Molecular Insights into Capsular Polysaccharide Secretion via ABC Transporters.

Jeremi Kazimierz Kuklewicz
Żory, Poland

M.S. Biotechnology, Uniwersytet Gdański, 2018
B.S. Biotechnology, Uniwersytet Wrocławski, 2015

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, 2023

Abstract

In all living things, complex carbohydrates play a vital role in numerous biological processes, from protein quality control to cell recognition, energy storage, cell wall formation and pathogenesis. Many of these processes are performed in topologically extracellular compartments or on the cell surface. From viruses to bacteria and higher vertebrates, a wide array of structurally diverse polysaccharides are exported through distinct secretion systems. These systems evolved to transport often long and charged sugar chains across hydrophobic membranes to their respective sites of action. During my doctoral work, I investigated two distinct and mechanistically unique polysaccharide secretion systems, hyaluronan synthase (HAS) and the bacterial capsular polysaccharides (CPS) ABC transporter KpsMT-KpsE. Both are involved in the export of hyaluronan (HA), an acidic heteropolysaccharide of alternating N-acetylglucosamine and glucuronic acid sugars. HA is ubiquitously found in the vertebrate extracellular matrix but also in some bacteria, playing crucial yet distinct roles in both. In higher organisms, HA modulates essential physiological processes in health and disease, including cell differentiation, tissue homeostasis, and angiogenesis, whereas in microbes, HA aids in capsule and biofilm formation, serves as nutrient storage and enhances virulence.

HAS is a remarkable small enzyme that combines bifunctional glycosyl-transferase activity with nascent polymer export. In Chapter 2, I present five cryo-electron microscopy structures of a viral HAS homolog at different states during substrate binding and initiation of polymer synthesis. Combined with biochemical analyses and molecular dynamics simulations, our data reveal how HAS selects its substrates, hydrolyzes the first substrate to prime the synthesis reaction, opens an HA-conducting transmembrane (TM) channel, ensures alternating substrate polymerization, and coordinates HA inside its TM pore. Our work proposes a detailed model for the formation of an acidic extracellular heteropolysaccharide and sheds unprecedented light on the multitasking enzyme responsible for the biosynthesis of an abundant and essential glycosaminoglycan in the human body.

In contrast, Gram-negative bacteria utilize a more complicated, multienzyme machinery to first synthesize and then export CPS. CPS are synthesized on a phospholipid anchor that facilitates its biosynthesis and secretion. In Chapter 3, I present my work on ABC transporter mediated CPS translocation, wherein I elucidate crucial steps of CPS recognition and secretion in Gram-negative bacteria. Through *in vivo* studies of HA encapsulation, I demonstrate that KpsMT exhibits broad substrate specificity, reveal that a KpsMT-KpsE complex is sufficient for CPS translocation across the inner bacterial membrane, and determine the distribution of CPS and its secretion system on the cell surface. Cryogenic electron microscopy analyses of a KpsMT-KpsE complex in six different states reveal a KpsE-encased ABC transporter, conformational rearrangements of KpsMT during ATP hydrolysis, and substrate recognition inside a lipid-exposed electropositive canyon. Our work provides the first atomistic model for CPS secretion.

`

Signature Page

`

Dedication Page

I dedicate this work to my partner Aleksandra for her support, and most importantly PATIENCE.

`

List of Abbreviations

β-ME, β-mercaptoethanol
ABC transporter, ATP-binding cassette transporter
ADP, adenosine diphosphate
AF2, AlphaFold2
ATP, adenosine triphosphate
AAA-ATPases, ATPases associated with diverse cellular activities-ATPases
ADP, atomic displacement parameters (Phenix)
APBS, Adaptive Poisson-Boltzmann Solver
C12E8, polyoxyethylene(8)-dodecyl ether
C15-PP, farnesyl-diphosphate
C5-PP, isopentenyl-diphosphate
C55-P, undecaprenyl-phosphate
CAZy, Carbohydrate Active EnZyme classification
CBD, carbohydrate-binding domain
CBM, carbohydrate-binding module
CDP, cytidine-diphosphate
CG, coarse-grained
CHAP, channel annotation package
CHARMM-GUI, CHARMM graphical user interface
CHARMM, Chemistry at HARvard Macromolecular Mechanics
CHS, cholesteryl hemisuccinate Tris salt
CHL, coupling helix-loop
CMP, cytidine monophosphate
CPA, common polysaccharide antigen
CPS, capsular polysaccharides
CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats
CryoEM, cryo electron microscopy
CTF, contrast transfer function
Cv, *Chlorella virus*
DAG, diacylglycerol
DCR, detector channel ratio
DDM, n-dodecyl-B-D-maltoside
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

`

ECM, extracellular matrix
EDTA, ethylenediaminetetraacetic acid
EH, external helix
ELISA, enzyme-linked immunosorbent assay
EM, electron microscopy
EMDB, Electron Microscopy Data Bank
ETL, *E. coli* Total Lipid
FB, filtration buffer
FFT, fast Fourier transform
FSC, Fourier shell correlation
GlcA, glucuronic acid
Gal, galactose
Galf, galactofuranose
GalNAc, N-acetyl-galactosamine
Galp, galactopyranose
GF, gel filtration
GH, gate helix
GIF, Gatan imaging filter
GlcNAc-P, N-acetylglucosamine-1-phosphate
GlcNAc, N-acetyl-glucosamine
GPU, graphics processing unit
GT, glycosyltransferase
H-loop, histidine loop
HA, hyaluronic acid
HAS, hyaluronic acid synthase
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
IMVs, inverted membrane vescicles
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
LMNG, lauryl maltose neopentyl glycol

`

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
MINFLUX, minimal photon FLUX
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
Nb, nanobody
NBD, nucleotide-binding domain
NCS, non-crystallographic symmetry
Ni-NTA, nickel-charged nitrilotriacetic acid
OD, optical density
OM, outer membrane;
OSA, O-specific antigen
OST, oligosaccharyltransferase;
Pa-Wzx, *Pseudomonas aeruginosa* Wzx
PBS, phosphate-buffered saline
PCP, polysaccharide co-polymerase
PCR, polymerase chain reaction
PDB, Protein Data Bank
PEG, polyethylene glycol
PEP, phosphoenolpyruvate
PG, phosphatidylglycerol
PG1, periplasmic gate helix 1
PG2, periplasmic gate helix 2
PGP, phosphatidylglycerol-phosphate
Pi, inorganic phosphate
PIPE, polymerase incomplete primer extension

`

pLDDT, predicted local distance difference test
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
RB, resuspension buffer
RMSD, root-mean-square deviation
RPM, rounds per minute
RSR, real-space refinement
RT, room temperature
S6, Superose 6 size exclusion chromatography resin
S200, Superdex 200 size exclusion chromatography resin
SB, solubilization buffer
SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC, size exclusion chromatography
SN2, substitution nucleophilic bi-molecular
SS, secretion system
TCEP, tris(2-carboxyethyl)phosphine
TB, terrific broth
TEV, Tobacco Etch Virus
TM, transmembrane
TMD, transmembrane domain
TMH, transmembrane helix
UDP, uridine diphosphate
Und-P, undecaprenyl-phosphate
Und-PP, undecaprenyl-diphosphate
Und, undecaprenol or undecaprenyl
UppS, undecaprenyl pyrophosphate synthase
VMD, visual molecular dynamics
WT, wild type
WTA, wall teichoic acid

`

List of Figures

`

`

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. I contributed to segments of the *Introduction* as well as the chapter on capsular polysaccharide biosynthesis. The review has been updated and reflects the current state of the literature.

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 farnesyldiphosphate (C15-PP) and the dephosphorylation of the undecaprenyl-diphosphate (Und-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].

`

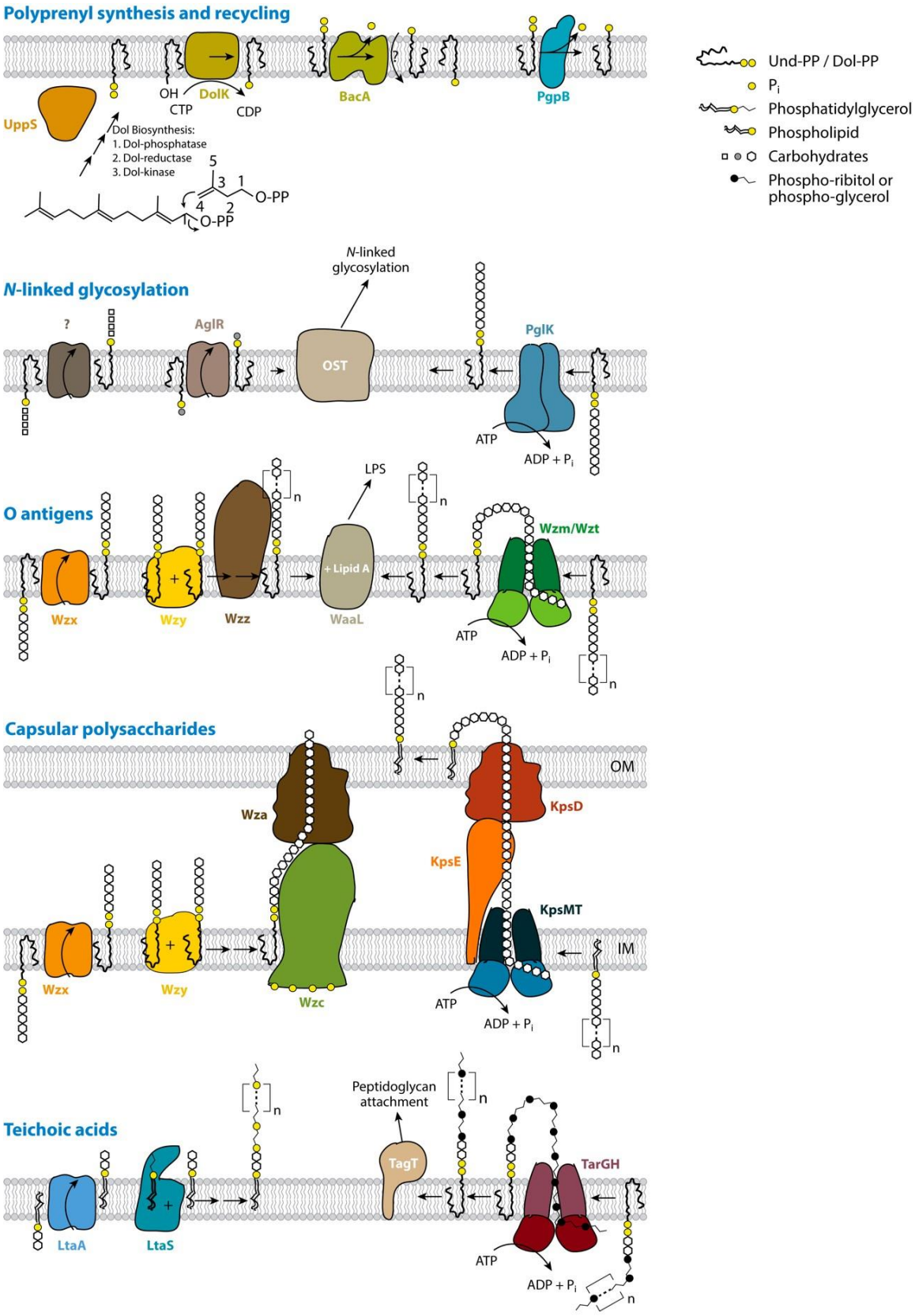

Polyprenyl synthesis and recycling
UppS
OH
CTP
DolK
CDP
BacA
PgpB
Dol Biosynthesis:
1. Dol-phosphatase
2. Dol-reductase
3. Dol-kinase
O-PP
O-PP
Und-PP / Dol-PP
Pi
Phosphatidylglycerol
Phospholipid
Carbohydrates
Phospho-ribitol or phospho-glycerol
N-linked glycosylation
N-linked glycosylation
?
AglR
OST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

Fig. 1-1

`

Figure 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.

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 homodimer 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-twofold 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 BacA's 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

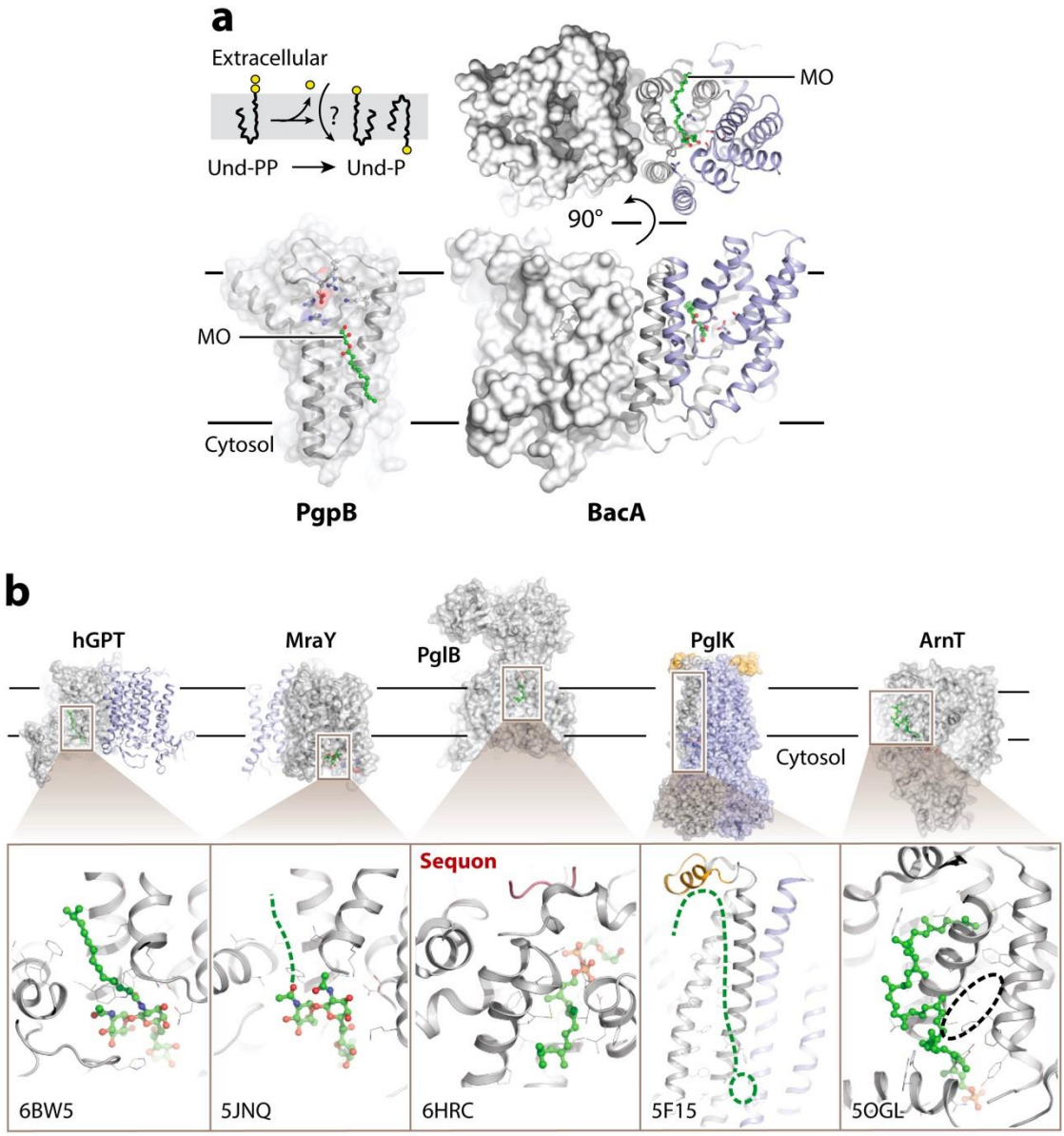


Figure 1-2 | Polyprenyl biosynthesis and recognition.
(a) Crystal structure of *Escherichia coli* BacA and *Bacillus subtilis* PgpB (PDB identifiers 6CB2 and 5JKI). MO is shown as green sticks. **(b)** Comparison of different polyprenyl interaction motifs. Lipids or lipid-like ligands are shown as ball-and-stick 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, undecaprenylphosphate; 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, C15PP, 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 polyprenyllinked 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 1phospho-GlcNAc to Dol-P, and MraY transfers the peptidoglycan precursor phosphoMurNAc-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 familyC 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 heptasaccharide 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 βphosphate [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 (Wzm/Wzt)–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 Wzm/Wzt 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-acetylglucosamine1-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-pentapeptideGlcNAc) flippase MurJ offer new insights into a potential Wzx translocation

`

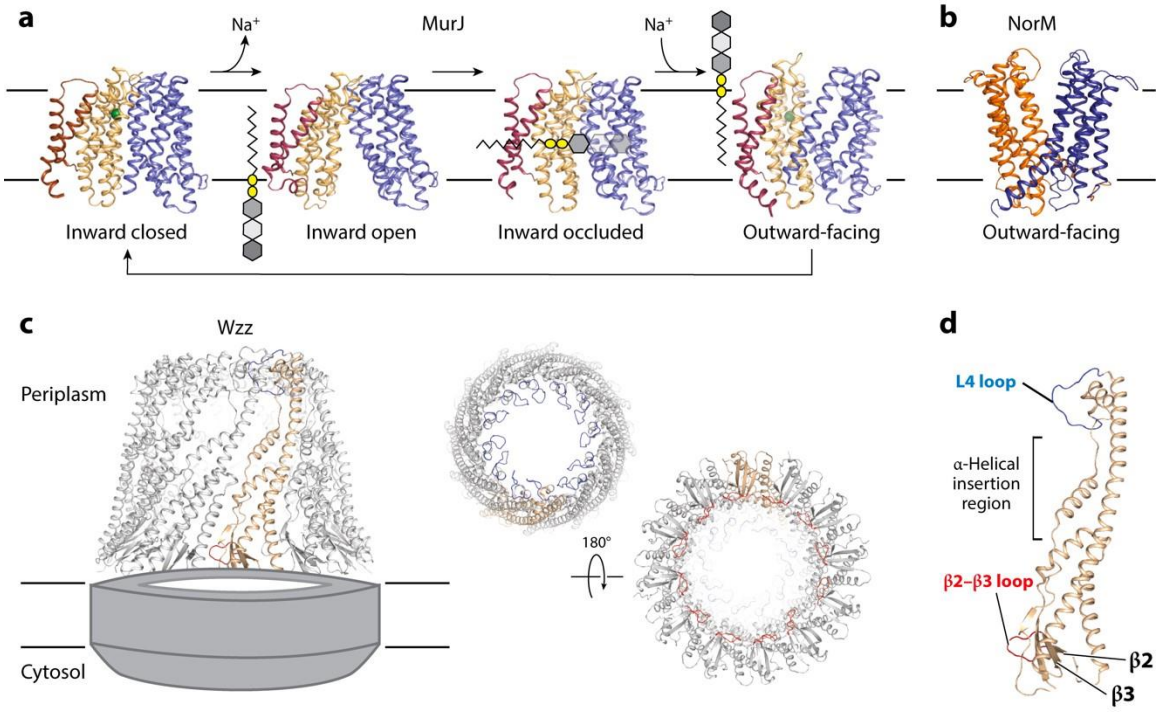


Figure 1-3 | Structural basis of translocation by Wzx flippases and the architecture of O antigen chain-length regulator Wzz. (a) Structures of the Wzx-like Lipid II flippase MurJ in each state of the putative translocation model: inward closed, inward open, inward occluded, and outward-facing (PDB identifiers 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.

