-type interaction with the highly conserved Gly54 of the Walker A motif (**SI, Fig. S3-7A**). In the case of the LPS ABC transporter LptB2FGC, replacing Tyr234 (equivalent to Wzt Tyr232) with alanine (LptB Y234A) abolished LPS transport, underscoring its importance for function. Notably, substituting LptB Tyr234 with phenylalanine (LptB Y234F) had only a minimal effect on LPS transport (226). Thus, the C-terminal helix of the NBD with its conserved aromatic residue may function as a general modulator of ATPase activity responding to different stimuli, such as O antigen binding to WzmWzt.

Accordingly, we previously observed the in-trans interaction between the CBD-truncated WzmWzt transporter and the isolated CBD dimer with a corresponding increase in ATPase activity (89). This interaction, in the absence of the NBD-CBD linkers, may stimulate ATP hydrolysis by stabilizing the hinge helix and thus the conserved Tyr232.

Following O antigen tethering at the CBD, its lipid anchor must interact with the cytosolic substrate entry site located at the intersection of the Wzt GH and the Wzm protomers to initiate transport. In a processive transport model, the pyrophosphate-GlcNAc head-group of the lipid-linked O antigen substrate would then reorient through the TM1-TM5 interprotomer seam to the periplasmic lateral gate, thereby loading the O antigen polysaccharide into the Wzm channel (**Fig. 3-6**) (34, 89).

Interestingly, our asymmetric cryo-EM structure shows only one of the two cytoplasmic and periplasmic channel portals in open conformations. These states likely reflect pre- and post-reorientation conformations in which a wider cytosolic entry site enables loading of the substrate's headgroup into the channel. The open periplasmic lateral gate is necessary for releasing the headgroup towards the periplasmic bilayer leaflet while also minimizing the physical translocation distance.

While previous apo and ATP-bound crystal structures suggested that WzmWzt exists in strictly symmetrical conformations, O antigen translocation likely requires an asymmetric function of the ABC transporter. This is because the translocating polymer would only interact with one CBD and enter and exit the TM channel through only one cytosolic and one periplasmic portal, respectively. The CBD O antigen cap binding sites, as well as the cytoplasmic channel entrances, are in direct proximity to the active sites of the NBDs. Therefore, polymer-induced asynchronous ATP hydrolysis by the NBDs seems plausible, perhaps resulting in staggered power strokes enabling processive translocation. However,

Fig. 3-6

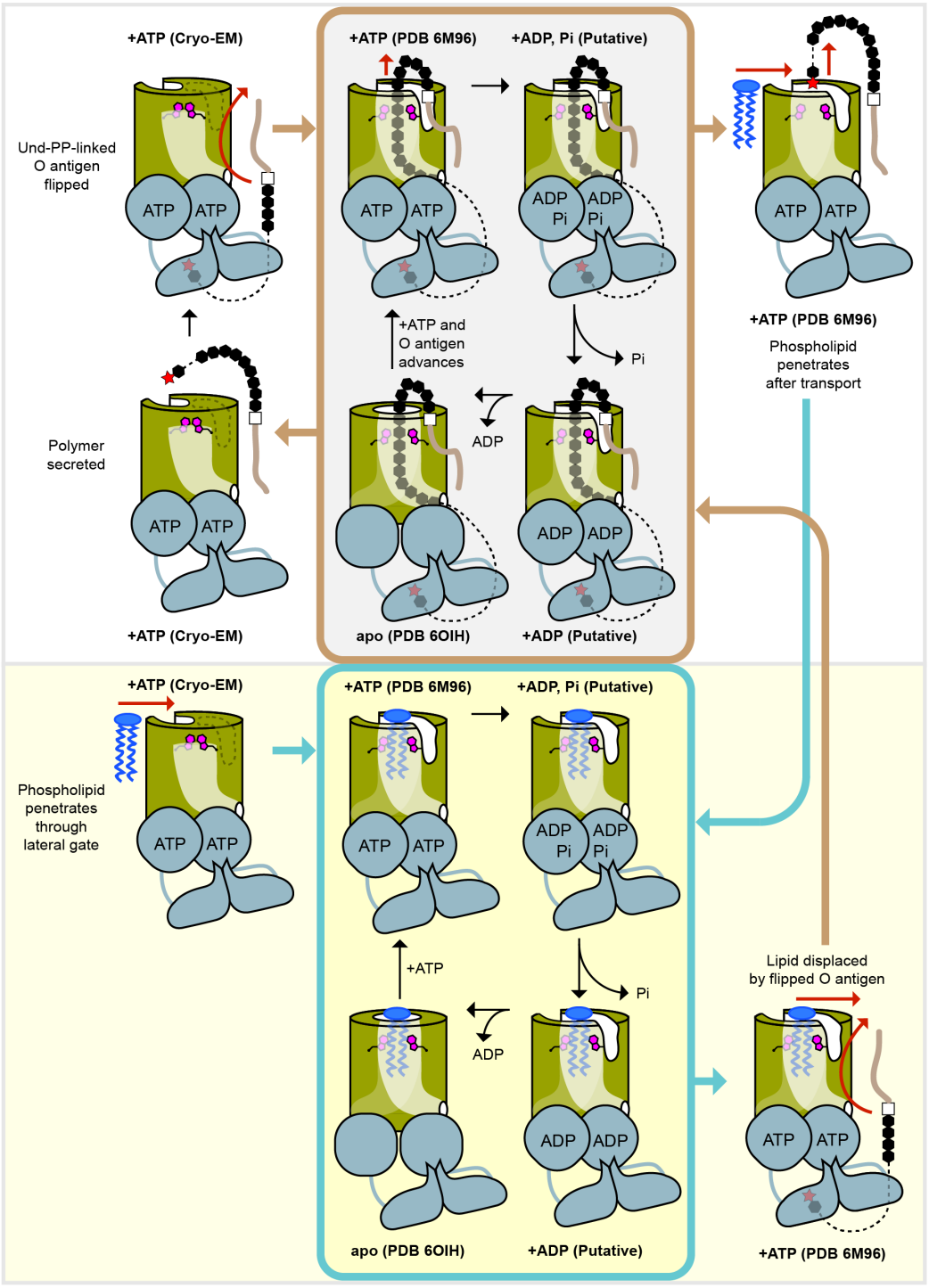

+ATP (Cryo-EM)
+ATP (PDB 6M96)
+ADP, Pi (Putative)
Und-PP-linked O antigen flipped
ATP
ADP Pi
+ATP and O antigen advances
Pi
ADP
Polymer secreted
+ATP (PDB 6M96)
Phospholipid penetrates after transport
apo (PDB 6OIH)
+ADP (Putative)
Phospholipid penetrates through lateral gate
Lipid displaced by flipped O antigen

Fig. 3-6 Revised dual-path model of O antigen translocation by WzmWzt. At least two possible paths are available for WzmWzt starting in a closed channel conformation, depending on channel occupancy: O antigen-occupied (active transport, upper panel) or lipid-occupied (resting state, lower panel). Upper panel, O antigen recruitment, channel insertion, and translocation. Lipid diffusion after O antigen release may seal the channel-forming transporter. Magenta cartoon tryptophans indicate aromatic belt residues. The O antigen cap and terminal sugar(s) are shown as a red star and black hexagon, respectively. Lower panel, ATP hydrolysis in a lipid-occupied state. Reorienting a new Und-PP-linked O antigen substrate displaces the lipid molecule, transitioning the transporter to the active transport path (upper panel). The transporter may also undergo ATP hydrolysis cycles with an aromatic belt-sealed transmembrane channel.

elucidating these processes requires structural insights into a stalled O antigen translocation intermediate.

During processive polymer translocation, ATP binding, hydrolysis, and ADP/Pi release must occur while maintaining a continuous TM channel that accommodates the polysaccharide. Our previous apo and ATP-bound crystal structures each contain such a channel sufficiently wide to fit an O antigen chain, representing two conformations within the polymer transport cycle (34, 89).

In a polymer-free state, however, the transporter could undergo the same conformational transitions with a lipid molecule sealing the channel, as previously suggested (34). Alternatively, WzmWzt could undergo ATP hydrolysis cycles with a closed aromatic belt, as observed in cryo-EM structures of WzmWzt (ATP-bound) and TarGH (apo) (220). These scenarios are not mutually exclusive. For example, transitioning from the aromatic belt-sealed to the lipid-plugged state would only require a small Wzm rotation around the CHL axis described above and the diffusion of a lipid molecule into the channel proper. Similar transitions are conceivable when moving from a channel-forming nucleotide-free state to either of the sealed ATP-bound states.

We thus propose a revised model of O antigen translocation by WzmWzt establishing two paths for the transporter, depending upon the occupancy of its channel (**Fig. 3-6**). In the absence of O antigen (for example after releasing it on the periplasmic side), a phospholipid can diffuse into the Wzm channel through the periplasmic lateral exit (lipid-occupied)

enabling basal ATP hydrolysis while maintaining the permeability barrier of the inner membrane. Following O antigen tethering at the CBD, reorienting the lipid-linked O antigen displaces the lipid plug and loads the polysaccharide into the channel (O antigen-occupied). Repeated rounds of ATP hydrolysis while maintaining the channel conformation drives processive translocation until transport is completed. The transporter then returns to a lipid- or aromatic belt-sealed resting state.

The commonalities between WzmWzt and other ABC transporters implicated in lipid-linked polysaccharide translocation in Gram-positive and mycobacteria suggest that our findings portray a generalized mechanism for how this unique class of ABC transporters moves high molecular weight polymers across biological membranes (105, 214, 215, 220).

3.4 Materials and methods

3.4.1 Cloning, expression, and purification

Cloning, expression and purification of the full-length Aa WzmWzt transporter was performed as previously described including cloning of the Walker B mutant (E167Q) by the Quikchange method (34).

3.4.2 MSP1E3D1 expression and purification

Expression and purification of the MSP1E3D1 membrane scaffold protein was adapted from a previously published protocol (227). *E. coli* BL21(DE3) were transformed with pMSP1E3D1 (Addgene plasmid #20066) in LB supplemented with 50 mg/L kanamycin and grown overnight at 37°C. Overnight cultures were diluted 1:100 into LB containing 50

mg/L kanamycin and grown at 37°C until an OD600 of 0.60. Protein expression was induced by supplementation with 1 mM IPTG followed by incubation at 37°C with agitation for 3 hours. Cells were harvested by centrifugation, resuspended in lysis buffer (40 mM Tris pH 8.0, 0.5 M NaCl, 10mM imidazole pH 7.5) and lysed in a microfluidizer. Insoluble material was pelleted by ultracentrifugation at 200,000 × g for 30 min in a Beckman Coulter Type 45 Ti rotor, and the supernatant was incubated at 4°C with Ni-NTA agarose resin (ThermoFisher Scientific). After draining the supernatant, the Ni-NTA resin was washed with four buffers: WB1 (lysis buffer with 1% Triton X-100), WB2 (lysis buffer with 50 mM sodium cholate), lysis buffer, and WB4 (lysis buffer with 50 mM imidazole pH 7.5). MSP1E3D1 was eluted with EB (40 mM Tris pH 8.0, 0.5 M NaCl, 0.5 M imidazole pH 7.5) and dialyzed overnight against DB1 (40 mM Tris-HCl pH 8.0, 100 mM NaCl, 10mM imidazole pH 7.5 and 0.5 mM EDTA pH 8.0). The MSP1E3D1 His-tag was cleaved by incubation overnight at 4°C with TEV protease in the presence of 1 mM DTT and removed by reverse IMAC Ni-NTA purification. The MSP1E3D1-containing fraction was dialyzed overnight against DB3 (20 mM Tris pH 7.5, 100 mM NaCl, 0.5 mM EDTA pH 8.0) for storage at -80°C.

3.4.3 Nanodisc reconstitution

Chloroform-solubilized *E. coli* Total Lipid Extract or POPG (Avanti Polar Lipids) was dried to a lipid film using an argon stream and maintained under vacuum for 24 hours. Lipids were then resuspended in LRB buffer comprising 20 mM Tris pH 7.5, 100 mM NaCl, and 100 mM sodium cholate. Final lipid concentrations of *E. coli* Total Lipid Extract and POPG after resuspension in LRB were 25 mM and 10 mM, respectively. The nanodisc

reconstitution mixture was prepared according to a 1:4:160 molar ratio of WzmWzt, MSP1E3D1 membrane scaffold protein, and lipid, respectively. Gel filtration buffer (20 mM Tris pH 7.5, 100 mM NaCl, 5 mM LDAO, 0.5 mM TCEP), lipid, sodium cholate (15 mM final concentration), and LDAO-solubilized WzmWzt were combined and incubated for 1 hour on ice to form mixed micelles. MSP1E3D1 was added and incubated up to 30 minutes on ice. Removal of detergents was initiated by the addition 0.2 mg/mL Bio-Beads SM2 (Bio-Rad) followed by incubation at 4°C for 1 hour. The same mass of Bio-Beads SM2 was added a second time and the mixture was incubated at 4°C overnight. The next day, the same mass of Bio-Beads SM2 was added followed by incubation at 4°C for 1 hour. After removal of Bio-Beads SM2, the nanodisc-reconstituted WzmWzt was purified on a Superdex 200 column equilibrated in nanodisc gel filtration buffer (20 mM Tris pH 7.5, 100 mM NaCl, 0.5mM TCEP). Fractions of interest were concentrated with a 100 kDa molecular weight cutoff filter for use in ATPase assays (WT AaWzmWzt) or for grid preparation (AaWzmWzt E167Q).

3.4.4 ATPase activity measurements

WzmWzt ATPase activity was measured using a NADH-coupled, ATP-regenerating assay as previously described (89, 228) with the following modifications. Nanodisc-reconstituted WzmWzt was combined with a buffer containing 50 mM HEPES pH 7.5, 4 mM phosphoenolpyruvate, 6 units lactate dehydrogenase, 6 units pyruvate kinase, 1 mM $MgCl_2$, and 0.3 mM NADH. The ATPase reaction was initiated by the addition of 1 mM ATP/Mg^{2+} and NADH oxidation was monitored as a decrease in absorbance at 340nm in a SpectraMax M5e plate reader (Molecular Devices). NADH oxidation rates for WzmWzt-

containing reactions were corrected for baseline NADH oxidation in the absence of transporter before conversion to the ATPase rate. All measurements were performed at 27°C over a period of 60 minutes and repeated to generate averages of three independent experiments.

3.4.5 Electron microscopy sample preparation

To prepare the cryo-EM sample, nanodisc-reconstituted AaWzmWzt E167Q was concentrated to 1.1 mg mL^{-1} and incubated with 2.5 mM ATP and 2.5 mM $MgCl_2$ on ice for 1 hour. A 2.5 µL sample was applied to Quantifoil holey carbon grids (R1.2/1.3, 300 mesh, copper) that were glow-discharged in the presence of amylamine. Grids were blotted for 7 seconds at 4°C, 100% humidity prior to being plunge-frozen in liquid ethane cooled by liquid nitrogen using a Vitrobot Mark IV (FEI).

3.4.6 Electron microscopy data acquisition and image processing

Cryo-EM data was collected at the Cryo-EM Facility of the University of Massachusetts Medical School on a Titan Krios (FEI) 300kV electron microscope equipped with a Gatan imaging filter (GIF) and K3 direct detection camera (Gatan). SerialEM software was used to collect 31-frame movies in super-resolution counting mode at 81,000X nominal magnification, corresponding to a calibrated pixel size of 0.53 Å/pixel.

A total of 5,938 movies were collected with defocus range of -1.2 to -2.2 µm and an estimated dose rate of 1.45 $e^-/Å^2$/frame. All movies were imported into Relion 3.0.8 (229, 230) for motion-correction, dose-weighting and 2-fold Fourier-cropping to a pixel size of

1.06 Å/pixel using MotionCor2. The dose-weighted summed images excluding the initial frame were used for all image processing except for CTF parameter estimation where non-dose-weighted summed images (excluding initial frame) were processed using CTFFind4 (231). Particles were autopicked using the Laplacian-of-Gaussian-based method from a subset of 4,935 micrographs and down-sampled to a pixel size of 4.24 Å/pixel for initial processing. Following several rounds of 2D classification, a particle set of 804,216 particles was imported into cryoSPARC (232) to generate an *ab initio* initial model. This initial model was imported into Relion to guide 3D classification and refinement.

An iterative process of 3D classification and 2D classification followed by 3D classification yielded a single class suitable for refinement. Re-extraction to 1.49 Å/pixel followed by refinement, post-processing, Bayesian polishing, and CTF refinement yielded a final map with 3.6 Å overall resolution based on 48,174 particles (MAP1).

In order to improve the density of the Wzt CBD, a particle pool of 178,053 re-extracted particles (1.49 Å/pixel) was subjected to signal subtraction to remove the Wzm TM and nanodisc region from particle images. These signal subtracted particles were 3D classified using a mask over the CBD dimer region only to yield a single class with improved CBD density consisting of 51,828 particles. Refinement of this particle set yielded a Wzt dimer map with an overall resolution of 3.6 Å (MAP2).

3.4.7 Model building and refinement

The crystal structure of the ATP-bound WzmWztN (PDB code 6M96) transporter was fit into the post-processed version of MAP1 in Chimera (233) and manually real-space refined in Coot (190, 234) followed by real-space refinement in Phenix (235) in an iterative manner to yield a final model of WzmWzt that excluded the Wzt CBD dimer. Separately, an initial model of a full Wzt dimer was prepared fitting the cryo-EM Wzt NBD dimer and the Aa Wzt CBD dimer crystal structure (PDB code 6O14) into MAP2 in Chimera. The full Wzt dimer model was then iteratively manually refined in Coot followed by real-space refinement in Phenix against a post-processed version of MAP2. The final Wzt model that was refined against MAP2 was then fit into the post-processed MAP1 in Chimera. The MAP2-based Wzt CBD dimer model was then used to complement the MAP1-based model of CBD-truncated WzmWzt to generate the final full-length WzmWzt model. This model was refined in Phenix against post-processed MAP1.

3.4.8 Figure preparation

All figures were prepared in the PyMOL Molecular Graphics System (Schrödinger, LLC http://pymol.org) and the UCSF Chimera package. Magnesium coordination was analyzed using the CheckMyMetal Metal Binding Site Validation Server (https://csgid.org/metal_sites) (236, 237).

3.5 Data availability

Cryo-EM data from this study has been deposited in the Electron Microscopy Data Bank (EMDB) with the entry ID EMD-22644. Coordinates for the atomic model have been deposited in the Protein Data Bank (PDB) with the entry ID 7K2T.

3.6 Acknowledgements

We thank Ruoya Ho for help with cryo-EM grid preparation and discussions as well as Kangkang Song and Chen Xu at the University of Massachusetts Medical School of Cryo-EM Facility for data collection on the Titan Krios microscope. C.C. was supported by the Cell and Molecular Biology NIH training grant NIH-5T32GM008136 and the University of Virginia Robert Wagner Fellowship. J.Z. is supported by NIH grant 5R01GM129666.

3.7 Supplementary Information

Fig. S3-1

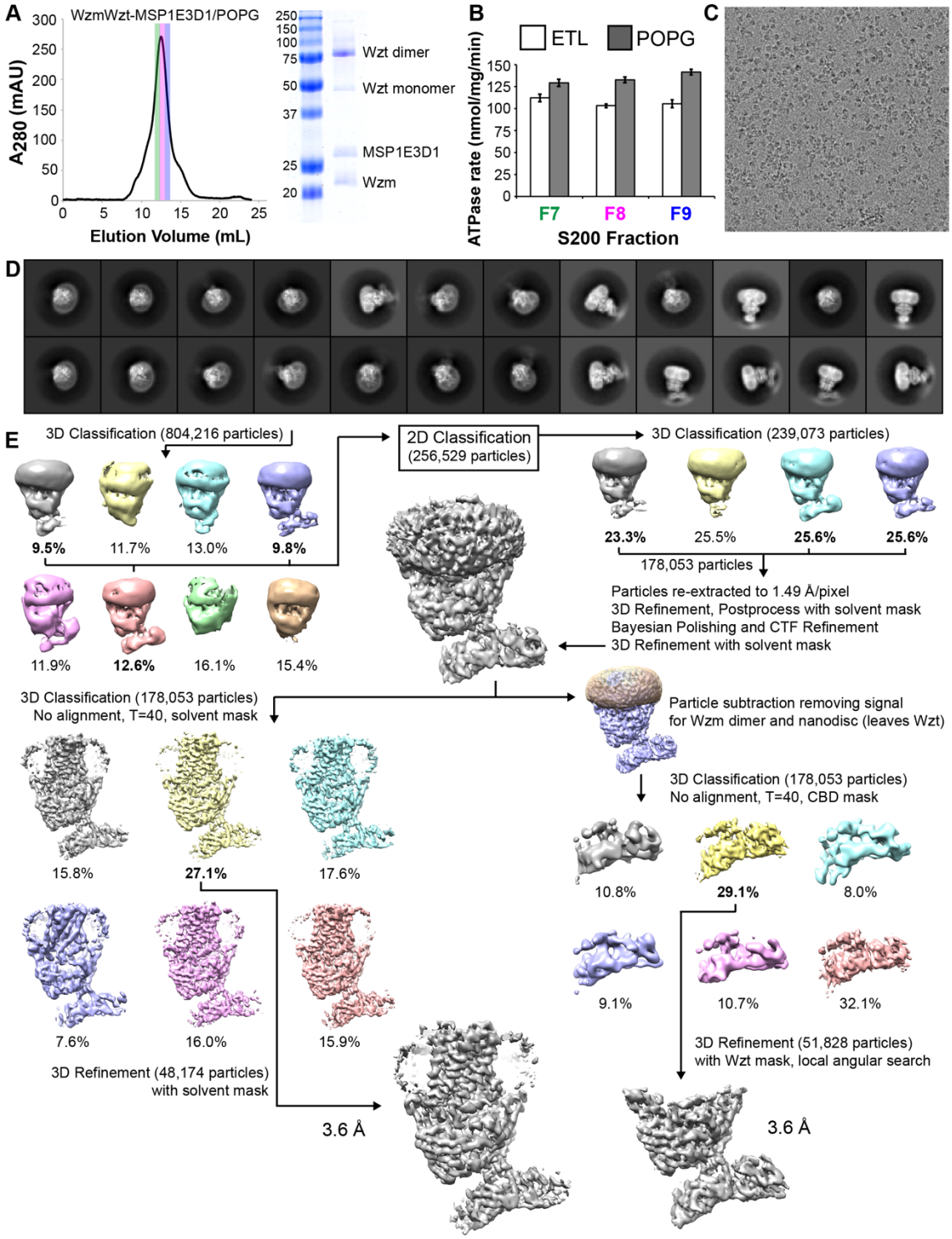


Fig. S3-1 Sample preparation and cryo-EM data processing workflow. **(A)** Left, representative S200 SEC profile of WzmWzt in nanodiscs. Right, Coomassie-blue-stained SDS-PAGE gels of WzmWzt in nanodiscs. **(B)** ATPase activity of wild-type WzmWzt in *E. coli* total lipid (ETL) and 1-palmitoyl-2-oleoyl-phosphatidylglycerol (POPG) nanodiscs. Each column represents the mean ± std. dev. from three separate assays for three S200 SEC main peak fractions, colored as in (A). **(C)** Representative cryo-EM micrograph of full-length WzmWzt particles in POPG nanodiscs. **(D)** Two-dimensional class averages of full-length WzmWzt in POPG nanodiscs. **(E)** Three- dimensional classification and refinement of cryo-EM particle images. Following two initial rounds of 3D classification (with one round of 2D classification in between to eliminate junk particles), two strategies were applied to data processing. Left: 3D classification and refinement of the full particles yielded a final, full-length WzmWzt map. Right: signal subtraction, 3D classification and refinement focusing on Wzt yielded a second Wzt map with an improved carbohydrate-binding domain region.

Fig. S3-2

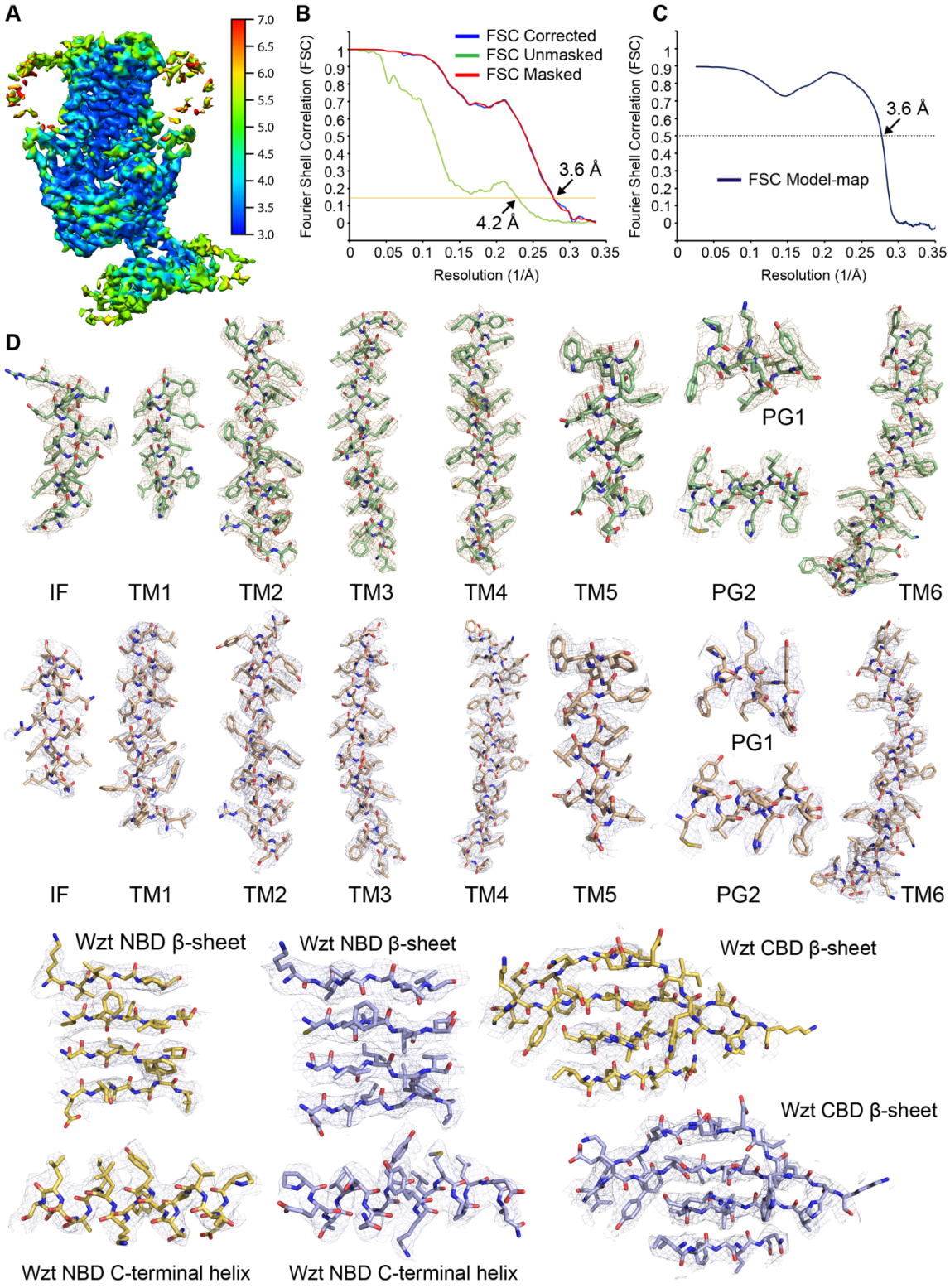

A
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
B
FSC Corrected
FSC Unmasked
FSC Masked
Fourier Shell Correlation (FSC)
3.6 Å
4.2 Å
Resolution (1/Å)
C
Fourier Shell Correlation (FSC)
3.6 Å
FSC Model-map
Resolution (1/Å)
D
PG1
IF
TM1
TM2
TM3
TM4
TM5
PG2
TM6
PG1
IF
TM1
TM2
TM3
TM4
TM5
PG2
TM6
Wzt NBD β-sheet
Wzt NBD β-sheet
Wzt CBD β-sheet
Wzt CBD β-sheet
Wzt NBD C-terminal helix
Wzt NBD C-terminal helix

Fig. S3-2 Model building for full-length WzmWzt in nanodiscs. (A) Local resolution of the final cryo-EM map of ATP/Mg^{2+}-bound full-length WzmWzt determined using ResMap. (B) Gold standard corrected, masked and unmasked FSC curves from Relion PostProcess with stated resolutions at FSC = 0.143 (yellow line) (C) FSC curve between the atomic model and the final map with stated resolution at FSC = 0.5 (dashed line). (D) Cryo-EM density (tan and gray mesh) overlayed on the atomic model of each Wzm protomer (wheat and green sticks) and select regions of each Wzt protomer (yellow and purple sticks). Cryo-EM density of Wzm shows clear side chain density for the IF, TM and PG helices. Cryo-EM density of Wzt show separation of β-strands and excellent density for the side chains of the NBD C-terminal helices. All cryo-EM density shown is contoured at 6σ and shown as wire mesh.