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**) 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 UndPP–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 αhelical 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 Wzm/Wzt

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 Wzm/Wzt 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**) 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.

`

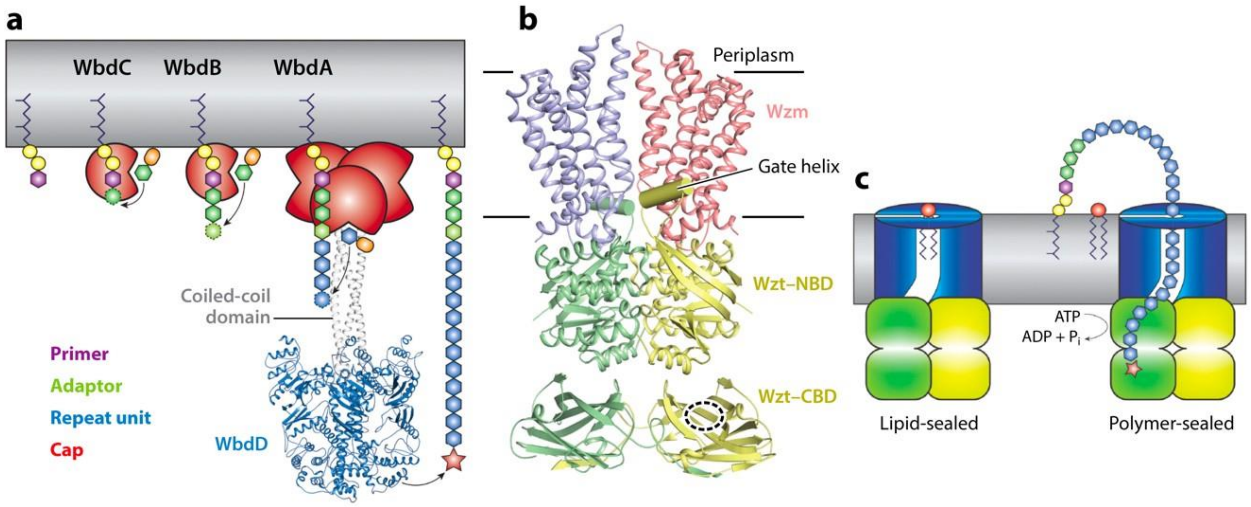


Figure 1-4 | ABC transporter–dependent O antigen biosynthesis.
(a) Biosynthesis and length control of Und-PP–linked O antigen. Glycosyltransferases are shown in red and labeled. The O antigen is represented with hexagons colored to distinguish the primer, adaptor, and repeat units and a red star indicating 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 Wzm–Wzt 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 Wzm/Wzt O antigen transporter. Abbreviations: ABC, ATP-binding cassette; CBD, carbohydrate-binding domain; NBD, nucleotide-binding domain; PDB, Protein Data Bank; Und-PP, undecaprenyl-diphosphate.

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 coexpressed [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 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* 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 membraneassociated 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 Wzm/Wzt 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 Cterminal β-strands. Each WztCBD subunit forms a β-sandwich with an immunoglobulinlike 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* Wzm–Wzt complex, nucleotidefree 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 UndPP–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 Wzm–Wzt 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 Wzm–Wzt 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 Wzm–Wzt 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 WaaL's function 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-β1-6glucosyl-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].

TagF's C-terminal GT-B domain binds the nucleotide moiety of a co-crystallized CDPglycerol substrate via a conserved DYSSVXXD/E motif. Interestingly, CDP-glycerol (**Fig. 1-5**) 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].

`

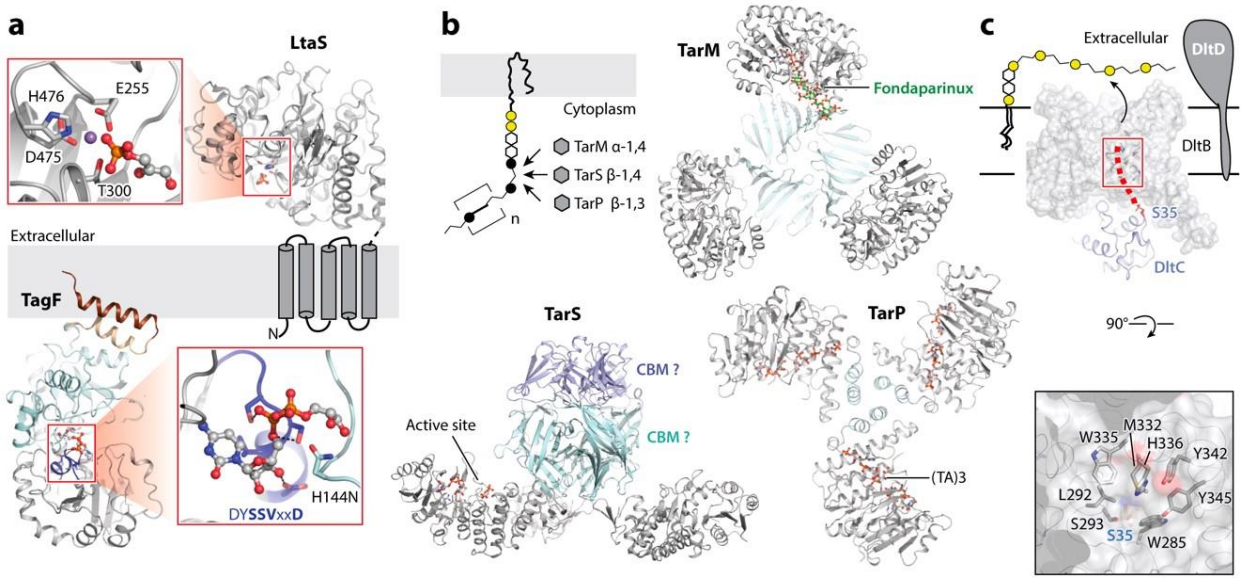


Figure 1-5 | Teichoic acid synthesis and modification.
(a) LtaS and TagF synthesize wall and lipoteichoic acids on opposing membrane sides (PDB identifiers 2W5S, 3L7K, respectively). The membrane-spanning segment of LtaS is indicated by gray cylinders. **(b)** TarM, TarS, and TarP glycosylate WTA with 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 ribitolphosphate 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.

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 Wzm/Wzt, 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 (Glc) 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's 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 ribitol's 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 Nterminal 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 polysaccharide copolymerase type 2a, Wzc, 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].

Recent crystallographic studies on a catalytically inactive mutant (K540M) of Wzc's 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].

Recently, Yang *et al.*[151] presented a near atomic cryoEM structure of a catalytically inactive mutant (K540M) of dephosphorylated Wzc from *E. coli* (**Fig. S3-22**). The study confirms that Wzc exists as an octamer, with each protomer composed of a periplasmic domain, a transmembrane domain comprising two alpha-helices, and a cytoplasmic domain with kinase activity. The assembled complex has approximate dimensions of ~146Å in width and ~165Å in height, forming a hollow enclosure in the membrane with eight oval-shaped portals bracketed by the transmembrane helices. It is suggested that this chamber may harbor Wzy, although no direct evidence supports this assumption. The overall fold of the periplasmic and transmembrane motifs can be superimposed with Wzz, showing a root mean square deviation (rmsd) of 2.1 Å over 130 Cα atoms.

`

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].

Accordingly, a model was proposed in which the Wzc complex regulates the polymerase activity of Wzy and opens Wza upon dephosphorylation by the Wzb phosphatase. Multi-phosphorylated Wzc is thought to be monomeric and not engaged in CPS translocation[151]. 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.

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 PG or 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**) 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. 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 KpsC's C-terminal domain attaches the second Kdo unit to the C7 position of the Kdo-lipid primer formed by KpsS. Subsequently, KpsC's 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.

KpsC's 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]. In recent crystallographic and biochemical studies, the acceptor binding mode of KpsC's

`

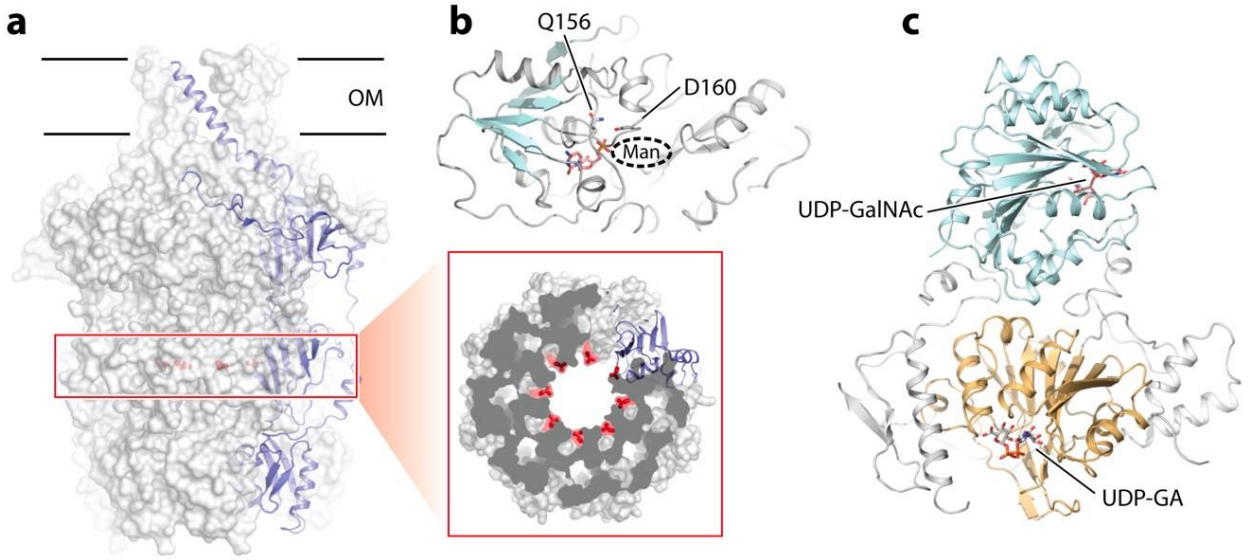


Figure 1-6 | Biosynthesis of capsular polysaccharides.
(a) Structure of the OM-integrated Wza (PDB identifier 2J58). One subunit of the octameric complex is shown as a blue diagram. The inset shows a slice through the periplasmic Wza barrel viewed from the periplasm, with residues contacting the translocating 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.

N-terminal domain was discovered and a double displacement reaction mechanism similar to WbbB was proposed [152, 153].

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].

Very recently Cifuente *et al.*[154] provided crystallographic insights into the multifunctional Bcs3 enzyme, the CPS polymerase from *Haemophilus influenzae* type b. Bcs3 comprises three catalytic functions, including ribose-5-phosphate transferase, ribose phosphatase, and ribitol-5-phosphate

`

transferase activities. In the form of a multimeric enzyme, Bcs3 is sufficient to synthesize the polyribosyl ribitol phosphate chain of *Haemophilus influenzae* type b.

Although multi-functional 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 Nacetylmannosamine (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].

KpsE belongs to the polysaccharide co-polymerase family 3 (PCP-3). The primary function of these proteins is likely to facilitate the translocation of the polysaccharide through the periplasm[37]. Previous studies predicted that KpsE either contains one[155] or two transmembrane[156] helixes, with the bulk of the protein in the periplasm. The oligomeric state of KpsE is uncertain, most likely dimeric or multimeric[157]. In a related study, CtrB, a homologous PCP-3 protein found in *Neisseria meningitidis,* has been shown to form conical oligomers approximately ~100Å in diameter and extending about ~125Å from the membrane when reconstituted in lipid bilayers [158]. If this is also the case *in vivo*, it would support the concept of a continuous periplasm spanning CPS secretion system.

`

KpsD is the OM subunit that likely partners with KpsE in forming the T1SS-like structure. Conducting structural and biochemical studies on KpsD has proven challenging due to its elusive nature. Significant fractions of KpsD have been detected not only in the OM but also in the cytoplasm and periplasmic space in *E. coli* K5 [157]. KpsD's correct membrane association seems to be dependent on the presence of KpsE, murein lipoprotein LPP and capsular polysaccharide [157, 158, 159].

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 Concluding remarks

Most, if not all, microbes decorate their cell surfaces with complex carbohydrates to form protective capsules, facilitate host tissue adhesion, evade host immune systems, or generate a 3-dimensional filtration system that aids in osmo-regulation and/or nutrient acquisition. Cell surface polysaccharides, such as teichoic acids, capsular polysaccharides and O antigens, are potent virulence factors, due to reducing the efficacies of the host's innate immune response or impeding antimicrobial treatments.

`

During the course of my graduate studies, I had the opportunity to assist in an exciting project focused on understanding how a viral hyaluronan synthase, Cv-HAS, synthesizes and translocates negatively charged HA across a hydrophobic membrane. Cv-HAS serves as a model system for HA biosynthesis in vertebrates where it forms an essential component of the extracellular matrix. By conducting collaborative structure-function experiments, along with biochemical and molecular dynamics analyses, we delineated the initial steps of HA biosynthesis. This research is described first in Chapter 2.

My main graduate work focused on studying the biosynthesis and transport of Gram-negative bacterial capsular polysaccharides using a multifaceted experimental approach. My research aims to shed light on the molecular and mechanistic basis for KpsMT-mediated CPS secretion. I correlate structural and functional analyses with *in vivo* encapsulation experiments to address these mechanistic questions. The mechanism of how the ABC transporter KpsMT recognizes a lipid-linked polysaccharide and then translocates it across the inner membrane is an open question. Additionally, the role of the auxiliary protein KpsE and how it complexes with KpsMT is enigmatic. In Chapter 3, I describe both aims of my thesis proposal. Aim 1 involves determining distinct conformations of the CPS ABC transporter KpsMT-KpsE using single particle cryo-electron microscopy. In Aim 2, I intend to engineer hyaluronan capsule formation in *E. coli* to serve as an *in vivo* secretion assay.

`

1.8 References

1. TouzÉ T, Mengin-Lecreulx D. 2008. Undecaprenyl phosphate synthesis. EcoSal Plus 3. https://doi.org/10.1128/ecosalplus.4.7.1.7
2. Manat G, Roure S, Auger R, Bouhss A, Barreteau H, et al. 2014. Deciphering the metabolism of undecaprenyl-phosphate: the bacterial cell-wall unit carrier at the membrane frontier. Microb. Drug Resist. 20:199–214
3. Eichler J, Guan Z. 2017. Lipid sugar carriers at the extremes: the phosphodolichols Archaea use in N-glycosylation. Biochim. Biophys. Acta Mol. Cell Biol. Lipids 1862:589–99
4. Surmacz L, Swiezewska E. 2011. Polyisoprenoids—Secondary metabolites or physiologically important superlipids? Biochem. Biophys. Res. Commun. 407:627–32
5. El Ghachi M, Howe N, Huang CY, Olieric V, Warshamanage R, et al. 2018. Crystal structure of undecaprenyl-pyrophosphate phosphatase and its role in peptidoglycan biosynthesis. Nat. Commun. 9:1078
6. Workman SD, Worrall LJ, Strynadka NCJ. 2018. Crystal structure of an intramembranal phosphatase central to bacterial cell-wall peptidoglycan biosynthesis and lipid recycling. Nat. Commun. 9:1159
7. Vergara-Jaque A, Fenollar-Ferrer C, Kaufmann D, Forrest LR. 2015. Repeat-swap homology modeling of secondary active transporters: updated protocol and prediction of elevator-type mechanisms. Front. Pharmacol. 6:183
8. El Ghachi M, Bouhss A, Blanot D, Mengin-Lecreulx D. 2004. The bacA gene of Escherichia coli encodes an undecaprenyl pyrophosphate phosphatase activity. J. Biol. Chem. 279:30106–13
9. El Ghachi M, Derbise A, Bouhss A, Mengin-Lecreulx D. 2005. Identification of multiple genes encoding membrane proteins with undecaprenyl pyrophosphate phosphatase (UppP) activity in Escherichia coli. J. Biol. Chem. 280:18689–95
10. Fernandez F, Rush JS, Toke DA, Han GS, Quinn JE, et al. 2001. The CWH8 gene encodes a dolichyl pyrophosphate phosphatase with a luminally oriented active site in the endoplasmic reticulum of Saccharomyces cerevisiae. J. Biol. Chem. 276:41455–64
11. Rush JS, Cho SK, Jiang S, Hofmann SL, Waechter CJ. 2002. Identification and characterization of a cDNA encoding a dolichyl pyrophosphate phosphatase located in the endoplasmic reticulum of mammalian cells. J. Biol. Chem. 277:45226–34
12. Ghachi ME, Howe N, Auger R, Lambion A, Guiseppi A, et al. 2017. Crystal structure and biochemical characterization of the transmembrane PAP2 type phosphatidylglycerol phosphate phosphatase from Bacillus subtilis. Cell. Mol. Life Sci. 74:2319–32
13. Fan J, Jiang D, Zhao Y, Liu J, Zhang XC. 2014. Crystal structure of lipid phosphatase Escherichia coli phosphatidylglycerophosphate phosphatase B. PNAS 111:7636–40
14. Sigal YJ, McDermott MI, Morris AJ. 2005. Integral membrane lipid phosphatases/phosphotransferases: common structure and diverse functions. Biochem. J. 387:281–93
15. Hartley MD, Imperiali B. 2012. At the membrane frontier: a prospectus on the remarkable evolutionary conservation of polyprenols and polyprenyl-phosphates. Arch. Biochem. Biophys. 517:83–97
16. ban Duijn G, Valtersson C, Chojnacki T, Verkleij AJ, Dallner G, de Kruijff B. 1986. Dolichyl phosphate induces non-bilayer structures, vesicle fusion and transbilayer movement of lipids: a model membrane study. Biochim. Biophys. Acta Biomembr. 861:211–23

`

17. Zhou GP, Troy FA II. 2005. 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 15:347–59
18. Wang X, Mansourian AR, Quinn PJ. 2008. The effect of dolichol on the structure and phase behaviour of phospholipid model membranes. Mol. Membr. Biol. 25:547–56
19. Yoo J, Mashalidis EH, Kuk ACY, Yamamoto K, Kaeser B, et al. 2018. GlcNAc-1-P-transferase-tunicamycin complex structure reveals basis for inhibition of N-glycosylation. Nat. Struct. Mol. Biol. 25:217–24
20. Hakulinen JK, Hering J, Branden G, Chen H, Snijder A, et al. 2017. MraY-antibiotic complex reveals details of tunicamycin mode of action. Nat. Chem. Biol. 13:265–67
21. Chung BC, Mashalidis EH, Tanino T, Kim M, Matsuda A, et al. 2016. Structural insights into inhibition of lipid I production in bacterial cell wall synthesis. Nature 533:557–60
22. Chung BC, Zhao J, Gillespie RA, Kwon DY, Guan Z, et al. 2013. Crystal structure of MraY, an essential membrane enzyme for bacterial cell wall synthesis. Science 341:1012–16
23. Kowarik M, Numao S, Feldman MF, Schulz BL, Callewaert N, et al. 2006. N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. Science 314:1148–50
24. Perez C, Gerber S, Boilevin J, Bucher M, Darbre T, et al. 2015. Structure and mechanism of an active lipid-linked oligosaccharide flippase. Nature 524:433–38
25. Perez C, Mehdipour AR, Hummer G, Locher KP. 2019. Structure of outward-facing PglK and molecular dynamics of lipid-linked oligosaccharide recognition and translocation. Structure 27:669–78.e5
26. Raetz CRH, Reynolds CM, Trent MS, Bishop RE. 2007. Lipid A modification systems in gram-negative bacteria. Annu. Rev. Biochem. 76:295–329
27. Needham BD, Trent MS. 2013. Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. Nat. Rev. Microbiol. 11:467–81
28. Petrou VI, Herrera CM, Schultz KM, Clarke OB, Vendome J, et al. 2016. Structures of aminoarabinose transferase ArnT suggest a molecular basis for lipid A glycosylation. Science 351:608–12
29. Napiorkowska M, Boilevin J, Sovdat T, Darbre T, Reymond JL, et al. 2017. Molecular basis of lipid-linked oligosaccharide recognition and processing by bacterial oligosaccharyltransferase. Nat. Struct. Mol. Biol. 24:1100–6
30. Lizak C, Gerber S, Numao S, Aebi M, Locher KP. 2011. X-ray structure of a bacterial oligosaccharyltransferase. Nature 474:350–55
31. Caffalette CA, Corey RA, Sansom MSP, Stansfeld PJ, Zimmer J. 2019. A lipid gating mechanism for the channel-forming O antigen ABC transporter. Nat. Comm. 10:824
32. Reeves P. 1995. Role of O-antigen variation in the immune response. Trends Microbiol. 3:381–86
33. Whitfield C, Trent MS. 2014. Biosynthesis and export of bacterial lipopolysaccharides. Annu. Rev. Biochem. 83:99–128
34. Liston SD, Mann E, Whitfield C. 2017. Glycolipid substrates for ABC transporters required for the assembly of bacterial cell-envelope and cell-surface glycoconjugates. Biochim. Biophys. Acta Mol. Cell Biol. Lipids 1862:1394–403
35. Raetz CRH, Whitfield C. 2002. Lipopolysaccharide endotoxins. Annu. Rev. Biochem. 71:635–700
36. Lam JS, Taylor VL, Islam ST, Hao Y, Kocincova D. 2011. Genetic and functional diversity of Pseudomonas aeruginosa lipopolysaccharide. Front. Microbiol. 2:118