Fig. S3-3

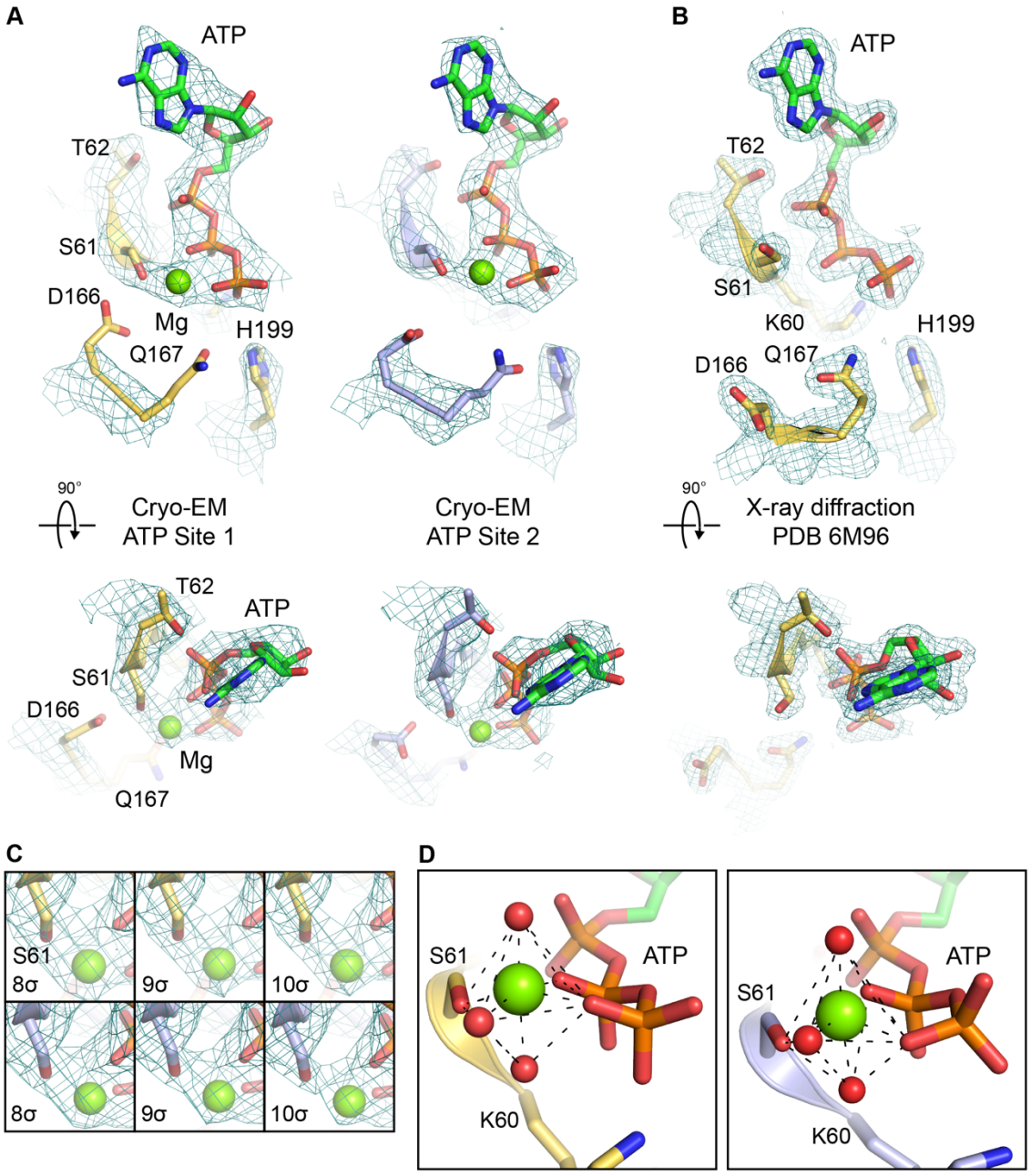

A
ATP
T62
S61
D166
Mg
Q167
H199
90°
Cryo-EM
ATP Site 1
Cryo-EM
ATP Site 2
B
ATP
T62
S61
K60
H199
D166
Q167
90°
X-ray diffraction
PDB 6M96
T62
ATP
S61
D166
Mg
Q167
C
S61
8σ
9σ
10σ
8σ
9σ
10σ
D
S61
ATP
K60
S61
ATP
K60

Fig. S3-3 Modeling ATP and magnesium into WzmWzt cryo-EM and X-ray diffraction maps. **(A)** Cryo-EM density (teal mesh) contoured at 10σ is overlayed on the atomic model of each ATP binding site within the Wzt NBDs. ATP-binding residues are shown as yellow and purple sticks for each Wzt protomer. **(B)** The atomic model of the ATP-bound WzmWzt crystal structure (PDB 6M96) is shown overlayed with the Sigma-A weighted 2Fo-Fc map (teal mesh) contoured at 2.0σ. ATP-binding residues are shown as yellow sticks **(C)** Cryo-EM density (teal mesh) contoured at 8σ, 9σ, and 10σ is shown for each Wzt NBD ATP binding site. Residue S61 is shown as sticks and colored as in (A) for each Wzt protomer. **(D)** Completion of octahedral coordination of the modeled magnesium ion (green sphere) by placement of three water molecules (red spheres, not resolved in cryo-EM map) around the magnesium. For all panels, ATP is shown as sticks.

Fig. S3-4

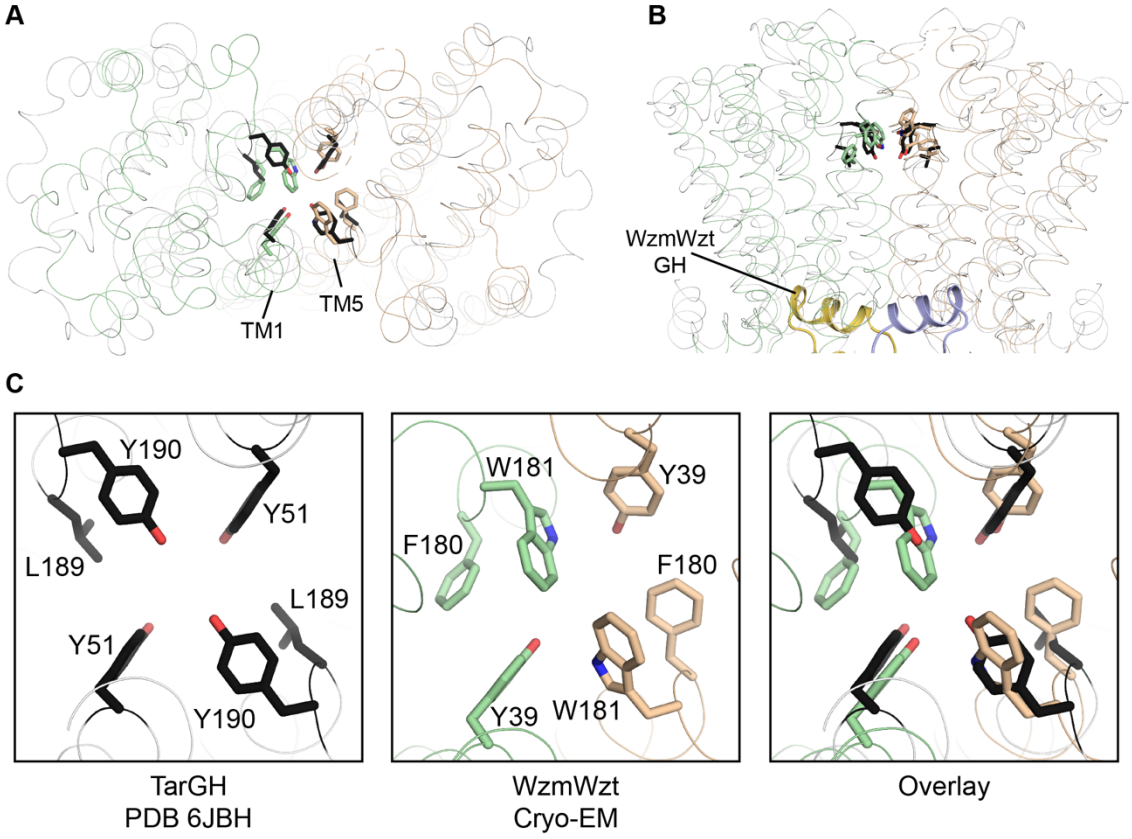


Fig. S3-4 Comparison of WzmWzt aromatic belt residues to corresponding residues in TarGH. **(A)** View from the periplasm of WzmWzt (cryo-EM) superimposed with TarGH (PDB 6JBH) shown in gray ribbon. WzmWzt aromatic belt residues are shown as sticks and the protein backbone is shown as ribbon, both colored green and wheat for different Wzm protomers. TarGH residues that close the TarGH channel and correspond to WzmWzt aromatic belt residues are shown as black sticks and the protein backbone is shown as gray ribbon. **(B)** Same overlay as in (A) rotated 90° and viewed from within the plan of the inner membrane. **(C)** Left, zoomed-in view of the TarGH residues that correspond in position to the WzmWzt aromatic belt residues. Center WzmWzt aromatic belt residues. Right, overlay of TarGH and WzmWzt residues showing comparable arrangement of channel-closing residues in both transporters. In all panels, superposition of TarGH on WzmWzt was done by aligning the TarG transmembrane dimer with the Wzm transmembrane dimer.

Fig. S3-5

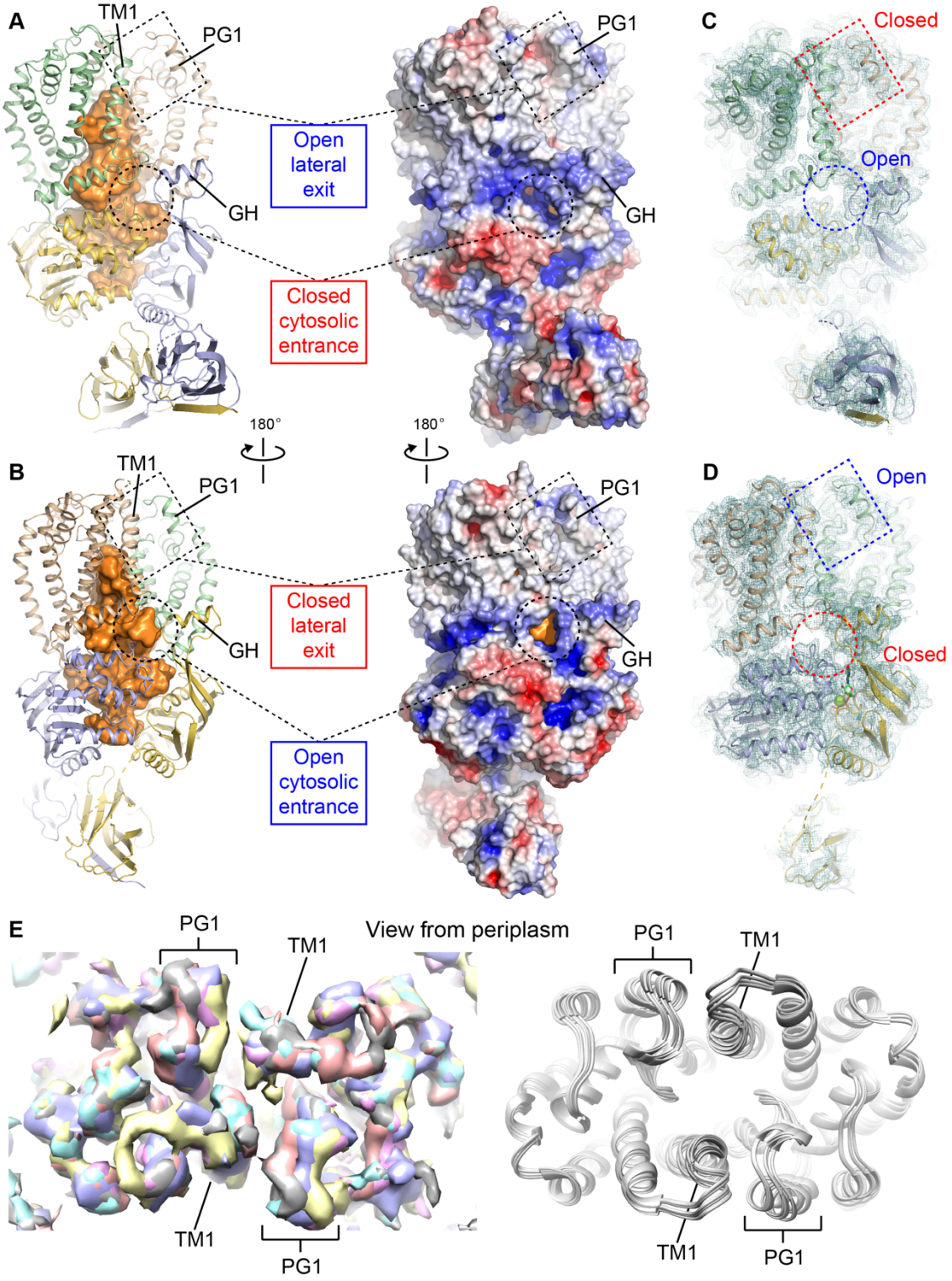

A
TM1
PG1
GH
Open lateral exit
Closed cytosolic entrance
PG1
GH
180°
180°
B
TM1
PG1
GH
Closed lateral exit
Open cytosolic entrance
PG1
GH
C
Closed
Open
D
Open
Closed
E
PG1
TM1
View from periplasm
TM1
PG1
PG1
TM1
TM1
PG1

Fig. S3-5 Asymmetry of the lateral exits and cytosolic entrances at the Wzm dimer interface. **(A), (B)** Left, surface representation of the WzmWzt channel calculated for the cryo-EM atomic model using a 1.4 Å-radius probe (orange surface) with a cartoon WzmWzt colored as in Fig. 1. Right, electrostatic surface potential of WzmWzt (cryo-EM) prepared in PyMOL using the APBS plugin (scale -5.0 (red) to +5.0 (blue) kT/e). (A) shows an open lateral exit between TM helix 1 and PG1 and a closed cytosolic channel entrance at the interface of the Wzm dimer and GH. In (B), WzmWzt is rotated 180° to show the opposite Wzm dimer interface with a closed lateral exit and open cytosolic channel entrance. **(C), (D)** Cryo-EM density (teal mesh) contoured at 7.0σ for a refined subclass of particles from gray, cyan and salmon maps in (E) showing an alternate conformation of the Wzm dimer and overlayed on a cartoon WzmWzt prepared from the refined atomic model from Fig. 1. Relative to (A) and (B), (C) and (D) show an opposite configuration of the lateral exits and cytosolic channel entrances. **(E)** Left, view from periplasm of cryo-EM maps from the final round of 3D classification for the full-length WzmWzt workflow showing six 3D classes. Right, atomic models of individual Wzm protomers from the ATP-bound WzmWzt crystal structure (PDB 6M96) fit into the 3D classes using Chimera to create Wzm dimers. Both maps and Wzm dimers show a range of distances between TM1 and PG1 at both Wzm dimer interfaces (CBD position is the same for all six maps).

Fig. S3-6

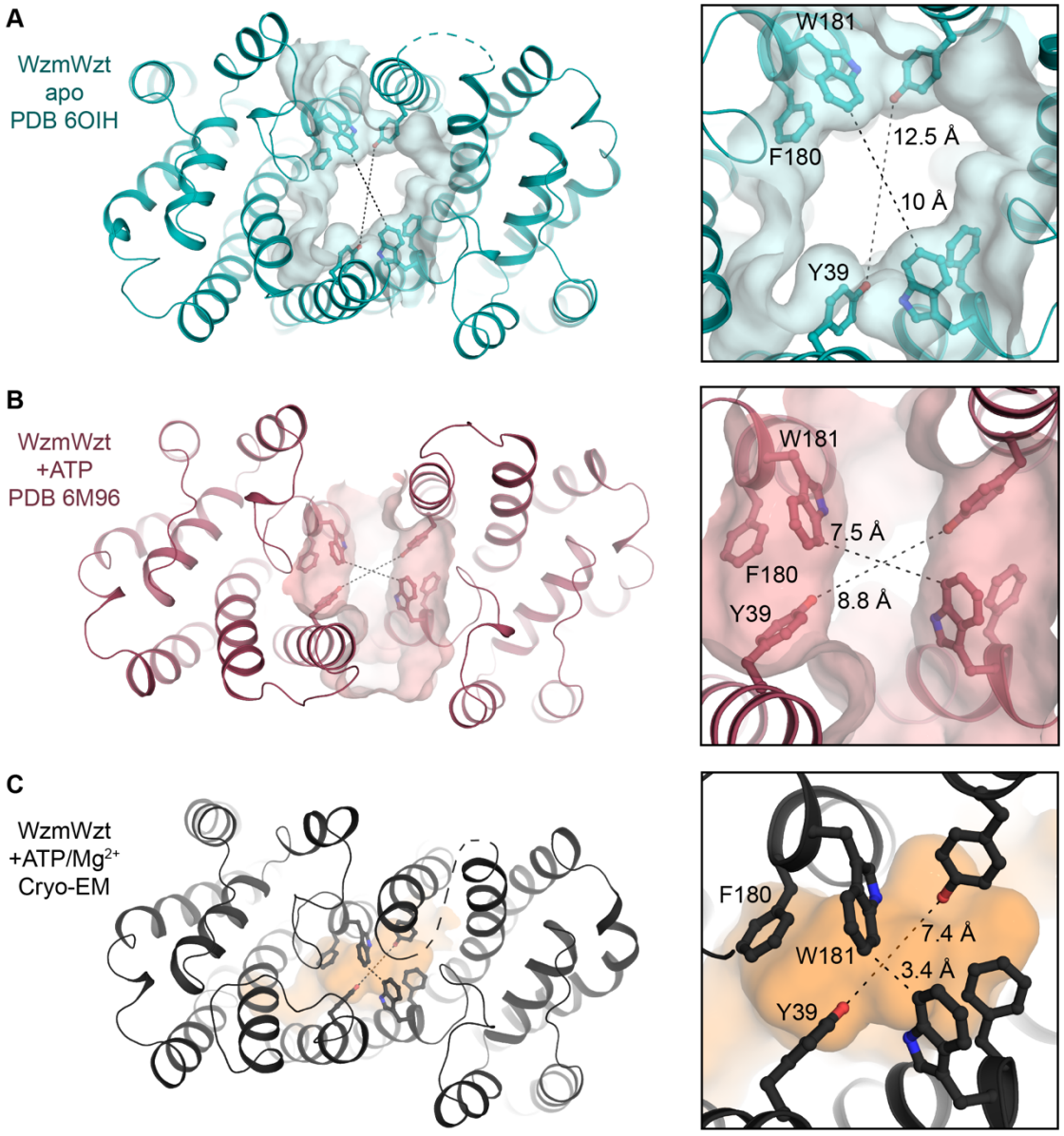

A
WzmWzt
apo
PDB 6OIH
W181
12.5 Å
F180
10 Å
Y39
B
WzmWzt
+ATP
PDB 6M96
W181
7.5 Å
F180
8.8 Å
Y39
C
WzmWzt
+ATP/Mg2+
Cryo-EM
F180
7.4 Å
W181
3.4 Å
Y39

Fig. S3-6 Changes in channel dimensions across three WzmWzt structures. **(A)** Atomic model of the apo-state WzmWzt crystal structure (PDB 6OIH) shown as cartoon (teal) with superimposed surface representation of the Wzm channel. **(B)** Atomic model of the ATP-bound WzmWzt crystal structure (PDB 6M96) shown as cartoon (red) with superimposed surface representation of the Wzm channel. **(C)** Atomic model of the ATP/Mg^{2+}-bound full-length WzmWzt cryo-EM structure shown as cartoon (black) with superimposed surface representation of the Wzm channel (orange). In all panels, insets show a zoomed-in view of the aromatic belt residues in ball-and-sticks of the two Wzm protomers. All surface representations of the WzmWzt channels were calculated using HOLLOW with a 1.4 Å-radius probe.

Fig. S3-7

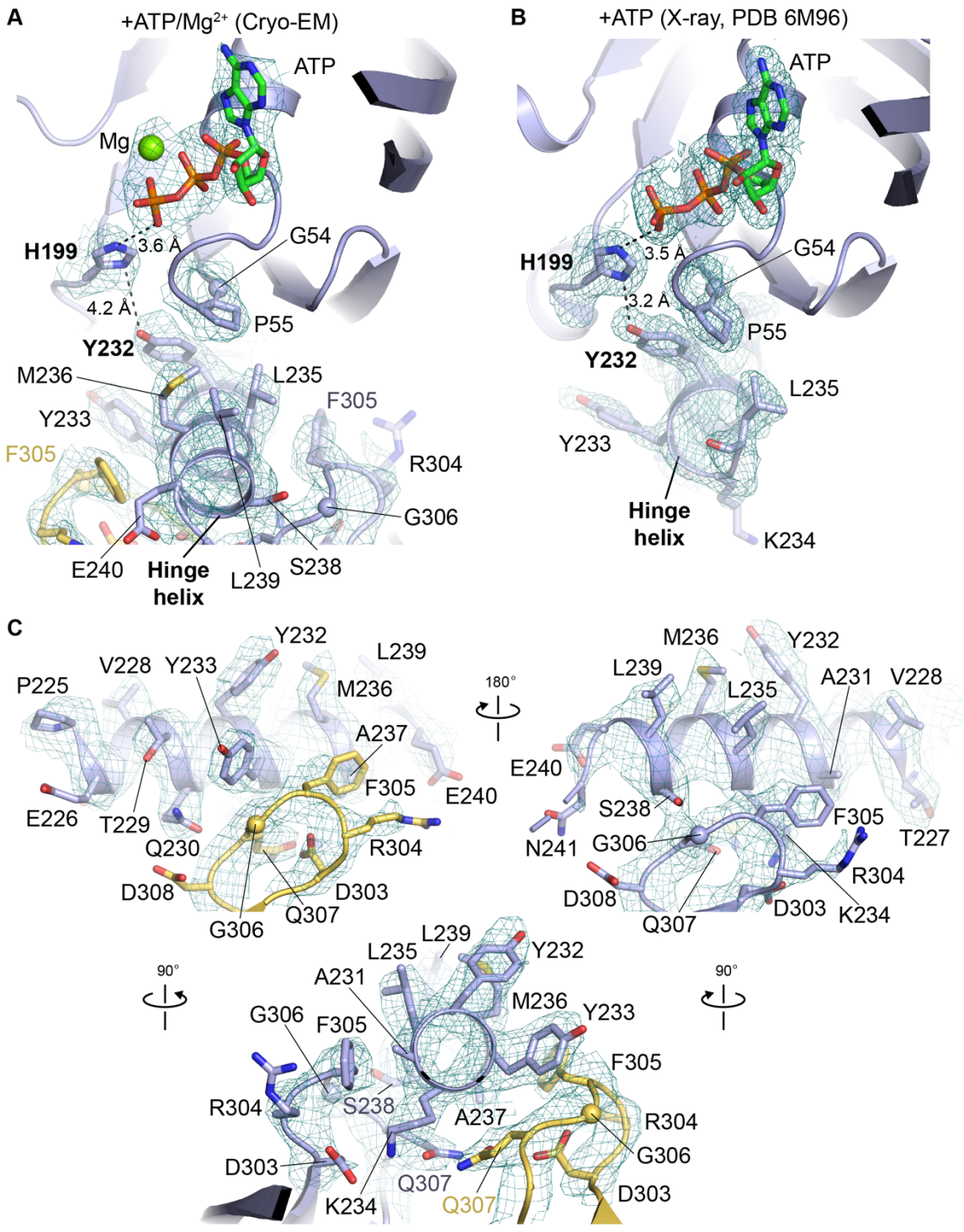

A
+ATP/Mg2+ (Cryo-EM)
ATP
Mg
H199
3.6 Å
4.2 Å
G54
P55
Y232
M236
L235
F305
Y233
F305
R304
G306
E240
Hinge helix
L239
S238
B
+ATP (X-ray, PDB 6M96)
ATP
H199
3.5 Å
3.2 Å
G54
P55
Y232
L235
Y233
Hinge helix
K234
C
Y232
V228
Y233
L239
P225
M236
A237
F305
E240
E226
T229
Q230
R304
D308
D303
Q307
G306
180°
M236
Y232
L239
L235
A231
V228
E240
S238
F305
T227
N241
G306
R304
D308
Q307
D303
K234
90°
90°
L235
L239
Y232
A231
M236
Y233
G306
F305
F305
R304
S238
A237
R304
G306
D303
Q307
Q307
D303
K234

Fig. S3-7 Interaction of hinge helix Tyr232 with H-loop His199 may couple O antigen CBD binding to ATP hydrolysis. **(A)** View of ATP/Mg^{2+}-bound WzmWzt (cryo-EM) from within the cytosol centered on the Wzt NBD C-terminal hinge helix (purple) with cryo-EM density (teal mesh) contoured at 8.5σ. Residues of the hinge helix and Wzt CBD central loops are shown as sticks in purple or yellow for different Wzt protomers. Magnesium is shown as a green sphere. **(B)** View of ATP-bound WzmWzt crystal structure (PDB 6M96) as in (A) with Sigma-A weighted 2Fo-Fc map contoured at 1σ. In both panels, ATP is shown as sticks. **(C)** Three additional views of the interaction between the Wzt CBD dimer central loops with the Wzt NBD hinge helix for ATP/Mg^{2+}-bound WzmWzt (cryo-EM). Cryo-EM density (teal mesh) contoured at 8.0σ. Residues of the hinge helix and Wzt CBD central loops colored as in (A).

Fig. S3-8

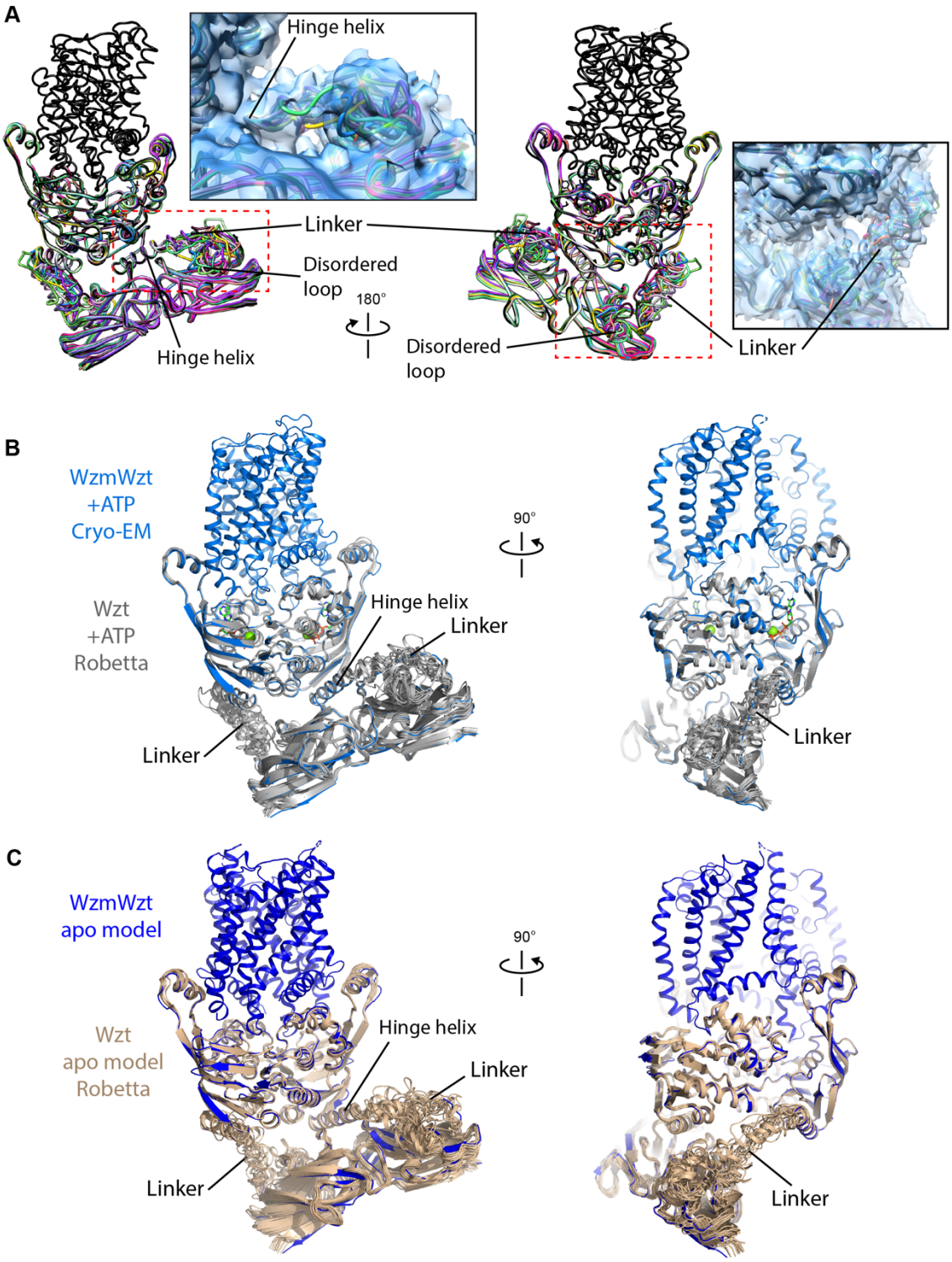

A
Hinge helix
Linker
Disordered loop
180°
Hinge helix
Disordered loop
Linker
B
WzmWzt +ATP Cryo-EM
Wzt +ATP Robetta
90°
Hinge helix
Linker
Linker
Linker
C
WzmWzt apo model
Wzt apo model Robetta
90°
Hinge helix
Linker
Linker
Linker

Fig. S3-8 Robetta structure predictions of the Wzt NBD-CBD linkers. **(A)** An ensemble of 15 Robetta structure predictions for each Wzt protomer (ribbon, various colors) are aligned to the cryo-EM WzmWzt NBD (black ribbon) and superimposed in Chimera. Inset (left and right): in essentially all 15 predicted models, the disordered NBD-CBD linker regions fall within the cryo- EM density in this same region when the unsharpened final map (blue surface) is contoured at a 2.8σ map threshold. **(B)** The same ensemble of 15 Robetta structure predictions (gray ribbon) is aligned with the corresponding region of the cryo-EM WzmWzt structure (blue cartoon) for comparison with (C). **(C)** Cryo-EM Wzt protomers were individually aligned with the Wzt NBD subunits in the apo state (PDB 6OIH) to generate 15 nucleotide-free Wzt Robetta models (wheat). The conformation shown in blue represents the starting model for Robetta structure prediction.

Fig. S3-9

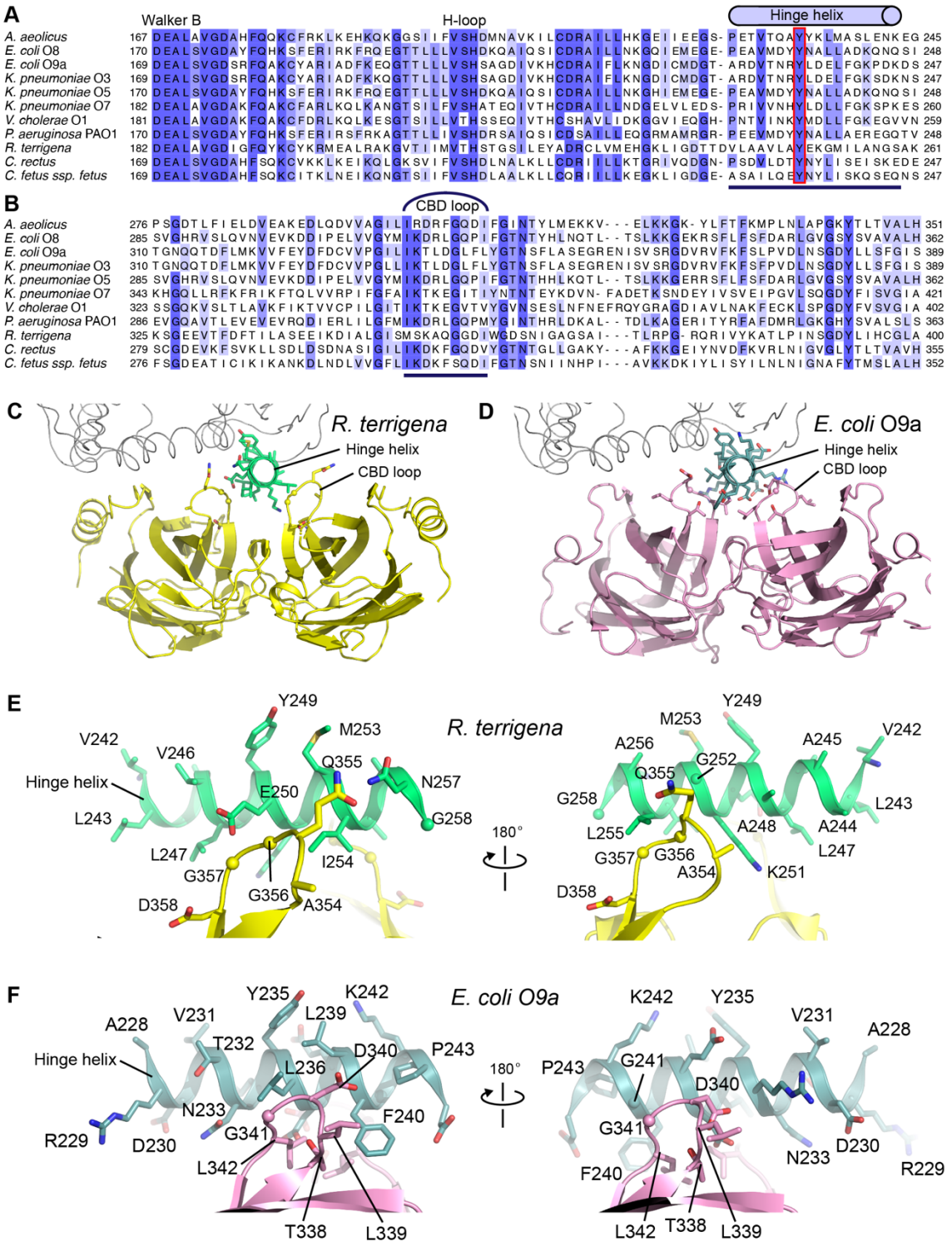

A
Walker B
H-loop
Hinge helix
A. aeolicus
E. coli O8
E. coli O9a
K. pneumoniae O3
K. pneumoniae O5
K. pneumoniae O7
V. cholerae O1
P. aeruginosa PAO1
R. terrigena
C. rectus
C. fetus ssp. fetus
B
CBD loop
C
R. terrigena
Hinge helix
CBD loop
D
E. coli O9a
Hinge helix
CBD loop
E
R. terrigena
Y249
M253
V242
V246
Q355
Hinge helix
E250
N257
L243
G258
L247
I254
G357
G356
D358
A354
180°
A256
A245
G252
L255
A248
A244
G357
G356
A354
K251
F
E. coli O9a
Y235
K242
A228
V231
L239
T232
D340
P243
Hinge helix
L236
N233
F240
R229
D230
G341
L342
T338
L339
180°
G241
V231
A228
N233
D230
R229
F240
L342
T338
L339

Fig. S3-9 Modeling NBD-CBD interactions in *R. terrigena* and *E. coli* O9a WzmWzt transporters. **(A)** Alignments of a select region of the Wzt nucleotide-binding domain (NBD). The Wzt NBD hinge helix is shown as a lavender cylinder. The conserved tyrosine Y232 at the center of the Wzt NBD hinge helix is highlighted with a red box. **(B)** Alignments of a select region of the Wzt carbohydrate-binding domain (CBD) as in (A). The Wzt CBD central loop is shown as a black arch. Sequences were aligned with CLUSTAL Omega and displayed in Jalview colored by sequence identity. **(C)** Model of *R. terrigena* WzmWzt prepared by aligning the *R. terrigena* Wzt CBD dimer crystal structure (yellow cartoon, PDB 5HNO) with the *A. aeolicus* Wzt CBD domain of the full-length structure (cryo-EM). A model of the *R. terrigena* Wzt hinge helix is shown as a green cartoon. **(D)** Model of *E. coli* O9a WzmWzt prepared by aligning the *E. coli* O9a Wzt CBD dimer crystal structure (magenta cartoon, PDB 2R5O) with the *A. aeolicus* Wzt CBD domain as in (C). A model of the *E. coli* O9a Wzt hinge helix is shown as a teal cartoon. **(E)** Two views of the *R. terrigena* Wzt model showing potential NBD-CBD interactions. The Wzt NBD hinge helix and CBD central loops are shown as green and yellow cartoons, respectively. **(F)** Two views of the *E. coli* O9a Wzt model showing potential NBD-CBD interactions. The Wzt NBD hinge helix and CBD central loops are shown as teal and magenta cartoons, respectively. In all panels, relevant residues are shown as sticks.

Table 2 Cryo-EM data collection, refinement and validation statistics

	WzmWzt E167Q (EMD-22644, PDB 7K2T)
Data collection and processing	
Microscope	Titan Krios
Camera	GIF, K3
Magnification	81,000
Voltage (kV)	300
Electron exposure ($e^-/Å^2$)	45.0252
Defocus range (μm)	-1.1 to -2.5
Pixel size (Å)	0.53
Symmetry imposed	C1
Initial particle images (no.)	5,490,947
Final particle images (no.)	48,174
Map resolution (Å)	3.6
FSC threshold	0.143
Map resolution range (Å)	3.3-7.0
Refinement	
Initial model used (PDB code)	6M96, 6O14
Model resolution (Å)	3.6
FSC threshold	0.5
Map sharpening *B* factor ($Å^2$)	-59.3
Model composition	
Non-hydrogen atoms	10,275
Protein residues	1,249
Ligand	ATP: 2, MG: 2
Mean *B* factors ($Å^2$)	
Protein	45.2
Ligand	36.6
R.m.s. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.668
Ramachandran plot	
Favored (%)	95.61
Allowed (%)	4.39
Disallowed (%)	0.00
Validation	
MolProbity score	1.81
Clashscore	9.48
Poor rotamers (%)	0.00
Overall correlation coefficients	
CC (mask)	0.84
CC (box)	0.69
CC (peaks)	0.64
CC (volume)	0.80
Mean CC for ligands	0.92

Table 3 CheckMyMetal (CMM) report for magnesium ions in cryo-EM WzmWzt model

	WzmWztFL Magnesium in ATP Site 1	Score	WzmWztFL Magnesium in ATP Site 2	Score
Metal	Mg		Mg	
Occupancy	1	Acceptable	1	Acceptable
B factor (env.)	24.6 (28.6)	Borderline	35.9 (32)	Acceptable
Ligands	O_6	Acceptable	O_6	Acceptable
Valence	2.0	Acceptable	2.2	Acceptable
nVECSUM	0.087	Acceptable	0.17	Borderline
Geometry	Octahedral	Acceptable	Octahedral	Acceptable
gRMSD(°)	13.1°	Acceptable	19.1°	Borderline
Vacancy	0	Acceptable	0	Acceptable

Legend	Not applicable	Outlier	Borderline	Acceptable

4 Chapter 4: Perspectives, future directions, and concluding remarks

4.1 Historical perspective

Twenty-seven years ago, the *Yersinia enterocolitica* O3 genes *rfbD* and *rfbE* of the *rfb* O antigen biosynthesis cluster were annotated as the transmembrane permease and ATP-binding components of an ABC transporter, respectively, marking the first example of an ABC transporter involved in O antigen export (238). Protein sequences of the gene products rfbD and rfbE were noted to be similar to the *E. coli* capsular polysaccharide ABC transporter components KpsM and KpsT, respectively (see above, **Section 1.6.6**). Further, insertion mutations in *rfbD* and *rfbE* resulted in intracellular accumulation of the O antigen, a homopolymer of 1→2-linked 6-deoxy-β-L-altropyranose (238, 239).

In the following year, analysis of the *K. pneumoniae* O1 *rfb* gene cluster revealed that the first two genes, *rfbA and rfbB,* encode an integral membrane protein and ATP-binding protein, respectively, again sharing homology with *E. coli* KpsM and KpsT (240). Using an *E. coli* K12 model system lacking O antigen, *in vivo* studies demonstrated that *rfbA* and rfbB are indeed the components of the ABC exporter that transports the *K. pneumoniae* O1 O antigen, a galactose homopolymer composed of D-galactan I [→3)-β-D-Gal*f*-(1→3)-α-D-Gal*p*-(1→] and D-galactan II [→3)-α-D-Gal*p*-(1→3)-β-D-Gal*p*-(1→], across the cytoplasmic membrane. Sequence analyses of *K. pneumonaie* O1 and *Y. enterocolitica* O3 *rfb* cluster ABC transporter components showed that *K. pneumonaie* O1 rfbAB and *Y. enterocolitica* O3 rfbDE transporters are homologs, distinct from the *E.* coli KpsMT CPS ABC transporter (240).

In 1996, confusing gene naming discrepancies across species lead to an effort to consolidate gene nomenclature for bacterial surface polysaccharide biosynthetic pathways. (241). Genes of a single function, such as the glycosyltransferases and transporters of the *rfb* gene clusters, were renamed using a *w*** naming convention. As a result of the proposed nomenclature changes, *K. pneumonaie* O1 rfbAB and *Y. enterocolitica* O3 rfbDE gene products were each renamed to WzmWzt in recognition of their identification as O antigen ABC transporters for their respective serotypes (241).

These two landmark publications, identifying the *K. pneumonaie* O1 and *Y. enterocolitica* O3 O antigen ABC transporters (238, 240) initiated the rich history of WzmWzt research, driven primarily by the laboratory of Chris Whitfield (*E. coli* and *K. pneumoniae* systems) and additionally by the laboratory of Joseph Lam (*P. aeruginosa*), both at the University of Guelph in Ontario, Canada.

The *A. aeolicus* WzmWzt system has created an opportunity to generate high-resolution structural information of the O antigen ABC transporter and has revealed the channel-forming architecture of WzmWzt in support of its role as a polysaccharide exporter. Despite the available structures for *A. aeolicus* WzmWzt [apo WzmWzt (89), ATP-bound WzmWzt (34), Wzt-CBD (216), ATP/Mg^{2+}-bound WzmWzt (cryo-EM)], as is evident in **Fig. 3-6**, there are many opportunities to use the *A. aeolicus* WzmWzt system to generate new data toward uncovering the mechanism of processive O antigen transport by WzmWzt.

4.2 Future directions for mechanistic studies of O antigen transport by WzmWzt

4.2.1 Biophysical studies in the absence of O antigen substrate

As presented in the revised model of WzmWzt O antigen translocation (**Fig. 3-6**), the transporter will likely undergo cycles of ATP hydrolysis both in the presence of O antigen substrate, as would be the case during active transport, as well as in the absence of O antigen substrate, between discrete translocation events.

The WzmWzt cryo-EM structure presented in Chapter 3 most likely represents a resting state transporter, in the absence of O antigen, and has captured an ATP- and Mg^{2+}-bound conformation. *A. aeolicus* WzmWzt reconstituted into lipid nanodiscs is therefore a system that enables the potential to gather structural information on additional resting state transporter conformations.

4.2.1.1 Cryo-EM structure of nucleotide-free (apo) full-length WzmWzt

The nucleotide-free (apo) full-length WzmWzt structure is a high priority aim of the overall WzmWzt project because it represents a position in the ATP hydrolysis cycle where the NBDs are expected to be fully open, a state opposite that of the completely closed NBDs in the ATP- and Mg^{2+}-bound cryo-EM WzmWzt structure. A fully open Wzt NBD dimer provides the maximum conformational change expected of the NBDs, and by interpolating between the fully open and fully closed conformations, intermediate, and likely transient positions of the NBDs can be modeled *in silico*.

Further, the conformation of the Wzm dimer in solution and in the absence of nucleotide has not yet been observed. In the case of the Gram-positive WTA transporter ortholog TarGH, the nucleotide-free cryo-EM structure in detergent shows a closed TM channel (220), as observed for the cryo-EM structure of ATP- and Mg^{2+}-bound full-length WzmWzt. What is the conformation of the Wzm dimer in solution in a nucleotide-free transporter? If the Wzm dimer is found in a closed conformation with aromatic belt residues in close proximity, this opens the possibility of a third ATP hydrolysis cycle, a closed-channel hydrolysis cycle, that would complement the O antigen-occupied and lipid-occupied cycles presented in **Fig. 3-6**. It is important to consider that the conformational dynamics of the Wzm dimer during ATP hydrolysis are not yet fully understood and that the channel's hydrophobic seal at the aromatic belt may be imperfect. A hybrid model that combines the elements of the lipid-occupied ATP hydrolysis cycle with the closed-channel ATP hydrolysis cycle would enable a lipid to serve a supporting role in closing the Wzm channel in addition to the aromatic belt. If, during the conformational dynamics of the Wzm dimer, the aromatic belt residues separate enough to allow water flux, a lipid occupying the lateral exit (as proposed for the non-protein density in **Fig. 3-2F**) could easily move to the center of the channel to inhibit water flux.

The conformation of the Wzm dimer in solution in a nucleotide-free transporter will also inform the Wzm-Wzt coupling centered at the CHL. For example, in a nucleotide-free structure, it would be interesting to observe the Wzm aromatic belt residues in close proximity while the NBDs are separated and open, as this may reveal additional flexibility of the Wzm TM2-6 rigid body at the CHL relative to its cognate Wzt NBD. Wzm-Wzt

coupling may then be more complex than initially proposed with a coupling mechanism specific to a channel-occupied pathway and a channel-closed pathway.

Lastly, the cryo-EM structure of ATP- and Mg^{2+}-bound full-length WzmWzt represents the only available positional information for the Wzt CBD dimer in the context of the complete WzmWzt transporter (**Figs. 3-1, 3-4, and 3-5**). The number and variety of contacts found in the Wzt NBD-CBD interaction between the Wzt NBD hinge helix and the Wzt CBD dimer central loops suggest thus far that the Wzt NBD-CBD interaction is strong, stable and persistent throughout the ATP hydrolysis cycle. However, the linker regions that connect each Wzt NBD C-terminus and Wzt CBD N-terminus have an unknown role in the opening and closing of the full Wzt dimer, and importantly, the rigidity or flexibility of these linker regions is unknown. A modeled nucleotide-free (apo) WzmWzt with predicted secondary structure for these linker regions suggests that the linkers must be flexible to allow the Wzt NBDs to open to the extent observed in the nucleotide-free WzmWzt crystal structure PDB 6OIH (**Chapter 3 SI, Fig. S3-8**). However, if the linker regions exhibit some rigidity, such as through helical secondary structure, this could be sufficient to perturb the interaction between the Wzt NBD hinge helix and the Wzt CBD dimer, causing transient dissociation of the Wzt CBD during ATP hydrolysis-induced NBD movements.

Determination of the structure of nucleotide-free (apo) full-length WzmWzt by cryo-EM is an important aim for answering these open questions about conformational changes of WzmWzt during the ATP hydrolysis cycle. The demonstration of the general protocol used

for determining the cryo-EM structure of ATP- and Mg^{2+}-bound WzmWzt in lipid nanodiscs substantiates the feasibility of an apo state structural study, making this a clear direction for future WzmWzt research.

4.2.1.2 Exploring the conformational space of WzmWzt during ATP hydrolysis

The nucleotide-free (89) and ATP-bound (34) crystal structures of WzmWzt established two positions in the conformational space for one ATP hydrolysis cycle of O antigen-occupied WzmWzt (**Fig. 3-6**). Similarly, the cryo-EM structure of ATP- and Mg^{2+}-bound WzmWzt (**Fig. 3-1**), combined with a potential nucleotide-free WzmWzt cryo-EM structure would establish two positions in a conformational space for the ATPase cycle of WzmWzt in the absence of O antigen substrate. The revised model of O antigen translocation by WzmWzt (**Fig. 3-6**) depicts four discrete conformational states in each O antigen-occupied and lipid-occupied ATP cycle, yet the total number of conformational states maybe be larger. Further, the same protein conformation may be observed for different nucleotide configurations, such as the ATP-bound state and the vanadate-trapped, post-hydrolysis ADP-Vi-bound state. Indeed, a structural study of the nanodisc-reconstituted peptide transporter TmrAB from *Thermus thermophilus* [a model system of the human transporter associated with antigen processing (TAP)] showed that TmrAB not only adopts two very distinct outward-facing conformations (termed open and occluded), but each conformation was observed twice, one with ATP bound at both NBD sites and another in a vanadate (Vi)-trapped, mixed ATP/ADP-Vi configuration. (221).

Applying an approach similar to that used for TmrAB would establish a matrix of nucleotide configurations to test in structural studies. First, trapping a post-hydrolysis state may be achieved by incubating WzmWzt with a combination of ATP, Mg^{2+} and Vi. It was previously shown that ADP does not replace ATP in the trapping reaction, suggesting that ATP hydrolysis to ADP-Pi occurs first and Vi rapidly replaces Pi, preserving a post-hydrolysis state (223, 242). One caveat to this sample condition is that in the case of both the TAP transporter TmrAB and three ABC importers ($MalFGK_2$, $BtuC_2D_2$ and human P-glycoprotein), the Vi-trapped state was found to be structurally equivalent to the ATP-bound, catalytically inactive Walker B mutant of the transporter (221, 223, 243, 244).

An alternative experiment to Vi-trapping that was successful in the case of TmrAB is the use of "turnover conditions" where the transporter is incubated with ATP and Mg^{2+} allowing free hydrolysis immediately prior to freezing the sample for cryo-EM data collection. Remarkably, four distinct conformations of TmrAB could be recovered from a single cryo-EM dataset through 3D classification of particle images, including two apo-state inward-facing conformations and two asymmetric conformations representative of post-Pi-release states (221). One interesting observation is that the two nucleotide-free, apo-state inward-facing conformations could be recovered at all, considering that active ATP hydrolysis is occurring at the time the sample is frozen for cryo-EM analysis. Additionally, the EM maps determined for these inward-facing conformations were of higher resolution and better quality than those determined from nucleotide-free protein samples. A second interesting observation is that the two asymmetric post-Pi-release conformations were modeled as ADP-bound transporters. It should be noted that TmrAB

is a heterodimeric ABC transporter with one protomer containing a degenerate ATP binding site that does not catalyze ATP. Therefore, ADP was modeled at only the catalytically active site, yet is representative of an ADP-bound intermediate state following Pi release (221). The success of the "turnover conditions" approach to yield structures of previously inaccessible positions in the conformational landscape of TmrAB establishes this experiment as an important future direction for testing the proposed model of O antigen translocation by WzmWzt.

4.2.1.3 Probing the effects of NBD asymmetry using heterodimeric Wzt constructs

The Wzt NBD-CBD linkers may induce asymmetry between the NBD protomers as they open following post-hydrolysis dissociation of Pi and ADP. Prior to this larger movement of NBD opening, bound ATP may be hydrolyzed at a different time for each NBD ATP-binding site, creating mixed nucleotide configurations for the two binding sites (ATP/ATP, ATP/ADP-Pi, ATP/ADP, ADP-Pi/ADP, ADP/ADP). These mixed nucleotide configurations may then induce conformational asymmetry in the NBD dimer, asymmetry that could then induce conformational changes in the Wzm dimer through coupling at the CHL.

The "turnover conditions" structural study proposed in **Section 4.2.1.2** could capture such mixed nucleotide configurations as ATP hydrolysis is occurring in the sample immediately prior to freezing onto grids for cryo-EM analysis. An alternative approach is to encourage staggered ATP hydrolysis by trapping ATP in one ATP-binding site while permitting ATP hydrolysis in the other ATP-binding site. This can be achieved by purifying a WzmWzt

transporter with a heterodimeric Wzt, with one Wzt protomer expressed as a WT construct and a second Wzt protomer co-expressed as a Walker B motif E167Q point mutant of the catalytic glutamate.

Under typical expression and purifications conditions, Wzm and Wzt are co-expressed from a pETDuet-based plasmid and natively form the tetrameric WzmWzt complex that is then purified via the His-tag on the C-terminus of Wzt using Ni-NTA-based IMAC (34, 89). Cloning of the Wzt Walker B mutant construct into a vector containing a different origin of replication, such as pACYCDuet or pET28, and replacing the C-terminal His-tag with an alternative tag of comparable length, such as a StrepII-tag, would generate a second Wzt construct for co-expression with WT His-tagged Wzt. Then, a two-column affinity purification method could be applied to capture a transporter population that contains one His-tag and one StrepII-tag, yielding a WzmWzt sample with a heterodimeric Wzt for analysis of ATPase activity and structure determination by crystallography or cryo-EM.

4.2.1.4 Studying WzmWzt dynamics using FRET

Förster or fluorescence resonance energy transfer (FRET) utilizes energy transfer from a donor fluorophore to an acceptor, which may or may not be fluorescent, thereby reporting on the distance between the positions of the donor and acceptor pair. Upon excitation with light, the donor transfers energy to the acceptor, and the efficiency of the energy transfer decreases with the sixth power of the distance between the donor and acceptor, serving as a sensitive measure of intermolecular distance (245).

As a complimentary approach to structural biology, when a donor and acceptor are conjugated to different positions within a protein, FRET can report on the dynamic movements of the protein as it undergoes conformational changes (246). In parallel with cryo-EM analysis of WzmWzt, FRET experiments could be utilized to study the dynamics of the regions within WzmWzt that may undergo conformational changes, dependent upon the distances involved. FRET is most effective when the distances of interest are close to the value of R_0, a donor- and acceptor-dependent distance at which the efficiency of energy transfer is 50% (245). Single molecule FRET (smFRET) studies of the glutamate transporter elevator mechanism using the homologous *Pyrococcus horikoshii* sodium/aspartate symporter Glt_{Ph}, employed two cyanine dyes, Cy3 and Cy5 (R_0 = ~60 Å), to monitor distance changes in the range of 40-60 Å (246, 247).

The Cy3 and Cy5 dyes require conjugation to cysteine side chains and three main criteria were used in the Glt_{Ph} study to select donor and acceptor positions for mutation to cysteine: sites should exhibit low sequence conservation to allow introduction of non-native cysteines for fluorophore conjugation, should be solvent exposed, and should have expected distances changes of > 20 Å (246). Using this same set of criteria, two candidate sites were selected for a smFRET experiment to study Wzt NBD opening and closing following the same overall protocol of the Glt_{Ph} smFRET experiment, including surface immobilization for TIRF microscopy (246). Wzt NBD residue Glu123 and Wzt CBD native cysteine Cys367 are separated by a ~45 Å distance in the cryo-EM structure of ATP and Mg^{2+}-bound full-length WzmWzt (**Fig. 4-1**). Upon ATP hydrolysis, dissociation of ADP and Pi allows the NBDs to open. The nucleotide-free apo-state *in silico* model of full-

Fig. 4-1

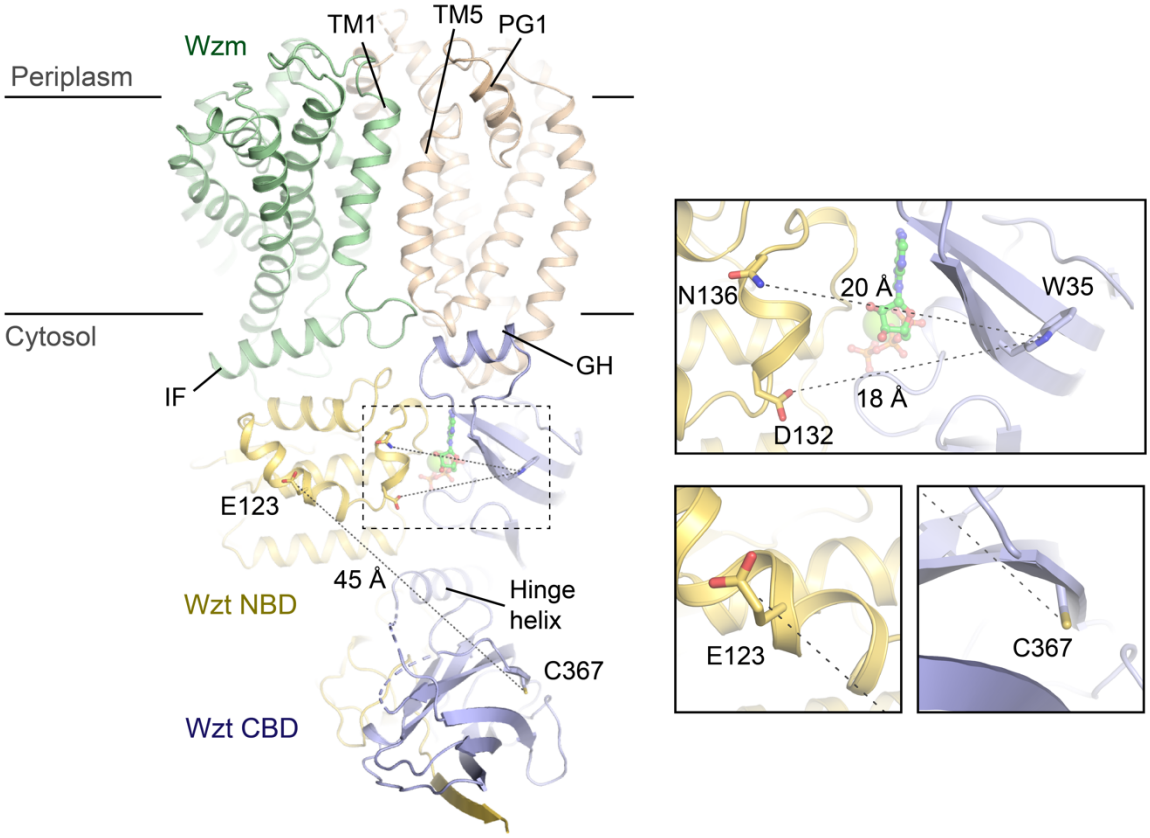


Fig. 4-1 Potential locations of FRET pairs to study WzmWzt protein dynamics. Atomic model of ATP/Mg^{2+}-bound full-length Aa WzmWzt colored as in Fig. 3-1. ATP is shown as sticks and magnesium is shown as a green sphere. IF: Interface helix, PG1: Periplasmic gate helix 1, GH: Gate helix. Inset (top), zoomed-in view of the NBD-NBD interface showing candidate side chains for a tmFRET pair. Inset (bottom), zoomed-in view of candidate side chains for chemical fluorophore FRET pair.

length WzmWzt based on the apo-state WzmWzt crystal structure (PDB 6OIH) suggests the Glu123-Cys367 distance to be ~48 Å (**Fig. S3-8**). If NBDs open wider in a nucleotide-free, closed-channel resting state, this new distance may be captured by the smFRET experiment considering that an interprotomer distance range of 45-60 Å range was expected for Glt_{Ph} and discrete FRET states could be observed in the smFRET experiment (246).