`

37. Whitfield C. 2006. Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. Annu. Rev. Biochem. 75:39–68
38. Wang L, Liu D, Reeves PR. 1996. C-terminal half of Salmonella enterica WbaP (RfbP) is the galactosyl-1-phosphate transferase domain catalyzing the first step of O-antigen synthesis. J. Bacteriol. 178:2598–604
39. Lehrer J, Vigeant KA, Tatar LD, Valvano MA. 2007. 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. 189:2618–28
40. Rocchetta HL, Burrows LL, Pacan JC, Lam JS. 1998. 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. 28:1103–19
41. Rush JS, Rick PD, Waechter CJ. 1997. Polyisoprenyl phosphate specificity of UDP-GlcNAc:undecaprenyl phosphate N-acetylglucosaminyl 1-P transferase from E. coli. Glycobiology 7:315–22
42. Al-Dabbagh B, Mengin-Lecreulx D, Bouhss A. 2008. Purification and characterization of the bacterial UDP-GlcNAc:undecaprenyl-phosphate GlcNAc-1-phosphate transferase WecA. J. Bacteriol. 190:7141–46
43. Al-Dabbagh B, Henry X, El Ghachi M, Auger G, Blanot D, et al. 2008. 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 47:8919–28
44. Al-Dabbagh B, Olatunji S, Crouvoisier M, El Ghachi M, Blanot D, et al. 2016. Catalytic mechanism of MraY and WecA, two paralogues of the polyprenyl-phosphate N-acetylhexosamine 1-phosphate transferase superfamily. Biochimie 127:249–57
45. Liu Y, Rodrigues JP, Bonvin AM, Zaal EA, Berkers CR, et al. 2016. New insight into the catalytic mechanism of bacterial MraY from enzyme kinetics and docking studies. J. Biol. Chem. 291:15057–68
46. Hong Y, Liu MA, Reeves PR. 2018. Progress in our understanding of Wzx flippase for translocation of bacterial membrane lipid-linked oligosaccharide. J. Bacteriol. 200:e00154-17
47. Paulsen IT, Beness AM, Saier MH Jr. 1997. Computer-based analyses of the protein constituents of transport systems catalysing export of complex carbohydrates in bacteria. Microbiology 143(Part 8):2685–99
48. Feldman MF, Marolda CL, Monteiro MA, Perry MB, Parodi AJ, Valvano MA. 1999. 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. 274:35129–38
49. Marolda CL, Tatar LD, Alaimo C, Aebi M, Valvano MA. 2006. Interplay of the Wzx translocase and the corresponding polymerase and chain length regulator proteins in the translocation and periplasmic assembly of lipopolysaccharide O antigen. J. Bacteriol. 188:5124–35
50. Marolda CL, Vicarioli J, Valvano MA. 2004. Wzx proteins involved in biosynthesis of O antigen function in association with the first sugar of the O-specific lipopolysaccharide subunit. Microbiology 150:4095–105

`

51. Islam ST, Lam JS. 2014. Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway. Can. J. Microbiol. 60:697–716
52. Liu MA, Morris P, Reeves PR. 2019. Wzx flippases exhibiting complex O-unit preferences require a new model for Wzx-substrate interactions. Microbiology 8:e00655
53. Islam ST, Fieldhouse RJ, Anderson EM, Taylor VL, Keates RA, et al. 2012. A cationic lumen in the Wzx flippase mediates anionic O-antigen subunit translocation in Pseudomonas aeruginosa PAO1. Mol. Microbiol. 84:1165–76
54. Islam ST, Eckford PD, Jones ML, Nugent T, Bear CE, et al. 2013. Proton-dependent gating and proton uptake by Wzx support O-antigen-subunit antiport across the bacterial inner membrane. mBio 4:e00678–13
55. Kuk AC, Mashalidis EH, Lee SY. 2017. Crystal structure of the MOP flippase MurJ in an inward-facing conformation. Nat. Struct. Mol. Biol. 24:171–76
56. Kuk ACY, Hao A, Guan Z, Lee SY. 2019. Visualizing conformation transitions of the Lipid II flippase MurJ. Nat. Commun. 10:1736
57. Whitfield C. 1995. Biosynthesis of lipopolysaccharide O antigens. Trends Microbiol. 3:178–85
58. Kim TH, Sebastian S, Pinkham JT, Ross RA, Blalock LT, Kasper DL. 2010. Characterization of the O-antigen polymerase (Wzy) of Francisella tularensis. J. Biol. Chem. 285:27839–49
59. Islam ST, Taylor VL, Qi M, Lam JS. 2010. Membrane topology mapping of the O-antigen flippase (Wzx), polymerase (Wzy), and ligase (WaaL) from Pseudomonas aeruginosa PAO1 reveals novel domain architectures. mBio 1:e00189-10
60. Woodward R, Yi W, Li L, Zhao G, Eguchi H, et al. 2010. In vitro bacterial polysaccharide biosynthesis: defining the functions of Wzy and Wzz. Nat. Chem. Biol. 6:418–23
61. Cuthbertson L, Mainprize IL, Naismith JH, Whitfield C. 2009. 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. 73:155–77
62. Collins RF, Kargas V, Clarke BR, Siebert CA, Clare DK, et al. 2017. Full-length, oligomeric structure of Wzz Determined by cryoelectron microscopy reveals insights into membrane-bound states. Structure 25:806–15.e3
63. Tocilj A, Munger C, Proteau A, Morona R, Purins L, et al. 2008. Bacterial polysaccharide co-polymerases share a common framework for control of polymer length. Nat. Struct. Mol. Biol. 15:130–38
64. Larue K, Kimber MS, Ford R, Whitfield C. 2009. Biochemical and structural analysis of bacterial O-antigen chain length regulator proteins reveals a conserved quaternary structure. J. Biol. Chem. 284:7395–403
65. Kalynych S, Cherney M, Bostina M, Rouiller I, Cygler M. 2015. Quaternary structure of WzzB and WzzE polysaccharide copolymerases. Protein Sci. 24:58–69
66. Kalynych S, Yao D, Magee J, Cygler M. 2012. Structural characterization of closely related O-antigen lipopolysaccharide (LPS) chain length regulators. J. Biol. Chem. 287:15696–705
67. Papadopoulos M, Morona R. 2010. Mutagenesis and chemical cross-linking suggest that Wzz dimer stability and oligomerization affect lipopolysaccharide O-antigen modal chain length control. J. Bacteriol. 192:3385–93
68. Huszczynski SM, Coumoundouros C, Pham P, Lam JS, Khursigara CM. 2019. Unique regions of the polysaccharide copolymerase Wzz2 from Pseudomonas aeruginosa are essential for O-specific antigen chain length control. J. Bacteriol. 201:e00165-19

`

69. Clarke BR, Cuthbertson L, Whitfield C. 2004. 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. 279:35709–18
70. Kos V, Cuthbertson L, Whitfield C. 2009. The Klebsiella pneumoniae O2a antigen defines a second mechanism for O antigen ATP-binding cassette transporters. J. Biol. Chem. 284:2947–56
71. Cuthbertson L, Kimber MS, Whitfield C. 2007. Substrate binding by a bacterial ABC transporter involved in polysaccharide export. PNAS 104:19529–34
72. Kos V, Whitfield C. 2010. A membrane-located glycosyltransferase complex required for biosynthesis of the D-galactan I lipopolysaccharide O antigen in Klebsiella pneumoniae. J. Biol. Chem. 285:19668–87
73. Clarke BR, Greenfield LK, Bouwman C, Whitfield C. 2009. 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. 284:30662–72
74. Kido N, Torgov VI, Sugiyama T, Uchiya K, Sugihara H, et al. 1995. 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. 177:2178–87
75. Greenfield LK, Richards MR, Li J, Wakarchuk WW, Lowary TL, Whitfield C. 2012. 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. 287:35078–91
76. Clarke BR, Bronner D, Keenleyside WJ, Severn WB, Richards JC, Whitfield C. 1995. Role of Rfe and RfbF in the initiation of biosynthesis of D-galactan I, the lipopolysaccharide O antigen from Klebsiella pneumoniae serotype O1. J. Bacteriol. 177:5411–18
77. Guan S, Clarke AJ, Whitfield C. 2001. Functional analysis of the galactosyltransferases required for biosynthesis of D-galactan I, a component of the lipopolysaccharide O1 antigen of Klebsiella pneumoniae. J. Bacteriol. 183:3318–27
78. Kido N, Sugiyama T, Yokochi T, Kobayashi H, Okawa Y. 1998. Synthesis of Escherichia coli O9a polysaccharide requires the participation of two domains of WbdA, a mannosyltransferase encoded within the wb* gene cluster. Mol. Microbiol. 27:1213–21
79. Greenfield LK, Richards MR, Vinogradov E, Wakarchuk WW, Lowary TL, Whitfield C. 2012. Domain organization of the polymerizing mannosyltransferases involved in synthesis of the Escherichia coli O8 and O9a lipopolysaccharide O-antigens. J. Biol. Chem. 287:38135–49
80. Liston SD, Clarke BR, Greenfield LK, Richards MR, Lowary TL, Whitfield C. 2015. Domain interactions control complex formation and polymerase specificity in the biosynthesis of the Escherichia coli O9a antigen. J. Biol. Chem. 290:1075–85
81. Hagelueken G, Huang H, Clarke BR, Lebl T, Whitfield C, Naismith JH. 2012. Structure of WbdD: a bifunctional kinase and methyltransferase that regulates the chain length of the O antigen in Escherichia coli O9a. Mol. Microbiol. 86:730–42
82. Hagelueken G, Clarke BR, Huang H, Tuukkanen A, Danciu I, et al. 2015. A coiled-coil domain acts as a molecular ruler to regulate O-antigen chain length in lipopolysaccharide. Nat. Struct. Mol. Biol. 22:50–56
83. Greenfield LK, Whitfield C. 2012. Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. Carbohydr. Res. 356:12–24
84. Bi Y, Mann E, Whitfield C, Zimmer J. 2018. Architecture of a channel-forming O-antigen polysaccharide ABC transporter. Nature 553(7688):361–65

`

85. Mann E, Mallette E, Clarke BR, Kimber MS, Whitfield C. 2016. The Klebsiella pneumoniae O12 ATP-binding cassette (ABC) transporter recognizes the terminal residue of its O-antigen polysaccharide substrate. J. Biol. Chem. 291:9748–61
86. Hashimoto H. 2006. Recent structural studies of carbohydrate-binding modules. Cell. Mol. Life Sci. 63:2954–67
87. Brown S, Santa Maria JP Jr., Walker S. 2013. Wall teichoic acids of gram-positive bacteria. Annu. Rev. Microbiol. 67:313–36
88. Morgan JLW, Strumillo J, Zimmer J. 2012. Crystallographic snapshot of cellulose synthesis and membrane translocation. Nature 493:181–86
89. Whitfield C, Amor PA, Köplin R. 1997. 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. 23:629–38
90. Abeyrathne PD, Lam JS. 2007. WaaL of Pseudomonas aeruginosa utilizes ATP in in vitro ligation of O antigen onto lipid A-core. Mol. Microbiol. 65:1345–59
91. Ruan X, Monjarás Feria J, Hamad M, Valvano MA. 2018. Escherichia coli and Pseudomonas aeruginosa lipopolysaccharide O-antigen ligases share similar membrane topology and biochemical properties. Mol. Microbiol. 110:95–113
92. Pérez JM, McGarry MA, Marolda CL, Valvano MA. 2008. Functional analysis of the large periplasmic loop of the Escherichia coli K-12 WaaL O-antigen ligase. Mol. Microbiol. 70:1424–40
93. Schild S, Lamprecht A-K, Reidl J. 2005. Molecular and functional characterization of O antigen transfer in Vibrio cholerae. J. Biol. Chem. 280:25936–47
94. Lairson LL, Henrissat B, Davies GJ, Withers SG. 2008. Glycosyltransferases: structures, functions, and mechanisms. Annu. Rev. Biochem. 77:521–55
95. Reichmann NT, Grundling A. 2011. Location, synthesis and function of glycolipids and polyglycerolphosphate lipoteichoic acid in Gram-positive bacteria of the phylum Firmicutes. FEMS Microbiol. Lett. 319:97–105
96. Lovering AL, Lin LY-C, Sewell EW, Spreter T, Brown ED, Strynadka NCJ. 2010. Structure of the bacterial teichoic acid polymerase TagF provides insights into membrane association and catalysis. Nature 17:582–89
97. Schertzer JW, Brown ED. 2008. Use of CDP-glycerol as an alternate acceptor for the teichoic acid polymerase reveals that membrane association regulates polymer length. J. Bacteriol. 190:6940–47
98. Schirner K, Marles-Wright J, Lewis RJ, Errington J. 2009. Distinct and essential morphogenic functions for wall- and lipo-teichoic acids in Bacillus subtilis. EMBO J. 28:830–42
99. Lu D, Wormann ME, Zhang X, Schneewind O, Grundling A, Freemont PS. 2009. Structure-based mechanism of lipoteichoic acid synthesis by Staphylococcus aureus LtaS. PNAS 106:1584–89
100. Cabacungan E, Pieringer RA. 1981. Mode of elongation of the glycerol phosphate polymer of membrane lipoteichoic acid of Streptococcus faecium ATCC 9790. J. Bacteriol. 147:75–79
101. Schirner K, Stone LK, Walker S. 2011. ABC transporters required for export of wall teichoic acids do not discriminate between different main chain polymers. ACS Chem. Biol. 6:407–12
102. Gerlach D, Guo Y, De Castro C, Kim SH, Schlatterer K, et al. 2018. Methicillin-resistant Staphylococcus aureus alters cell wall glycosylation to evade immunity. Nature 563:705–9

`

103. Sobhanifar S, Worrall LJ, Gruninger RJ, Wasney GA, Blaukopf M, et al. 2015. Structure and mechanism of Staphylococcus aureus TarM, the wall teichoic acid α-glycosyltransferase. PNAS 112:E576–85
104. Koc C, Gerlach D, Beck S, Peschel A, Xia G, Stehle T. 2015. 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. 290:9874–85
105. Sobhanifar S, Worrall LJ, King DT, Wasney GA, Baumann L, et al. 2016. Structure and mechanism of Staphylococcus aureus TarS, the wall teichoic acid β-glycosyltransferase involved in methicillin resistance. PLOS Pathog. 12:e1006067
106. Reichmann NT, Cassona CP, Grundling A. 2013. Revised mechanism of D-alanine incorporation into cell wall polymers in Gram-positive bacteria. Microbiology 159:1868–77
107. Whitfield GB, Marmont LS, Howell PL. 2015. Enzymatic modifications of exopolysaccharides enhance bacterial persistence. Front. Microbiol. 6:471
108. Perego M, Glaser P, Minutello A, Strauch MA, Leopold K, Fischer W. 1995. Incorporation of D-alanine into lipoteichoic acid and wall teichoic acid in Bacillus subtilis. Identification of genes and regulation. J. Biol. Chem. 270:15598–606
109. Ma D, Wang Z, Merrikh CN, Lang KS, Lu P, et al. 2018. Crystal structure of a membrane-bound O-acyltransferase. Nature 562:286–90
110. Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, et al. 2002. N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. Science 298:1790–93
111. Breitling J, Aebi M. 2013. N-linked protein glycosylation in the endoplasmic reticulum. Cold Spring Harb. Perspect. Biol. 5:a013359
112. Jarrell KF, Ding Y, Meyer BH, Albers SV, Kaminski L, Eichler J. 2014. N-linked glycosylation in Archaea: a structural, functional, and genetic analysis. Microbiol. Mol. Biol. Rev. 78:304–41
113. Nothaft H, Szymanski CM. 2010. Protein glycosylation in bacteria: sweeter than ever. Nat. Rev. Microbiol. 8:765–78
114. Horwitz MA, Silverstein SC. 1980. Influence of the Escherichia coli capsule on complement fixation and on phagocytosis and killing by human phagocytes. J. Clin. Invest. 65:82–94
115. Cress BF, Englaender JA, He W, Kasper D, Linhardt RJ, Koffas MAG. 2014. Masquerading microbial pathogens: Capsular polysaccharides mimic host–tissue molecules. Microbiol. Rev. 38:660–97
116. Sachdeva S, Palur RV, Sudhakar KU, Rathinavelan T. 2017. E. coli group 1 capsular polysaccharide exportation nanomachinary as a plausible antivirulence target in the perspective of emerging antimicrobial resistance. Front. Microbiol. 8:70
117. Taylor CM, Roberts IS. 2005. Capsular polysaccharides and their role in virulence. Contrib. Microbiol. 12:55–66
118. Hyams C, Camberlein E, Cohen JM, Bax K, Brown JS. 2010. The Streptococcus pneumoniae capsule inhibits complement activity and neutrophil phagocytosis by multiple mechanisms. Infect. Immun. 78:704–15
119. Whitfield C. 2010. Glycan chain-length control. Nat. Chem.. Biol. 6:403–4
120. Wugeditsch T, Paiment A, Hocking J, Drummelsmith J, Forrester C, Whitfield C. 2001. Phosphorylation of Wzc, a tyrosine autokinase, is essential for assembly of group 1 capsular polysaccharides in Escherichia coli. J. Biol. Chem. 267:2361–71