A major caveat of the cyanine dye approach is the need to generate cysteine-free constructs, ensuring that the dye is conjugated only to the cysteine at the target site. Further, due to the dimeric nature of Wzt, a heterodimer must be generated according to the description in **Section 4.2.1.3**, such that the transporter has one FRET pair, with one dye per protomer. One of the target sites proposed for WzmWzt was the native CBD cysteine Cys367 to mitigate the number of non-native manipulations of the construct. However, the stability and structural integrity of a Glu123Cys mutant in the background of an otherwise cysteine-free WzmWzt would have to be established prior to any smFRET experiments.

If it is determined that the approach of using traditional FRET dyes, such as Cy3 and Cy5, will not be appropriate for WzmWzt, another mode of FRET called transition metal ion FRET (tmFRET) that uses transition metals such as nickel, copper or cobalt as non-fluorescent acceptors may be an appropriate alternative experiment (248). An especially interesting application of tmFRET utilized a fluorescent, non-natural amino acid called 3-(6-acetylnaphthalen-2-ylamino)-2-aminopropanoic acid (Anap) as a fluorescent donor paired with cobalt as a non-fluorescent acceptor (249). The R_0 distances for tmFRET pairs

are substantially shorter (~10-20 Å) than traditional FRET R_0 distances (~30-60 Å) and provide an opportunity to monitor distance changes for positions that are closer together than classical FRET experiments. Binding of the acceptor transition metal can be achieved by mutating two non-conserved residues of an α-helix one turn apart to histidines thereby trapping the metal (248). Two candidate Wzt residues for histidine mutation, Asp132, Asn136, and one candidate residue for Anap incorporation are proposed for tmFRET experiments to analyze NBD opening and closing based on an interprotomer distance of ~15-20 Å (**Fig. 4-1**). As before, a heterodimeric Wzt expression and purification protocol and a TIRF-based smFRET experimental setup would be employed to analyze Wzt NBD dynamics.

Ultimately, one caveat of smFRET is the timescale involved in NBD opening and closing during ATP hydrolysis and whether distinct states can be determined from the data. Success in observing clear discrete FRET states has come when the target protein dwells in individual states for periods of time on the order of seconds (246, 250, 251). However, a recent study of bovine multidrug resistance protein 1 (MRP1), a multidrug ABC exporter, used smFRET to monitor conformational changes of the transporter during ATP hydrolysis at a 300 millisecond time resolution (252). The success of the MRP1 study suggests that this protocol could be applied to WzmWzt to gather data on transporter dynamics to complement the available structural data.

4.2.2 Biochemical and biophysical studies in the presence of O antigen substrate or substrate analog

Our current understanding of how the lipid-linked O antigen interacts with WzmWzt during transport is grounded in the synergy of detailed *in vivo* and *in vitro* experiments of lipid-linked O antigen transport and high-resolution structural data. These combined biophysical and biochemical approaches have enabled mapping residues critical for efficient O antigen transport to the WzmWzt structure uncovering numerous mechanistic details (76, 89, 90, 212).

One of the greatest challenges thus far in WzmWzt research is generating data in the presence of a lipid-linked O antigen substrate, or substrate analog such as a lipid-linked oligosaccharide or a lipid-free O antigen polysaccharide. However, generating new data toward understanding how the lipid-linked O antigen interacts with WzmWzt to initiate transport and move through WzmWzt as it is translocated across the inner membrane represents the cutting edge of the research field. The following sections will explore various future directions for WzmWzt that may enable uncovering the nature of these transporter-substrate interactions.

4.2.2.1 Reconstitution of WzmWzt into nanodiscs doped with isoprenyl-linked substrate analogs

The reconstitution of WzmWzt into lipid nanodiscs, presented in **Chapter 3,** is a method that can be applied to future biochemical and biophysical studies to co-reconstitute WzmWzt in a lipid environment with an isoprenyl-linked O antigen substrate or minimal

substrate analog, such as an isoprenyl-PP-GlcNAc. The advantage of this approach is two-fold, the isoprenyl moiety is surrounded by phospholipids, more closely mimicking the environment of a native membrane, and the substrate is restricted to the immediate vicinity of the protein, eliminating the limitations of diffusion and formation of mixed micelles. Further, if an interaction exists between the membrane-facing turns of the transmembrane helices and the substrate's isoprenyl lipid anchor, it is unclear if the chosen detergent disrupts the hydrophobic interactions between protein and isoprenyl lipid, such as could be the case for a detergent such as LDAO.

Such an interaction between the transmembrane helices and the isoprenyl units of the substrate's lipid anchor was observed for the bacterial *N*-linked glycosylation oligosaccharyltransferase (OST) PglB. *Campylobacter lari* PglB was co-crystallized with its two substrates, a hexapeptide and a synthetic, non-hydrolyzable lipid-linked oligosaccharide (LLO) analog nerylneryl-PPC-GlcNAc (32). Within the final map, density was observed for the entire nerylneryl-PPC-GlcNAc molecule, revealing that the four isoprene units of the nerylneryl tail fit into a hydrophobic groove within the PglB transmembrane helices. The native LLO is anchored to Und-P (11 isoprene units) with the PglB crystal structure suggesting that at least four isoprene units contribute to the binding of the LLO donor to the PglB active site (32).

Reconstitution of WzmWzt into lipid nanodiscs doped with isoprenyl-linked substrate analogs will also enable structural studies of the interactions of the PP-GlcNAc priming

sugar headgroup of O antigens, with the WzmWzt cytosolic channel entrance, an initial step in O antigen transport (34, 89).

The crystal structure of peptide- and LLO-bound PglB showed that residues of the LLO binding site form contacts with a PP-GlcNAc headgroup that are entirely electrostatic (Arg) or polar (Ser, Gln, and Tyr) in nature (32). In WzmWzt, a similar cluster of residues on Wzt (Lys12, Tyr14, Arg20, His31) located at the interface of the Wzt GH, Wzm TM1, Wzm TM5, and Wzm TM6 were proposed to be the interaction site of the isoprenyl-linked O antigen's PP-GlcNAc headgroup, and to serve as a cytosolic gate for entry to the Wzm channel (**Fig. 2-1**) (34). The importance of the conserved GH residue Tyr14 in *A. aeolicus* WzmWzt was demonstrated with *in vivo* translocation assays using the *E. coli* O9a model system. Mutations of *E. coli* Tyr15 (*E. coli* O9a Tyr15 is equivalent to *A. aeolicus* Tyr14) had a direct effect on O antigen transport, substantiating the role for the cytosolic gate as a substrate entry site (89).

A third important aspect of determining the structure of WzmWzt in the presence of a substrate or substrate analog is the potential for induction of a conformational change due to interactions with the substrate molecule. Indeed, in the hexapeptide acceptor- and LLO-bound PglB crystal structure, an external loop (EL5) is highly ordered and exhibits secondary structure at the substrate binding sites (32). A previous crystal structure of PglB co-crystallized with a hexapeptide acceptor, but without the nerylneryl-PPC-GlcNAc LLO substrate analog, showed an ordered region of EL5 at the occupied hexapeptide binding site, but a completely disordered region of EL5 at the unoccupied LLO binding site (33).

Similarly, the structure of the orthologous archaeal OST AglB determined from one of two crystal forms showed that EL5 is completely disordered in the absence of substrates (253). Structural studies of WzmWzt in the presence of substrate will thus be important to assess whether conformational changes are induced upon interaction with substrate, such as a change in the position or structure of the Wzt GH at the Wzm TM1-TM5 interface.

For forthcoming structural studies, co-reconstitution of WzmWzt and an isoprenyl-linked O antigen substrate or substrate analog into the nanodisc lipid environment is an important sample preparation method. Insertion of a substrate into the immediate lipid environment surrounding WzmWzt creates a unique opportunity to simultaneously gather data on three open questions, isoprenyl lipid-Wzm interactions, PP-GlcNAc and Wzt GH interactions, and movement or conformational changes of domains, such as the Wzt GH, induced by substrate. Synthetic isoprenyl-linked-PP-GlcNAc substrate analogs can be synthesized and purified, and methods to extract isoprenyl-linked substrates, such as Lipid II, from native sources have been developed, establishing multiple routes for developing a WzmWzt-substrate co-reconstitution procedure for future structural and biochemical studies (32, 254).

4.2.2.2 Time-resolved cryo-EM

Recent developments in alternative cryo-EM sample preparation procedures have decreased the time from sample application to the EM grid to plunge freezing down to the millisecond timescale. Further improvements of the same approach have led to the

development of rapid on-grid mixing immediately prior to freezing, allowing the capture of sample states occurring 50-500 milliseconds after component mixing (255).

In the case of a "turnover conditions" structural study or heterodimeric Wzt structural study, as there is on-going ATP hydrolysis with turnover on the order of the microsecond timescale, time-resolved cryo-EM at the millisecond timescale may not yield different results as compared with traditional blotting-plunge freezing cryo-EM sample preparation in a substrate-free sample.

In a sample containing WzmWzt co-reconstituted into lipid nanodiscs with an isoprenyl-linked O antigen substrate, however, a time-resolved cryo-EM experiment could be advantageous. As described in **Section 4.2.2.1**, a number of interactions may be observed in a structural study of WzmWzt in the presence of a substrate, yet this discussion focused on pre-translocation interactions between WzmWzt and substrate. Using a sample comprised of WzmWzt in the presence of a substrate, by adding ATP to begin hydrolysis and immediately freezing the sample, it could be possible to gather structural data on the ATP hydrolysis-induced transport events that occur following the initial substrate interaction, toward capturing a WzmWzt structure mid-translocation. For example, if the substrate is an isoprenyl-pyrophosphate-linked oligosaccharide or polysaccharide, a time-resolved cryo-EM experiment could be crucial for observing the sugar moiety within in the Wzm channel.

4.3 Concluding remarks

Gram-negative bacteria employ cell surface polysaccharides to evade host immune responses, enhance survival in extreme environments, and protect themselves against antimicrobial agents. The outer membrane of Gram-negatives is rich in LPS, comprised of the endotoxin lipid A, a short oligosaccharide, and a long terminal polysaccharide known as the O antigen. O antigens extend the cellular envelope creating a barrier that obscures the cell membrane to protect it from complement component deposition and insertion, thereby evading the innate immune response of the host. Understanding O antigen biosynthesis and how the specialized ABC transporter WzmWzt translocates the O antigen for assembly of mature LPS provides a molecular basis for manipulating this fundamental process in Gram-negative pathogens.

More broadly, Gram-negatives, Gram-positives, and mycobacteria all expose a vast array of complex carbohydrates on their surfaces, with only a small fraction biochemically characterized to date. Accordingly, diverse biosynthesis and secretion pathways evolved to ensure the deposition of the polymers on the cell surface, yet their underlying enzymatic and transporter mechanisms are only beginning to emerge. In addition, how complex carbohydrates interact with other cell-surface polymers to form native cell walls is an important future research direction requiring integration of biology, physics, and material sciences.

5 Appendices

5.1 Appendix 1: WzmWzt expression and purification

5.1.1 Materials

Table 4 WzmWzt expression and purification materials list

Solution	Suggested Volume	Composition
Lysogeny Broth (LB)	12 L	10 g/L tryptone, 5 g/L yeast extract, 10 g/L sodium chloride
100mg/mL IPTG	13 mL	100mg/mL IPTG in Milli-Q water
100mg/mL ampicillin	13 mL	100mg/mL ampicillin in Milli-Q water
Running Buffer (RB)	500 mL	20 mM Tris HCl pH 7.5, 100 mM NaCl, and 5 mM β-mercaptoethanol
Solubilization Buffer (SB)	250 mL (reserve 35 mL to equilibrate Ni-NTA)	50 mM sodium phosphate pH 7.2, 100 mM NaCl, 20 mM imidazole, 5 mM β-ME, and 2% (w/v) C12E8 (Anatrace)
500mM PMSF	1 mL	500mM PMSF in DMSO
Wash Buffer 1 (WB1)	50 mL	RB with 22 mM imidazole, 5 mM LDAO (Anatrace)
Wash Buffer 2 (WB2)	50 mL	RB buffer with 40 mM imidazole, 5 mM LDAO
Wash Buffer 3 (WB3)	50 mL	RB buffer with 1.5 M NaCl, 22 mM imidazole, 5 mM LDAO
Elution Buffer (EB)	50 mL	RB with 300 mM imidazole, 5 mM LDAO
Dialysis Buffer (DB)	500 mL	20 mM Tris pH 7.5, 100 mM NaCl, 5 mM β-ME, 5 mM LDAO
Gel Filtration Buffer (GF)	200 mL	20 mM Tris HCl pH 7.5, 100 mM NaCl, 0.5 mM TCEP, 5 mM LDAO

5.1.2 Protocol

Freshly transform E. coli C43 cells with pETDuet-WzmWzt plasmid (~50ng).

Inoculate LB (50-100 ml) supplemented with 100 mg/L ampicillin in a baffled flask, and shake overnight at 37°C.

Inoculate 12L of LB medium supplemented with 100 mg/L ampicillin and shake the flasks at 220 rpm at 37°C until the OD600 value reaches 0.6-0.8.

Induce protein production by the addition of 100 mg/L IPTG (from 100 mg/mL stock in Milli-Q water).

Let the culture shake for 3-4 hours at 220 rpm and 37°C.

Harvest cells by centrifugation at 4500 rpm for 15 minutes in 1L bottles at 4°C.

Decant the supernatant and scrape the wet cell paste into a 50mL vial. The cell pellet can be stored at -80°C.

Alternatively, prepare RB (see Materials) buffer and 500mM PMSF.

Resuspend the cell pellet in 300mL of Running Buffer (RB) and pass through the microfluidizer to begin lysing cells. After the first pass through the microfluidizer, add 600uL 500mM PMSF (target 1mM PMSF).

Pass sample through the microfluidizer three additional times (four total) to complete cell lysis.

Transfer sample to six Ti45 tubes and ultracentrifuge the lysate at 42,000 rpm for 60 min at 4°C using the Ti45 rotor.

After centrifugation, discard the supernatant and scrape the membrane pellet into a 50mL vial. Flash freeze in liquid N_2 and store at -80°C.

When needed, thaw 12L membrane pellet in 30°C water bath for 10 minutes.

Resuspend the membrane pellet in 200-215mL solubilization buffer (SB). Add 400uL 500mM PSMF to target 1mM PMSF.

Solubilize the membrane pellet for a minimum of 1 hour (maximum 3 hours) at 4°C with rocking.

Transfer sample to four Ti45 tubes and ultracentrifuge the sample at 42,000 rpm for 30 min at 4°C using the Ti45 rotor.

Prepare a bed volume of 3.5mL Ni-NTA beads by thoroughly washing with Milli-Q water.

Equilibrate the Ni-NTA with 10 column volumes (CV) of SB buffer.

After ultracentrifuging the solubilization sample, combine the supernatant with the Ni-NTA resin and rock the sample for 1 hour at 4°C to batch bind.

Transfer the supernatant/Ni-NTA slurry to a glass column and drain the flowthrough.

Add 50mL WB1, mix and drain.

Add 50mL WB2, mix and drain.

Add 50mL WB3, mix and drain.

Add 35mL EB, collect the elution sample.

Transfer sample to Amicon-15 concentrator(s) with 100 kDa molecular weight cutoff and concentrate to 10mL (for two gel filtration runs on 10/300 S200 column).

Prepare Dialysis Buffer (DB).

Transfer the 10mL sample to a SnakeSkin (ThermoFisher) dialysis bag prepared with clamps.

Dialyze the protein sample against 150mL DB for 1 hour.

Dialyze the protein sample against 350mL DB overnight. Removal of imidazole will precipitate free Wzt. A large amount of white precipitate will appear over time. Intact WzmWzt is unaffected.

Transfer sample to 1mL vials and centrifuge to remove precipitation.

Transfer sample to Amicon-15 concentrator(s) with 100 kDa molecular weight cutoff and concentrate to 1000uL (for two gel filtration runs on 10/300 S200 column).

Prepare the GF buffer and equilibrate a 10/300 S200 column.

Filter each 500uL concentrated protein sample with a Costar Spin-X filter.

Inject protein sample onto 10/300 S200 column and collect 0.5mL fractions under the main peak.

Run SDS-PAGE to verify WzmWzt location and purity.

Transfer fraction pool to Amicon-4 concentrator with 100 kDa molecular weight cutoff and concentrate to desired protein concentration.

5.1.3 Representative SEC chromatogram of full-length WT WzmWzt in 5mM LDAO detergent

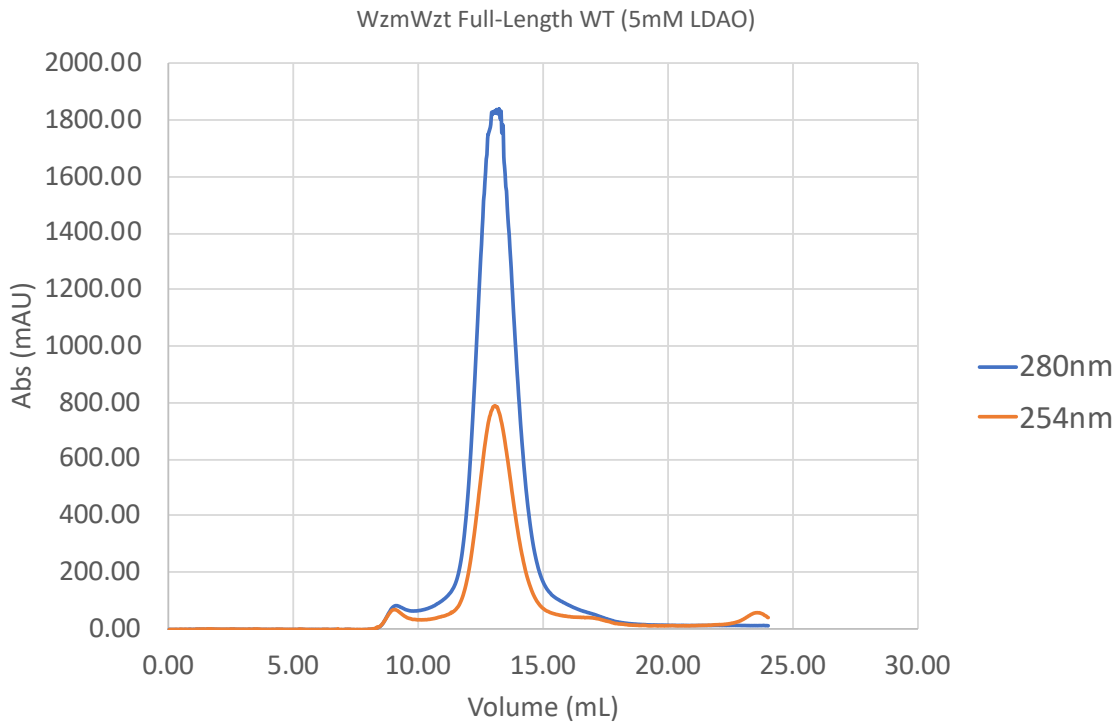


Fig. A-1 SEC purification of full-length WT WzmWzt in 5mM LDAO detergent. Representative chromatogram from purification of full-length WT WzmWzt in 5mM LDAO using a Superdex 200 (S200) 10/300 SEC column. The void volume (V_0) for this column is approximately 8 mL. The absorbance traces at 280 nm and 254 nm are shown in blue and red, respectively.

5.1.4 Representative SDS-PAGE of full-length WT WzmWzt in 5mM LDAO detergent

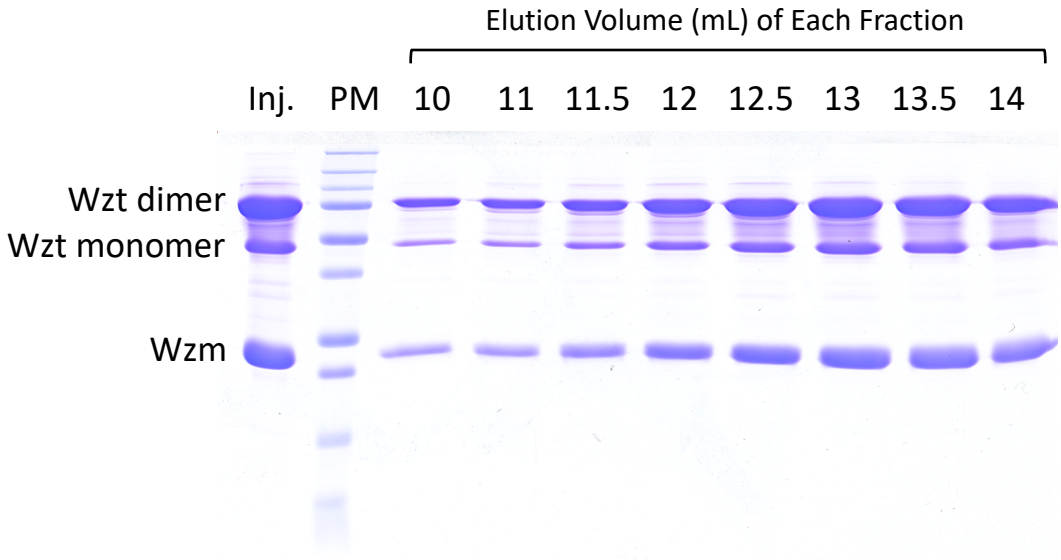


Fig. A-2 SDS-PAGE of full-length WT WzmWzt purified in 5mM LDAO detergent. Representative Coomassie-stained SDS-PAGE of full-length WT WzmWzt when purified on a S200 10/300 column in 5mM LDAO detergent as shown in Figure A-1. The sample to the left of the protein marker (PM) is the sample injecting into the S200 column. The samples to the right of the PM are 0.5 mL fractions collected from the eluate.

5.2 Appendix 2: Membrane Scaffold Protein (MSP) expression and purification [adapted from (227)]

See next page.

5.2.1 Materials

Table 5 MSP expression and purification materials list

Solution	Suggested Volume	Composition
Lysogeny Broth (LB)	12 L	10 g/L tryptone, 5 g/L yeast extract, 10 g/L sodium chloride
100mg/mL IPTG	13 mL	100mg/mL IPTG in Milli-Q water
50mg/mL kanamycin	13 mL	50mg/mL kanamycin in Milli-Q water
Basal Buffer	500 mL (Reserve 50mL for WB3)	40mM Tris pH 8.0, 500mM NaCl, 10mM imidazole pH 7.5
Basal Buffer + 1% Triton X-100	250 mL (Reserve 50mL for WB1)	40mM Tris pH 8.0, 500mM NaCl, 10mM imidazole pH 7.5, 1% Triton X-100
Lysozyme stock	15 mL	10 mg/mL lysozyme
500mM PMSF	1 mL	500mM PMSF in DMSO (store at RT, do not put on ice)
1M MgCl2	As needed	1M MgCl2 (optional)
DNase powder	Small amount, enough to fit on spatula tip	Dnase powder (optional)
Wash Buffer 1 (WB1)	50 mL	40mM Tris pH 8.0, 500mM NaCl, 10mM imidazole pH 7.5, 1% Triton X-100
Wash Buffer 2 (WB2)	50 mL	40mM Tris pH 8.0, 500mM NaCl, 10mM imidazole pH 7.5, 50mM sodium cholate (50mL Basal Buffer + 1.08g sodium cholate)
Wash Buffer 3 (WB3)	50 mL	40mM Tris pH 8.0, 500mM NaCl, 10mM imidazole pH 7.5
Wash Buffer 4 (WB4)	50 mL	40mM Tris pH 8.0, 500mM NaCl, 50mM imidazole pH 7.5 (50mL Basal Buffer + 1mL 2M imidazole pH 7.5)
Elution Buffer (EB)	50 mL	40mM Tris pH 8.0, 500mM NaCl, 500mM imidazole pH 7.5
Dialysis Buffer 1 (DB1)	2000 mL	40 mM Tris-HCl pH 8.0, 100 mM NaCl, 10mM imidazole pH 7.5, 0.5mM EDTA pH 8.0
Dialysis Buffer 2 (DB2)	2000 mL	20 mM Tris pH 8.0, 500mM NaCl, 10mM imidazole pH 7.5
Reverse IMAC Equil/Wash Buffer (RIWB)	100 mL	20mM Tris pH 8.0, 500mM NaCl, 10mM imidazole pH 7.5
Reverse IMAC Elution Buffer (RIEB)	50 mL	20mM Tris pH 8.0, 500mM NaCl, 500mM imidazole pH 7.5

5.2.2 Protocol

5.2.2.1 Expression of MSP

Freshly transform E. coli BL21 (DE3) cells with pET28a-MSP plasmid (~50ng).

Inoculate LB (50-100 ml) supplemented with 50 mg/L kanamycin in a baffled flask, and shake overnight at 37°C.

Inoculate 12L of LB medium supplemented with 50 mg/L kanamycin and shake the flasks at 220 rpm at 37°C until the OD600 value reaches 0.6–0.8.

Induce protein production by the addition of 1 mM IPTG (from 1M stock in Milli-Q H2O).

Let the culture shake for 3–4 hours at 220 rpm and 37°C. NOTE: MSP degradation is a problem after 3-4 hours of induction, limit induction time to 3-4 hours maximum to limit proteolysis.

Harvest cells by centrifugation at 4500 rpm for 15 minutes in 1L bottles at 4°C.

Decant the supernatant and scrape the wet cell paste into a 50mL vial. The expected yield should be ≥2 g of wet cell paste per liter of culture. Store the cell pellet at -80°C.

5.2.2.2 Purification of MSP

Prepare 500mL Basal Buffer (40mM Tris pH 8.0, 500mM NaCl, 10mM imidazole pH 7.5). Reserve 150mL for Wash Buffer 1, 2 and 3. Store at 4°C.

Prepare 250mL Basal Buffer + 1% Triton X-100. Reserve 50mL for Wash Buffer 1 (WB1). Store at 4°C.

Prepare 15mL of 10mg/mL lysozyme stock in Milli-Q H2O. Store at 4°C.

Prepare 1mL of 500mM PMSF in DMSO. Store at room temperature (RT); do NOT store at 4°C, it will freeze.

Thaw the frozen cell pellet at RT for 30 min or in a 30°C for 10 minutes then keep on ice.

Resuspend the cell pellet in ~4-5 ml of Basal Buffer per gram of cell paste using a Dounce homogenizer (e.g. adding 100mL-125mL Basal Buffer to 26 grams wet cell paste will yield a final resuspension volume of ~130mL).

When ready to microfluidize, add 1mM PMSF (e.g. 300uL of 500mM PSMF to 130mL) and 1mg/mL lysozyme (e.g. 12.5mL of 10mg/mL lysozyme to 130mL) to the cell resuspension and mix.

Microfluidize the cell resuspension. After 1 pass through the microfluidizer add 1mM PMSF again. Perform two additional passes.

After the final pass through the microfluidizer, dilute the sample 1:1 with Basal Buffer + 1% Triton X-100. Do NOT put Triton X-100 through the microfluidizer. Dilution after microfluidizing prevents putting Triton X-100 detergent through the microfluidizer protection other users' samples from mixing with detergent.

Add remaining 500mM PMSF.

OPTIONAL: Add 5mM magnesium chloride, and spatula tip of DNase powder to degrade DNA.

Ultracentrifuge the lysate at 42,000 rpm for 30 min at 4°C using the Ti45 rotor. Save the lysate supernatant for batch binding to Ni-NTA beads.

Prepare Ni-NTA beads (~1 mL of resin/5 g of wet cell paste) and equilibrate with Basal Buffer + 1% Triton X-100.

Apply the lysate supernatant to the Ni-NTA beads and batch bind for 60 min at 4°C.

Prepare Wash Buffer 2 (WB2; 50mL Basal Buffer + 1.08g sodium cholate).

Prepare Wash Buffer 4 (WB4; 50mL Basal Buffer + 1mL 2M imidazole pH 7.5).

Prepare Elution Buffer (EB; 40mM Tris pH 8.0, 500mM NaCl, 500mM imidazole pH 7.5).

Transfer the lysate supernatant/Ni-NTA bead slurry to a glass column. Completely drain the flow through, but do not let the beads dry out.

Add 10 column volumes (CV) e.g. 50mL WB1 and stir the beads. Drain WB1.

Add 10 CV (50mL) WB2 and stir the beads. Drain WB2.

Add 10 CV (50mL) WB3 and stir the beads. Drain WB3.

Add 10 CV (50mL) WB4 and stir the beads. Drain WB4.

Elute the protein with 4 CV (20mL) EB.

Dialyze the elution sample overnight against 1.5-2.0L of Dialysis Buffer 1 (DB1; 40 mM Tris-HCl pH 8.0, 100 mM NaCl, 10mM imidazole pH 7.5, 0.5 mM EDTA pH 8.0) using a 6- to 8-kDa MWCO membrane.

If you will NOT cleave the histidine tag, you may concentrate the post-dialysis MSP sample up to 500–600 μM, save in 20-30uL aliquots, and flash-freeze it in liquid N_2. Can be stored at -80°C for up to 1 year. NOTE: If you do not cleave the His tag, then, the previous dialysis step can use a buffer more appropriate to your target protein buffer such as Tris pH 7.5).

5.2.2.3 Cleavage of the MSP histidine tag

To cleave off the His tag, add TEV protease (following a ratio of 1 unit A280 absorbance for TEV to 100 units A280 absorbance for MSP protein) and 1 mM DTT, and incubate the protein with agitation (rocking) for 4–12 h at 4°C.

Dialyze the protein solution overnight against 2L of Dialysis Buffer 2 (DB2; 20 mM Tris pH 8.0, 500mM NaCl, 10mM imidazole pH 7.5*) using a 6- to 8-kDa MWCO membrane.

NOTE: MSP will stick to the Ni-NTA resin if there is no imidazole during batch binding. *Also note that MSP2N2 with its additional repeats may bind more strongly to Ni-NTA than smaller MSP constructs and thus may require a higher background imidazole concentration than 10mM to prevent non-specific binding.

Prepare Ni-NTA resin (~same column volume as during initial purification) and equilibrate with Reverse IMAC Equilibration/Wash Buffer (RIWB; 20mM Tris pH 8.0, 500mM NaCl, 10mM imidazole pH 7.5).

Apply the post-dialysis protein sample to the Ni-NTA resin and batch bind for 60 min at 4°C.

Transfer the protein sample/Ni-NTA bead slurry to a glass column. Completely drain AND COLLECT the flow through containing MSP, but do not let the beads dry out. Save the flowthrough sample.

Wash the beads with RIWB and collect wash fractions.

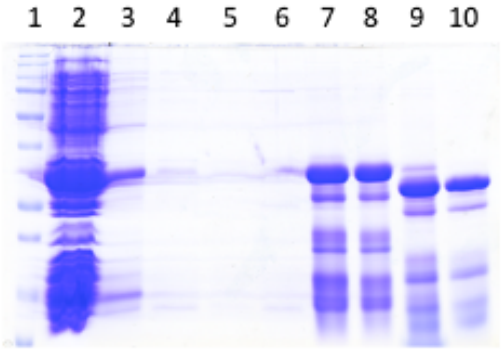
Elute any protein bound to the Ni-NTA resin with Reverse IMAC Elution Buffer (RIEB; 20mM Tris pH 8.0, 500mM NaCl, 500mM imidazole pH 7.5).

Save wash and elution fractions to check for residual MSP by SDS-PAGE.

Dialyze MSP-containing flowthrough sample overnight against 2L of Dialysis Buffer 3 (DB3; 20mM Tris pH 7.5, 100mM NaCl, 0.5mM EDTA pH 8.0).

Concentrate the post-dialysis MSP sample up to 500–600 μM, save in 20-30uL aliquots, and flash-freeze it in liquid N_2. Can be stored at -80°C for up to 1 year.

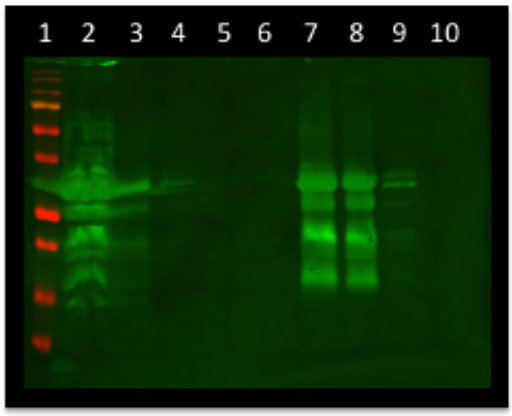
5.2.2.4 Representative SDS-PAGE for MSP1E3D1 protein preparation



1. PM
2. Pre-NiNTA supernatant
3. WB1 FT
4. WB2 FT
5. WB3 FT
6. WB4 FT
7. Elution
8. Post-dialysis/Pre-TEV cleavage
9. Post-TEV cleavage
10. Reverse IMAC FT (final sample)

Fig. A-3 SDS-PAGE for MSP1E3D1 protein preparation. Representative Coomassie-stained SDS-PAGE of samples taken during purification of MSP1E3D1 using the protocol outlined in section 5.2.2.2.

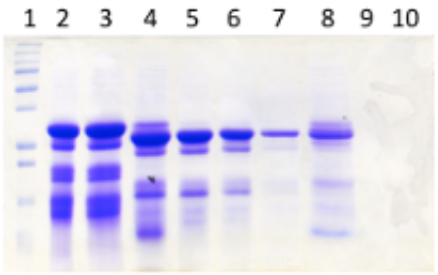
5.2.2.5 Representative Western blot for MSP1E3D1 protein preparation



1. PM
2. Pre-NiNTA supernatant
3. WB1 FT
4. WB2 FT
5. WB3 FT
6. WB4 FT
7. Elution
8. Post-dialysis/Pre-TEV cleavage
9. Post-TEV cleavage
10. Reverse IMAC FT (final sample)

Fig. A-4 Western blot of MSP1E3D1 protein preparation samples. Representative Western blot of duplicate SDS-PAGE from Figure A-3 showing bands for histidine-tagged MSP1E3D1 in green.

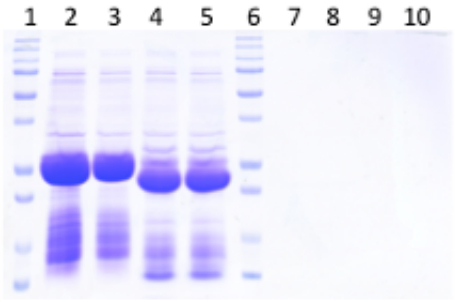
5.2.2.6 Representative SDS-PAGE for MSP1E3D1 histidine tag cleavage



1. PM
2. Post-Dialysis 1, Pre-TEV cleavage (conc.)
3. Post-Dialysis 1, Pre-TEV cleavage (dil.)
4. Post-TEV cleavage
5. Reverse IMAC FT
6. Reverse IMAC Wash 1
7. Reverse IMAC Wash 2
8. Reverse IMAC Elution

Fig. A-5 SDS-PAGE of MSP1E3D1 histidine tag cleavage samples. Representative Coomassie-stained SDS-PAGE of samples taken during TEV protease cleavage of the histidine tag at the N-terminus of the MSP1E3D1 construct.

5.2.2.7 Representative SDS-PAGE for MSP1D1 histidine tag cleavage



1. PM
2. Post-Dialysis 1, Pre-TEV cleavage (conc.)
3. Post-Dialysis 1, Pre-TEV cleavage (dil.)
4. Post-TEV cleavage
5. Post-TEV cleavage
6. PM

Fig. A-6 SDS-PAGE of MSP1D1 histidine tag cleavage samples. Representative Coomassie-stained SDS-PAGE of samples taken during TEV protease cleavage of the histidine tag at the N-terminus of the MSP1D1 construct.

5.3 Appendix 3: Lipid stock preparation protocol

5.3.1 Materials

Table 6 Lipid stock preparation materials list

Solution	Suggested Volume	Composition
Lipid in chloroform	As needed	Avanti polar lipids
Lipid Resuspension Buffer (LRB)	As needed, e.g. 15 mL	20mM Tris pH 7.5, 100mM NaCl, 100mM sodium cholate (for WzmWzt lipids)

5.3.2 Protocol

Determine the target concentration of the lipid stock solution that you will prepare, typically concentrations in the range of 10mM to 25mM lipid will be soluble in 100mM sodium cholate. Stock lipid concentration will generally depend on your later reconstitution conditions (i.e. will the volume of lipid stock used increase the final cholate concentration beyond the reconstitution target of 15mM sodium cholate? If so, use a more concentrated lipid stock).

Calculate the molar concentration of lipid from the mass-based concentration (e.g. 25mg/mL) of the commercial chloroform-solubilized lipid. Determine the volume of chloroform-solubilized lipid that will be required to achieve the desired mass of lipid after removal of chloroform.

Transfer the desired volume of chloroform-solubilized lipid to one or more glass vials or round bottom glass test tubes for removal of chloroform by nitrogen or argon stream. Try to minimize the volume of chloroform in the glass container to avoid trapping chloroform later due to an excessive mass of lipid in a single glass container. Additionally, keep the volume low to avoid running the lipid film too far up the walls of the glass container or spilling the chloroform-solubilized lipids out.

Using a gentle stream of gas, evaporate the chloroform while maintaining a rippling surface of the liquid. As the chloroform cools due to evaporative cooling, keep it warm with your hand and move or rotate the glass container to distribute the lipid film along the walls of the glass container. Do this until you see a white or very thin glassy film (appearance may be lipid-dependent; a very thick glassy film may trap solvent, so minimizing the start volume/mass is an important consideration).

Once all of the chloroform has been evaporated, place the open glass container in a vacuum chamber, seal and apply vacuum to attempt to remove any residual solvent. Overnight or longer is appropriate. While *E. coli* Total Lipids and POPG are not light sensitive, be mindful of light sensitivity of the lipid used.

Prepare the Lipid Resuspension Buffer (LRB; 100mM sodium cholate buffer that will be used to dissolve the lipid). The buffer can contain the buffer system and sodium chloride concentration that matches the final gel filtration buffer (e.g. 20mM Tris pH 7.5 and 100mM sodium chloride). The sodium cholate concentration will always be 100mM to solubilize the lipids regardless of the lipid concentration.

After sufficient time under vacuum, retrieve the dried lipids for resuspension in 100mM sodium cholate.

Add the target volume of LRB to resuspend the lipid at its target concentration.

Gently vortex the container to begin mixing the lipid into the LRB.

Use a combination of vortexing, warming the glass container with hot faucet water (be careful not to get tap water into the vial, just use the faucet stream to warm the bottom of the container), placing the container on ice, and if necessary, bath sonication (do not use probe sonication) to dissolve the lipid into the LRB.

After using the above approaches (e.g. vortexing, warming, vortexing, cooling, vortexing, etc.), a properly dissolved lipid will be clear and colorless at room temperature.

Transfer the dissolved lipid to the final container such as a screw cap glass vial and carefully flush the headspace with nitrogen or argon. Tighten the cap and wrap with parafilm.

Store the lipid stock solutions at -20°C, or -80°C if space allows.

5.3.3 Example *E. coli* lipid extract mass to molarity conversion

Table 7 *E.coli* total lipid extract mass to molarity conversion table

E. coli **Total Lipid Extract Phospholipid Profile (Avanti Polar Lipids Cat. #100500)**						
	% (wt/wt)	Per 1mL		g/mol	mmol	mmol/L
PE	57.5	14.4	mg	718.0	0.020021	20.0
PG	15.1	3.8	mg	771.0	0.00489631	4.9
CA	9.8	2.5	mg	1430.0	0.00171334	1.7
Unknown	17.6	4.4	mg	973.0	0.00452219	4.5
Total	100.0	25.0	mg			

0.03115284	mmol/mL
31.2	mmol/L

Table 8 *E. coli* polar lipid extract mass to molarity conversion table

E. coli **Polar Lipid Extract Phospholipid Profile (Avanti Polar Lipids Cat. #100600)**						
	% (wt/wt)	Per 1mL		g/mol	mmol	mmol/L
PE	67.0	16.8	mg	718.0	0.02332882	23.3
PG	23.2	5.8	mg	771.0	0.00752281	7.5
CA	9.8	2.5	mg	1430.0	0.00171334	1.7
Total	100.0	25.0	mg			

0.03256497	mmol/mL
32.6	mmol/L

5.4 Appendix 4: Membrane protein reconstitution into lipid nanodiscs

5.4.1 Materials

Table 9 Nanodisc reconstitution materials list

Solution	Suggested Volume	Composition
WzmWzt Gel Filtration Buffer	As needed	20 mM Tris HCl pH 7.5, 100 mM NaCl, 0.5 mM TCEP, 5 mM LDAO
0.5M EDTA pH 8.0	As needed	Preservative for overnight incubation, optional.
500mM sodium cholate	As needed	500mM sodium cholate in Milli-Q water
Lipid stock solution	As needed	
WzmWzt stock solution	As needed	WzmWzt solubilized in Gel Filtration Buffer
MSP stock solution	As needed	~500uM MSP in storage buffer
Nanodisc Gel Filtration Buffer	250	20mM Tris pH 7.5, 100mM NaCl, 0.5mM TCEP (does NOT contain detergent)

5.4.2 Protocol

Based on the approximate size of the transmembrane domain, choose one or more membrane scaffold protein (MSP) constructs (e.g. MSP1D1, MSP1E3D1, MSP2N2) to screen in reconstitution studies.

Express and purify the appropriate MSPs.

Prepare the appropriate lipid stock solutions.

Determine the molar ratios of membrane protein, membrane scaffold protein (MSP) and lipid.

Molar ratios of membrane protein to MSP begin at 1:2 as there are generally two copies of MSP per disc for a monomeric protein (except for MSP2N2 where a single copy may be long enough to wrap twice). The protein:MSP ratio can be increased to ensure

reconstitution or generate excess free nanodiscs (probably will be observed in gel filtration purification at ratios of 1:8 or higher).

Molar ratios of MSP to lipid depend on the MSP construct being used. Review the literature to assist in determining a good MSP:lipid ratio operating range. Ranges include 1:30 for MSP1D1 up to 1:150 or 1:200 for MSP2N2. MSP:lipid ratio is project-specific and a range of ratios must be screened.

Molar concentrations will be dictated by the available material i.e. how much of the target membrane protein is available following purification in detergent. A rough guideline is a concentration of 5-10mM membrane protein as the basis for reconstitution.

Once the molar concentration of the membrane protein is chosen, the other molar concentrations are calculated from the previously determined molar ratios. For example, for a smaller nanodisc using MSP1D1, there may be a ratio of 1:2:80 protein:MSP:lipid (protein:MSP 1:2 ratio and MSP:lipid 1:40 ratio). If using a 10mM target membrane protein concentration, MSP will be 20mM and lipid will be 800mM.

Calculate the volumes required for the reconstitution mixture based on a 500uL final volume (this allows direct injection onto a 10/300 S200 or Superose column without concentrating the sample).

Prepare a concentrated sodium cholate stock solution (e.g. 500mM sodium cholate in water or buffer). This will be used to increase the concentration of the sodium cholate in the reconstitution mixture to a target of 15mM (the lipid stock will only contribute a portion of the required sodium cholate).

Prepare Bio-Rad Bio-Beads SM-2 Resin in Milli-Q water by transferring the Bio-Beads to a clean 50mL vial. Fill the vial with Milli-Q water and allow the Bio-Beads to settle, decant as much water as possible without losing beads. Repeat to wash.

Prepare or gather the following solutions: membrane protein sample, gel filtration buffer (same as membrane protein purification buffer), sodium cholate stock, lipid stock, MSP sample, Bio-Rad Bio-Beads SM-2 Resin in Milli-Q water.

Based on the previous calculations, prepare a nanodisc reconstitution mixture starting with gel filtration buffer.

Add sodium cholate stock.

Add lipid stock.

The solution is now 15mM sodium cholate and contains the desire amount of lipid.

Add the membrane protein to target its final concentration in the reconstitution mixture.

Allow the reconstitution mixture to sit on ice for 30-60 minutes to enable the formation of mixed micelles.

Add MSP and let sit for 5-60 minutes on ice.

Add the first batch of Bio-Beads (either determine the amount by volume of slurry in Milli-Q water and spot on Whatman paper to remove excess water or by mass of the beads after removing water with a pipette, e.g. 100mg), and incubate the reconstitution mixture at 4°C with tumbling for ~1 hour. NOTE: Bio-Beads remove detergent from solution enabling the lipids to interact with the membrane protein and the MSP to interact with the lipids, thus forming the nanodisc. Bio-Beads are added stepwise to enable gradual removal of detergent from solution.

After 1 hour, add the second batch of Bio-Beads, and incubate the reconstitution mixture at 4°C with tumbling overnight.

The following day, add the third batch of Bio-Beads, and incubate the reconstitution mixture at 4C with tumbling for ~1 hour or as needed.

Prepare a nanodisc gel filtration buffer (the same buffer, pH and salt concentration as used during protein purification except NO DETERGENT is used) and equilibrate a 10/300 S200 or Superose column.

After successful reconstitution, the mixture will be clear and colorless.

To remove the Bio-Beads, first transfer as much of the mixture as possible either first to a new 1.5mL vial or directly onto a Corning Spin-X filter. Spin at 4,000-5,000 rpm for 5-6 minutes at 4°C. Spin again if any volume was trapped on top of the filter.

After filtration, spin the sample again at 15,000 rpm for 10-20 minutes at 4°C to pellet aggregates (there may still be material that you can pellet after filtration -- this was observed with *E. coli* Total Lipids).

Inject the nanodisc reconstitution protein onto the 10/300 S200 or Superose column to purify.

Collect fractions for your sample.

Verify sample purity and the approximate ratio of membrane protein to MSP by SDS-PAGE. A late eluting peak which could be indicative of an empty nanodisc peak (on an S200 10/300 column, this will be at 14-15mL, if at all).

Nanodisc reconstitution protein can be concentrated prior to use.

5.5 Appendix 5: Detailed cryo-EM data processing procedures

5.5.1 EM data acquisition

Cryo-EM data was collected at the UMASS Medical School Cryo-EM facility on a Titan Krios (FEI) electron microscope operating at 300 kV and equipped with a Gatan imaging filter (GIF) and K3 direct detection camera (Gatan). SerialEM software was used to collect 31-frame movies in super-resolution counting mode at 81,000X nominal magnification, corresponding to a calibrated pixel size of 0.53Å/pixel.

A total of 5,938 movies were collected from two grids over ~44 hours. The defocus range was -1.2 to 2.2 um and an estimated dose rate of 1.45 e^- /A^2/frame for a total dose of 45e^- /A^2. The data was provided as a collection of TIFF files (*.tif) and a gain reference file (SuperRef_2-g2_00000.dm4) and uploaded to the local server called “Server 3”. Raw TIFF and gain reference files are located in the directory:
/srv3/home/lab_srv3/chris/EM_data/umasskrios20190905/Chris_20190912/rawTIFF

The following are the important parameters for raw images (.tif files):
Defocus range: -1.2 to -2.2
Method: One shot per hole, 4 holes (image shift)
Magnification = 81000
ExposureDose = 45.0252
PixelSpacing = 0.53 A
SpotSize = 6
ExposureTime = 2.4767 sec/image

Binning = 0.5 (super resolution mode)

NumSubFrames = 31 frames

FrameDosesAndNumber = 1.4524 31

The following is reproduced from the exact *.mdoc file

T = SerialEM: UMass_Krios Camera -> 0:Ceta 1:GIF-K3 04-Sep-19 12:56:14

Voltage = 300

[FrameSet = 0]

TiltAngle = -0.0154981

StagePosition = 633.133 523.703

StageZ = -154.146

Magnification = 81000

Intensity = 0.114146

ExposureDose = 45.0252

PixelSpacing = 0.53

SpotSize = 6

Defocus = -2.31964

ImageShift = -0.421342 -1.69177

RotationAngle = 177.263

ExposureTime = 2.4767

Binning = 0.5

CameraIndex = 1

DividedBy2 = 0

OperatingMode = 1

UsingCDS = 0

MagIndex = 30

LowDoseConSet = 4

CountsPerElectron = 1

TargetDefocus = -2.2

SubFramePath = X:\Chris_20190905\2-g2_03533.tif

NumSubFrames = 31

FrameDosesAndNumber = 1.4524 31

DateTime = 06-Sep-19 20:17:19

NavigatorLabel = 782

FilterSlitAndLoss = 20 0

MultishotHoleAndPosition = 4 0

GainReference = SuperRef_2-g2_00000.dm4

Please note that prior to import into either cryoSPARC or Relion, the gain reference file must be converted from *.dm4 to *.mrc using the program **dm2mrc** (dm2mrc is part of the IMOD package that is included in SBGRID; any workstation that can load SBGRID will run **dm2mrc**).

5.5.2 Summary of cryo-EM data processing in Relion 3.0.8

An iterative process of 3D classification, 2D classification followed by 3D classification yielded a single class suitable for refinement. Re-extraction to 1.49 Å/pixel followed by refinement, post-processing, Bayesian polishing, and CTF refinement yielded a final map with 3.6 Å overall resolution based on 48,174 particles (MAP1).

In order to improve the density of the Wzt CBD, a particle pool of 178,053 re-extracted particles (1.49 Å/pixel) was subjected to signal subtraction to remove the Wzm TM and nanodisc region from particle images. These signal subtracted particles were 3D classified using a mask over the CBD dimer region only to yield a single class with improved CBD density consisting of 51,828 particles. Refinement of this particle set yielded a Wzt dimer map with an overall resolution of 3.6 Å (MAP2).

5.5.3 Cryo-EM data processing Relion 3.0.8 procedure

NOTE: Generally, unless otherwise noted below, the default parameters from the Relion 3.0 tutorial was followed for data processing.

5.5.3.1 Creating the project directory and importing data

A Relion job was created in the directory:

/srv3/home/lab_srv3/users/chris/relion/umass090519.

All 5,938 TIFF (*.tif) files were imported into Relion 3.0.8 using the Import job (job001) with the Node type “2D micrograph movies”.

5.5.3.2 Motion correction

Gain correction, beam-induced motion correction, dose-weighting and 2-fold Fourier-cropping to a pixel size of 1.06 Å/pixel was performed using UCSF MotionCor2 loaded through Relion (not Relion's own implementation). In the MotionCor2 job (job008) on the I/O tab, the initial frame was excluded and non-dose-weighted summed images were saved. On the Motion tab, the settings were B-factor of 500, 5x5 patches, Binning factor 2, No Gain rotation, No Gain flip.

MotionCorr

I/O Tab

Input movies STAR file: Import/job001/movies.star

First frame for corrected sum: 2

First frame for corrected sum: -1

Last frame for corrected sum: 0.53

Pixel size (Å): 0.53

Voltage (kV): 300

Dose per frame ($e^-/Å^2$): 1.4524

Pre-exposure ($e^-/Å^2$): 0

Do DW? Yes

Save nonDW? Yes

Motion Tab

Bfactor: 500

Number of patches X, Y: 5 x 5

Group frames: 1

Binning factor: 2

Gain-reference image: Movies/SuperRef_2-g2_00000_GainRef.mrc

Gain rotation: No rotation (0)

Gain flip: No flipping (0)

Defect file: Not used

Use RELION's own implementation? No

MOTIONCOR2 executable:

/programs/x86_64-linux/motioncor2/1.2.3/MotionCor2_1.2.3-Cuda100

Which GPUs to use: 0:1:2:3

Other

Running Tab

Number of MPI procs: 4

Number of threads: 4

5.5.3.3 CTF estimation and micrograph curation

The dose-weighted summed images which are outputted from this job were used for all image processing except for CTF parameter estimation. Rather, the non-dose-weighted summed images (excluding initial frame) were processed using CTFFind4 for CTF parameter estimation (job009). Spherical aberration for the UMASS Medical School Titan Krios is 2.7mm. Magnified pixel size (Angstrom) is now the binned pixel size of 1.06 Å/pixel.

CTF Estimation

I/O Tab

Input micrographs STAR file: MotionCorr/job008/corrected_micrographs.star

Use micrograph without dose-weighting? Yes

Spherical aberration (mm): 2.7

Voltage (kV): 300

Amplitude contrast: 0.1

Magnified pixel size (Angstrom): 1.06

Amount of astigmatism (A): 100

Searches Tab

FFT box size (pix): 512

Minimum resolution (A): 30

Maximum resolution (A): 5

Minimum defocus value (A): 5000

Maximum defocus value (A): 50000

Defocus step size (A): 500

Estimate phase shifts? No

Phase shift - Min, Max, Step (deg): Not used

CTFFIND-4.1 Tab

Use CTFFIND-4.1? Yes

CTFFIND-4.1 executable: /programs/x86_64-linux/ctffind4/4.1.13/bin/ctffind

Use exhaustive search? No

Estimate CTF on window size (pix) -1

Gctf Tab

Use Gctf instead? No

Running Tab

Number of MPI procs: 5

The micrographs outputted by the CTFFind4 job were checked for quality by sorting against Defocus, CTFMaxResolution, and CTFFigureOfMerit and reviewing the Fourier transform. Any obvious ice contaminations were removed (significant ice rings or defects in the symmetry of the Thon rings). The main cutoff was eliminating micrographs above a CTFMaxResolution of 5.0 Å. Micrograph subsets were created using the Select job with the previous job micrograph.star as input (job010, job011, job013, job017).

5.5.3.4 Particle picking

Particles were autopicked using the Laplacian-of-Gaussian-based method from a subset of 4,935 micrographs (Relion job018). The main parameters changed for autopicking were in

the Laplacian tab LoG filter settings: Min. diameter for LoG filter (A): 100 and Max. diameter for LoG filter (A): 140.

Autopick

I/O Tab

Input micrographs for autopick: Select/job017/micrographs.star

Pixel size in micrographs (A): -1

2D references: Not used

OR: provide a 3D reference? No

OR: use Laplacian-of-Gaussian? Yes

Laplacian Tab

Min. diameter for LoG filter (A): 100

Max. diameter for LoG filter (A): 140

Are the particles white? No

Max. resolution to consider (A): 20

Adjust default threshold: 0

References Tab (Not used)

autopicking Tab

Picking threshold: 0.05

Min. inter-particle distance (A): 100

Max. stddev noise: 1.1

Min. avg. noise: -999

Write FOM maps? No

Read FOM maps? No

Shrink factor: 1

Use GPU acceleration? No

Running Tab

Number of MPI procs: 5

5.5.3.5 Particle extraction

These settings resulted in autopicking 5,490,947 particles and these particles were extracted (Relion Extract job020) with 4X downsampling.

Note that 4X downsampling translates to a new pixel size of 1.06 Å/pixel * 4 = 4.24 Å/pixel

Extract

I/O Tab

micrograph STAR file: CtfFind/job009/micrographs_ctf.star

Input coordinates: AutoPick/job018/coords_suffix_autopick.star

OR re-extract refined particles? No

Refined particles STAR file: Not used

Reset the refined offsets to zero? Not used

OR: re-center refined coordinates? Not used

Recenter on - X, Y, Z (pix) Not used

Manually set pixel size? Not used

extract Tab

Particle box size (pix): 256

Invert contrast? Yes

Normalize particles? Yes

Diameter background circle (pix): 180

Stddev for white dust removal: -1

Stddev for black dust removal: -1

Rescale particles? Yes

Re-scaled size (pixels) 64

Running Tab

Number of MPI procs: 5

5.5.3.6 2D classification rounds 1 through 7

Seven rounds of 2D classification were performed. The parameters for the initial round (job026) are listed below and only the changes to these are highlighted thereafter.

2D Class (Round 1 job026, starting particle number 5,490,947)

I/O Tab

Input images STAR file: Extract/job020/particles.star

Continue from here: Not used

CTF Tab

Do CTF-correction? Yes

Have data been phase-flipped? No

Ignore CTFs until first peak? No

Optimisation Tab

Number of classes: 200

Regularisation parameter T: 2

Number of iterations: 25

Use fast subsets? Yes

Mask diameter (A): 190

Mask individual particles with zeros? Yes

Limit resolution E-step to (A): -1

Sampling Tab

Perform image alignment? Yes

In-plane angular sampling: 6

Offset search range (pix): 5

Offset search step (pix): 1

Compute Tab

Use parallel disc I/O? Yes

Number of pooled particles 10

Pre-read all particles into RAM? No

Copy particles to scratch dir.: /srv3/home/scratch

Combine iterations through disc? No

Use GPU acceleration? Yes

Which GPUs to use: 0:1:2:3

Running Tab

Number of MPI procs: 5

Number of threads: 10

2D Class (Round 2 job030, starting particle number 5,487,307)

I/O Tab

Input images STAR file: Select/job029/particles.star

Optimisation Tab

Number of classes: 250

Compute Tab

Number of pooled particles: 30

2D Class (Round 3 job043, starting particle number 3,160,370)

I/O Tab

Input images STAR file: Select/job042/particles.star

Optimisation Tab

Number of classes: 200

Compute Tab

Number of pooled particles: 50

2D Class (Round 4 job048, starting particle number 1,673,933)

I/O Tab

Input images STAR file: Select/job047/particles.star

Optimisation Tab

Number of classes: 200

2D Class (Round 5 job052, starting particle number 1,181,981)

I/O Tab

Input images STAR file: Select/job051/particles.star

Optimisation Tab

Number of classes: 200

2D Class (Round 6 job062, starting particle number 1,008,049)

I/O Tab

Input images STAR file: Select/job060/particles.star

Optimisation Tab

Number of classes: 200

2D Class (Round 7 job067, starting particle number 899,159)

I/O Tab

Input images STAR file: Select/job066/particles.star

Optimisation Tab

Number of classes: 100

5.5.3.7 CryoSPARC data processing for initial model building

Following the seventh round of 2D classification, 804,216 particles were selected for import into cryoSPARC to build an *ab initio* initial model. Initial model building in Relion generally failed to yield a satisfactory initial model from the data.