`

121. Collins RF, Beis K, Dong C, Botting CH, McDonnell C, et al. 2007. The 3D structure of a periplasm-spanning platform required for assembly of group 1 capsular polysaccharides in Escherichia coli. PNAS 104:2390–95
122. Bechet E, Gruszczyk J, Terreux R, Gueguen-Chaignon V, Vigouroux A, et al. 2010. Identification of structural and molecular determinants of the tyrosine-kinase Wzc and implications in capsular polysaccharide export. Mol. Microbiol. 77:1315–25
123. Drummelsmith J. 2000. 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
124. Dong C, Beis K, Nesper J, Brunkan-LaMontagne AL, Clarke BR, et al. 2006. Wza the translocon for E. coli capsular polysaccharides defines a new class of membrane protein. Nature 444:226–29
125. Nickerson NN, Mainprize IL, Hampton L, Jones ML, Naismith JH, Whitfield C. 2014. Trapped translocation intermediates establish the route for export of capsular polysaccharides across Escherichia coli outer membranes. PNAS 111:8203–8
126. Jann K, Jann B. 1992. Capsules of Escherichia coli, expression and biological significance. Can. J. Microbiol. 38:705–10
127. Willis LM, Stupak J, Richards MR, Lowary TL, Li J, Whitfield C. 2013. Conserved glycolipid termini in capsular polysaccharides synthesized by ATP-binding cassette transporter-dependent pathways in Gram-negative pathogens. PNAS 110:7868–73
128. Willis LM, Whitfield C. 2013. KpsC and KpsS are retaining 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) transferases involved in synthesis of bacterial capsules. PNAS 110:20753–58
129. Kohlbrenner WE, Fesik SW. 1985. Determination of the anomeric specificity of the Escherichia coli CTP:CMP-3-deoxy-D-manno-octulosonate cytidyltransferase by 13C NMR spectroscopy. J. Biol. Chem. 260:14695–700
130. Satola SW, Schirmer PL, Farley MM. 2003. Complete sequence of the cap locus of Haemophilus influenzae serotype b and nonencapsulated b capsule-negative variants. Infect. Immun. 71:3639–44
131. Frosch M, Müller A. 1993. Phospholipid substitution of capsular polysaccharides and mechanisms of capsule formation in Neisseria meningitidis. Mol. Microbiol. 8:483–93
132. Jelakovic S, Jann K, Schulz GE. 1996. The three-dimensional structure of capsule-specific CMP: 2-keto-3-deoxy-manno-octonic acid synthetase from Escherichia coli. FEBS Lett. 391:157–61
133. Doyle L, Ovchinnikova OG, Myler K, Mallette E, Huang BS, et al. 2019. Biosynthesis of a conserved glycolipid anchor for Gram-negative bacterial capsules. Nat. Chem. Biol. 15:632–40
134. Liston SD, Ovchinnikova OG, Whitfield C. 2016. Unique lipid anchor attaches Vi antigen capsule to the surface of Salmonella enterica serovar Typhi. PNAS 113:6719–24
135. Hu X, Chen Z, Xiong K, Wang J, Rao X, Cong Y. 2017. Vi capsular polysaccharide: synthesis, virulence, and application. Crit. Rev. Microbiol. 43:440–52
136. DeAngelis PL, White CL. 2002. Identification and molecular cloning of a heparosan synthase from Pasteurella multocida Type D. J. Biol. Chem. 277:7209–13
137. Jing W, De Angelis PL. 2000. Dissection of the two transferase activities of the Pasteurella multocida hyaluronan synthase: Two active sites exist in one polypeptide. Glycobiology 10:883–89
138. Osawa T, Sugiura N, Shimada H, Hirooka R, Tsuji A, et al. 2009. Crystal structure of chondroitin polymerase from Escherichia coli K4. Biochem. Biophys. Res. Comm. 378:10–14

`

139. Wasteson Å. 1971. A method for the determination of the molecular weight and molecular-weight distribution of chondroitin sulphate. J. Chromatogr. A 59:87–97
140. Sugiura N, Tawada A, Sugimoto K, Watanabe H. 2002. Molecular cloning and characterization of chondroitin polymerase from Escherichia coli strain K4. J. Biol. Chem. 227:21567–75
141. Jing W, DeAngelis PL. 2003. Analysis of the two active sites of the hyaluronan synthase and the chondroitin synthase of Pasteurella multocida. Glycobiology 13:661–71
142. Mandawe J, Infanzon B, Eisele A, Zaun H, Kuballa J, et al. 2018. Directed evolution of hyaluronic acid synthase from Pasteurella multocida towards high-molecular-weight hyaluronic acid. ChemBioChem. 19:1414–23
143. Otto NJ, Green DE, Masuko S, Mayer A, Tanner ME, et al. 2012. Structure/function analysis of Pasteurella multocida heparosan synthases: toward defining enzyme specificity and engineering novel catalysts. J. Biol. Chem. 287:7203–12
144. Sugiura N, Baba Y, Kawaguchi Y, Iwatani T, Suzuki K, et al. 2010. 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. 285:1597–606
145. Liu J, Yang A, Liu J, Ding X, Liu L, Shi Z. 2014. KfoE encodes a fructosyltransferase involved in capsular polysaccharide biosynthesis in Escherichia coli K4. Biotechnol. Lett. 36:1469–77
146. Neuberger A, Du D, Luisi BF. 2018. Structure and mechanism of bacterial tripartite efflux pumps. Res. Microbiol. 169:401–13
147. Nsahlai CJ, Silver RP. 2003. Purification and characterization of KpsT, the ATP-binding component of the ABC-capsule exporter of Escherichia coli K1. FEMS Microbiol. Lett. 224:113–18
148. Pavelka MS, Hayes SF, Silver RP. 1994. 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. 269:20149–58
149. Lo RYC, McKerral LJ, Hills TL, Kostrzynska M. 2001. Analysis of the capsule biosynthetic locus of Mannheimia (Pasteurella) haemolytica A1 and proposal of a nomenclature system. Infect. Immun. 69:4458–64
150. Liston SD, McMahon SA, Le Bas A, Suits MDL, Naismith JH, Whitfield C. 2018. Periplasmic depolymerase provides insight into ABC transporter-dependent secretion of bacterial capsular polysaccharides. PNAS 115:E4870–79
151. Yang Y, Liu J, Clarke BR, Seidel L, Bolla JR, Ward PN, Zhang P, Robinson CV, Whitfield C, Naismith JH. 2021. The molecular basis of regulation of bacterial capsule assembly by Wzc. Nat Commun. Jul 16;12(1):4349.
152. Doyle L, Ovchinnikova OG, Huang BS, Forrester TJB, Lowary TL, Kimber MS, Whitfield C. 2023. Mechanism and linkage specificities of the dual retaining β-Kdo glycosyltransferase modules of KpsC from bacterial capsule biosynthesis. J Biol Chem. May;299(5):104609.
153. Forrester T.J.B., Ovchinnikova O.G., Li Z., Kitova E.N., Nothof J.T., Koizumi A., et al. 2022. The retaining β-Kdo glycosyltransferase WbbB uses a double-displacement mechanism with an intermediate adduct rearrangement step. *Nat. Commun.* 13:6277.
154. Cifuente, J.O., Schulze, J., Bethe, A. et al. 2023. A multi-enzyme machine polymerizes the Haemophilus influenzae type b capsule. Nat Chem Biol 19, 865–877

`

155. Phoenix DA, Brandenburg K, Harris F, Seydel U, Hammerton T, Roberts IS. 2001. An investigation into the membrane-interactive potential of the Escherichia coli KpsE C-terminus. Biochem Biophys Res Commun. Jul 27;285(4):976-80
156. Rosenow, C., Esumeh, F., Roberts, I. S. & Jann, K. 1995. Characterization and localization of the KpsE protein of Escherichia coli K5, which is involved in polysaccharide export. Journal of Bacteriology 177
157. Arrecubieta C, Hammarton TC, Barrett B, Chareonsudjai S, Hodson N, Rainey D, Roberts IS. 2001. The transport of group 2 capsular polysaccharides across the periplasmic space in Escherichia coli. Roles for the KpsE and KpsD proteins. J Biol Chem. Feb 9;276(6):4245-50
158. Larue K, Ford RC, Willis LM, Whitfield C. 2011. Functional and structural characterization of polysaccharide co-polymerase proteins required for polymer export in ATP-binding cassette transporter-dependent capsule biosynthesis pathways. J Biol Chem. May 13;286(19):16658-68.
158. Diao J, Bouwman C, Yan D, Kang J, Katakam AK, Liu P, Pantua H, Abbas AR, Nickerson NN, Austin C, Reichelt M, Sandoval W, Xu M, Whitfield C, Kapadia SB. 2017. Peptidoglycan Association of Murein Lipoprotein Is Required for KpsD-Dependent Group 2 Capsular Polysaccharide Expression and Serum Resistance in a Uropathogenic Escherichia coli Isolate. mBio. May 23;8(3):e00603-17.
159. Sande C, Bouwman C, Kell E, Nickerson NN, Kapadia SB, Whitfield C. 2019. Structural and Functional Variation in Outer Membrane Polysaccharide Export (OPX) Proteins from the Two Major Capsule Assembly Pathways Present in Escherichia coli. J Bacteriol. Jun 21;201(14):e00213-19.

`

2. Chapter 2: Structure, substrate-recognition, and initiation of hyaluronan synthase

Finn P. Maloney[1]†, Jeremi Kuklewicz[1]†, Robin A. Corey[2], Yunchen Bi[1,3], Ruoya Ho[1], Lukasz Mateusiak[4], Els Pardon[5,6], Jan Steyaert[5,6], Phillip J. Stansfeld[7], Jochen Zimmer[1]

† These authors contributed equally

[1] Department of Molecular Physiology and Biological Physics, University of Virginia School of Medicine; 480 Ray C. Hunt Dr., Charlottesville, VA 22903, USA

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

[3] Present address: CAS and Shandong Province Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Center for Ocean Mega-Science, Chinese Academy of Sciences; Qingdao, China, and Laboratory for Marine Biology and Biotechnology, Pilot National Laboratory for Marine Science and Technology (Qingdao); Qingdao, China

[4] Laboratory for In Vivo Cellular and Molecular Imaging, ICMI-BEFY, Vrije Universiteit Brussel; Laarbeeklaan 103, 1090 Brussels, Belgium

[5] VIB-VUB Center for Structural Biology, VIB; Pleinlaan 2, 1050 Brussels, Belgium

[6] Structural Biology Brussels, Vrije Universiteit Brussel; VUB, Pleinlaan 2, 1050 Brussels, Belgium

[7] School of Life Sciences & Department of Chemistry, University of Warwick; Gibbet Hill Campus, CV4 7AL, UK

Nature 604, 195-201, doi: 10.1038/s41586-022-04534-2 (2022)

`

In this chapter, I will present a recently published article presenting five novel structures of a chlorella virus hyaluronan synthase, Cv-HAS, at different states during substrate binding and initiation of polymer synthesis. Combined with biochemical analyses and molecular dynamics simulations, we delineate that HA biosynthesis starts with the hydrolysis of a UDP-GlcNAc substrate to liberate a GlcNAc monosaccharide primer. This primer, upon diffusion to the enzyme's acceptor positions, can be elongated in subsequent glycosyl transfer reaction. Further, we identify conformational changes in the enzyme's TM segment upon priming that open an electropositive TM channel for HA secretion. I joined this project in January of 2019 and the article was published in April 2022. My main contribution is the determination of two substrate-bound and two primed Cv-HAS structures.

`

2.1 Introduction

Hyaluronan (HA) is a ubiquitous acidic glycosaminoglycan of the vertebrate extracellular matrix (ECM), particularly enriched in connective tissues, the vasculature, and cartilage[1]. HA modulates a broad range of tissue remodeling processes, including wound healing, embryological development, angiogenesis, and tumorigenesis[5]. It is a linear polysaccharide of alternating β-1,3 and β-1,4-linked N-acetylglucosamine (GlcNAc) and glucuronic acid (GlcA) units, respectively, (**Fig. 2-1a and b**) that is typically megadaltons in size[2,6]. High molecular weight HA (>10^6 kDa) associates with healthy tissue homeostasis, while low molecular weight species exhibit pro-inflammatory and angiogenic properties[2].

HA is synthesized from UDP-activated monosaccharides. Type-I HASs contain a single catalytic domain and secrete the nascent HA polymer during synthesis through a TM channel formed by their membrane-embedded segment[7,3,4]. Type-II HAS, however, is limited to select bacteria and is a bifunctional cytosolic glycosyltransferase (GT)[8,9]. Thus, Type-I HASs couple HA synthesis with translocation, utilizing a single catalytic domain to transfer two different donor sugars and to form substrate-specific glycosidic linkages.

Vertebrates express three HAS isoforms (**Fig. S2-1**) that differ in tissue expression and catalytic activity; HAS-2 is essential[10,11]. HAS contains a cytosolic GT domain flanked by two N-terminal and four C-terminal TM helices (TMH). How HAS selects its substrates, catalyzes regio- and stereospecific glycosyl transfer, secretes HA, and controls polymer length are important unresolved questions.

Chlorella viruses (Cv) contain HAS enzymes homologous to vertebrate HAS that synthesize HA *in vivo* and *in vitro*[12,13]. Cv infect unicellular green algal endosymbionts of the ciliate *Paramecium*

`

bursaria; HA production by the infected algae may benefit the endosymbiotic relationship. Cv-HAS shares with human HAS-2 ~45% sequence similarity, the same predicted number and distribution of TMHs, as well as a conserved GT domain (**Fig. S2-1**).

We determined cryo-electron microscopy (EM) structures of Cv-HAS in apo, UDP-bound, UDP-GlcNAc-bound, and GlcNAc-primed states. Our analyses provide the structural and mechanistic basis for substrate selectivity and alternating polymerization, demonstrate that only GlcNAc can prime HA biosynthesis, and suggest a model for processive HA synthesis and translocation. These insights are corroborated by functional studies and molecular dynamics (MD) simulations.

2.2 Results

Cv-HAS was expressed in *E. coli* and purified as described in the Methods. For EM analysis, Cv-HAS was bound to two camelid nanobodies produced *in vivo*, Nb872 and Nb881 (**Fig. 2-1c and d**)[14]. Nb872, binding to the extracellular TMH5-6 loop, supports *in vitro* catalytic activity similar to un-complexed Cv-HAS and thermo-stabilizes the enzyme (**Fig. S2-2a and b**). Nb881 binds the periphery of the GT domain and reduces HA biosynthesis for unknown reasons (**Fig. S2-2 b-c**).

Cryo-EM analyses of Cv-HAS were performed in *E. coli* lipid nanodiscs, see Methods. Samples were analyzed bound to UDP, a product and competitive inhibitor of Cv-HAS[15], and the substrate UDP-GlcNAc. Three-dimensional sorting and variability analyses of the 'UDP-GlcNAc-bound' dataset revealed additional states: nucleotide-free 'apo', substrate-bound, and monosaccharide-bound 'primed' (**Fig. S2-2- 4**). The cryo-EM maps generated in cryoSPARC[16] range in estimated

`

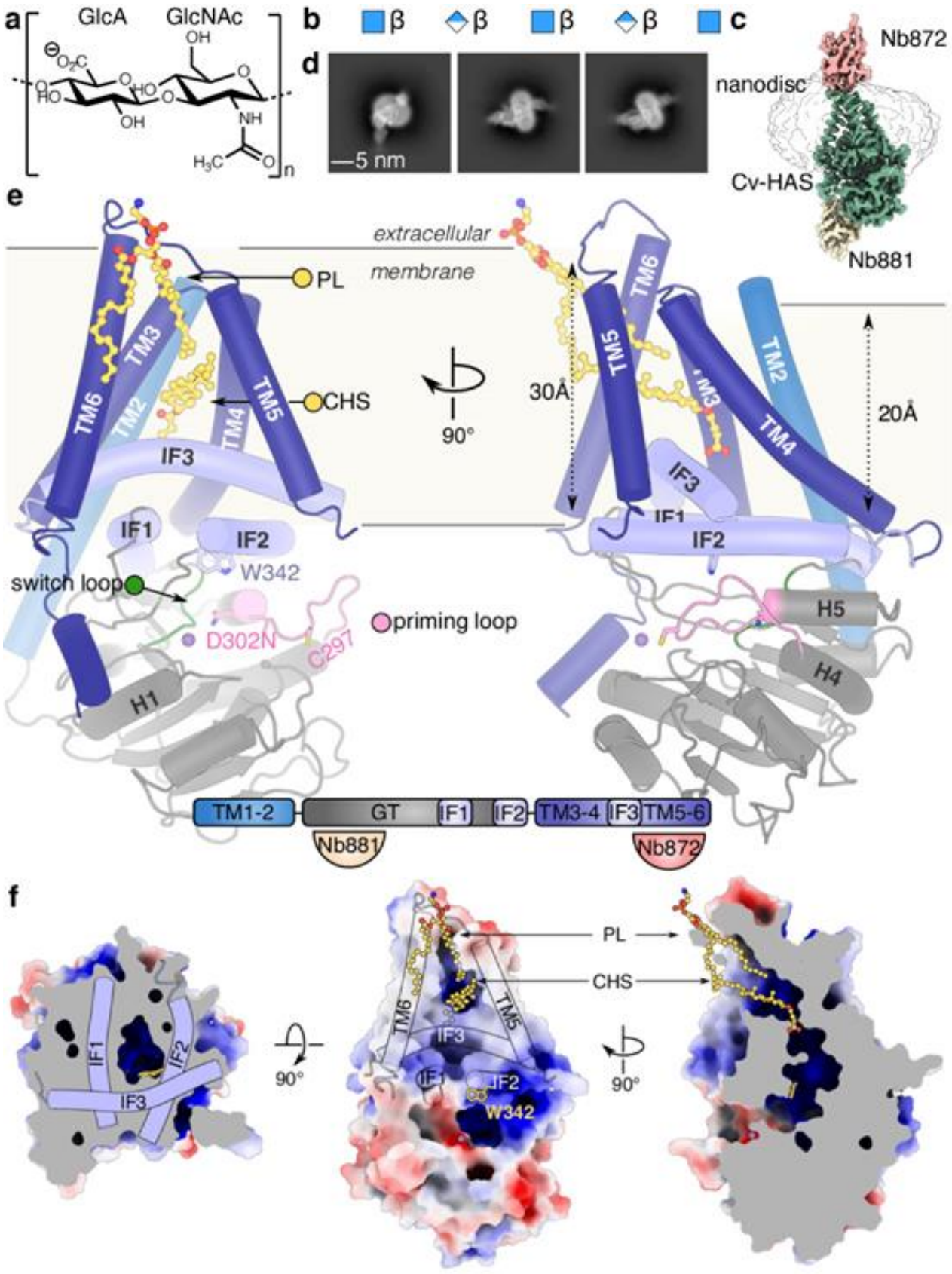

a
GlcA
GlcNAc
b
β
β
β
β
c
d
—5 nm
Nb872
nanodisc
Cv-HAS
Nb881
e
extracellular
membrane
PL
CHS
TM6
TM3
TM2
TM4
TM5
IF3
IF1
IF2
W342
switch loop
D302N
C297
priming loop
H1
H5
H4
90°
30Å
20Å
TM1-2
GT
IF1
IF2
TM3-4
IF3
TM5-6
Nb881
Nb872
f
PL
CHS
IF1
IF2
IF3
W342
90°

Fig. 2-1

`

Figure 2-1 | Structure of hyaluronan synthase. (**a** and **b**) HA GlcA-GlcNAc disaccharide repeat unit and representation of an HA polymer (blue square: GlcNAc, blue/white diamond: GlcA). (**c**) Volume of the Cv-HAS-Nb872/881 complex with Nb872 and 881 in salmon and yellow, respectively, Cv-HAS in green, with the nanodisc at lower contour represented as an outline. (**d**) Representative 2D class averages of nanobody-bound Cv-HAS. (**e**) Apo conformation of Cv-HAS. The glycosyltransferase (GT), interface (IF), and TM regions are colored green, light blue, and blue, respectively. Manganese ions are shown as purple spheres. Bottom panel: Schematic of Cv-HAS' domain organization and Nb interactions. (**f**) Surface electrostatics of Cv-HAS. Partially ordered lipid and detergent molecules occupying the membrane exposed channel's opening are shown as yellow ball-and-sticks. The electrostatic potential was calculated using the APBS plugin in PyMol (red – blue: -5 to +5 kT)[30].

resolutions from 3.1-2.7 Å (**Table 1**) and resolve essentially all residues, with the exceptions of residues 1–37 (including TMH1), 452–464 connecting IF3 with TMH5, and 553-561 at the C-terminus.