All TIFF (*.tif) files were imported into cryoSPARC using the Import Movies job type (J1). The relevant parameters are as follows:

Movies data path:

/srv3/home/lab_srv3/chris/EM_data/umasskrios20190905/Chris_20190912/rawTIFF/*.tif

Gain Reference data path:

/srv3/home/lab_srv3/chris/EM_data/umasskrios20190905/Chris_20190912/rawTIFF/SuperRef_2-g2_00000_GainRef.mrc

Raw pixel size (Å): 0.53

Accelerating Voltage (kV): 300

Spherical Aberration (mm): 2.7

Total exposure dose ($e^-/Å^2$): 45.0252

After importing the raw data (movies) into cryoSPARC, the existing processed micrographs from Relion were imported using the Import Micrographs job (J111). References to the J1 imported movies were J1.imported_movies. Import Micrographs parameters were as follows:

Import Micrographs

Micrographs data path:

/srv3/home/lab_srv3/chris/relion/umass090519/MotionCorr/job008/Movies/2-g?_0????.mrc

Length of movie path suffix to cut: 4

Length of mic. path suffix to cut: 4

Pixel size (Å): 1.06

Accelerating Voltage (kV): 300

Spherical Aberration (mm): 2.7

Total exposure dose ($e^-/Å^2$): 45.0252

With the Relion micrographs imported into cryoSPARC, the 804,216 particles picked in Relion were imported using the Import Particles job (J245) with the following parameters:

Import Particles

Particle Meta Path:

/srv3/home/lab_srv3/chris/relion/umass090519/Select/SelectRound7/particles.star

Particle Data Path:

/srv3/home/lab_srv3/chris/relion/umass090519/Extract/job020/Movies/

Accelerating Voltage (kV): 300

Spherical Aberration (mm): 2.7

Amplitude Contrast (fraction): 0.1

Pixel Size (Å): 4.24

Data Sign: light-on-dark

With the Relion particles imported into cryoSPARC, the Ab Initio job (J246) was used to generate an initial model from the particle set specifying 5 classes (dependency on J245.imported_particles). The best volume generated was class 01: cryosparc_P2_J246_class_01_final_volume.mrc

This volume was processed in relion_image_handler to ensure that the box size was set to 64 pixels and the pixel size was set to 4.24 Å/pixel generating the following file:

cryosparc_P2_J246_class_01_box64.mrc

The initial model was then imported into Relion (Import job076) using the Import job type with Node type 3D reference (.mrc) and the input file:

/srv3/home/lab_srv3/users/chris/cryosparc/cryosparc_P2_J246_class_01_box64.mrc

5.5.3.8 3D classification round 1

The initial model imported from cryoSPARC was used as the reference map for the first round of 3D classification (Relion job079) with the following settings:

Class3D (Round 1 job079, starting particle number 804,216)

I/O Tab

Input Images STAR file: Select/job068/particles.star

Continue from here (iter 26-40): Class3D/job079/run_it025_optimiser.star

Reference map: Import/job076/cryosparc_P2_J246_class_01_box64.mrc

Reference mask (optional): Not used

Reference Tab

Abs. greyscale? No

Initial low-pass filter (A): 35

Symmetry: C1

CTF Tab

Do CTF-correction? Yes

Has reference been CTF-corrected? No

Phase-flipped? No

Ignore CTF until 1st peak? No

Optimisation Tab

Number of classes: 8

Reg. Parameter T: 4

Number Iter.: 25 initially, 40 when continued

Use fast subsets: No

Mask Diameter (A): 190

Mask with zeroes? Yes

Limit resolution E-step to (A): 10

Sampling Tab

Perform image alignment? Yes

Angular sampling interval: 7.5 degrees

Offset search range (pix): 5

Offset search step (pix): 1

Perform local angular searches? No

Local angular search range: Not used

Compute Tab

Use parallel disc I/O? No

Number of pooled particles: 30

Skip padding? No

Pre-read all particles into RAM? No

Copy particles to scratch directory: /srv3/home/scratch

Combine iterations through disc? No

Use GPU acceleration? Yes

Which GPUs to use: 0:1:2:3

Running Tab

Number of MPI procs: 5

Number of threads: 10

5.5.3.9 2D classification round 8

After the first round of 3D classification, class 1, class 4, and class 6 were selected (Select job080) totaling 256,529 particles. This particle subset was then subjected to an intermediate round of 2D classification (Relion Class2D job081) to review the 2D classes for quality using the following parameters:

Class2D (Round 8 job081, starting particle number 256,529)

I/O Tab

Input images STAR file: Select/job080/particles.star

Continue from here: Class2D/job081/run_it025_optimiser.star

CTF Tab

Do CTF-correction? Yes

Have data been phase-flipped? No

Ignore CTFs until first peak? No

Optimisation Tab

Number of classes: 75

Regularisation parameter T: 2

Number of iterations: 25 initially, 40 when continued

Use fast subsets? No

Mask diameter (A): 190

Mask individual particles with zeros? Yes

Limit resolution E-step to (A): 8

Sampling Tab

Perform image alignment? Yes

In-plane angular sampling: 6

Offset search range (pix): 5

Offset search step (pix): 1

Compute Tab

Use parallel disc I/O? No

Number of pooled particles: 50

Pre-read all particles into RAM? No

Copy particles to scratch dir.: /srv3/home/scratch

Combine iterations through disc? No

Use GPU acceleration? Yes

Which GPUs to use: 0:0:1:1:2:2:3:3

Running Tab

Number of MPI procs: 9

Number of threads: 6

5.5.3.10 3D classification round 2

After 2D classification, 239,073 particles were selected (Select job086) and input into a second round of 3D classification (Relion job087) with the following settings:

Class 3D (Round 2 job087, starting particle number 239,073)

I/O Tab

Input Images STAR file: Select/job086/particles.star

Continue from here (iter 26-40): Class3D/job087/run_it025_optimiser.star

Reference map: Import/job076/cryosparc_P2_J246_class_01_box64.mrc

Reference mask (optional): Not used

Reference Tab

Abs. greyscale? No

Initial low-pass filter (A): 35

Symmetry: C1

CTF Tab

Do CTF-correction? Yes

Has reference been CTF-corrected? No

Phase-flipped? No

Ignore CTF until 1st peak? No

Optimisation Tab

Number of classes: 4

Reg. Parameter T: 4

Number Iter.: 25 initially, 40 when continued

Use fast subsets? No

Mask Diameter (A): 190

Mask with zeroes? Yes

Limit resolution E-step to (A): 8

Sampling Tab

Perform image alignment? Yes

Angular sampling interval: 7.5 degrees

Offset search range (pix): 5

Offset search step (pix): 1

Perform local angular searches? No

Local angular search range: Not used

Compute Tab

Use parallel disc I/O? No

Number of pooled particles: 30

Skip padding? No

Pre-read all particles into RAM? No

Copy particles to scratch directory: /srv3/home/scratch

Combine iterations through disc? No

Use GPU acceleration? Yes

Which GPUs to use: 0:1:2:3

Running Tab

Number of MPI procs: 5

Number of threads: 10

After the second round of 3D classification, class 1, class 3, and class 4 were selected totaling 178,053 particles and the particles were re-grouped into 50 groups (Relion Select job122).

5.5.3.11 New reference map

In order to prepare for working with re-extracted particles which will be at a higher resolution (lower Å/pixel) and larger box size, a new reference map needs to be generated. One option was to use the initial model used for 3D classification followed by using relion_image_handler to change the pixel size and box size. However, in this case the ~10 Å class 4 volume from the second round of 3D classification was used as the new reference map. Using this map allowed the volume to be low pass filtered to 50 Å to remove more high-resolution features than was possible with the 35 Å low-pass filtered initial model from cryoSPARC, because it would have degraded to a featureless blob. As such, job087 class 4 volume, (Class3D/job087/run_it025_class004.mrc) was converted to a new pixel size, 1.490625 Å, and box size, 256 pixels, based on the expected re-extraction parameters and imported into Relion (Relion Import job090).

In the Relion tutorial, there is the suggestion that if CTF refinement is to be done as a later step, it may be beneficial to re-extract at a larger box size (e.g. 360 pixels) and then downscale to the normal box size of 256 pixels to work with smaller images (to speed up processing). Thus, the particles were re-extracted using a box size of 360 pixels, then downsampled to a box size of 256 pixels to yield a final pixel size of (360*1.06 Å)/256 = 1.490625 Å.

In order to build the new reference map, Class3D/job087/run_it025_class004.mrc was converted in relion_image_handler from box size 64 pixels and pixel size 4.24 Å/pixel to a box size of 256 pixels and pixel size 1.490625 Å/pixel. Then, this new volume

Class3D/job087/run_it025_class004_box256.mrc was imported into Relion to serve as an initial model for subsequent processing (Relion Import job090).

5.5.3.12 Particle re-extraction

The 178,053 particle subset was then re-extracted (Relion Extract job123) with the following parameters:

I/O Tab

micrograph STAR file: CtfFind/job009/micrographs_ctf.star

Input coordinates: Not used

OR re-extract refined particles? Yes

Refined particles STAR file: Select/job122/particles.star

Reset the refined offsets to zero? No

OR: re-center refined coordinates? Yes

Recenter on - X, Y, Z (pix): 0 0 0

Manually set pixel size? No

extract Tab

Particle box size (pix): 360

Invert contrast? Yes

Normalize particles? Yes

Diameter background circle (pix): 180

Stddev for white dust removal: -1

Stddev for black dust removal: -1

Rescale particles? Yes

Re-scaled size (pixels): 256

Running Tab

Number of MPI procs: 17

5.5.3.13 Initial 3D refinement

The re-extracted set of 178,053 particles (now at a box size of 256 pixels and pixel size of 1.490625 Å/pixel) was refined against the new reference map (Relion Refine3D job124) using the following settings:

I/O Tab

Input images STAR file: Extract/job123/particles.star

Continue from here: Not used

Reference map: Import/job090/run_it025_class004_box256.mrc

Reference mask (optional): Not used

Reference Tab

Ref. map is on absolute greyscale? No

Initial low-pass filter? 50

Symmetry: C1

CTF Tab

Do CTF-correction? Yes

Has ref. been CTF-corrected? Yes

Have data been phase-flipped? No

Ignore CTFs until first peak? No

Optimisation Tab

Mask diameter (A): 190

Mask individual particles with zeros? Yes

Use solvent-flattened FSCs? No

Auto-sampling Tab

Initial angular sampling: 7.5 degrees

Initial offset range (pix): 5

Initial offset step (pix): 1

Local searches from auto-sampling: 1.8 degrees

Compute Tab

Use parallel disc I/O? No

Number of pooled particles: 30

Skip padding? No

Pre-read all particles into RAM? No

Copy particles to scratch directory /srv3/home/scratch

Combine iterations through disc? No

Use GPU acceleration? Yes

Which GPUs to use: 0:1:2:3

Running Tab

Number of MPI procs: 5

Number of threads: 12

The initial 3D refinement resulted in a map with a final resolution (without masking) of 6.4678 Å.

5.5.3.14 Initial mask creation

Mask creation is most successful when done as an iterative process involving Relion and Chimera as described in the Relion tutorial. That is to say, take the Refine3D output, load it into Chimera, and using the Volume Viewer, change the threshold to the point where the volume is largest, but does not have any stray density outside of the core density of the volume (no random pieces of density outside of where the protein should be). This value is then the "Initial binarization threshold" value in the Relion MaskCreate job. It is advisable to save a mask at this threshold, import it into Relion and run it through MaskCreate, then view the mask in Chimera to check if it will be continuous and encompass the volume from the Refine3D job. This is important because the mask is intended to be a solvent mask downweighing data outside of the mask to improve the quality of the data within the mask (the volume from Refine3D). Based on Chimera, the output from Refine3D was manipulated to create a solvent mask in Relion (MaskCreate job125) using the following parameters:

MaskCreate job125

I/O Tab

Input 3D map: Refine3D/job124/run_class001.mrc

Mask Tab

Lowpass filter map (A): 18

Pixel size (A): 1.49

Initial binarisation threshold: 0.012

Extend binary map this many pixels: 6

Add a soft-edge of this many pixels: 6

Running Tab

Number of threads: 32

5.5.3.15 Post-processing with the solvent mask

Post-processing (Relion PostProcess job126) was performed to check improvement of the map with the application of a solvent mask:

PostProcess job126

I/O Tab

One of the 2 unfiltered half-maps: Refine3D/job124/run_half1_class001_unfil.mrc

Solvent mask: MaskCreate/job125/mask.mrc

Calibrated pixel size (A): 1.49063

Sharpen Tab

MTF of the detector (STAR file):

/srv3/home/lab_srv3/users/chris/relion/ mtf_k3_standard_300kV_FL2.star

Estimate B-factor automatically? Yes

Lowest resolution for auto-B fit (A): 10

Use your own B-factor? No

User provided B-factor: Not used

Filter Tab

Skip FSC-weighting? No

Ad-hoc low-pass filter (A): Not used

After post-processing, the final resolution reported was 4.05959 Å, upon applying the mask and sharpening the map.

5.5.3.16 Bayesian polishing

Now that the particles have been assigned their best orientations (best Euler angles) from Refine3D job124 and the resulting volume has been post-processed with a solvent mask in PostProcess job126, these particles can be polished to potentially improve resolution. NOTE that in the Relion 3.0 software and tutorial there is no resolution guideline for when to or to not do Bayesian polishing so polishing was done to try to push the resolution inside 4.0 Å which would bring the resolution closer to the ~3 Å target for CTF refinement.

As discussed in the Relion tutorial, the Bayesian polishing was first trained on the dataset to generate optimized parameters (Relion Polish job131):

Polish (Training, job131)

I/O Tab

Micrographs (from MotionCorr): MotionCorr/job008/corrected_micrographs.star

Particles (from Refine3D or CtfRefine): Refine3D/job124/run_data.star

Postprocess STAR file: PostProcess/job126/postprocess.star

First movie frame: 1

Last movie frame: -1

Train Tab

Train optimal parameters? **Yes**

Fraction of Fourier pixels for testing: 0.5

Use this many particles: 10,000

Polish Tab

Perform particle polishing? **No**

Optimised parameter file: Not used

OR use your own parameters? Not used

Sigma for velocity (A/dose): Not used

Sigma for divergence (A): Not used

Sigma for acceleration (A/dose): Not used

Minimum resolution for B-factor fit (A): Not used

Maximum resolution for B-factor fit (A): Not used

Running Tab

Number of MPI procs: 5

Number of threads: 10

The Bayesian polishing training run generated Polish/job131/opt_params.txt to load into the true Bayesian polishing job (Polish job132). Setting up the Polish job (Polish job132) is nearly identical to the training job (Polish job131) with the following exceptions:

Train Tab

Train optimal parameters? **No**

Polish Tab

Perform particle polishing? **Yes**

Optimised parameter file: Polish/job131/opt_params.txt

OR use your own parameters? No

Sigma for velocity (A/dose): 0.2

Sigma for divergence (A): 5000

Sigma for acceleration (A/dose): 2

Minimum resolution for B-factor fit (A): 20

Maximum resolution for B-factor fit (A): -1

5.5.3.17 3D refinement after polishing round 1

Polishing generates a new particle STAR file: Polish/job132/shiny.star. To check the potential improvement in resolution, a Refine3D job (Relion Refine3D job133) was run using Polish/job132/shiny.star as the Input images STAR file. Refine3D job133 was cloned from Refine3D job124 (first 3D refinement, without a mask), so all parameters are identical to Refine3D job124. Refine3D job133 also did not include a mask. Auto-refine final resolution (without masking) is: 5.53043 Å.

A mask was generated from the resulting volume of Refine3D job133 (Relion MaskCreate job134) using the following parameters:

I/O Tab

Input 3D map: Refine3D/job133/run_class001.mrc

Mask Tab

Lowpass filter map (A): 18

Pixel size (A): 1.49

Initial binarisation threshold: 0.012

Extend binary map this many pixels: 6

Add a soft-edge of this many pixels: 6

Running Tab

Number of threads: 32

A PostProcess job (PostProcess job135) was run on the Refine3D job133 output with the new solvent mask applied (Relion PostProcess job135). The final resolution estimate was 3.93403 Å.

PostProcess job135

I/O Tab

One of the 2 unfiltered half-maps: Refine3D/job133/run_half1_class001_unfil.mrc

Solvent mask: MaskCreate/job134/mask.mrc

Calibrated pixel size (A): 1.49063

Sharpen Tab

MTF of the detector (STAR file):

/srv3/home/lab_srv3/users/chris/relion/ mtf_k3_standard_300kV_FL2.star

Estimate B-factor automatically? Yes

Lowest resolution for auto-B fit (A): 10

Use your own B-factor? No

User provided B-factor Not used

Filter Tab

Skip FSC-weighting? No

Ad-hoc low-pass filter (A): Not used

5.5.3.18 3D refinement with solvent mask followed by post-processing

After the observation that Bayesian polishing brought the resolution inside 4.0 Å after the solvent mask is applied during the PostProcess job, the polished particles (Polish/job132/shiny.star) were refined again with the solvent mask applied during the refinement job (Relion Refine3D job136). Refine3D job136 was cloned from Refine3D job133, yet from this point forward all Refine3D jobs will include a reference mask and use solvent-flattened FSCs so the parameters will be reproduced here in full:

Refine3D job136 (mask from MaskCreate job125)

I/O Tab

Input images STAR file: Polish/job132/shiny.star

Continue from here: Not used

Reference map: Import/job090/run_it025_class004_box256.mrc

Reference mask (optional): MaskCreate/job125/mask.mrc

Reference Tab

Ref. map is on absolute greyscale? No

Initial low-pass filter? 50

Symmetry: C1

CTF Tab

Do CTF-correction? Yes

Has ref. been CTF-corrected? Yes

Have data been phase-flipped? No

Ignore CTFs until first peak? No

Optimisation Tab

Mask diameter (A): 190

Mask individual particles with zeros? Yes

Use solvent-flattened FSCs? Yes

Auto-sampling Tab

Initial angular sampling: 7.5 degrees

Initial offset range (pix): 5

Initial offset step (pix): 1

Local searches from auto-sampling: 1.8 degrees

Compute Tab

Use parallel disc I/O? No

Number of pooled particles: 30

Skip padding? No

Pre-read all particles into RAM? No

Copy particles to scratch directory: /srv3/home/scratch

Combine iterations through disc? No

Use GPU acceleration? Yes

Which GPUs to use: 0:1:2:3

Running Tab

Number of MPI procs: 5

Number of threads: 12

Results of the Refine3D job 136: Auto-refine final resolution (already with masking) is 3.66923 Å. The subsequent post-processing (Relion PostProcess job137) gives a sharpened

map at the same resolution as the solvent mask has already been applied during 3D refinement (though with the added benefit of improved particle alignment): 3.66924 Å.

PostProcess job137

I/O Tab

One of the 2 unfiltered half-maps: Refine3D/job136/run_half1_class001_unfil.mrc

Solvent mask: MaskCreate/job125/mask.mrc

Calibrated pixel size (A): 1.49063

Sharpen Tab

MTF of the detector (STAR file):

/srv3/home/lab_srv3/users/chris/relion/ mtf_k3_standard_300kV_FL2.star

Estimate B-factor automatically? Yes

Lowest resolution for uto-B fit (A): 10

Use your own B-factor? No

User provided B-factor: Not used

Filter Tab

Skip FSC-weighting? No

Ad-hoc low-pass filter (A): Not used

5.5.3.19 CTF refinement

The resolution had been brought in well beyond 4.0 Å, and an attempt at CTF Refinement was tested (Relion CtfRefine job138):

CtfRefine job138

I/O Tab

Particles (from Refine3D): Refine3D/job136/run_data.star

Postprocess STAR file: PostProcess/job137/postprocess.star

Fit Tab

Minimum resolution for fits (A): 30

Perform CTF parameter fitting? Yes

Fit per-particle defocus? Yes

Range for defocus fit (A): 2000

Fit per-micrograph astigmatism? No

Fit per-particle astigmatism? No

Fit per-micrograph phase-shift? No

Perform beamtilt estimation? No

Running Tab

Number of MPI procs: 4

Number of threads: 8

5.5.3.20 3D refinement after CTF refinement

The CTF refinement job output a new particle set:

CtfRefine/job138/particles_ctf_refine.star and this was used for another Refine3D job (Relion Refine3D job139). This job was cloned from Refine3D job136 and is identical with the following exception:

Refine3D job139

I/O Tab

Input images STAR file: CtfRefine/job138/particles_ctf_refine.star

The results of Refine3D job139: Auto-refine final resolution (already with masking) is 3.56636 Å. The subsequent post-processing job (PostProcess job140) was cloned from PostProcess job137 and so its parameters are identical with the following exception:

I/O Tab

One of the 2 unfiltered half-maps: Refine3D/job139/run_half1_class001_unfil.mrc

Results of the post-processing job: Final resolution 3.56637 Å.

5.5.3.21 Bayesian polishing round 2

The Relion tutorial describes that Bayesian polishing and CTF refinement can be done in either order or even iteratively, so the Bayesian polishing was performed again on the particles after having been improved by an initial round of polishing and CTF refinement. Bayesian polishing was again first trained (Relion Polish job145) using the post-CtfRefine Refine3D particles (Refine3D/job139/run_data.star) and its associated PostProcess job (PostProcess/job140/postprocess.star). Bayesian polishing training (Polish job145) was cloned from the previous training (Polish job131), and all parameters are identical with the following exceptions:

Polish job145 (training round 2)

I/O Tab

Micrographs (from MotionCorr): MotionCorr/job008/corrected_micrographs.star

Particles (from Refine3D or CtfRefine): Refine3D/job139/run_data.star

Postprocess STAR file: PostProcess/job140/postprocess.star

Bayesian polishing was the run using the optimized parameters file (Polish/job145/opt_params.txt) from training (Relion Polish job147). Polish job147 was cloned from Polish job132), and all parameters are identical with the following exceptions:

Polish job147 (polishing round 2)

I/O Tab

Micrographs (from MotionCorr): MotionCorr/job008/corrected_micrographs.star

Particles (from Refine3D or CtfRefine): Refine3D/job139/run_data.star

Postprocess STAR file: PostProcess/job140/postprocess.star

Polish Tab

Optimised parameter file: Polish/job145/opt_params.txt

5.5.3.22 3D refinement after polishing round 2 (consensus refinement)

A 3D refinement job was performed using the Polish/job147/shiny.star particles (Relion Refine3D job148) cloning the job from Refine3D job139 with the following exception:

I/O Tab

Input images STAR file: Polish/job147/shiny.star

The 3D refinement (Relion Refine3D job148) gave the following result: Auto-refine final resolution (already with masking) is: 3.46909 Å. The subsequent PostProcess (Relion PostProcess job149) reported a final resolution of: 3.4691 Å.

PostProcess job149

I/O Tab

One of the 2 unfiltered half-maps: Refine3D/job148/run_half1_class001_unfil.mrc

For later use, a mask was created from the volume generated by Refine3D job148 (MaskCreate job203).

MaskCreate job203

I/O Tab

Input 3D map: Refine3D/job148/run_class001.mrc

Mask Tab

Lowpass filter map (A): 18

Pixel size (A): 1.49

Initial binarisation threshold: 0.0151

Extend binary map this many pixels: 0

Add a soft-edge of this many pixels: 5

Running Tab

Number of threads: 32

5.5.3.23 3D classification without alignment

Despite the fact that the data was improved to an overall resolution of ~3.5 Å according to Relion, there existed some heterogeneity in the data, especially with respect to the WzmWzt TM1 helix. In general, once a 3D refinement has been performed (Refine3D), this can be considered a "consensus refinement" in that it represents all of the heterogeneity in the particle set. Because the Refine3D assigns the best alignment of each particle, it is possible to take the heterogeneous data and tease apart different conformations by doing 3D classification without re-aligning the particles and instead using their original Refine3D

alignments (3D classification without alignment). This approach was done using Refine3D job148 as the consensus refinement for a new 3D classification (Relion Class3D job225), and, in this specific case, a high regularization parameter T=40 was used to force separating out conformations that were very similar into different classes:

Class3D job225 (No alignment, T=40)

I/O Tab

Input Images STAR file: Refine3D/job148/run_data.star

Continue from here (iter 26-40): Not used

Reference map: Import/job090/run_it025_class004_box256.mrc

Reference mask (optional): MaskCreate/job125/mask.mrc

Reference Tab

Abs. greyscale? No

Initial low-pass filter (A): 50

Symmetry: C1

CTF Tab

Do CTF-correction? Yes

Has reference been CTF-corrected? Yes

Phase-flipped? No

Ignore CTF until 1st peak? No

Optimisation Tab

Number of classes: 6

Reg. Parameter T: 40

Number Iter.: 25

Use fast subsets: No

Mask Diameter (A): 190

Mask with zeroes? Yes

Limit resolution E-step to (A): -1

Sampling Tab

Perform image alignment? **No**

Angular sampling interval: Not used

Offset search range (pix): Not used

Offset search step (pix): Not used

Perform local angular searches? Not used

Local angular search range: Not used

Compute Tab

Use parallel disc I/O? No

Number of pooled particles: 30

Skip padding? No

Pre-read all particles into RAM? No

Copy particles to scratch directory: /srv3/home/scratch

Combine iterations through disc? No

Use GPU acceleration? No

Which GPUs to use: Not used

Running Tab

Number of MPI procs: 5

Number of threads: 10

5.5.3.24 3D refinement of class 2 (MAP1)

Of the 6 requested classes, class 2 contained the most particles (27% or 48,174 particles) and was the best class overall by review of the volume density in Chimera resulting in class 2 being selected for 3D refinement (Refine3D job227, cloned from Refine3D job148).

Refine3D job227

I/O Tab

Input images STAR file: Select/job226/particles.star

The results of Refine3D job227: Auto-refine final resolution (already with masking) is 3.6 Å. The subsequent post-processing job (PostProcess job229, cloned from job149) resulted in a final resolution of 3.60001 Å.

PostProcess job229 (MAP1)

One of the 2 unfiltered half-maps: Refine3D/job227/run_half1_class001_unfil.mrc

5.5.3.25 Particle subtraction

The final volume in Refine3D job227 and PostProcess job229 was very good overall, however some density was lacking in the extremities of the Wzt CBD, potentially due to flexibility. A procedure was developed to improve alignment of the particles on the Wzt CBD and give a better map despite the small size of the Wzt CBD dimer region (~36 kDa and its distance from the center of the particle box. This procedure utilized particle subtraction to focus on a region large enough to allow good particle alignment [~100 kDa according to Relion literature (256)] but small enough to improve alignment on just the

Wzt CBD. The full-length Wzt dimer (NBD and CBD) is roughly a ~90 kDa, thus making this region a good candidate for improved particle alignment due to its size. The plan was therefore to subtract the signal of the nanodisc and Wzm dimer from all particles leaving a dataset consisting only of signal for the full-length Wzt dimer. This signal subtraction is achieved by making a **difference mask** representing the region to be removed (i.e. nanodisc and Wzm dimer) which is then overlayed on the volume of the full molecule (Refine3D job148) and subtracted leaving signal only for the Wzt dimer in all particles from Refine3D job148.

A difference mask (Relion Import job234) was created by subtracting individual masks made from the Refine3D job148 volume (full transporter) and a new volume of the Wzt dimer. This new volume of just a Wzt dimer was created by using Chimera to fit the crystal structures of the truncated WzmWztN (PDB 6M96) and the Wzt CBD dimer (PDB 6O14) into the Refine3D job148 volume. The Wzm dimer model was then removed and the remaining chains are combined into a single model of a Wzt dimer (NBD and CBD for each protomer). Then a Chimera molmap was created from the Wzt model at 4.0 Å and this molmap was imported into Relion (Import job230) and used to make a Wzt dimer mask in Relion (MaskCreate job231).

MaskCreate job231

I/O Tab

Input 3D map: Import/job230/UMASS_job_148_run_class1_NBDCBDmask1.mrc

Mask Tab

Lowpass filter map (A): 15

Pixel size (A): 1.49

Initial binarisation threshold: 0.055

Extend binary map this many pixels: 0

Add a soft-edge of this many pixels: 5

Running Tab

Number of threads: 32

In order to make the difference mask (nanodisc and Wzm dimer), the mask previously created from Refine3D job148 (MaskCreate job203), was used as the full-length volume mask (full transporter). Subtraction of the Wzt dimer mask (MaskCreate job231) from the full-length transporter mask (MaskCreate job203) creates a **difference mask** of just the nanodisc and the Wzm dimer. Subtraction of volumes is performed using the relion_image_handler command-line program:

relion_image_handler --i MaskCreate/job203/mask.mrc --angpix 1.49063 --subtract MaskCreate/job231/mask.mrc --o MaskCreate/job231/mask_diffmask.mrc

The new **difference mask** was imported into Relion for particle subtraction:

Import job234

Input files: ../../masksRelion/UMASS_job231_mask_diffmask3edit.mrc

Node type: 3D Reference (.mrc)

Using the new difference mask, the nanodisc and Wzm dimer were subtracted from all particles in the Refine3D job148 run_data.star file (Relion Subtract job235) and outputs a new particle set: Subtract/job235/subtracted.star

Subtract job235

I/O Tab

Input particles: Refine3D/job148/run_data.star

Subtract partial signal? Yes

Map to be projected: Refine3D/job148/run_class001.mrc

Mask to apply to this map: Import/job234/UMASS_job231_mask_diffmask3edit.mrc

OR revert to original particles? No

CTF Tab

Do apply CTFs? Yes

Have data been phase-flipped? No

Ignore CTFs until first peak? No

5.5.3.26 3D refinement of signal subtracted particle set fails to build a satisfactory 3D volume

The 3D refinement of the subtracted particles (Refine3D job238) constructed a poor Wzt 3D volume (resolution 10.9 Å). This suggested that the particles are difficult to align into a single 3D volume perhaps because the subtracted signal leaves a particle that is too small or too off-center or other scenario that makes it difficult to align/reconstruct when Auto-sampling settings are at their defaults for full re-alignment of particles:

Auto-sampling Tab

Initial angular sampling: 7.5 degrees

Initial offset range (pix): 5

Initial offset step (pix): 1

Local searches from auto-sampling: 1.8 degrees

However, in a later 3D refinement (Refine3D job283), that also used the Subtract/job235/subtracted.star file as input, a smaller initial angular sampling angle was used, and the resolution was dramatically different (resolution 3.4 Å).

Refine3D job283 (local angular search)

I/O Tab

Input images STAR file: Subtract/job235/subtracted.star

Continue from here: Not used

Reference map: Import/job090/run_it025_class004_box256.mrc

Reference mask (optional): MaskCreate/job231/mask.mrc

Reference Tab

Ref. map is on absolute greyscale? No

Initial low-pass filter? 50

Symmetry: C1

CTF Tab

Do CTF-correction? Yes

Has ref. been CTF-corrected? Yes

Have data been phase-flipped? No

Ignore CTFs until first peak? No

Optimisation Tab

Mask diameter (A): 190

Mask individual particles with zeros? Yes

Use solvent-flattened FSCs? Yes

Auto-sampling Tab

Initial angular sampling: **1.8 degrees**

Initial offset range (pix): 5

Initial offset step (pix): 1

Local searches from auto-sampling: 1.8 degrees

Compute Tab

Use parallel disc I/O? No

Number of pooled particles: 30

Skip padding? No

Pre-read all particles into RAM? No

Copy particles to scratch directory: /srv3/home/scratch

Combine iterations through disc? No

Use GPU acceleration? Yes

Which GPUs to use: 0:1:2:3

Running Tab

Number of MPI procs: 5

Number of threads: 12

The results of Refine3D job283 support the previous thought that trying to broadly re-align the particles (initial angular sampling of Euler angles = 7.5 degrees) is difficult to do, perhaps due to their small size after signal subtraction. Thus, decreasing the initial search angle (initial angular sampling of Euler angles = 1.8 degrees) better preserves their original Euler angles. This observation suggested the possibility that re-alignment may be unnecessary and that the particles may already be at optimal Euler angles when refined as the complete particles, before signal subtraction.

In order to test this possibility, the procedure of 3D Classification, No Alignment, T=40 (Class3D job225) that was previously applied to the full transporter volume was applied to this subtracted particle set.

5.5.3.27 Wzt signal subtracted particle set: 3D classification, no alignment, T=40

In order to try to generate a particle subset with the best alignment on the Wzt CBD dimer, the Subtract/job235/subtracted.star particle set was used as input into a 3D classification job to classify the signal-subtracted particles without alignment, regularization parameter T=40, and using a Wzt CBD dimer-only mask (Relion Class3D job282).

The goal in using a Wzt CBD dimer-only mask was to attempt to generate classes based on particles that would already be aligned within the masked Wzt CBD dimer region). Toward this end, a Wzt CBD-only mask was created by fitting the atomic model of the *Aquifex aeolicus* Wzt CBD dimer (PDB 6O14) into the map from Refine3D job148. This is important for correct alignment of the mask. A Chimera molmap was created from the

Wzt CBD dimer model at 4.0 Å, and this molmap was imported into Relion (Import job275) and used to make a Wzt CBD dimer mask in Relion (MaskCreate job278).

Import job275

Input files: ../../masksRelion/UMASS_job148_CBD4Aformask.mrc

Node type: 3D Reference (.mrc)

MaskCreate job278

I/O Tab

Input 3D map: Import/job275/UMASS_job148_CBD4Aformask.mrc

Mask Tab

Lowpass filter map (A): 18

Pixel size (A): 1.49

Initial binarisation threshold: 0.03

Extend binary map this many pixels: 2

Add a soft-edge of this many pixels: 6

Running Tab

Number of threads: 32

Visual inspection in Chimera of the six classes generated by Class3D job282 determined that class002 was the best map of the Wzt CBD dimer. Relion Select job286 was used to pick class002 (51,828 particles, re-grouped into 21 groups).

5.5.3.28 3D refinement of Wzt CBD dimer particle set (MAP2)

Two Refine3D jobs were performed, each using a different mask. Relion Refine3D job287 used a Wzt CBD dimer-only mask (Relion MaskCreate job278, the same Wzt CBD dimer-only mask used for Relion Class3D job282). Refine3D job287 was cloned from Refine3D job283 and the parameters are identical with the following exceptions:

Refine3D job287

I/O Tab

Input images STAR file: Select/job286/particles.star

Continue from here: Not used

Reference map: Import/job090/run_it025_class004_box256.mrc

Reference mask (optional): MaskCreate/job278/mask.mrc

Refine3D job287 resulted in a poor alignment despite local angular searches (final resolution 13.2 Å). In order to determine whether the relatively small size of the Wzt CBD dimer only mask was the source of particle alignment problems, Refine3D job289 instead used a full Wzt dimer mask (Relion MaskCreate job231, same mask as Refine3D job283). Refine3D job289 was cloned from Refine3D job283 and is identical with the following exception:

Refine3D job289

I/O Tab

Input images STAR file: Select/job286/particles.star

Refine3D job289 resulted in Wzt dimer map with an overall resolution of 3.6 Å and the resulting volume appeared to be an accurate and complete map by visual inspection in Chimera. Relion PostProcess job290 is the corresponding post-processing job for Refine3D job289 and resulted in a sharpened map with an overall resolution of 3.6 Å (MAP2). PostProcess job290 was cloned from PostProcess job229 and is identical with the following exceptions:

PostProcess job290

I/O Tab

One of the 2 unfiltered half-maps: Refine3D/job289/run_half1_class001_unfil.mrc

Solvent mask: MaskCreate/job231/mask.mrc

5.5.3.29 Model building and real-space refinement

ATP-bound WzmWztN (PDB code 6M96, no CBD) half-transporters were fit into the post-processed map from Relion PostProcess job229 (MAP1) in Chimera (Fit In Map function). The map and the updated WzmWztN model were then brought into Coot for manual real space refinement going through the entire sequence of both Wzm protomers and both WztN protomers. An initial real space refinement in Phenix (RSR20) generated a model that was later used for the NBD dimer with MAP2. Phenix real space refinement was operated with default settings except rigid_body and morphing were applied and Use NCS was deselected. A subsequent real space refinement in Phenix (RSR34, job cloned from RSR20) generated the final WzmWztN model used for later building a full WzmWzt model.

In order to create the full Wzt dimer model, the Wzt NBD dimer from WzmWztN RSR20 was fit into the post-processed MAP2 (Relion PostProcess job290) in Chimera. The Aa Wzt CBD dimer crystal structure (PDB code 6O14) was then fit into the post-processed job290 map, MAP2. The two models were combined using Chimera and output as a single Wzt dimer atomic model.

MAP2 and the initial full Wzt dimer model were loaded into Coot for manual real space refinement. No density for a final β-strand of the CBD jellyroll surface was observed in MAP2, despite this final β-strand being observed in the Aa Wzt CBD dimer crystal structure. Thus, for both Wzt CBD protomers, Trp362 was fit into the corresponding density above the putative O antigen cap binding pocket, and the residues of the short segment leading to the disordered loop were removed up to Lys351 due to poor density for this region. Ultimately, this full Wzt dimer model was real space refined in Phenix (RSR46, same settings as RSR20, RSR34) against post-processed job290 map, MAP2.

To build the complete full-length WzmWzt model, the WzmWztN model (RSR34) and postprocessed MAP1 were loaded into Chimera. The full Wzt dimer model (RSR46) was then loaded into Chimera and also fit into postprocessed MAP1. Fitting the full Wzt dimer (RSR46) model into postprocessed MAP1 superimposed the NBDs of the full Wzt dimer on the WzmWztN NBDs, aligning the two Wzt NBD dimers and placing the RSR46 Wzt CBD dimer into the corresponding density in postprocessed MAP1. The MAP2-based Wzt NBD dimer (RSR46) was then deleted in Chimera, leaving a full-length WzmWzt model comprised of the MAP1-based WzmWztN transporter model (RSR34) and the MAP2-

based Wzt CBD dimer (RSR46). The full-length WzmWzt model was then refined in Phenix against postprocessed MAP1 using the real-space refine program with default settings with only ADP selected in order to avoid disparate B-factors for WzmWztN and the Wzt CBD dimer and secondary structure restraints are not used. Finally, the resulting atomic coordinate file was loaded into Coot and the separate NBD and CBD chains were merged to create full-length Wzt protomers as single chains such that PyMol renders the two Wzt NBD-CBD linker regions when 'set cartoon_gap_cutoff, 15' is configured. This chain change resulted in the final full-length WzmWzt atomic model coordinate file.

5.5.3.30 Data availability

MAP1 was deposited in the Electron Microscopy Data Bank (EMDB) with the entry ID EMD-22644. Coordinates for the final, full-length WzmWzt atomic model have been deposited in the Protein Data Bank (PDB) with the entry ID 7K2T.

6 References

1. C. Whitfield, Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. *Annu Rev Biochem* **75**, 39-68 (2006).
2. C. Whitfield, M. S. Trent, Biosynthesis and export of bacterial lipopolysaccharides. *Annu Rev Biochem* **83**, 99-128 (2014).
3. S. Brown, J. P. Santa Maria, Jr., S. Walker, Wall teichoic acids of gram-positive bacteria. *Annu Rev Microbiol* **67**, 313-336 (2013).
4. E. T. Touz, D. Mengin-Lecreulx, Undecaprenyl Phosphate Synthesis. *EcoSal Plus* **3** (2008).
5. G. Manat *et al.*, Deciphering the metabolism of undecaprenyl-phosphate: the bacterial cell-wall unit carrier at the membrane frontier. *Microb Drug Resist* **20**, 199-214 (2014).
6. J. Eichler, Z. Guan, Lipid sugar carriers at the extremes: The phosphodolichols Archaea use in N-glycosylation. *Biochim Biophys Acta Mol Cell Biol Lipids* **1862**, 589-599 (2017).
7. L. Surmacz, E. Swiezewska, Polyisoprenoids - Secondary metabolites or physiologically important superlipids? *Biochem Biophys Res Commun* **407**, 627-632 (2011).
8. M. El Ghachi *et al.*, Crystal structure of undecaprenyl-pyrophosphate phosphatase and its role in peptidoglycan biosynthesis. *Nat Commun* **9**, 1078 (2018).
9. S. D. Workman, L. J. Worrall, N. C. J. Strynadka, Crystal structure of an intramembranal phosphatase central to bacterial cell-wall peptidoglycan biosynthesis and lipid recycling. *Nat Commun* **9**, 1159 (2018).
10. A. Vergara-Jaque, C. Fenollar-Ferrer, D. Kaufmann, L. R. Forrest, Repeat-swap homology modeling of secondary active transporters: updated protocol and prediction of elevator-type mechanisms. *Front Pharmacol* **6**, 183 (2015).
11. M. El Ghachi, A. Bouhss, D. Blanot, D. Mengin-Lecreulx, The bacA gene of Escherichia coli encodes an undecaprenyl pyrophosphate phosphatase activity. *J Biol Chem* **279**, 30106-30113 (2004).
12. M. El Ghachi, A. Derbise, A. Bouhss, D. Mengin-Lecreulx, Identification of multiple genes encoding membrane proteins with undecaprenyl pyrophosphate phosphatase (UppP) activity in Escherichia coli. *J Biol Chem* **280**, 18689-18695 (2005).
13. F. Fernandez *et al.*, The CWH8 gene encodes a dolichyl pyrophosphate phosphatase with a luminally oriented active site in the endoplasmic reticulum of Saccharomyces cerevisiae. *J Biol Chem* **276**, 41455-41464 (2001).
14. J. S. Rush, S. K. Cho, S. Jiang, S. L. Hofmann, C. J. Waechter, Identification and characterization of a cDNA encoding a dolichyl pyrophosphate phosphatase located in the endoplasmic reticulum of mammalian cells. *J Biol Chem* **277**, 45226-45234 (2002).
15. M. E. Ghachi *et al.*, Crystal structure and biochemical characterization of the transmembrane PAP2 type phosphatidylglycerol phosphate phosphatase from Bacillus subtilis. *Cell Mol Life Sci* **74**, 2319-2332 (2017).

16. J. Fan, D. Jiang, Y. Zhao, J. Liu, X. C. Zhang, Crystal structure of lipid phosphatase Escherichia coli phosphatidylglycerophosphate phosphatase B. *Proc Natl Acad Sci U S A* **111**, 7636-7640 (2014).
17. Y. J. Sigal, M. I. McDermott, A. J. Morris, Integral membrane lipid phosphatases/phosphotransferases: common structure and diverse functions. *Biochem J* **387**, 281-293 (2005).
18. M. D. Hartley, B. Imperiali, At the membrane frontier: a prospectus on the remarkable evolutionary conservation of polyprenols and polyprenyl-phosphates. *Arch Biochem Biophys* **517**, 83-97 (2012).
19. G. van Duijn *et al.*, Dolichyl phosphate induces non-bilayer structures, vesicle fusion and transbilayer movement of lipids: a model membrane study. *Biochim Biophys Acta* **861**, 211-223 (1986).
20. G. P. Zhou, F. A. Troy, 2nd, NMR study of the preferred membrane orientation of polyisoprenols (dolichol) and the impact of their complex with polyisoprenyl recognition sequence peptides on membrane structure. *Glycobiology* **15**, 347-359 (2005).
21. X. Wang, A. R. Mansourian, P. J. Quinn, The effect of dolichol on the structure and phase behaviour of phospholipid model membranes. *Mol Membr Biol* **25**, 547-556 (2008).
22. J. Yoo *et al.*, GlcNAc-1-P-transferase-tunicamycin complex structure reveals basis for inhibition of N-glycosylation. *Nat Struct Mol Biol* **25**, 217-224 (2018).
23. J. K. Hakulinen *et al.*, MraY-antibiotic complex reveals details of tunicamycin mode of action. *Nat Chem Biol* **13**, 265-267 (2017).
24. B. C. Chung *et al.*, Structural insights into inhibition of lipid I production in bacterial cell wall synthesis. *Nature* **533**, 557-560 (2016).
25. B. C. Chung *et al.*, Crystal structure of MraY, an essential membrane enzyme for bacterial cell wall synthesis. *Science* **341**, 1012-1016 (2013).
26. M. Kowarik *et al.*, N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science* **314**, 1148-1150 (2006).
27. C. Perez *et al.*, Structure and mechanism of an active lipid-linked oligosaccharide flippase. *Nature* **524**, 433-438 (2015).
28. C. Perez, A. R. Mehdipour, G. Hummer, K. P. Locher, Structure of Outward-Facing PglK and Molecular Dynamics of Lipid-Linked Oligosaccharide Recognition and Translocation. *Structure* **27**, 669-678 e665 (2019).
29. C. R. Raetz, C. M. Reynolds, M. S. Trent, R. E. Bishop, Lipid A modification systems in gram-negative bacteria. *Annu Rev Biochem* **76**, 295-329 (2007).
30. B. D. Needham, M. S. Trent, Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. *Nat Rev Microbiol* **11**, 467-481 (2013).
31. V. I. Petrou *et al.*, Structures of aminoarabinose transferase ArnT suggest a molecular basis for lipid A glycosylation. *Science* **351**, 608-612 (2016).
32. M. Napiorkowska *et al.*, Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. *Nat Struct Mol Biol* **24**, 1100-1106 (2017).
33. C. Lizak, S. Gerber, S. Numao, M. Aebi, K. P. Locher, X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474**, 350-355 (2011).

34. C. A. Caffalette, R. A. Corey, M. S. P. Sansom, P. J. Stansfeld, J. Zimmer, A lipid gating mechanism for the channel-forming O antigen ABC transporter. *Nat Commun* **10**, 824 (2019).
35. M. Wacker *et al.*, N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science* **298**, 1790-1793 (2002).
36. J. Breitling, M. Aebi, N-linked protein glycosylation in the endoplasmic reticulum. *Cold Spring Harb Perspect Biol* **5**, a013359 (2013).
37. K. F. Jarrell *et al.*, N-linked glycosylation in Archaea: a structural, functional, and genetic analysis. *Microbiol Mol Biol Rev* **78**, 304-341 (2014).
38. H. Nothaft, C. M. Szymanski, Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* **8**, 765-778 (2010).
39. P. Reeves, Role of O-antigen variation in the immune response. *Trends Microbiol* **3**, 381-386 (1995).
40. S. D. Liston, E. Mann, C. Whitfield, Glycolipid substrates for ABC transporters required for the assembly of bacterial cell-envelope and cell-surface glycoconjugates. *Biochim Biophys Acta Mol Cell Biol Lipids* **1862**, 1394-1403 (2017).
41. C. R. Raetz, C. Whitfield, Lipopolysaccharide endotoxins. *Annu Rev Biochem* **71**, 635-700 (2002).
42. J. S. Lam, V. L. Taylor, S. T. Islam, Y. Hao, D. Kocincova, Genetic and Functional Diversity of Pseudomonas aeruginosa Lipopolysaccharide. *Front Microbiol* **2**, 118 (2011).
43. L. Wang, D. Liu, P. R. Reeves, C-terminal half of Salmonella enterica WbaP (RfbP) is the galactosyl-1-phosphate transferase domain catalyzing the first step of O-antigen synthesis. *J Bacteriol* **178**, 2598-2604 (1996).
44. J. Lehrer, K. A. Vigeant, L. D. Tatar, M. A. Valvano, Functional characterization and membrane topology of Escherichia coli WecA, a sugar-phosphate transferase initiating the biosynthesis of enterobacterial common antigen and O-antigen lipopolysaccharide. *J Bacteriol* **189**, 2618-2628 (2007).
45. H. L. Rocchetta, L. L. Burrows, J. C. Pacan, J. S. Lam, Three rhamnosyltransferases responsible for assembly of the A-band D-rhamnan polysaccharide in Pseudomonas aeruginosa: a fourth transferase, WbpL, is required for the initiation of both A-band and B-band lipopolysaccharide synthesis. *Mol Microbiol* **28**, 1103-1119 (1998).
46. J. S. Rush, P. D. Rick, C. J. Waechter, Polyisoprenyl phosphate specificity of UDP-GlcNAc:undecaprenyl phosphate N-acetylglucosaminyl 1-P transferase from E.coli. *Glycobiology* **7**, 315-322 (1997).
47. B. Al-Dabbagh, D. Mengin-Lecreulx, A. Bouhss, Purification and characterization of the bacterial UDP-GlcNAc:undecaprenyl-phosphate GlcNAc-1-phosphate transferase WecA. *J Bacteriol* **190**, 7141-7146 (2008).
48. B. Al-Dabbagh *et al.*, Catalytic mechanism of MraY and WecA, two paralogues of the polyprenyl-phosphate N-acetylhexosamine 1-phosphate transferase superfamily. *Biochimie* **127**, 249-257 (2016).
49. B. Al-Dabbagh *et al.*, Active site mapping of MraY, a member of the polyprenyl-phosphate N-acetylhexosamine 1-phosphate transferase superfamily, catalyzing

the first membrane step of peptidoglycan biosynthesis. *Biochemistry* **47**, 8919-8928 (2008).
50. Y. Liu *et al.*, New Insight into the Catalytic Mechanism of Bacterial MraY from Enzyme Kinetics and Docking Studies. *J Biol Chem* **291**, 15057-15068 (2016).
51. Y. Hong, M. A. Liu, P. R. Reeves, Progress in Our Understanding of Wzx Flippase for Translocation of Bacterial Membrane Lipid-Linked Oligosaccharide. *J Bacteriol* **200** (2018).
52. I. T. Paulsen, A. M. Beness, M. H. Saier, Jr., Computer-based analyses of the protein constituents of transport systems catalysing export of complex carbohydrates in bacteria. *Microbiology (Reading)* **143 (Pt 8)**, 2685-2699 (1997).
53. M. F. Feldman *et al.*, The activity of a putative polyisoprenol-linked sugar translocase (Wzx) involved in Escherichia coli O antigen assembly is independent of the chemical structure of the O repeat. *J Biol Chem* **274**, 35129-35138 (1999).
54. C. L. Marolda, L. D. Tatar, C. Alaimo, M. Aebi, M. A. Valvano, 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 (2006).
55. C. L. Marolda, J. Vicarioli, M. A. Valvano, Wzx proteins involved in biosynthesis of O antigen function in association with the first sugar of the O-specific lipopolysaccharide subunit. *Microbiology (Reading)* **150**, 4095-4105 (2004).
56. S. T. Islam, J. S. Lam, Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway. *Can J Microbiol* **60**, 697-716 (2014).
57. M. A. Liu, P. Morris, P. R. Reeves, Wzx flippases exhibiting complex O-unit preferences require a new model for Wzx-substrate interactions. *Microbiologyopen* **8**, e00655 (2019).
58. S. T. Islam *et al.*, A cationic lumen in the Wzx flippase mediates anionic O-antigen subunit translocation in Pseudomonas aeruginosa PAO1. *Mol Microbiol* **84**, 1165-1176 (2012).
59. S. T. Islam *et al.*, Proton-dependent gating and proton uptake by Wzx support O-antigen-subunit antiport across the bacterial inner membrane. *mBio* **4**, e00678-00613 (2013).
60. A. C. Kuk, E. H. Mashalidis, S. Y. Lee, Crystal structure of the MOP flippase MurJ in an inward-facing conformation. *Nat Struct Mol Biol* **24**, 171-176 (2017).
61. A. C. Y. Kuk, A. Hao, Z. Guan, S. Y. Lee, Visualizing conformation transitions of the Lipid II flippase MurJ. *Nat Commun* **10**, 1736 (2019).
62. C. Whitfield, Biosynthesis of lipopolysaccharide O antigens. *Trends Microbiol* **3**, 178-185 (1995).
63. T. H. Kim *et al.*, Characterization of the O-antigen polymerase (Wzy) of Francisella tularensis. *J Biol Chem* **285**, 27839-27849 (2010).
64. S. T. Islam, V. L. Taylor, M. Qi, J. S. Lam, Membrane topology mapping of the O-antigen flippase (Wzx), polymerase (Wzy), and ligase (WaaL) from Pseudomonas aeruginosa PAO1 reveals novel domain architectures. *mBio* **1** (2010).
65. R. Woodward *et al.*, In vitro bacterial polysaccharide biosynthesis: defining the functions of Wzy and Wzz. *Nat Chem Biol* **6**, 418-423 (2010).

66. L. Cuthbertson, I. L. Mainprize, J. H. Naismith, C. Whitfield, Pivotal roles of the outer membrane polysaccharide export and polysaccharide copolymerase protein families in export of extracellular polysaccharides in gram-negative bacteria. *Microbiol Mol Biol Rev* **73**, 155-177 (2009).
67. R. F. Collins *et al.*, Full-length, Oligomeric Structure of Wzz Determined by Cryoelectron Microscopy Reveals Insights into Membrane-Bound States. *Structure* **25**, 806-815 e803 (2017).
68. A. Tocilj *et al.*, Bacterial polysaccharide co-polymerases share a common framework for control of polymer length. *Nat Struct Mol Biol* **15**, 130-138 (2008).
69. K. Larue, M. S. Kimber, R. Ford, C. Whitfield, Biochemical and structural analysis of bacterial O-antigen chain length regulator proteins reveals a conserved quaternary structure. *J Biol Chem* **284**, 7395-7403 (2009).
70. S. Kalynych, M. Cherney, M. Bostina, I. Rouiller, M. Cygler, Quaternary structure of WzzB and WzzE polysaccharide copolymerases. *Protein Sci* **24**, 58-69 (2015).
71. S. Kalynych, D. Yao, J. Magee, M. Cygler, Structural characterization of closely related O-antigen lipopolysaccharide (LPS) chain length regulators. *J Biol Chem* **287**, 15696-15705 (2012).
72. M. Papadopoulos, R. Morona, 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 (2010).
73. S. M. Huszczynski, C. Coumoundouros, P. Pham, J. S. Lam, C. M. Khursigara, Unique Regions of the Polysaccharide Copolymerase Wzz2 from Pseudomonas aeruginosa Are Essential for O-Specific Antigen Chain Length Control. *J Bacteriol* **201** (2019).
74. B. R. Clarke, L. Cuthbertson, C. Whitfield, Nonreducing terminal modifications determine the chain length of polymannose O antigens of Escherichia coli and couple chain termination to polymer export via an ATP-binding cassette transporter. *J Biol Chem* **279**, 35709-35718 (2004).
75. V. Kos, L. Cuthbertson, C. Whitfield, The Klebsiella pneumoniae O2a antigen defines a second mechanism for O antigen ATP-binding cassette transporters. *J Biol Chem* **284**, 2947-2956 (2009).
76. L. Cuthbertson, M. S. Kimber, C. Whitfield, Substrate binding by a bacterial ABC transporter involved in polysaccharide export. *Proc Natl Acad Sci U S A* **104**, 19529-19534 (2007).
77. V. Kos, C. Whitfield, A membrane-located glycosyltransferase complex required for biosynthesis of the D-galactan I lipopolysaccharide O antigen in Klebsiella pneumoniae. *J Biol Chem* **285**, 19668-19687 (2010).
78. B. R. Clarke, L. K. Greenfield, C. Bouwman, C. Whitfield, Coordination of polymerization, chain termination, and export in assembly of the Escherichia coli lipopolysaccharide O9a antigen in an ATP-binding cassette transporter-dependent pathway. *J Biol Chem* **284**, 30662-30672 (2009).
79. N. Kido *et al.*, Expression of the O9 polysaccharide of Escherichia coli: sequencing of the E. coli O9 rfb gene cluster, characterization of mannosyl

transferases, and evidence for an ATP-binding cassette transport system. *J Bacteriol* **177**, 2178-2187 (1995).
80. L. K. Greenfield *et al.*, Biosynthesis of the polymannose lipopolysaccharide O-antigens from Escherichia coli serotypes O8 and O9a requires a unique combination of single- and multiple-active site mannosyltransferases. *J Biol Chem* **287**, 35078-35091 (2012).
81. B. R. Clarke *et al.*, A bifunctional O-antigen polymerase structure reveals a new glycosyltransferase family. *Nat Chem Biol* 10.1038/s41589-020-0494-0 (2020).
82. S. Guan, A. J. Clarke, C. Whitfield, Functional analysis of the galactosyltransferases required for biosynthesis of D-galactan I, a component of the lipopolysaccharide O1 antigen of Klebsiella pneumoniae. *J Bacteriol* **183**, 3318-3327 (2001).
83. N. Kido, T. Sugiyama, T. Yokochi, H. Kobayashi, Y. Okawa, Synthesis of Escherichia coli O9a polysaccharide requires the participation of two domains of WbdA, a mannosyltransferase encoded within the wb* gene cluster. *Mol Microbiol* **27**, 1213-1221 (1998).
84. L. K. Greenfield *et al.*, Domain organization of the polymerizing mannosyltransferases involved in synthesis of the Escherichia coli O8 and O9a lipopolysaccharide O-antigens. *J Biol Chem* **287**, 38135-38149 (2012).
85. S. D. Liston *et al.*, Domain interactions control complex formation and polymerase specificity in the biosynthesis of the Escherichia coli O9a antigen. *J Biol Chem* **290**, 1075-1085 (2015).
86. G. Hagelueken *et al.*, Structure of WbdD: a bifunctional kinase and methyltransferase that regulates the chain length of the O antigen in Escherichia coli O9a. *Mol Microbiol* **86**, 730-742 (2012).
87. G. Hagelueken *et al.*, A coiled-coil domain acts as a molecular ruler to regulate O-antigen chain length in lipopolysaccharide. *Nat Struct Mol Biol* **22**, 50-56 (2015).
88. L. K. Greenfield, C. Whitfield, Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. *Carbohydr Res* **356**, 12-24 (2012).
89. Y. Bi, E. Mann, C. Whitfield, J. Zimmer, Architecture of a channel-forming O-antigen polysaccharide ABC transporter. *Nature* **553**, 361-365 (2018).
90. E. Mann, E. Mallette, B. R. Clarke, M. S. Kimber, C. Whitfield, The Klebsiella pneumoniae O12 ATP-binding Cassette (ABC) Transporter Recognizes the Terminal Residue of Its O-antigen Polysaccharide Substrate. *J Biol Chem* **291**, 9748-9761 (2016).
91. H. Hashimoto, Recent structural studies of carbohydrate-binding modules. *Cell Mol Life Sci* **63**, 2954-2967 (2006).
92. J. L. Morgan, J. Strumillo, J. Zimmer, Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature* **493**, 181-186 (2013).
93. C. Whitfield, P. A. Amor, R. Koplin, Modulation of the surface architecture of gram-negative bacteria by the action of surface polymer:lipid A-core ligase and by determinants of polymer chain length. *Mol Microbiol* **23**, 629-638 (1997).