2.2.1 Architecture of Cv-HAS

The Cv-HAS GT domain adopts a GT-A fold[17] and packs against the five resolved TMHs (TMH2-6) as well as three amphipathic interface helices (IF1-3) (**Fig. 2-1e**). It contacts the TMHs via IF1 and IF2 (residues 238-263 and 331-357, respectively), of which IF2 contains a QxxRW motif characteristic of membrane-embedded processive family-2 GTs, according to the CAZy classification[18] (residues 339-342). IF3, a C-terminal extension of TMH4, runs perpendicular and on top of the IF1/2 pair, framing the entrance to the TM channel.

The TM architecture resembles a teepee with helices straddling the cytosolic IF helices. TMH2 extends past the cytosolic water/lipid interface to interact with the GT domain. Past TMH6, 13 C-terminal residues (540-552) form a short α-helix that packs into a groove on the GT domain formed by its central β-sheet and helix 1, together with IF1 (**Fig. 2-1e**).

`

TMH1 is disordered in all cryo-EM maps. A MapPred co-evolution analysis[19] suggests evolutionarily-coupled residues within TMH1 and TMH2 that position the helix in a groove between TMH2 and 4 (**Fig. S2-5a**). Direct interactions with TMH2 are also supported by AlphaFold2[20] and RoseTTAfold[21] Cv-HAS models, although they position TMH1 differently around TMH2 (**Fig. S2-5b and c**).

TMH1 is also disordered when using two cytosolic Nbs for structure determination (**Fig. S2-5d**), indicating that its flexibility is not induced by Nb872. The helix is not essential for function as TMH1-truncated Cv-HAS produces HA *in vitro,* albeit at reduced levels compared to the wild type enzyme (**Fig. S2-5e**).

Helices 4 and 5 of the GT domain are connected by an 11-residue long 'priming loop' that forms one wall of the active site (residue 291–301). The loop precedes the putative base catalyst Asp302 within the GDD motif that facilitates glycosyl transfer. It contains a conserved cysteine at its tip (Cys297) and extends roughly along the GT's central β-sheet towards the nucleotide-binding pocket (**Fig. 2-1e**). As demonstrated initially by the structure of the soluble enzyme SpsA and later cellulose synthase, the catalytic pocket of family-2 GTs includes characteristic acidic motifs that form a single substrate binding pocket[17,22,23]. For Cv-HAS, these include the nucleotide binding 'DGD' (121-123), the cation binding 'DSD' (201-203), and the putative catalytic 'GDD' motif (300-302) (**Fig. 2-1e, Fig. S2-1**).

2.2.2 Cv-HAS forms a lipid-filled channel

Cv-HAS contains a short membrane-embedded region about 20 and 30 Å thick near TMH2 and above the active site (proximal to the TMH5-6 face), respectively (**Fig. 2-1e and f**). It forms a

`

curved channel with a cytosolic entrance above the active site formed by Trp342 of the QxxRW motif within IF2 (**Fig. 2-1f**). Halfway across the membrane, IF3, TMH5, and TMH6 create a lateral channel opening towards the lipid bilayer. Two lipid molecules, assigned as a CHS molecule and a phospholipid tail based on their shapes, occupy and seal the portal (**Fig. 2-1f and Fig. S2-6**). Atomistic MD simulations in a POPE bilayer corroborate that these lipid acyl chains indeed prevent water flux across the membrane (**Fig. S2-6b**).

2.2.3 Substrate-bound conformation of Cv-HAS

Cv-HAS structures bound to UDP-GlcNAc were determined using an inactive enzyme in which Asp302 was replaced with Asn (**Fig. S2-1 and S2-7**). Unexpectedly, the substrate was partially hydrolyzed during sample preparation, capturing views of the active site before and after priming (**Fig. 2-2 a-c**). Nucleotide binding displaces the priming loop from the active site towards the cytosolic water-lipid interface (**Fig. 2-2d, Fig. S2-7d and f**). This conformation is stabilized by the loop's Phe292, interacting with Arg348 and Tyr352 of IF2, as well as Tyr299 occupying a hydrophobic pocket at the interface of IF2, IF3 and TMH5. Cys297 fits into a crevice at the water/lipid interface formed by Thr298 and Lys344 as well as Arg348 of IF2. Cys297 is critical for function; replacing it with Ala renders Cv-HAS inactive (**Fig. 2-2e**). Relative to the apo state, nucleotide binding induces a ~10 degree rigid body rotation of the GT domain towards the membrane, fostering the priming loop-IF2 interaction and narrowing the active site cleft (**Fig. 2-2a**).

`

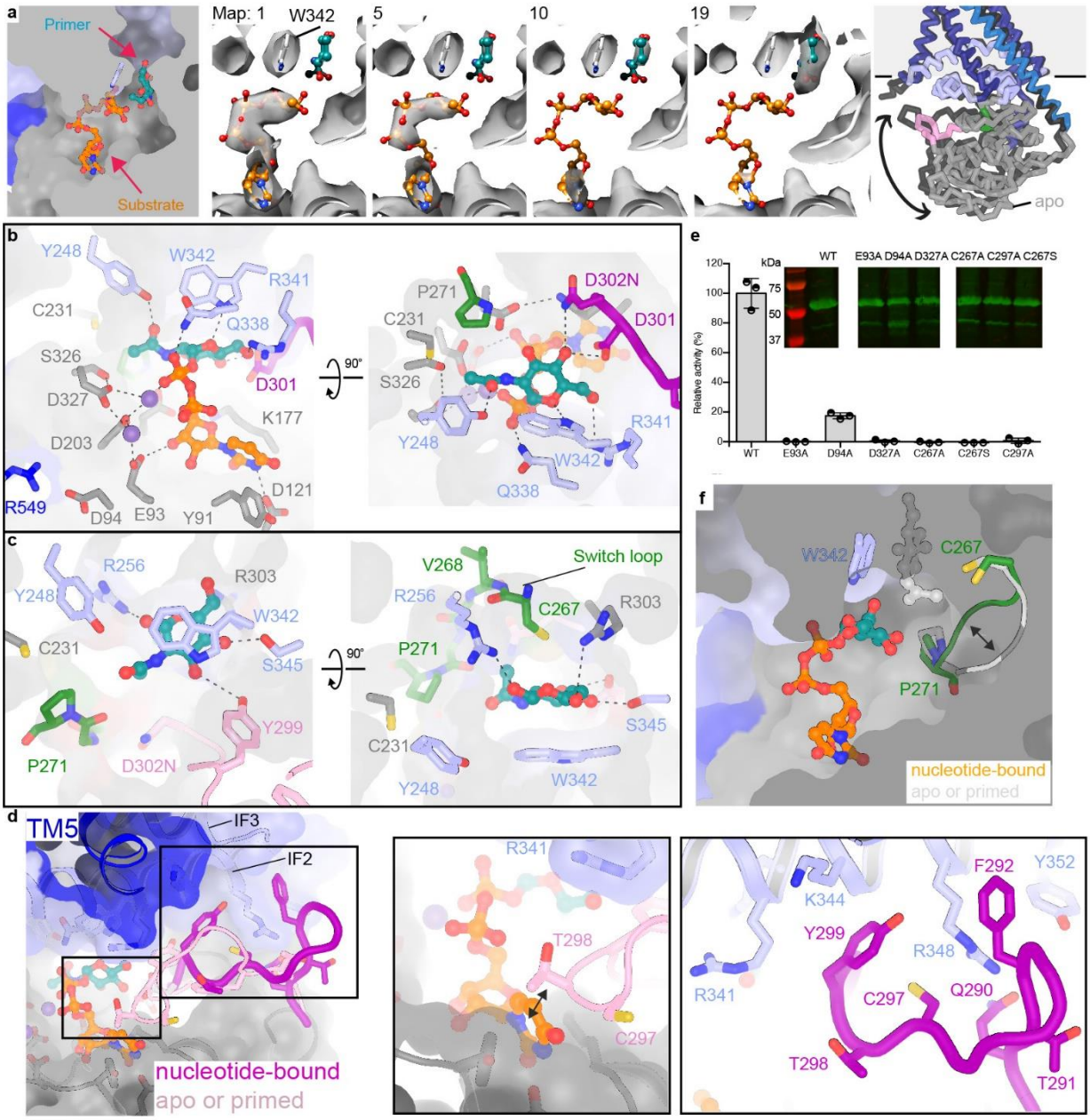


Fig. 2-2

Cv-HAS requires manganese for catalytic activity[13], indicating that it binds Mn^{2+}-complexed substrates. The UDP and UDP-GlcNAc-bound structures reveal similar UDP coordination (**Fig. S2-7a and b**). The UDP and UDP-GlcNAc-bound structures reveal similar UDP coordinations (**Fig. S2-8 a-c**). For UDP alone, the position of the diphosphate subtly changes compared to UDP-

Figure 2-2 | Substrate-bound and primed Cv-HAS conformations. (**a**) 3D variability analysis of UDP-GlcNAc-incubated Cv-HAS particles. Shown are representative volumes of the generated ensemble (Supplementary Video 1). Left panel: Substrate and acceptor locations mapped onto one structure; Right panel: Rotation of the GT domain towards the TM region upon substrate binding. (**b** and **c**) Substrate coordination at the active site. Dashed lines indicate distances from ~2.5 to ~4.2Å. (**d**) Priming loop retraction in response to substrate binding. The loop is shown as a cartoon with the flanking residues shown as sticks and semi-transparent surfaces. (**e**) Contribution of conserved residues to catalytic activity normal to wild type (WT) levels and obtained by quantifying the production of ^{3}H-labeled HA. Inset: Western blot of the inverted membrane vesicles used for activity measurements. Error bars represent deviations from the means with n=3 independent experiments. (**f**) Conformational changes of the switch loop. The primer and substrate are shown as ball-and-sticks in the same structure.

GlcNAc. Here, the β-phopsphate shifts towards Arg341 and Trp342, which changes the coordination of the associated Mn^{2+} ion. This could promote UDP release following glycosyl transfer. Asp121 of the 'DGD' motif contacts the uracil's N3 ring nitrogen, while the ribose's C3 hydroxyl group interacts with Glu93 from the conserved 'EDP' motif (**Fig. 2-2b, Fig. S2-1**). The UDP-GlcNAc diphosphate hydrogen-bonds with Gln338 of the QxxRW motif while the associated Mn^{2+} ion is coordinated by Asp203 and Asp327. A second Mn^{2+} ion is coordinated by Glu93 as well as Asp327 and Asp203 (**Fig. 2-2b**). Replacing Glu93 or Asp327 with alanine renders Cv-HAS inactive, thus both Mn^{2+} ions are functionally important (**Fig. 2-2e**). The additional Mn^{2+} ion may position the DxD motif (Asp201-Asp203) for substrate binding. Mutating Asp94 of the EDP motif to Ala reduces catalytic activity to about 20% relative to wild type. Asp94 does not directly coordinate Mn^{2+}, but interacts with Arg549 near Cv-HAS' C-terminus (**Fig. 2-2b**).

`

In the substrate-bound state, the donor sugar ring oxygen is proximal to Trp342's N1 hydrogen (**Fig. 2-2b**), right below the acceptor-binding site. The GlcNAc acetamido group occupies a pocket formed by Asp203, Cys231, Tyr248, Pro271, Asp327, and Ser326, and its carbonyl oxygen hydrogen bonds with the side chain hydroxyl of Tyr248 (**Fig. 2-2b**). The sugar's C4 hydroxyl interacts with Asp301 and Asp302, while its C6 hydroxyl points towards Arg341, positioned within a positively charged pocket containing Lys177. This pocket likely accommodates the C6 carboxylate when the substrate is UDP-GlcA (see below).

Compared to the nucleotide-free or primed conformations (see below), UDP or substrate binding repositions a partially conserved CVGGP loop (switch loop, residues 267-271) at the back of the nucleotide-binding pocket (**Fig. 2-2f**). The loop moves toward the membrane interface upon nucleotide binding, likely stabilized by unresolved water molecules. Cys267 at the beginning of the switch loop is necessary for function (**Fig. 2-2e**). In the inserted state, the priming loop's tip overlaps with the binding site for the uridine moiety of the substrate. Accordingly, the loop must reposition during substrate binding. Interestingly, the conserved cysteine residue at its tip is necessary for function. Because cysteines can be reversibly modified with, for example, hydrophobic acyl chains, it is conceivable that such modification displaces the loop from the active site to support catalytic activity. S-acylation is increasingly detected as a post-translational protein modification.

2.2.4 Priming of Cv-HAS

Substrate hydrolysis likely generated the GlcNAc monosaccharide primer that diffused to the acceptor-binding site at the TM channel entrance (**Fig. 2-2a, Fig. S2-7e**). The primer's acetamido group is surrounded by Pro271, Arg256, Cys231, Tyr248 as well as backbone atoms from the

`

switch loop (**Fig. 2-2c**). The switch loop's 'down' conformation positions the Cys267 sulfhydryl group ~ 4.2 Å away from the ring oxygen (**Fig. 2-2f**), while CH-π stacking interactions with Trp342 stabilize the sugar ring. Further, the conserved Ser345 (**Fig. S2-1**) is in hydrogen bonding distance to the primer's C5 hydroxyl, while the conserved Arg303 binds the C6 hydroxyl group. These interactions position the primer's C3 hydroxyl towards the active site and the putative base catalyst (D302N) (**Fig. 2-2c**).

Upon priming, the priming loop re-inserts into the catalytic pocket (**Fig. 2-2d, Fig. S2-7f**) and the GT domain relaxes away from the membrane (**Fig. 2-2a**). While an inserted loop position is also observed in the apo state, the loop is better resolved in the primed state, suggesting that the primer (or HA polymer) is stabilizing. Accordingly, the priming loop's Tyr299, a conserved Phe in vertebrate HAS, rotates towards the primer-binding site in close proximity to GlcNAc's C3 hydroxyl group (**Fig. 2-2c**).

2.2.5 Substrate hydrolysis initiates HA biosynthesis

The *in situ* generation of the GlcNAc primer suggests substrate hydrolysis as a general initiation mechanism of HAS-related enzymes. Accordingly, supplied monosaccharides should also prime HA biosynthesis. To test this model, wild type Cv-HAS was incubated with ^{14}C-labeled GlcNAc or GlcA and unlabeled substrates. The synthesized polysaccharide revealed that the GlcNAc monosaccharide is readily incorporated into HA, while GlcA fails to prime the synthesis reaction (**Fig. S2-8**).

Additionally, we used substrate hydrolysis experiments to verify the initiation mechanism. Cv-HAS hydrolyzes its substrates when exposed to only one of them without forming a homo-

`

polysaccharide (**Fig. S2-8a, lane 1**). This allows monitoring substrate hydrolysis rates in real time based on enzyme-coupled NADH oxidation (**Fig. S2-9**)[7]. We observe that GlcNAc monosaccharide increases the UDP-GlcA hydrolysis rate ~30-fold and decreases UDP-GlcNAc hydrolysis ~2.5-fold. Conversely, hydrolysis rates of both substrates were insignificantly affected by comparable concentrations of GlcA (**Fig. S2-9e and f**). Thus, a GlcNAc primer facilitates the binding and/or turnover of UDP-GlcA to form an HA disaccharide, while a GlcA primer does not enhance UDP-GlcNAc hydrolysis.

2.2.6 HAS priming creates a continuous TM channel

Strikingly, we resolve two different TM architectures of the primed state – one similar to the apo and substrate-bound states, and another with the N-terminus of TMH2 tilted away from TMH3 (**Fig. 2-3 and Fig. S2-7h and i**), creating a continuous TM channel. The newly formed channel overlaps with the curved, lipid-plugged channel above the active site to form a common vestibule, yet diverges halfway across the membrane (**Fig. 2-3c**). The newly created channel is lined with hydrophilic and apolar residues and its dimensions are sufficient to accommodate an HA polymer about 5 disaccharides long (**Fig. 2-3d**).

TMH2 kinks around its contact point with IF1, mediated by Gln57 and Phe60 of TMH2 and Val254, Ala258 and Leu261 of IF1 (**Fig. S2-7g**). The conserved residues Arg256 and Gln259 on the opposite side of IF1 line the channel lumen and Arg256 contacts the primer's C1 hydroxyl group (**Fig. 2-2c**). TMH2 could be further stabilized in this 'open' conformation by the predicted location of TMH1 (**Fig. S2-5 a-c**). Thus, HAS priming creates an exit path for the nascent HA at the interface of TMH1-2 and 3-4, similar to cellulose synthase (**Fig. S2-6c**)[23].

`

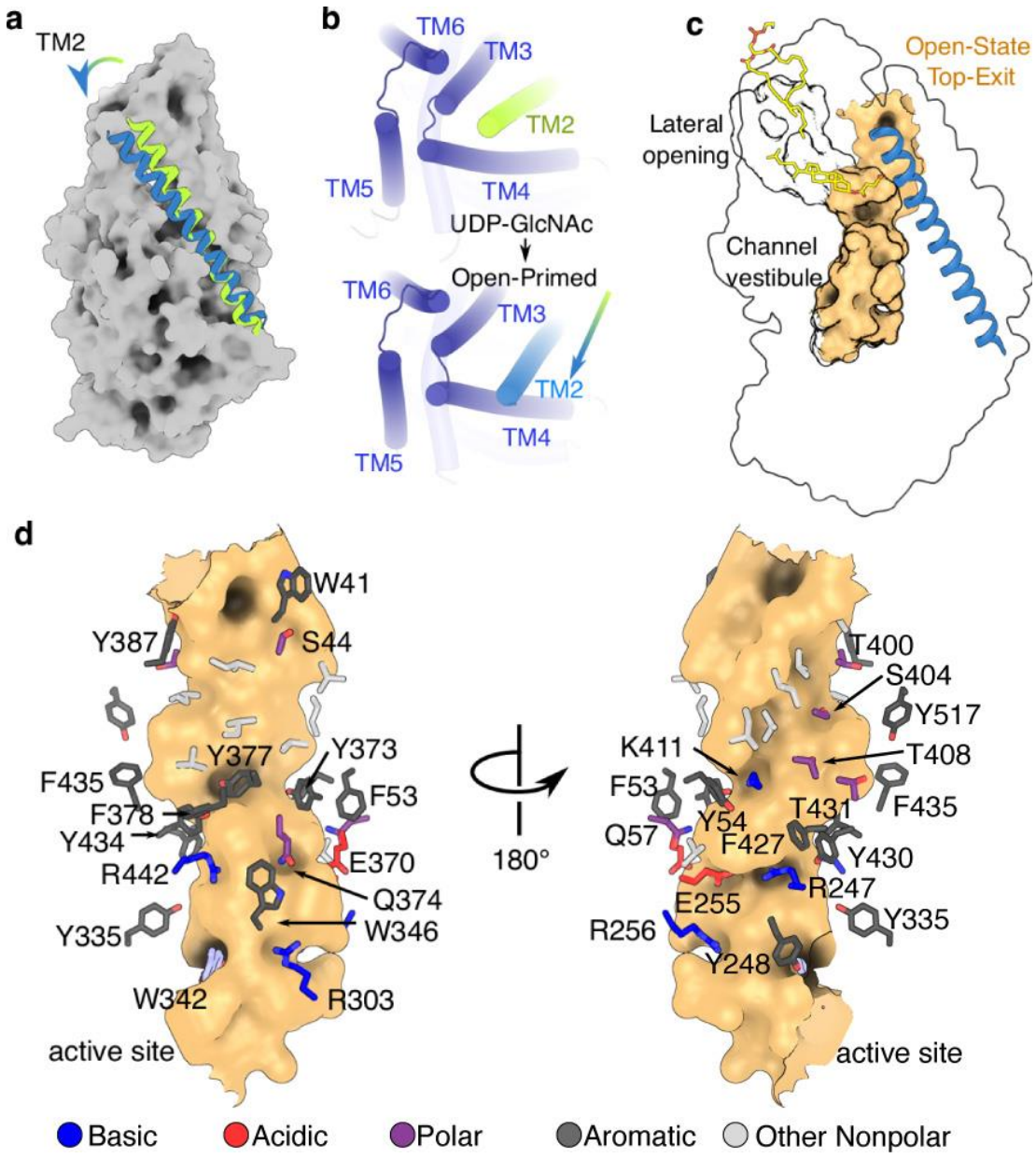


Figure 2-3 | Structural rearrangements upon GlcNAc priming. (**a**) Movement of TMH2 (green= substrate-bound, blue= primed-open) upon substrate hydrolysis. (**b**) Movement of TMH2. TMH1 is shown at its predicted location as a green cylinder. (**c**) Partial overlap of the lipid-plugged channel with the continuous TM channel formed upon priming (shown as an orange surface). The predicted location of TMH1 is indicated as a green cartoon helix. The channels were rendered in HOLLOW using a 1.2 Å probe[31]. (**d**) Residues lining the open TM-channel.