94. P. D. Abeyrathne, J. S. Lam, WaaL of Pseudomonas aeruginosa utilizes ATP in in vitro ligation of O antigen onto lipid A-core. *Mol Microbiol* **65**, 1345-1359 (2007).
95. X. Ruan, J. Monjaras Feria, M. Hamad, M. A. Valvano, Escherichia coli and Pseudomonas aeruginosa lipopolysaccharide O-antigen ligases share similar membrane topology and biochemical properties. *Mol Microbiol* **110**, 95-113 (2018).
96. J. M. Perez, M. A. McGarry, C. L. Marolda, M. A. Valvano, Functional analysis of the large periplasmic loop of the Escherichia coli K-12 WaaL O-antigen ligase. *Mol Microbiol* **70**, 1424-1440 (2008).
97. S. Schild, A. K. Lamprecht, J. Reidl, Molecular and functional characterization of O antigen transfer in Vibrio cholerae. *J Biol Chem* **280**, 25936-25947 (2005).
98. L. L. Lairson, B. Henrissat, G. J. Davies, S. G. Withers, Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **77**, 521-555 (2008).
99. N. T. Reichmann, A. Grundling, Location, synthesis and function of glycolipids and polyglycerolphosphate lipoteichoic acid in Gram-positive bacteria of the phylum Firmicutes. *FEMS Microbiol Lett* **319**, 97-105 (2011).
100. A. L. Lovering *et al.*, Structure of the bacterial teichoic acid polymerase TagF provides insights into membrane association and catalysis. *Nat Struct Mol Biol* **17**, 582-589 (2010).
101. J. W. Schertzer, E. D. Brown, Use of CDP-glycerol as an alternate acceptor for the teichoic acid polymerase reveals that membrane association regulates polymer length. *J Bacteriol* **190**, 6940-6947 (2008).
102. K. Schirner, J. Marles-Wright, R. J. Lewis, J. Errington, Distinct and essential morphogenic functions for wall- and lipo-teichoic acids in Bacillus subtilis. *EMBO J* **28**, 830-842 (2009).
103. D. Lu *et al.*, Structure-based mechanism of lipoteichoic acid synthesis by Staphylococcus aureus LtaS. *Proc Natl Acad Sci U S A* **106**, 1584-1589 (2009).
104. E. Cabacungan, R. A. Pieringer, Mode of elongation of the glycerol phosphate polymer of membrane lipoteichoic acid of Streptococcus faecium ATCC 9790. *J Bacteriol* **147**, 75-79 (1981).
105. K. Schirner, L. K. Stone, S. Walker, ABC transporters required for export of wall teichoic acids do not discriminate between different main chain polymers. *ACS Chem Biol* **6**, 407-412 (2011).
106. D. Gerlach *et al.*, Methicillin-resistant Staphylococcus aureus alters cell wall glycosylation to evade immunity. *Nature* **563**, 705-709 (2018).
107. S. Sobhanifar *et al.*, Structure and mechanism of Staphylococcus aureus TarM, the wall teichoic acid alpha-glycosyltransferase. *Proc Natl Acad Sci U S A* **112**, E576-585 (2015).
108. C. Koc *et al.*, Structural and enzymatic analysis of TarM glycosyltransferase from Staphylococcus aureus reveals an oligomeric protein specific for the glycosylation of wall teichoic acid. *J Biol Chem* **290**, 9874-9885 (2015).
109. S. Sobhanifar *et al.*, Structure and Mechanism of Staphylococcus aureus TarS, the Wall Teichoic Acid beta-glycosyltransferase Involved in Methicillin Resistance. *PLoS Pathog* **12**, e1006067 (2016).

110. N. T. Reichmann, C. P. Cassona, A. Grundling, Revised mechanism of D-alanine incorporation into cell wall polymers in Gram-positive bacteria. *Microbiology (Reading)* **159**, 1868-1877 (2013).
111. G. B. Whitfield, L. S. Marmont, P. L. Howell, Enzymatic modifications of exopolysaccharides enhance bacterial persistence. *Front Microbiol* **6**, 471 (2015).
112. M. Perego *et al.*, Incorporation of D-alanine into lipoteichoic acid and wall teichoic acid in Bacillus subtilis. Identification of genes and regulation. *J Biol Chem* **270**, 15598-15606 (1995).
113. D. Ma *et al.*, Crystal structure of a membrane-bound O-acyltransferase. *Nature* **562**, 286-290 (2018).
114. M. A. Horwitz, S. C. Silverstein, Influence of the Escherichia coli capsule on complement fixation and on phagocytosis and killing by human phagocytes. *J Clin Invest* **65**, 82-94 (1980).
115. B. F. Cress *et al.*, Masquerading microbial pathogens: capsular polysaccharides mimic host-tissue molecules. *FEMS Microbiol Rev* **38**, 660-697 (2014).
116. S. Sachdeva, R. V. Palur, K. U. Sudhakar, T. Rathinavelan, E. coli Group 1 Capsular Polysaccharide Exportation Nanomachinary as a Plausible Antivirulence Target in the Perspective of Emerging Antimicrobial Resistance. *Front Microbiol* **8**, 70 (2017).
117. C. M. Taylor, I. S. Roberts, Capsular polysaccharides and their role in virulence. *Contrib Microbiol* **12**, 55-66 (2005).
118. C. Hyams, E. Camberlein, J. M. Cohen, K. Bax, J. S. Brown, The Streptococcus pneumoniae capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. *Infect Immun* **78**, 704-715 (2010).
119. C. Whitfield, Glycan chain-length control. *Nat Chem Biol* **6**, 403-404 (2010).
120. T. Wugeditsch *et al.*, Phosphorylation of Wzc, a tyrosine autokinase, is essential for assembly of group 1 capsular polysaccharides in Escherichia coli. *J Biol Chem* **276**, 2361-2371 (2001).
121. R. F. Collins *et al.*, The 3D structure of a periplasm-spanning platform required for assembly of group 1 capsular polysaccharides in Escherichia coli. *Proc Natl Acad Sci U S A* **104**, 2390-2395 (2007).
122. E. Bechet *et al.*, Identification of structural and molecular determinants of the tyrosine-kinase Wzc and implications in capsular polysaccharide export. *Mol Microbiol* **77**, 1315-1325 (2010).
123. J. Drummelsmith, C. Whitfield, Translocation of group 1 capsular polysaccharide to the surface of Escherichia coli requires a multimeric complex in the outer membrane. *EMBO J* **19**, 57-66 (2000).
124. C. Dong *et al.*, Wza the translocon for E. coli capsular polysaccharides defines a new class of membrane protein. *Nature* **444**, 226-229 (2006).
125. N. N. Nickerson *et al.*, Trapped translocation intermediates establish the route for export of capsular polysaccharides across Escherichia coli outer membranes. *Proc Natl Acad Sci U S A* **111**, 8203-8208 (2014).
126. K. Jann, B. Jann, Capsules of Escherichia coli, expression and biological significance. *Can J Microbiol* **38**, 705-710 (1992).

127. L. M. Willis *et al.*, Conserved glycolipid termini in capsular polysaccharides synthesized by ATP-binding cassette transporter-dependent pathways in Gram-negative pathogens. *Proc Natl Acad Sci U S A* **110**, 7868-7873 (2013).
128. L. M. Willis, C. Whitfield, KpsC and KpsS are retaining 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) transferases involved in synthesis of bacterial capsules. *Proc Natl Acad Sci U S A* **110**, 20753-20758 (2013).
129. W. E. Kohlbrenner, S. W. Fesik, Determination of the anomeric specificity of the Escherichia coli CTP:CMP-3-deoxy-D-manno-octulosonate cytidylyltransferase by 13C NMR spectroscopy. *J Biol Chem* **260**, 14695-14700 (1985).
130. S. W. Satola, P. L. Schirmer, M. M. Farley, Complete sequence of the cap locus of Haemophilus influenzae serotype b and nonencapsulated b capsule-negative variants. *Infect Immun* **71**, 3639-3644 (2003).
131. M. Frosch, A. Muller, Phospholipid substitution of capsular polysaccharides and mechanisms of capsule formation in Neisseria meningitidis. *Mol Microbiol* **8**, 483-493 (1993).
132. S. Jelakovic, K. Jann, G. E. Schulz, The three-dimensional structure of capsule-specific CMP: 2-keto-3-deoxy-manno-octonic acid synthetase from Escherichia coli. *FEBS Lett* **391**, 157-161 (1996).
133. L. Doyle *et al.*, Biosynthesis of a conserved glycolipid anchor for Gram-negative bacterial capsules. *Nat Chem Biol* **15**, 632-640 (2019).
134. S. D. Liston, O. G. Ovchinnikova, C. Whitfield, Unique lipid anchor attaches Vi antigen capsule to the surface of Salmonella enterica serovar Typhi. *Proc Natl Acad Sci U S A* **113**, 6719-6724 (2016).
135. X. Hu *et al.*, Vi capsular polysaccharide: Synthesis, virulence, and application. *Crit Rev Microbiol* **43**, 440-452 (2017).
136. P. L. DeAngelis, C. L. White, Identification and molecular cloning of a heparosan synthase from Pasteurella multocida type D. *J Biol Chem* **277**, 7209-7213 (2002).
137. W. Jing, P. L. DeAngelis, Dissection of the two transferase activities of the Pasteurella multocida hyaluronan synthase: two active sites exist in one polypeptide. *Glycobiology* **10**, 883-889 (2000).
138. T. Osawa *et al.*, Crystal structure of chondroitin polymerase from Escherichia coli K4. *Biochem Biophys Res Commun* **378**, 10-14 (2009).
139. A. Wasteson, A method for the determination of the molecular weight and molecular-weight distribution of chondroitin sulphate. *J Chromatogr* **59**, 87-97 (1971).
140. T. Ninomiya *et al.*, Molecular cloning and characterization of chondroitin polymerase from Escherichia coli strain K4. *J Biol Chem* **277**, 21567-21575 (2002).
141. W. Jing, P. L. DeAngelis, Analysis of the two active sites of the hyaluronan synthase and the chondroitin synthase of Pasteurella multocida. *Glycobiology* **13**, 661-671 (2003).
142. J. Mandawe *et al.*, Directed Evolution of Hyaluronic Acid Synthase from Pasteurella multocida towards High-Molecular-Weight Hyaluronic Acid. *Chembiochem* **19**, 1414-1423 (2018).

143. N. J. Otto *et al.*, Structure/function analysis of Pasteurella multocida heparosan synthases: toward defining enzyme specificity and engineering novel catalysts. *J Biol Chem* **287**, 7203-7212 (2012).
144. N. Sugiura *et al.*, Glucuronyltransferase activity of KfiC from Escherichia coli strain K5 requires association of KfiA: KfiC and KfiA are essential enzymes for production of K5 polysaccharide, N-acetylheparosan. *J Biol Chem* **285**, 1597-1606 (2010).
145. J. Liu *et al.*, KfoE encodes a fructosyltransferase involved in capsular polysaccharide biosynthesis in Escherichia coli K4. *Biotechnol Lett* **36**, 1469-1477 (2014).
146. A. Neuberger, D. Du, B. F. Luisi, Structure and mechanism of bacterial tripartite efflux pumps. *Res Microbiol* **169**, 401-413 (2018).
147. C. J. Nsahlai, R. P. Silver, Purification and characterization of KpsT, the ATP-binding component of the ABC-capsule exporter of Escherichia coli K1. *FEMS Microbiol Lett* **224**, 113-118 (2003).
148. M. S. Pavelka, Jr., S. F. Hayes, R. P. Silver, Characterization of KpsT, the ATP-binding component of the ABC-transporter involved with the export of capsular polysialic acid in Escherichia coli K1. *J Biol Chem* **269**, 20149-20158 (1994).
149. R. Y. Lo, L. J. McKerral, T. L. Hills, M. Kostrzynska, Analysis of the capsule biosynthetic locus of Mannheimia (Pasteurella) haemolytica A1 and proposal of a nomenclature system. *Infect Immun* **69**, 4458-4464 (2001).
150. S. D. Liston *et al.*, Periplasmic depolymerase provides insight into ABC transporter-dependent secretion of bacterial capsular polysaccharides. *Proc Natl Acad Sci U S A* **115**, E4870-E4879 (2018).
151. I. S. Roberts, The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annu Rev Microbiol* **50**, 285-315 (1996).
152. J. Poole, C. J. Day, M. von Itzstein, J. C. Paton, M. P. Jennings, Glycointeractions in bacterial pathogenesis. *Nat Rev Microbiol* **16**, 440-452 (2018).
153. B. B. Finlay, G. McFadden, Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. *Cell* **124**, 767-782 (2006).
154. C. Whitfield, C. M. Szymanski, M. Aebi, "Eubacteria" in Essentials of Glycobiology*,* rd *et al.*, Eds. (Cold Spring Harbor (NY), 2015), 10.1101/glycobiology.3e.021, pp. 265-282.
155. R. J. Mostowy, K. E. Holt, Diversity-Generating Machines: Genetics of Bacterial Sugar-Coating. *Trends Microbiol* **26**, 1008-1021 (2018).
156. I. Hug, M. F. Feldman, Analogies and homologies in lipopolysaccharide and glycoprotein biosynthesis in bacteria. *Glycobiology* **21**, 138-151 (2011).
157. J. M. May, D. J. Sherman, B. W. Simpson, N. Ruiz, D. Kahne, Lipopolysaccharide transport to the cell surface: periplasmic transport and assembly into the outer membrane. *Philos Trans R Soc Lond B Biol Sci* **370** (2015).
158. S. Kalynych, X. Ruan, M. A. Valvano, M. Cygler, Structure-guided investigation of lipopolysaccharide O-antigen chain length regulators reveals regions critical for modal length control. *J Bacteriol* **193**, 3710-3721 (2011).

159. L. M. Matano *et al.*, Antibiotic That Inhibits the ATPase Activity of an ATP-Binding Cassette Transporter by Binding to a Remote Extracellular Site. *J Am Chem Soc* **139**, 10597-10600 (2017).
160. M. Kumari, R. B. Sunoj, P. V. Balaji, Exploration of CH...pi mediated stacking interactions in saccharide: aromatic residue complexes through conformational sampling. *Carbohydr Res* **361**, 133-140 (2012).
161. P. C. Smith *et al.*, ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol Cell* **10**, 139-149 (2002).
162. L. Cuthbertson, J. Powers, C. Whitfield, The C-terminal domain of the nucleotide-binding domain protein Wzt determines substrate specificity in the ATP-binding cassette transporter for the lipopolysaccharide O-antigens in Escherichia coli serotypes O8 and O9a. *J Biol Chem* **280**, 30310-30319 (2005).
163. K. P. Locher, Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat Struct Mol Biol* **23**, 487-493 (2016).
164. A. L. Davidson, E. Dassa, C. Orelle, J. Chen, Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol Mol Biol Rev* **72**, 317-364, table of contents (2008).
165. S. B. Long, P. J. Casey, L. S. Beese, Reaction path of protein farnesyltransferase at atomic resolution. *Nature* **419**, 645-650 (2002).
166. J. Kubler-Kielb, C. Whitfield, E. Katzenellenbogen, E. Vinogradov, Identification of the methyl phosphate substituent at the non-reducing terminal mannose residue of the O-specific polysaccharides of Klebsiella pneumoniae O3, Hafnia alvei PCM 1223 and Escherichia coli O9/O9a LPS. *Carbohydr Res* **347**, 186-188 (2012).
167. W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38 (1996).
168. J. L. Trick *et al.*, Functional Annotation of Ion Channel Structures by Molecular Simulation. *Structure* **24**, 2207-2216 (2016).
169. G. Klesse, S. Rao, M. S. P. Sansom, S. J. Tucker, CHAP: A Versatile Tool for the Structural and Functional Annotation of Ion Channel Pores. *J Mol Biol* **431**, 3353-3365 (2019).
170. L. Monticelli *et al.*, The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J Chem Theory Comput* **4**, 819-834 (2008).
171. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. de Vries, The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B* **111**, 7812-7824 (2007).
172. R. Gaudet, D. C. Wiley, Structure of the ABC ATPase domain of human TAP1, the transporter associated with antigen processing. *EMBO J* **20**, 4964-4972 (2001).
173. C. Orelle *et al.*, Dynamics of alpha-helical subdomain rotation in the intact maltose ATP-binding cassette transporter. *Proc Natl Acad Sci U S A* **107**, 20293-20298 (2010).
174. J. C. Cordova *et al.*, Stochastic but highly coordinated protein unfolding and translocation by the ClpXP proteolytic machine. *Cell* **158**, 647-658 (2014).

175. N. Monroe, H. Han, P. S. Shen, W. I. Sundquist, C. P. Hill, Structural basis of protein translocation by the Vps4-Vta1 AAA ATPase. *Elife* **6** (2017).
176. S. N. Gates *et al.*, Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104. *Science* **357**, 273-279 (2017).
177. J. L. Morgan *et al.*, Observing cellulose biosynthesis and membrane translocation in crystallo. *Nature* **531**, 329-334 (2016).
178. J. L. W. Morgan, J. F. Acheson, J. Zimmer, Structure of a Type-1 Secretion System ABC Transporter. *Structure* **25**, 522-529 (2017).
179. I. B. Holland, L. Schmitt, J. Young, Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway (review). *Mol Membr Biol* **22**, 29-39 (2005).
180. H. Jeong *et al.*, Pseudoatomic Structure of the Tripartite Multidrug Efflux Pump AcrAB-TolC Reveals the Intermeshing Cogwheel-like Interaction between AcrA and TolC. *Structure* **24**, 272-276 (2016).
181. J. Y. Lee *et al.*, Crystal structure of the human sterol transporter ABCG5/ABCG8. *Nature* **533**, 561-564 (2016).
182. H. Qian *et al.*, Structure of the Human Lipid Exporter ABCA1. *Cell* **169**, 1228-1239 e1210 (2017).
183. B. Wang, M. Dukarevich, E. I. Sun, M. R. Yen, M. H. Saier, Jr., Membrane porters of ATP-binding cassette transport systems are polyphyletic. *J Membr Biol* **231**, 1-10 (2009).
184. W. Kabsch, Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010).
185. N. Collaborative Computational Project, The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-763 (1994).
186. P. A. Karplus, K. Diederichs, Linking crystallographic model and data quality. *Science* **336**, 1030-1033 (2012).
187. A. J. McCoy *et al.*, Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
188. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).
189. J. Painter, E. A. Merritt, Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* **62**, 439-450 (2006).
190. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
191. B. K. Ho, F. Gruswitz, HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct Biol* **8**, 49 (2008).
192. O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**, 455-461 (2010).
193. H. J. C. v. d. S. Berendsen, D.; van Drunen, R., GROMACS: A message-passing parallel molecular dynamics implementation. *Comp Phys Comm* **91**, 43-56 (1995).

194. R. B. Best *et al.*, Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J Chem Theory Comput* **8**, 3257-3273 (2012).
195. S. Jo, T. Kim, V. G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29**, 1859-1865 (2008).
196. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *J Chem Phys* **126**, 014101 (2007).
197. M. R. Parrinello, A., Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys* **52**, 7182-7190 (1981).
198. P. J. Moynihan *et al.*, The hydrolase LpqI primes mycobacterial peptidoglycan recycling. *Nat Commun* **10**, 2647 (2019).
199. P. J. Stansfeld *et al.*, MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes. *Structure* **23**, 1350-1361 (2015).
200. P. J. Stansfeld, M. S. Sansom, From Coarse Grained to Atomistic: A Serial Multiscale Approach to Membrane Protein Simulations. *J Chem Theory Comput* **7**, 1157-1166 (2011).
201. C. A. Caffalette, J. Kuklewicz, N. Spellmon, J. Zimmer, Biosynthesis and Export of Bacterial Glycolipids. *Annu Rev Biochem* **89**, 741-768 (2020).
202. O. Schneewind, D. Missiakas, Lipoteichoic acids, phosphate-containing polymers in the envelope of gram-positive bacteria. *J Bacteriol* **196**, 1133-1142 (2014).
203. M. Jankute, J. A. Cox, J. Harrison, G. S. Besra, Assembly of the Mycobacterial Cell Wall. *Annu Rev Microbiol* **69**, 405-423 (2015).
204. S. Shankar-Sinha *et al.*, The Klebsiella pneumoniae O antigen contributes to bacteremia and lethality during murine pneumonia. *Infect Immun* **72**, 1423-1430 (2004).
205. G. L. Murray, S. R. Attridge, R. Morona, Altering the length of the lipopolysaccharide O antigen has an impact on the interaction of Salmonella enterica serovar Typhimurium with macrophages and complement. *J Bacteriol* **188**, 2735-2739 (2006).
206. C. D. Clay, S. Soni, J. S. Gunn, L. S. Schlesinger, Evasion of complement-mediated lysis and complement C3 deposition are regulated by Francisella tularensis lipopolysaccharide O antigen. *J Immunol* **181**, 5568-5578 (2008).
207. E. M. Goebel, D. N. Wolfe, K. Elder, S. Stibitz, E. T. Harvill, O antigen protects Bordetella parapertussis from complement. *Infect Immun* **76**, 1774-1780 (2008).
208. N. Ho, A. N. Kondakova, Y. A. Knirel, C. Creuzenet, The biosynthesis and biological role of 6-deoxyheptose in the lipopolysaccharide O-antigen of Yersinia pseudotuberculosis. *Mol Microbiol* **68**, 424-447 (2008).
209. M. Caboni *et al.*, An O antigen capsule modulates bacterial pathogenesis in Shigella sonnei. *PLoS Pathog* **11**, e1004749 (2015).
210. I. Orskov, F. Orskov, B. Jann, K. Jann, Serology, chemistry, and genetics of O and K antigens of Escherichia coli. *Bacteriol Rev* **41**, 667-710 (1977).
211. J. D. King, S. Berry, B. R. Clarke, R. J. Morris, C. Whitfield, Lipopolysaccharide O antigen size distribution is determined by a chain extension complex of variable stoichiometry in Escherichia coli O9a. *Proc Natl Acad Sci U S A* **111**, 6407-6412 (2014).

212. E. Mann *et al.*, Substrate recognition by a carbohydrate-binding module in the prototypical ABC transporter for lipopolysaccharide O-antigen from Escherichia coli O9a. *J Biol Chem* **294**, 14978-14990 (2019).
213. S. Okuda, E. Freinkman, D. Kahne, Cytoplasmic ATP hydrolysis powers transport of lipopolysaccharide across the periplasm in E. coli. *Science* **338**, 1214-1217 (2012).
214. M. Braibant, P. Gilot, J. Content, The ATP binding cassette (ABC) transport systems of Mycobacterium tuberculosis. *FEMS Microbiol Rev* **24**, 449-467 (2000).
215. P. Dianiskova *et al.*, Investigation of ABC transporter from mycobacterial arabinogalactan biosynthetic cluster. *Gen Physiol Biophys* **30**, 239-250 (2011).
216. Y. Bi, J. Zimmer, Structure and Ligand-Binding Properties of the O Antigen ABC Transporter Carbohydrate-Binding Domain. *Structure* **28**, 252-258 e252 (2020).
217. C. Thomas, R. Tampé, Structural and Mechanistic Principles of ABC Transporters. *Annual Review of Biochemistry* **89**, 605-636 (2020).
218. N. M. I. Taylor *et al.*, Structure of the human multidrug transporter ABCG2. *Nature* **546**, 504-509 (2017).
219. I. Manolaridis *et al.*, Cryo-EM structures of a human ABCG2 mutant trapped in ATP-bound and substrate-bound states. *Nature* **563**, 426-430 (2018).
220. L. Chen *et al.*, Cryo-electron Microscopy Structure and Transport Mechanism of a Wall Teichoic Acid ABC Transporter. *mBio* **11**, e02749-02719 (2020).
221. S. Hofmann *et al.*, Conformation space of a heterodimeric ABC exporter under turnover conditions. *Nature* **571**, 580-583 (2019).
222. J. Zaitseva, S. Jenewein, T. Jumpertz, I. B. Holland, L. Schmitt, H662 is the linchpin of ATP hydrolysis in the nucleotide-binding domain of the ABC transporter HlyB. *EMBO J* **24**, 1901-1910 (2005).
223. M. L. Oldham, J. Chen, Snapshots of the maltose transporter during ATP hydrolysis. *Proc Natl Acad Sci U S A* **108**, 15152-15156 (2011).
224. D. E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* **32**, W526-531 (2004).
225. S. Ovchinnikov *et al.*, Improved de novo structure prediction in CASP11 by incorporating coevolution information into Rosetta. *Proteins* **84 Suppl 1**, 67-75 (2016).
226. B. W. Simpson *et al.*, Combining Mutations That Inhibit Two Distinct Steps of the ATP Hydrolysis Cycle Restores Wild-Type Function in the Lipopolysaccharide Transporter and Shows that ATP Binding Triggers Transport. *mBio* **10** (2019).
227. F. Hagn, M. L. Nasr, G. Wagner, Assembly of phospholipid nanodiscs of controlled size for structural studies of membrane proteins by NMR. *Nat Protoc* **13**, 79-98 (2018).
228. K. Kiianitsa, J. A. Solinger, W. D. Heyer, NADH-coupled microplate photometric assay for kinetic studies of ATP-hydrolyzing enzymes with low and high specific activities. *Anal Biochem* **321**, 266-271 (2003).
229. S. H. Scheres, Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol* **579**, 125-157 (2016).

230. J. Zivanov *et al.*, New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7** (2018).
231. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221 (2015).
232. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017).
233. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612 (2004).
234. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
235. D. Liebschner *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877 (2019).
236. H. Zheng *et al.*, Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. *Nat Protoc* **9**, 156-170 (2014).
237. H. Zheng *et al.*, CheckMyMetal: a macromolecular metal-binding validation tool. *Acta Crystallogr D Struct Biol* **73**, 223-233 (2017).
238. L. Zhang, A. al-Hendy, P. Toivanen, M. Skurnik, Genetic organization and sequence of the rfb gene cluster of Yersinia enterocolitica serotype O:3: similarities to the dTDP-L-rhamnose biosynthesis pathway of Salmonella and to the bacterial polysaccharide transport systems. *Mol Microbiol* **9**, 309-321 (1993).
239. J. Hoffman, B. Lindberg, R. R. Brubaker, Structural studies of the O-specific side-chains of the lipopolysaccharide from Yersinia enterocolitica Ye 128. *Carbohydr Res* **78**, 212-214 (1980).
240. D. Bronner, B. R. Clarke, C. Whitfield, Identification of an ATP-binding cassette transport system required for translocation of lipopolysaccharide O-antigen side-chains across the cytoplasmic membrane of Klebsiella pneumoniae serotype O1. *Mol Microbiol* **14**, 505-519 (1994).
241. P. R. Reeves *et al.*, Bacterial polysaccharide synthesis and gene nomenclature. *Trends Microbiol* **4**, 495-503 (1996).
242. E. Janas *et al.*, The ATP hydrolysis cycle of the nucleotide-binding domain of the mitochondrial ATP-binding cassette transporter Mdl1p. *J Biol Chem* **278**, 26862-26869 (2003).
243. M. Yang *et al.*, Single-molecule probing of the conformational homogeneity of the ABC transporter BtuCD. *Nat Chem Biol* **14**, 715-722 (2018).
244. B. Verhalen *et al.*, Energy transduction and alternating access of the mammalian ABC transporter P-glycoprotein. *Nature* **543**, 738-741 (2017).
245. P. R. Selvin, Fluorescence resonance energy transfer. *Methods Enzymol* **246**, 300-334 (1995).
246. N. Akyuz, R. B. Altman, S. C. Blanchard, O. Boudker, Transport dynamics in a glutamate transporter homologue. *Nature* **502**, 114-118 (2013).
247. R. Roy, S. Hohng, T. Ha, A practical guide to single-molecule FRET. *Nat Methods* **5**, 507-516 (2008).

248. J. W. Taraska, M. C. Puljung, N. B. Olivier, G. E. Flynn, W. N. Zagotta, Mapping the structure and conformational movements of proteins with transition metal ion FRET. *Nat Methods* **6**, 532-537 (2009).
249. T. K. Aman, S. E. Gordon, W. N. Zagotta, Regulation of CNGA1 Channel Gating by Interactions with the Membrane. *J Biol Chem* **291**, 9939-9947 (2016).
250. N. Akyuz *et al.*, Transport domain unlocking sets the uptake rate of an aspartate transporter. *Nature* **518**, 68-73 (2015).
251. D. Ciftci *et al.*, Single-molecule transport kinetics of a glutamate transporter homolog shows static disorder. *Sci Adv* **6**, eaaz1949 (2020).
252. L. Wang *et al.*, Characterization of the kinetic cycle of an ABC transporter by single-molecule and cryo-EM analyses. *Elife* **9** (2020).
253. S. Matsumoto *et al.*, Crystal structures of an archaeal oligosaccharyltransferase provide insights into the catalytic cycle of N-linked protein glycosylation. *Proc Natl Acad Sci U S A* **110**, 17868-17873 (2013).
254. Y. Qiao *et al.*, Lipid II overproduction allows direct assay of transpeptidase inhibition by beta-lactams. *Nat Chem Biol* **13**, 793-798 (2017).
255. V. P. Dandey *et al.*, Time-resolved cryo-EM using Spotiton. *Nat Methods* **17**, 897-900 (2020).
256. T. Nakane, D. Kimanius, E. Lindahl, S. H. Scheres, Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* **7** (2018).