`

2.2.7 Dynamic coordination of HA inside the TM channel

We performed atomistic MD simulations to gain insights into HA coordination by Cv-HAS. Simulations were performed with a HA hexasaccharide manually placed within the channel vestibule, with the terminal non-reducing end sugar occupying the acceptor-binding site next to Trp342 (**Fig. 2-2c**).

We identify significant register-dependent differences in the positional stability of HA. While oligosaccharides starting with GlcNAc at the acceptor site were stable over the course of 5 simulations, polymers starting with GlcA were pulled into the TM channel by roughly one sugar unit (**Fig. 2-4a**). After this motion, the GlcA-ending polymer is coordinated similarly compared to the GlcNAc-ending chain, yet one sugar unit farther inside the channel (**Fig. 2-4 a-e, Table 2**).

HA primarily interacts with conserved polar and charged side chains. Arg303 and Arg256 bind GlcA carboxylates in the terminal or the second position (**Fig. 2-4 b-e**), while Arg247 and Arg442 coordinate GlcA at the third or fourth positions. The GlcNAc groups are mainly coordinated by polar and aromatic residues, including Tyr248, which coordinates GlcNAc, but not GlcA, when in the terminal position. Trp346 interacts with GlcNAc when at the second or third position. All of these HA-coordinating interactions are also seen in simulations seeded from different initial HA poses albeit with some localized variability (**Table 2**).

MD simulations of UDP-GlcA and HA-bound Cv-HAS also provide insights into the likely coordination of the UDP-GlcA substrate. While the UDP moiety is coordinated similarly to the experimental UDP-GlcNAc-bound state (**Fig. 2-2b**), the substrate's carboxylate is coordinated by Lys177 as well as Arg341 of the QxxRW motif, which bridges the nucleotide's α-phosphate and GlcA's carboxylate (**Fig. 2-4f and Table 2**).

`

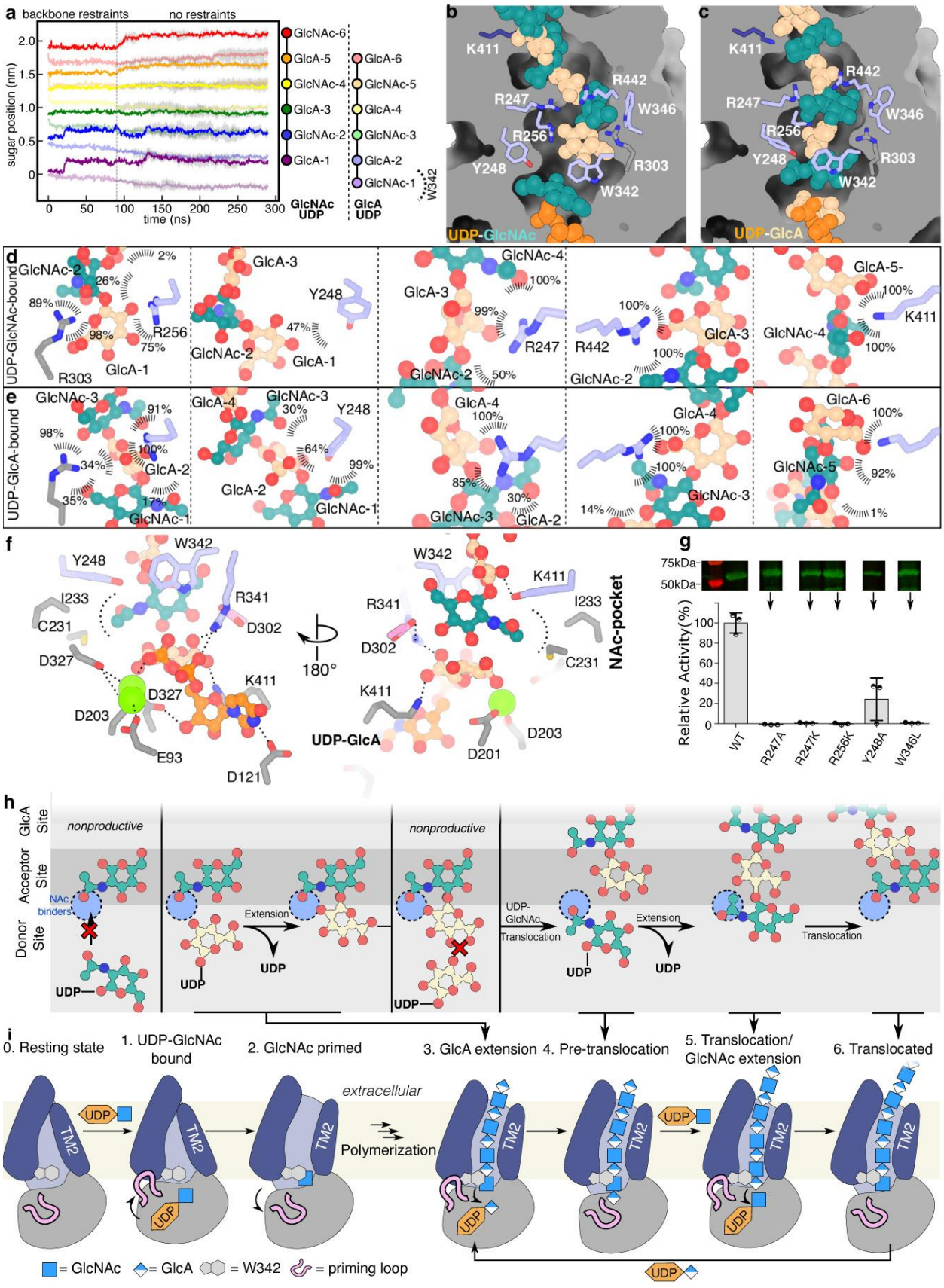

a backbone restraints
no restraints
sugar position (nm)
time (ns)
GlcNAc-6
GlcA-5
GlcNAc-4
GlcA-3
GlcNAc-2
GlcA-1
GlcA-6
GlcNAc-5
GlcA-4
GlcNAc-3
GlcA-2
GlcNAc-1
W342
GlcNAc UDP
GlcA UDP
b
K411
R442
R247
W346
R256
Y248
R303
W342
UDP-GlcNAc
c
K411
R442
R247
W346
R256
Y248
R303
W342
UDP-GlcA
d
UDP-GlcNAc-bound
GlcNAc-2
GlcA-3
GlcNAc-4
GlcA-5-
K411
R256
R303
GlcA-1
Y248
R247
R442
GlcA-3
GlcNAc-4
e
UDP-GlcA-bound
GlcNAc-3
GlcA-4
GlcNAc-3
Y248
GlcA-4
GlcA-4
GlcA-6
GlcA-2
GlcNAc-1
GlcA-2
GlcNAc-3
GlcNAc-3
GlcNAc-5
f
W342
Y248
I233
R341
C231
D302
D327
K411
D327
D203
E93
D121
180°
W342
K411
R341
I233
D302
C231
NAc-pocket
K411
D203
UDP-GlcA
D201
g
75kDa
50kDa
Relative Activity(%)
WT
R247A
R247K
R256K
Y248A
W346L
h
GlcA Site
Acceptor Site
Donor Site
nonproductive
NAc binders
UDP
Extension
UDP
nonproductive
UDP
UDP-GlcNAc Translocation
UDP
Extension
UDP
Translocation
i
0. Resting state
1. UDP-GlcNAc bound
2. GlcNAc primed
3. GlcA extension
4. Pre-translocation
5. Translocation/ GlcNAc extension
6. Translocated
extracellular
TM2
UDP
Polymerization
= GlcNAc
= GlcA
= W342
= priming loop

Fig. 2-4

Figure 2-4 | Dynamics of HA-bound HAS. (**a**) MD simulations of Cv-HAS bound to either UDP-GlcNAc and a GlcA-ending HA hexasaccharide or UDP-GlcA and a GlcNAc-ending polymer. (**b-c**) Representative coordination of HA at different registers after 300 ns MD simulations. (**d**) Coordination of HA inside the TM channel in a UDP-GlcNAc bound state. Values represent contact probabilities in percent over 300 ns of simulation, averaged over 5 independent repeats. The N-acetyl binding pocket is highlighted in blue. (**e**) As in panel (d) but for an UDP-GlcA bound state and HA polymer with different register. (**f**) Coordination of UDP-GlcA at the active site. Shown is a representative pose from a 300 ns simulation. The contact analysis is shown in Table 2. (**g**) Mutagenesis of HA coordinating residues and corresponding catalytic activity. Inset: Anti-His-tag Western blot of IMVs used for activity assays. Full scans are provided in Supplementary Fig. 1b. Error bars represent deviations from the means with n=3 independent experiments. (**h** and **i**) Model of HA elongation and membrane translocation. **(h)** HA's terminal sugar unit determines the selectivity of the catalytic pocket. GlcNAc or GlcA acceptors only allow binding UDP-GlcA or UDP-GlcNAc substrate, respectively. **(i)** UDP-GlcNAc hydrolysis and outward movement of TMH2 to prime HAS. Right: Processive primer extension and translocation leads to HA secretion. HA translocation is likely coupled to substrate binding. The priming loop (shown in pink) retracts from the active site to enable substrate binding, while the GT contracts and relaxes upon substrate binding and HA-extension, respectively.

HA biosynthesis experiments with Cv-HAS carrying substitutions at HA coordinating residues reveal their functional importance. Replacing Arg247 with Ala or Lys and Arg256 with Lys inactivates the enzyme, supporting a critical role for GlcA coordination. Substituting Trp346 with Leu similarly abolishes activity, while an Ala substitution of Tyr248 supports activity at roughly 20% of wild type (**Fig. 2-4g**).

2.3 Discussion

Cellulose, chitin and HA synthases are membrane-embedded GTs that synthesize and secrete high molecular weight polysaccharides. Unlike cellulose and chitin, HA is a heteropolysaccharide of

`

alternating GlcNAc and GlcA units linked via different glycosidic linkages. Our analyses provide molecular insights into how HAS combines these tasks.

Cv-HAS elongates HA's non-reducing end, similar to cellulose and chitin synthases[13,23,24]. Considering the similarities of vertebrate HASs to Cv-HAS, we propose that Cv-HAS mechanistically and structurally represents the vertebrate enzymes. This is supported by the agreement of AlphaFold2 and RoseTTAfold models of human HAS-2 with the Cv-HAS structure (**Fig. S2-5 a-c**)[20]. The only discrepancies exist within the predicted location of TMH1, which appears to be flexible.

Initiating HA biosynthesis requires the formation of a priming carbohydrate acceptor, likely by substrate hydrolysis (**Fig. S2-8 and 9**). A similar mechanism has been proposed for cellulose synthase[25,26] and experimentally tested for chitin synthase[27]. The observation that GlcA fails to prime the biosynthesis reaction is consistent with its instability at the acceptor-binding site in MD simulations.

HAS extends GlcNAc-ending HA polymers only with GlcA to ensure an alternating HA sequence. Our 3D variability analysis of Cv-HAS incubated with UDP-GlcNAc reveals that the enzyme either binds UDP-GlcNAc at the active site or GlcNAc at the acceptor position, but not simultaneously. This is likely due to clashes between the GlcNAc primer and substrate and corroborated by the reduced rate of UDP-GlcNAc hydrolysis in the presence of saturating GlcNAc monosaccharide (**Fig. S2-9**). To investigate this phenomenon at the structural level, we performed atomistic MD simulations of HAS with GlcNAc both in the donor (UDP-GlcNAc) and acceptor (HA) positions. Quantifying the occupancy of the acetamido binding pocket reveals that either the donor or acceptor GlcNAc, but not both, are able to bind the pocket (**Fig. S2-10a and b**).

`

We further propose that the GlcNAc and GlcA acceptors are differently coordinated and positioned at the acceptor site. This would ensure that only the C4 hydroxyl of GlcA and the C3 hydroxyl of GlcNAc serve as acceptors during glycosyl transfer due to their proximities to the base catalyst Asp302 of the GDD motif (**Fig. 2-4h**).

Compared to cellulose synthase, HAS' TM channel is wider and rich in positively charged residues, reflecting differences of the conducted polysaccharides. Structural insights into cellulose synthase are limited to polysaccharide associated states, i.e. describing enzymes with an open TM channel, whereas our HAS analyses are predominantly of the closed state. The overall similarity of the HAS and cellulose synthase TM architectures (**Fig. S2-6c and 10c**) reinforce our assignment of the extracellular HAS channel exit at the TMH2-4 interface.

Our cryo EM structures do not resolve a conserved loop connecting IF3 with TMH5, right above the catalytic pocket (**Fig. 2-1**). The loop contains a WGTR/SG motif (**Fig. S2-1**), which is important for function of bacterial HAS[28]. It resembles cellulose synthase's gating loop, which transiently inserts into the catalytic pocket for substrate binding (**Fig. S2-10c**)[25]. A similar process in HAS could reposition the substrate towards the acceptor sugar to facilitate HA elongation.

Following glycosyl transfer, cellulose synthase translocates the extended polysaccharide upon substrate binding through conformational changes of a conserved α-helix of the GT domain[29]. The same helix exists in HAS, with the GDD motif capping its N-terminus. A similar step-by-step HA translocation mechanism seems plausible (**Fig. 2-4h and i**). Because HA is negatively charged, the membrane's electrochemical potential and the conserved positive charges lining the TM channel likely also contribute to translocation. Our MD simulations suggest that an unfavorable

`

register with GlcA at the acceptor position may be a transient intermediate that facilitates polymer translocation upon UDP-GlcNAc binding (**Fig. 2-4h and i**).

2.4 Methods

2.4.1 Construct Design and Mutagenesis

The gene for *Paramecium bursaria Chlorella virus CZ-2* hyaluronan synthase (Cv-HAS) was cloned as described[13] and sub-cloned after restriction enzyme digest with NcoI/XhoI into a pET28a vector modified to encode a C-terminal 10x histidine tag (pHAScz2). Mutagenesis was performed by PCR with overlapping primers (Supplementary Table 2) on pHAScz2 using KOD Hot Start polymerase (Novagen). Constructs for deletions of the N-terminal TMH were generated by amplifying the vector and remaining Cv-HAS gene with primers containing complementary overhangs.

2.4.2 Protein Expression and Purification

Cv-HAS was expressed in *E.coli* C43. Cultures were grown in terrific broth (TB) supplemented with 4% v/v glycerol and 1XM buffer[32]. Each culture was inoculated from an overnight culture in lysogeny broth (LB), grown to an OD_{600} of ~0.8 at 30 °C, cooled 1 hour to 20 °C, then expression was induced with 100 mg/L Isopropyl β-D-1-thiogalactopyranoside (IPTG). Constructs were expressed ~18 hours at 20 °C, then cell pellets were harvested by centrifugation.

Cell pellets were resuspended in a buffer of 10% glycerol, 100mM NaCl, and 20mM Tris pH 7.5 (RB), then incubated for 1 hour at 4 °C with 1 mg/mL lysozyme (Alfa Aesar). Cell suspensions

`

were processed by 3 passes through a microfluidizer, then intact cells and aggregated material was removed by low-speed centrifugation for 25 min at 12,500 RPM in a JA-20 rotor (Beckman). Membranes were isolated from the resulting lysates by centrifugation for 2 hours at 200,000 g in a Ti45 rotor (Beckman), then harvested and flash-frozen in liquid N_2 and stored at -80 °C.

For *in surfo* samples, membranes were thawed and resuspended in 300 mM NaCl, 20 mM Tris 7.5, 10% glycerol, 5mM β-mercaptoethanol (BME) (SB), supplemented with 40 mM imidazole, 1% lauryl maltose neopentyl glycol (LMNG), and 0.2% cholesterol hemisuccinate (CHS), then incubated 1 hour with agitation. Aggregated material was removed by centrifugation at 200,000 g for 30 min, then the supernatant was incubated with Ni-NTA resin for 1 hour. Resin was washed with SB supplemented with 80 mM imidazole and the protein was eluted with SB supplemented with 320 mM imidazole. The eluted fraction was concentrated and purified by size-exclusion chromatography (SEC) using an S200-increase column equilibrated in RB. Peak fractions were harvested and concentrated, followed by glycerol removal by a second SEC in 100 mM NaCl and 20 mM Tris pH 8.0 (FB) (all steps performed at 4°C).

Purification of nanodisc samples was similar, except membranes were solubilized in SB supplemented with 40 mM imidazole, 1 % n-Dodecyl β-D-maltoside (DDM), and 0.1 % CHS, the Ni-NTA wash and elution buffers were supplemented with 0.02 % DDM/0.002 % CHS, and the first SEC buffer was 100 mM NaCl, 20 mM Tris pH 7.5, 0.02 % DDM and 0.002 % CHS. Following the first SEC, protein was concentrated and reconstituted into nanodiscs in a 1:3:30 molar ratio of HAS:MSPE3D1:*E.coli* total lipid extract (solubilized in 60 mM DDM). To produce nanobody complexes, selected nanobodies were added in two-fold molar excess compared to HAS to the nanodisc reconstitution mixture. After a 30 minute incubation, BioBeads were added

`

stepwise to a final volume of 0.5 mL and incubated overnight to sequester detergent. Properly formed HAS nanodiscs were purified by SEC in FB, then fractions were screened by SDS-PAGE and negative stain EM for presence of both MSP and HAS and particle quality (all steps performed at 4°C).

2.4.3 *In vitro* HA Synthesis Assays

HA synthesis was assayed by incorporation of ^{3}H- or ^{14}C-labeled sugars into HA produced by Cv-HAS in inverted membrane vesicles (IMVs), as described [13]. Wild type or mutated HAS constructs were expressed as described above. Lysates resulting from the low-speed spin were floated on a 2 M sucrose cushion, then spun for 2 hours at 200k x g in a Ti45 rotor. IMVs were harvested, diluted in RB, and centrifuged again at 200,000 g for 1.5 hours. The resulting IMV pellet was harvested, homogenized with a dounce in RB, and flash-frozen in liquid N_2.

Cv-HAS levels in each IMV isolate were quantified by Western blotting. 200 µL IMVs were solubilized by addition of 800 µL SB containing 1 % DDM, 0.1 % CHS, and 20 mM Imidazole, then incubated 1 hour with Ni-NTA resin. The resin was washed twice in 250 µL SB containing 0.03 % DDM, 0.003 % CHS, and 20 mM Imidazole. The washed resin was then resuspended in Laemmli buffer and 1 M imidazole, then analyzed on a 12.5 % polyacrylamide gel, and transferred at 100 V to a nitrocellulose membrane. The membrane was washed 5 min in dH_2O, blocked twice for 15 minutes with 5 % milk in Tris-buffered saline-tween buffer (TBST), and incubated with primary antibody for 1 hour (all at room temperature). The membrane was rinsed 15 mins in TBST, then incubated for 1 hour with IRDye800-conjugated secondary antibody, then rinsed again for 15

`

mins in TBST. Blots were imaged using an Odyssey Licor scanner at 700 and 800nm, and analyzed using imageJ to determine relative protein content across multiple Cv-HAS variants.

Prior to each assay, IMVs were diluted to normalize Cv-HAS concentration. 10 µL of each IMV were mixed with 10 µL 2X reaction buffer, containing 40 mM $MnCl_2$, 80 mM Tris pH 7.5, 150 mM NaCl, 10 mM TCEP, 10 mM UDP-GlcNAc, 10 mM UDP-GlcA, and 0.2 µCi of the appropriate radiolabeled sugar (listed below with each experiment). Reaction mixtures were incubated 2 hours at 30 °C then analyzed.

For descending paper chromatography, the radiolabel was ^{3}H-UDP-GlcNAc and each reaction was quenched with 2% SDS. 20 µL of each reaction mixture was spotted onto Whatman 2mm filter paper, developed with 65% 1 M $(NH_4)_2SO_4$/35 % ethanol, and the origins were counted using a Beckman S Ls liquid scintillation counter. Activity of each variant was repeated in triplicate, from which background was subtracted as determined by a control reaction lacking UDP-GlcA.

For autoradiography comparison of HA sizes produced across Cv-HAS variants, the radiolabel was ^{14}C-UDP-GlcA. Reactions were quenched by addition of loading buffer, then run at 100 V for 2 hours on a 12.5 % polyacrylamide gel. The gel was fixed using 30 % methanol/3 % glycerol for 15 mins and dried onto a Whatman 2mm filter paper using a Biorad model 583 gel dryer for 2 hours. The dried gel was exposed for ~1 week to a Kodak Biomax MS film. Autoradiography of the initiation experiments was identical, except the radiolabels were ^{14}C-labeled GlcNAc or GlcA (not UDP activated). Hyaluronate lyase digestions were performed by adding 1 mg/mL final concentration of recombinantly expressed *S. pneumoniae* hyaluronate lyase[33] to the relevant reaction mixture.

`

2.4.4 Nanobody generation

Camelid nanobodies against Cv-HAS were generated *in vivo* by immunizing a male llama weekly over a period of 6 weeks with purified wild type and catalytically inactive (D201A) Cv-HAS. Animal protocols were approved by the committee for ethical treatment of laboratory animals, Vrije Universiteit Brussel (VUB). Approximately 1 mg of total protein reconstituted into *E. coli* total lipid proteoliposomes was used for immunization. The isolation of high affinity Cv-HAS binders and nanobody cloning into the pMESy4 expression vector followed previously described procedures[14]. Briefly, Cv-HAS wild type, Cv-HAS wild type with synthesized hyaluronan, and Cv-HAS D201A reconstituted into E. coli total lipid proteoliposome were each solid phase coated and used as target for the biopanning in 25 mM MES pH 6.5, 0.2 M NaCl, 10 % glycerol, 5 mM beta-mercaptoethanol, 3.5 mM $MnCl_2$. After rescue of the eluted phage, individual colonies were screened in ELISA using the same buffer conditions. Thirty-eight different Cv-HAS-specific nanobody families were discovered belonging to 28 families based on their CDR3 sequences.

2.4.5 Nanobody purification

Constructs for each nanobody gene cloned in pMESy4 were expressed in *E.coli* WK6 cultures grown in TB supplemented with 4% v/v glycerol and 1XM buffer. Periplasmic fractions of the harvested cell pellets were extracted by osmotic shock using a Tris-sucrose-EDTA buffer (TES). Periplasmic extracts were incubated with Ni-NTA resin, then nonspecifically bound material was removed by successive washes in 20 mM Tris pH 7.5 containing 1 M NaCl and 20 mM Imidazole, then 0.1 M NaCl and 40 mM Imidazole, and 0.1 M NaCl, 320 mM Imidazole to elute. Eluted protein was further purified by SEC using an S75 column equilibrated in FB, then concentrated and flash-frozen in liquid N_2.

`

2.4.6 Thermo-stability assays

Stability measurements were based on monitoring Cv-HAS' enzymatic activity following incubation at elevated temperatures. The activity was estimated by quantifying the release of UDP in real time using an enzyme-coupled reaction that oxidizes NADH, as previously described[7]. Cv-HAS was individually incubated with the 38 nanobodies in a thermo-cycler at temperatures ranging from 30 to 66°C for 2 hours in 1:3 molar ratio. Biosynthesis reactions were performed in 120 µL volumes containing 1 mM PEP, 0.75 mM NADH, and 1 U of pyruvate kinase and lactate dehydrogenase each.

The oxidation of NADH was monitored at room temperature based on its absorbance at 340 nm in a microplate spectrophotometer in 30 s intervals for 1 hour. Reaction rates were calculated based on Lambert-Beer's law using an NADH extinction coefficient at 340 nm of 6220 $(M\ cm)^{-1}$. All experiments were performed in triplicate and error bars represent the deviations from the means.

2.4.7 Substrate hydrolysis assays

Hydrolysis of UDP-GlcNAc or UDP-GlcA was measured using an enzyme-coupled assay as described above. 1 µM Cv-HAS in nanodiscs was incubated with 2 mM substrate and the indicated concentration of GlcNAc or GlcA monosaccharide. Background hydrolysis was measured by omitting Cv-HAS. For each combination of substrate and monosaccharide, Cv-HAS catalyzed and background hydrolysis rates were measured in triplicate. Net hydrolysis rates were normalized separately for each substrate based on reactions with no monosaccharide.

`

2.4.8 EM Sample Preparation and Data Collection

Protein samples were concentrated using a 50 kDa filter (Amicon) to 3-4 mg/mL. For relevant samples $MnCl_2$ and UDP or UDP-GlcNAc were added to a final concentration of 2 mM. 3 µL samples in detergent were applied to glow-discharged C-flat holey carbon grids (Cu 1.2/1.3, 300 mesh), blotted for 10 seconds at 4°C and 100 % humidity, then plunge frozen in liquid ethane using a Vitrobot Mark IV (FEI). 3 µL samples in nanodiscs were applied to Quantifoil holey carbon grids (Cu 1.2/1.3, 300 mesh) glow discharged with amylamine, blotted for 4 s at 4 °C and 100 % humidity, then plunge frozen in liquid ethane.

Cryo-EM data were collected at the University of Virginia Molecular Electron Microscopy Core (MEMC) on a Titan Krios (FEI) 300-kV electron microscope using a Gatan Imaging Filter (GIF) and a K3 direct detection camera. Movies were collected in EPU (Thermofisher) at a magnification of 81,000x with an energy filter width of 10 eV, using counting mode with a total dose of 51 e^-/Å over 40 frames, and with a target defocus of -1.0 to -2.0 µm. Several earlier (unreported) datasets were collected under similar conditions at the National Cryo-Electron Microscopy Facility (NCEF) at the National Cancer Institute (NCI).

2.4.9 EM Data Processing and Model Building

All data processing steps were performed in cryoSPARC[16], unless otherwise noted. Movies were imported and gain corrected, then subjected to patch-based motion-correction and contrast transfer function estimation (CTFFIND4)[34]. Initially, particles were manually picked from accepted micrographs to generate references for template-based particle picking. References for later datasets were based on 2D classes from earlier sets. Particles identified from template-based

`

picking were extracted using 2-fold Fourier cropping for initial rounds of 2D classification and/or *ab initio* reconstruction followed by heterogeneous refinement to remove junk particles. Accepted particles were re-extracted at full scale then subjected to further rounds of classification. High-resolution reconstructions from clean particle sets were generated by non-uniform (NU) refinement, then masked local refinement (using NU algorithm), followed by local CTF estimation and refinement of the corrected particles. Finally, 3D variability analysis and subsequent clustering was used to separate populations within the high-resolution reconstructions belonging to distinct catalytic states. Clusters belonging to apo, UDP-bound, substrate bound, or primed states were individually pooled and used to generate final reconstructions from NU or local refinement. Two maps for the apo state were derived by 3D variability analyses from the dataset collected for wild-type Cv-HAS in the presence of UDP and from the dataset for D302N Cv-HAS collected in the presence of UDP-GlcNAc. The wild-type apo model (CC mask 0.68) and D302N apo model (CC mask 0.75) are virtually identical with an RMSD of 0.132 Å. Reported is the model derived from the UDP-GlcNAc dataset, which extends to 3.1 Å versus 3.3 Å.

The initial atomic model was built into a ~3.5 Å map, derived from collection on an apo Cv-HAS sample *in surfo*, starting from a poly-alanine model of the RsBcsA GT domain. The model was generated from iterative rounds of building in Coot[35] and real-space refinement in PHENIX [36]. Register was assigned based on well-resolved map regions and on the location of highly conserved catalytic motifs. TM helices were built *de novo* with register interpreted from density for bulky residues and based on connectivity to the GT domain. The initial model was refined and register errors corrected using subsequent higher resolution maps *in surfo* and in nanodiscs to generate the final models. All figures were prepared using PyMol and Chimera[30,37].

`

2.4.10 Molecular Dynamics Simulations

The coordinates of Cv-HAS, without nanobodies, obtained in nanodiscs were used for building atomistic MD simulations. Where included, the substrate coordinates were taken from the resolved cryo-EM densities for UDP-GlcNAc, and the terminal residue of the HA chain placed according to the position of the primer monosaccharide. Three different HA poses were manually constructed in Coot[35] for each register, to provide a broad degree of HA conformational sampling. The systems were described with the CHARMM36m force field [38] and built into model POPE membranes with TIP3P waters and K^+ and Cl^- to 150 mM, using CHARMM-GUI[39,40]. The bound Mn^{2+} were swapped to Mg^{2+}. The final systems had approximately 300 lipids, 150 ions, and 27500 waters, for a total of ca. 125,000 atoms and a box size of 10 x 10 x 12.5 nm. All titratable side chains were set to their default protonation state, based on analysis run with propKa3.1 [41,42].

Each system was minimized using the steepest descents method, then equilibrated in 6 rounds with restraints initially applied to protein backbone and sidechain atoms, lipid phosphate atoms and selected protein, lipid and sugar dihedrals, as per the standard CHARMM-GUI output. Where the HA chain was present, to optimise its positioning in the channel additional equilibration simulations of 90 ns were run with 50 kJ/mol/nm^2 positional restraints applied to the protein backbone and to the bound substrate, but not to the HA chain. Production simulations were then seeded using the output frame of this final equilibration, with 3 or 5 repeats seeded for each system. Simulations were run using 2 fs time steps, with V-rescale thermostat at 303.15 K with a tau t of 1 ps, and semi-isotropic Parrinello-Rahman pressure coupling with tau p of 5 ps and a reference of 1 bar [40,43]. All simulations were run in Gromacs 2019.4[43-45]. In total, ca. 9 μs of data were generated.

`

Lipid and solvent densities in the apo MD simulations were computed over the 5 repeats for each system using the VolMap tool of VMD [46,47]. HA-residue contact analysis was determined for each frame of the simulation based on a distance cut-off of 0.4 nm between each residue in the system and the specified sugar, and was run using MDAnalysis [47]. Quantification of the occupancy of the binding pocket using in was performed by counting the number of frames that the acetamido group of the donor or acceptor GlcNAc was within 0.4 nm of any atom in Cys-231, and was run using the Gromacs tool gmx mindist. Analysis of the HA chain position in relation to the channel was calculated using the Gromacs tool gmx distance. The distance in the z-dimension between the centre-of-mass of each sugar and the centre-of-mass of the protein was computed, and plotted using NumPy[48] and Matplotlib[49].

2.5 Data availability

Raw EM movies and maps have been deposited at the PDB and EM data banks under accession codes 7SP7/EMD-25367, 7SP6/EMD-25366, 7SP8/EMD-25368, 7SP9/ EMD-25369, and 7SPA/EMD-25370 for the UDP-bound, D302N apo, UDP-GlcNAc-bound, primed (closed), and primed (open) states, respectively.

2.6 Acknowledgments

We thank K. Dryden and M. Purdy from the MEMC at the University of Virginia as well as Adam Wier and support staff at the NCI-NCEF. E.P. and J.S. acknowledge the support of Instruct-ERIC, part of the European Strategy Forum on Research Infrastructures (ESFRI), and the Research Foundation - Flanders (FWO) for supporting the nanobody discovery and thank Eva Beke for

`

technical assistance. J.Z. and F.M. were supported by NIH grant R21AI148853. R.H. was supported by NIH grant R01GM101001 (awarded to J.Z.). R.A.C. and P.J.S. are supported by Wellcome (208361/Z/17/Z). P.J.S.'s lab is supported by awards from the BBSRC (BB/P01948X/1, BB/R002517/1 and BB/S003339/1) and MRC (MR/S009213/1). P.J.S. acknowledges the University of Warwick Scientific Computing Research Technology Platform for computational access.

`

2.7 References

1 Girish, K. S. & Kemparaju, K. The magic glue hyaluronan and its eraser hyaluronidase: a biological overview. *Life Sci* **80**, 1921-1943, (2007).

2 Cyphert, J. M., Trempus, C. S. & Garantziotis, S. Size Matters: Molecular Weight Specificity of Hyaluronan Effects in Cell Biology. *Int J Cell Biol* **2015**, 563818, (2015).

3 Vigetti, D. *et al.* Hyaluronan: biosynthesis and signaling. *Biochim Biophys Acta* **1840**, 2452-2459, (2014).

4 DeAngelis, P. Hyaluronan synthases: fascinating glycosyltransferases from vertebrates, bacterial pathogens, and algal viruses. *Cell Mol Life Sci* **56**, 670-682, (1999).

5 Sironen, R. *et al.* Hyaluronan in human malignancies. *Exp Cell Res* **317**, 383-391, (2011).

6 Cowman, M. K., Lee, H.-G., Schwertfeger, K. L., McCarthy, J. B. & Turley, E. A. The Content and Size of Hyaluronan in Biological Fluids and Tissues. *Front Immunol* **6**, 261, (2015).

7 Hubbard, C., McNamara, J., Azumaya, C., Patel, M. & Zimmer, J. The hyaluronan synthase catalyzes the synthesis and membrane translocation of hyaluronan. *J Mol Biol* **418**, 21-31, (2012).

8 Weigel, P. H. & Deangelis, P. L. Hyaluronan synthases: a decade-plus of novel glycosyltransferases. *J Biol Chem* **282**, 36777-36781, (2007).

9 DeAngelis, P. L., Jing, W., Drake, R. R. & Achyuthan, A. M. Identification and molecular cloning of a unique hyaluronan synthase from Pasteurella multocida. *J Biol Chem* **273**, 8454-8458, (1998).

10 Camenisch, T. D. *et al.* Disruption of hyaluronan synthase-2 abrogates normal cardiac morphogenesis and hyaluronan-mediated transformation of epithelium to mesenchyme. *J Clin Invest* **106**, 349-360, (2000).

11 Itano, N. *et al.* Three isoforms of mammalian hyaluronan synthases have distinct enzymatic properties. *J Biol Chem* **274**, 25085-25092, (1999).

12 DeAngelis, P., Jing, W., Graves, M., Burbank, D. & Van Etten, J. Hyaluronan synthase of chlorella virus PBCV-1. *Science* **278**, 1800-1803, (1997).

13 Blackburn, M. R. *et al.* Distinct reaction mechanisms for hyaluronan biosynthesis in different kingdoms of life. *Glycobiology* **28**, 108-121, (2018).

14 Pardon, E. *et al.* A general protocol for the generation of Nanobodies for structural biology. *Nat Protoc* **9**, 674-693, (2014).

15 Tlapak-Simmons, V. L., Baron, C. A. & Weigel, P. H. Characterization of the purified hyaluronan synthase from Streptococcus equisimilis. *Biochemistry* **43**, 9234-9242, (2004).

16 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, (2017).

17 Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **77**, 521-555, (2008).

18 Cantarel, B., Coutinho, P., Rancurel, C. & Bernard, T. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37**, D233-238, (2009).

19 Wu, Q. *et al.* Protein contact prediction using metagenome sequence data and residual neural networks. *Bioinformatics* **36**, 41-48, (2020).

20 Tunyasuvunakool, K., Adler, J., Wu, Z., Jumper, J. & Hassabis, D. Highly accurate protein structure prediction for the human proteome. *Nature*, https://doi.org/10.1038/s41586-41021-03828-41581, (2021).

`

21 Baek, M. Accurate prediction of protein structures and interactions using a 3-track network. *bioRxiv* https://doi.org/10.1101/2021.06.14.448402 (2021).
22 Charnock, S. J. & Davies, G. J. Structure of the nucleotide-diphospho-sugar transferase, SpsA from Bacillus subtilis, in native and nucleotide-complexed forms. *Biochemistry* **38**, 6380-6385, (1999).
23 Morgan, J., Strumillo, J. & Zimmer, J. Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature* **493**, 181-186, (2013).
24 Purushotham, P., Ho, R. & Zimmer, J. Architecture of a catalytically active homotrimeric plant cellulose synthase complex. *Science* **369**, 1089-1094, (2020).
25 Morgan, J. L. W., McNamara, J. T. & Zimmer, J. Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. *Nature Struct Mol Biol* **21**, 489-496, (2014).
26 McManus, J., Yang, H., Wilson, L., Kubicki, J. & Tien, M. Initiation, Elongation, and Termination of Bacterial Cellulose Synthesis. *ACS Omega* **3**, 2690-2698, (2018).
27 Orlean, P. & Funai, D. Priming and elongation of chitin chains: Implications for chitin synthase mechanism. *Cell Surf* **5**, 100017, (2019).
28 Yang, J. *et al.* Key Role of the Carboxyl Terminus of Hyaluronan Synthase in Processive Synthesis and Size Control of Hyaluronic Acid Polymers. *Biomacromolecules* **18**, 1064-1073, (2017).
29 Morgan, J. L. *et al.* Observing cellulose biosynthesis and membrane translocation in crystallo. *Nature* **531**, 329-334, (2016).
30 PyMol. *PyMol*, <https://pymol.org/2/> (https://pymol.org/2/).
31 Ho, B. & Gruswitz, F. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct Biol* **8**, 49, (2008).
32 Studier, F. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**, 207-234, (2005).
33 Kelly, S. J., Taylor, K. B., Li, S. & Jedrzejas, M. J. Kinetic properties of Streptococcus pneumoniae hyaluronate lyase. *Glycobiology* **11**, 297-304, (2001).
34 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221, (2015).
35 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, (2004).
36 Adams, P. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, (2010).
37 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, (2004).
38 Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature* **14**, 71-73, (2016).
39 Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29**, 1859-1865, (2008).
40 Lee, J. *et al.* CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theo Comput* **12**, 405-413, (2016).
41 Sondergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK(a) Values. *J Chem Theor Comp* **7**, 2284-2295, (2011).

`

42 Olsson, M. H. M., Sondergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK(a) Predictions. *J Chem Theor Comp* **7**, 525-537, (2011).
43 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* **126**, 014101, (2007).
44 Parrinello, M., Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* **52**, 7182-7190, (1981).
45 Berendsen, H. J. C., van der Spoel, D., van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comp Phys Comm* **91**, 43-56, (1994).
46 Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38, (1996).
47 Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comp Sci Engineering* **9**, 90-95, (2007).
48 Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J Comp Chem* **32**, 2319-2327, (2011).
49 Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357-362, (2020).
50 Bernsel, A., Viklund, H., Hennerdal, A. & Elofsson, A. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res* **37**, W465-468, (2009).
52 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* **126**, 014101, (2007).

`

2.8 Supplementary information

Table 1

State	Apo (WT) (detergent)	UDP (WT)	D302N: Apo	UDP-GlcNAc	Primed (closed)	Primed (open)	Cytosolic Nbs
PDB		7SP7	7SP6	7SP8	7SP9	7SPA	
EMDB		EMD-25367	EMD-25366	EMD-25368	EMD-25369	EMD-25370	
Data collection/ processing							
Magnification	81,000x	81,000x		81,000x			81,000x
Voltage (kV)	300	300		300			300
Electron Exposure (e–/Å^2)	51	51		48			50
Defocus range (μm)	-2.2 to -1.4	-2.0 to -1.0		-2.0 to -1.0			-2.25 to -1.0
Pixel size (Å^2)	1.08	1.08		1.08			1.08
Symmetry imposed	P1	P1		P1			P1
Initial Particles	1,548,653	1,897,858		4,391,341			2,065,933
Final Particles	114152	174,023	91931	105,056	107,417	104,995	387,019
Map resolution (Å)	3.4	3.1	3.1	2.7	2.9	2.8	5.7
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Map Resolution Range (Å)	10-3.3	50.6-2.74	51.1-2.83	45.5-2.41	42.2-2.58	26.4-2.48	--
Refinement							
Initial Model	4P00	Apo (det)	UDP	UDP	UDP	Closed	
Model resolution (Å)		3.4	3.5	2.9	3.2	3.1	
FSC threshold		0.5	0.5	0.5	0.5	0.5	
Map sharpening *B* factor (Å^2)	113.2	82.9	61.3	55.7	56.6	56.7	305.0
Model composition							
Non-hydrogen Atoms		6052	5957	6090	5966	6017	
Protein Residues		739	733	742	737	737	
Ligands		5	3	5	2	3	
B factors (Å^2)							
Protein		78.87	96.12	66.52	54.16	49.08	
Ligand		83.23	104.60	74.96	65.46	59.66	
R.M.S. deviations							
Bond lengths (Å)		0.001	0.001	0.001	0.001	0.002	
Bond angles (°)		0.352	0.419	0.406	0.399	0.409	
Validation							
MolProbity Score		1.12	1.29	1.30	1.29	1.34	
Clash Score		3.34	5.33	5.47	5.43	6.2	
Poor rotomers (%)		0	0	0.16	0	0	
Ramachandran plot							
Favored (%)		98.63	98.48	98.50	98.76	98.62	
Allowed (%)		1.37	1.52	1.50	1.24	1.38	
Outliers (%)		0	0	0	0	0	

Table 1 | Data collection, processing, and refinement statistics.

EM statistics are listed for wild type (WT) and catalytically inactive (D302N) Cv-HAS.

`

Table 2 **HA Chain Contacts**

UDP-GlcNAc bound	GlcA5	GlcNAc4	GlcA3	GlcNAc2	GlcA1
303ARG	0.00	0.00	0.00	1.00	0.38
256ARG	0.00	0.00	0.23	1.00	0.94
247ARG	0.92	1.00	0.99	0.89	0.00
442ARG	0.00	0.91	0.95	0.67	0.06
248TYR	0.00	0.00	0.00	0.99	0.39
346TRP	0.00	0.18	0.68	0.99	0.00
411LYS	1.00	0.97	0.00	0.00	0.00

UDP-GA bound	GlcNAc5	GlcA4	GlcNAc3	GlcA2	GlcNAc1
303ARG	0.00	0.00	0.00	0.94	0.97
256ARG	0.00	0.00	0.08	0.98	1.00
247ARG	0.16	1.00	1.00	0.54	0.87
442ARG	0.00	0.00	0.99	0.63	0.32
248TYR	0.00	0.00	0.00	0.01	1.00
346TRP	0.00	0.00	0.48	0.93	0.69
411LYS	1.00	1.00	0.00	0.00	0.00

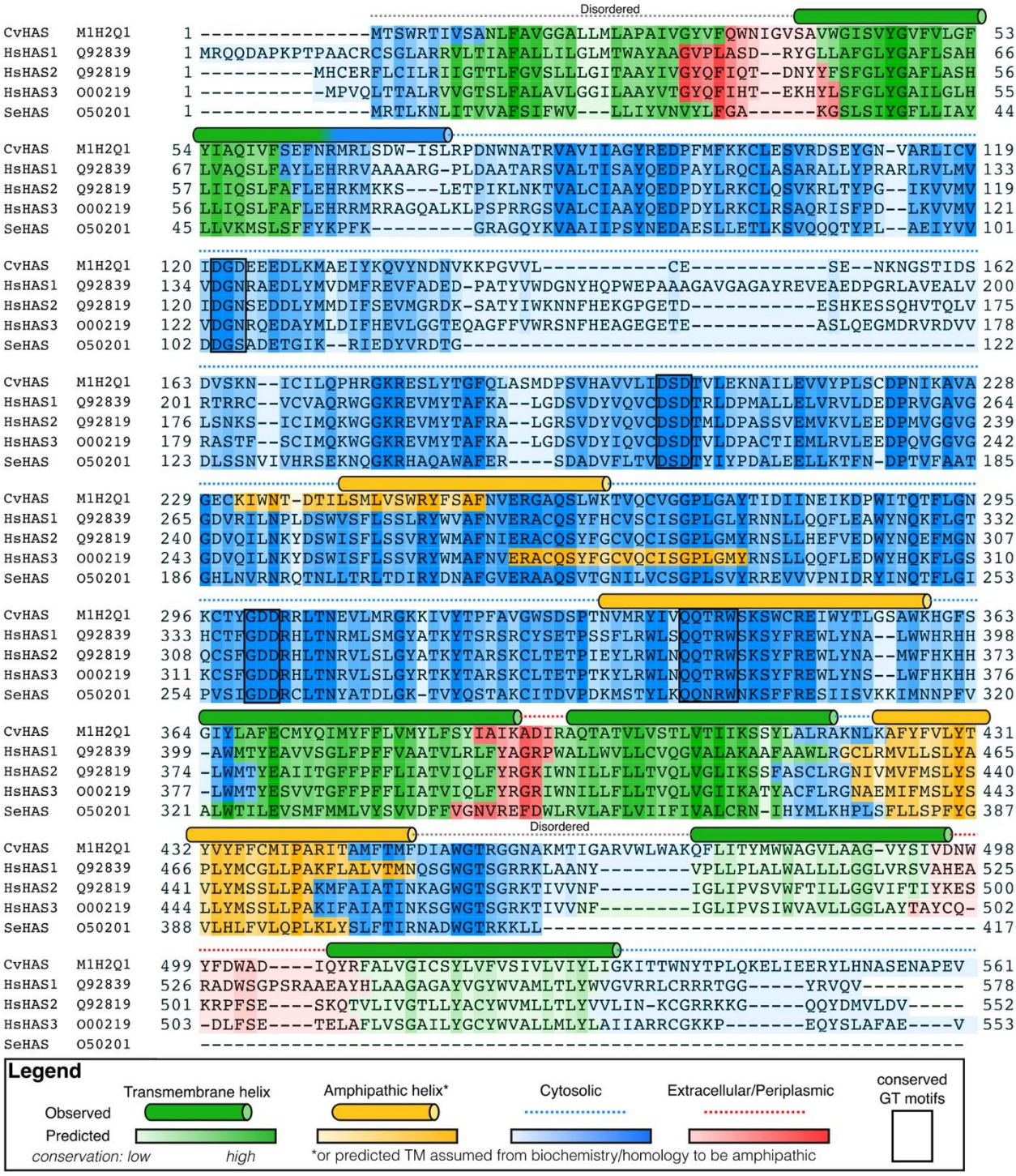
Donor-Acceptor Contacts

UDP-GlcA bound:	GlcNAc1	UDP-GlcA	UDP-GlcNAc bound:	GlcA1	UDP-GlcNAc
177LYS	0.04	1		0	1
201ASP	0	1		0	1
231CYS	0.91	0.06		0.01	0.8
233ILE	0.55	0		0	0.98
248TYR	0.99	0		0.47	0.99
252PHE	0.58	0		0.15	0.01
256ARG	0.17	0		0.75	0
267CYS	0.95	0		0.69	0.01
271PRO	0.67	0.9		0	0.78
301ASP	0.05	0.05		0.05	0.8
302ASP	0.79	0.8		0	0.81
303ARG	0.35	0		0.98	0.02
341ARG	0.05	0.65		0	0.16

Table 2 | Contact analysis of Cv-HAS with modeled ligands.
Top two: contact information for key HA-coordinating residues highlighted in the text. Data are from MD simulations seeded with different initial starting HA poses from those used in the main figures. Bottom: contact information for active site residues that contact the acceptor or UDP-donor.

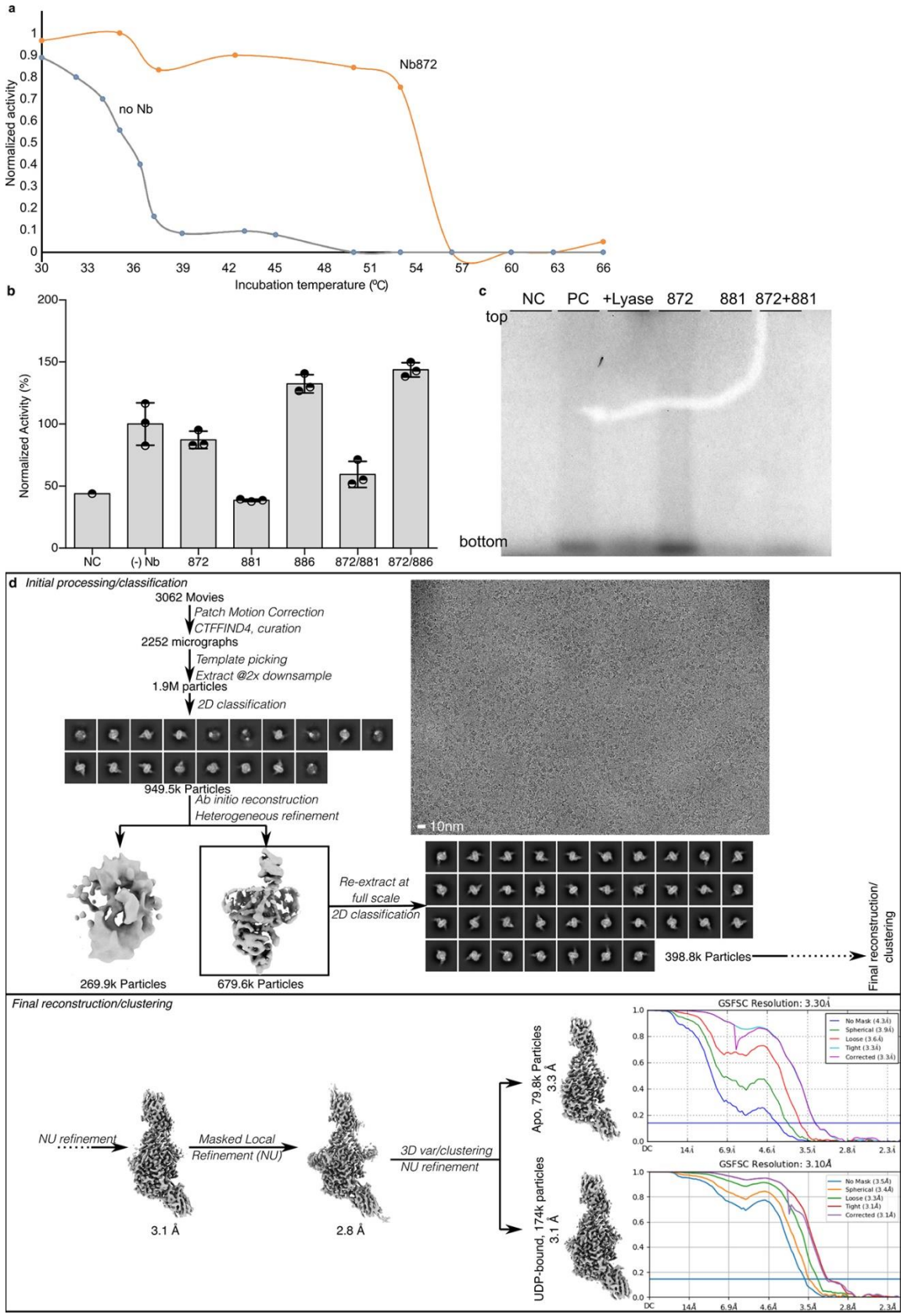
`

Fig. S2-1



Supplemental Figure 2-1 | Sequence alignment of HAS orthologues.
Comparison of HAS primary sequences from *Chlorella virus* (Cv), *Homo sapiens* (Hs) and *Streptococcus equisimilis* (Se). Topology predictions were performed using TopCons[50]. Cylinders indicate secondary structure elements observed in Cv-HAS.

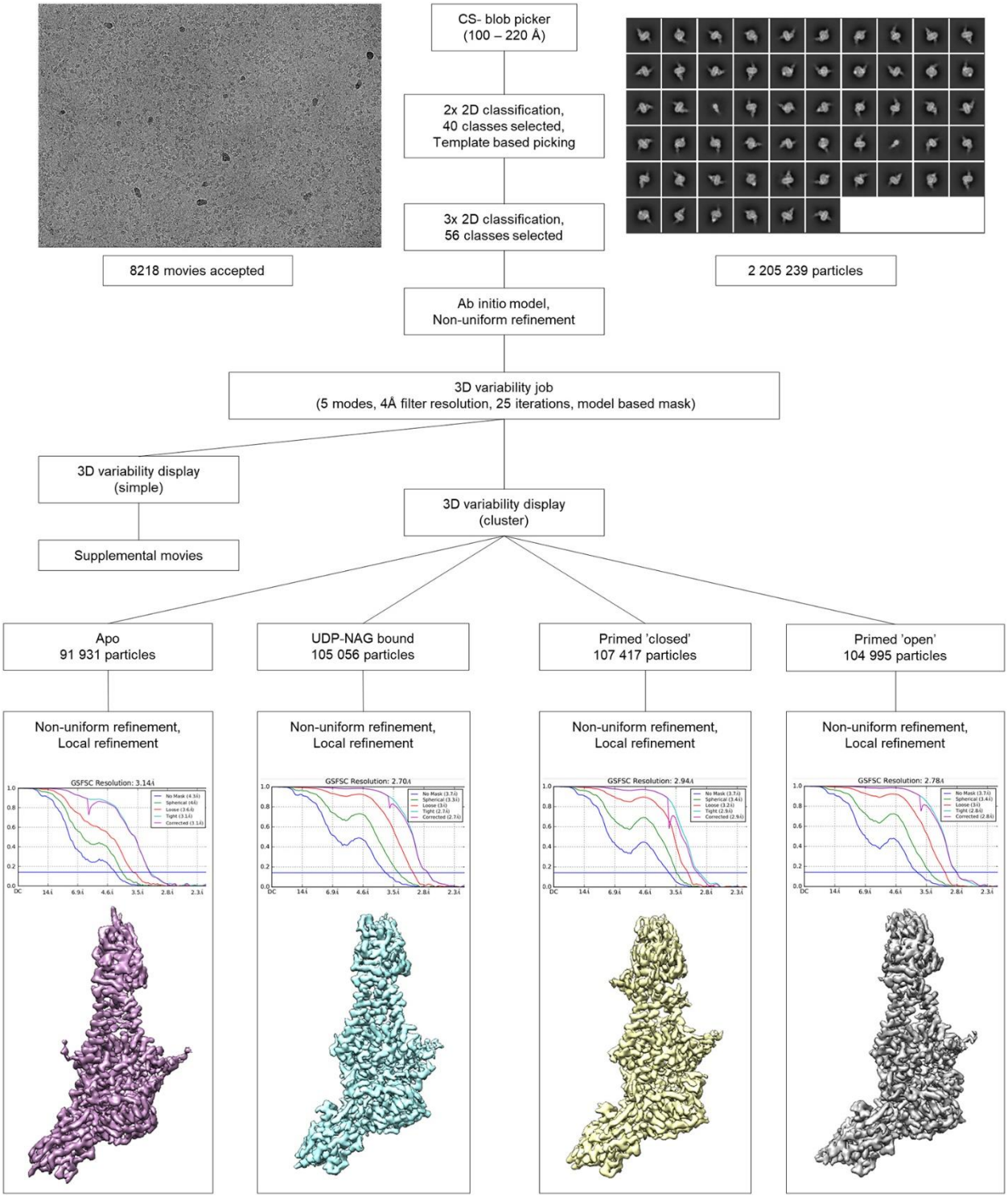
Fig. S2-2


a
Normalized activity
Nb872
no Nb
Incubation temperature (°C)
b
Normalized Activity (%)
NC
(-) Nb
872
881
886
872/881
872/886
c
NC
PC
+Lyase
872
881
872+881
top
bottom
d Initial processing/classification
3062 Movies
Patch Motion Correction
CTFFIND4, curation
2252 micrographs
Template picking
Extract @2x downsample
1.9M particles
2D classification
949.5k Particles
Ab initio reconstruction
Heterogeneous refinement
10nm
269.9k Particles
679.6k Particles
Re-extract at full scale
2D classification
398.8k Particles
Final reconstruction/clustering
Final reconstruction/clustering
NU refinement
3.1 Å
Masked Local Refinement (NU)
2.8 Å
3D var/clustering
NU refinement
Apo, 79.8k Particles
3.3 Å
UDP-bound, 174k particles
3.1 Å
GSFSC Resolution: 3.30Å
GSFSC Resolution: 3.10Å

Supplemental Figure 2-2 | Identification, data collection, and processing of Cv-HAS bound to two Nanobodies and UDP. (**a**) Increased melting temperature of Cv-HAS in the presence of Nb872. Protein melting was measured based on enzymatic activity detected by quantifying the release of UDP in real time. (**b**) HA biosynthesis in the presence of the indicated nanobodies and based on quantification of ^{3}H-labelled HA by scintillation counting. Data is normalized relative to product yields in the absence of nanobodies. Error bars represent deviations from the means with n = 3 independent experiments. (**c**) Representative autoradiography of ^{14}C-labelled HA produced in the presence of the indicated nanobodies. The experiment has been repeated at least 4 times with essentially identical results. NC: Negative control in the absence of UDP-GlcNAc substrate (for panel b) or UDP-GlcA (for panel c). PC: Positive control in the absence of nanobody. Lyase: Hyaluronan lyase treatment prior to SDS-PAGE. (**d**) This workflow produced the UDP-bound Cv-HAS structure.

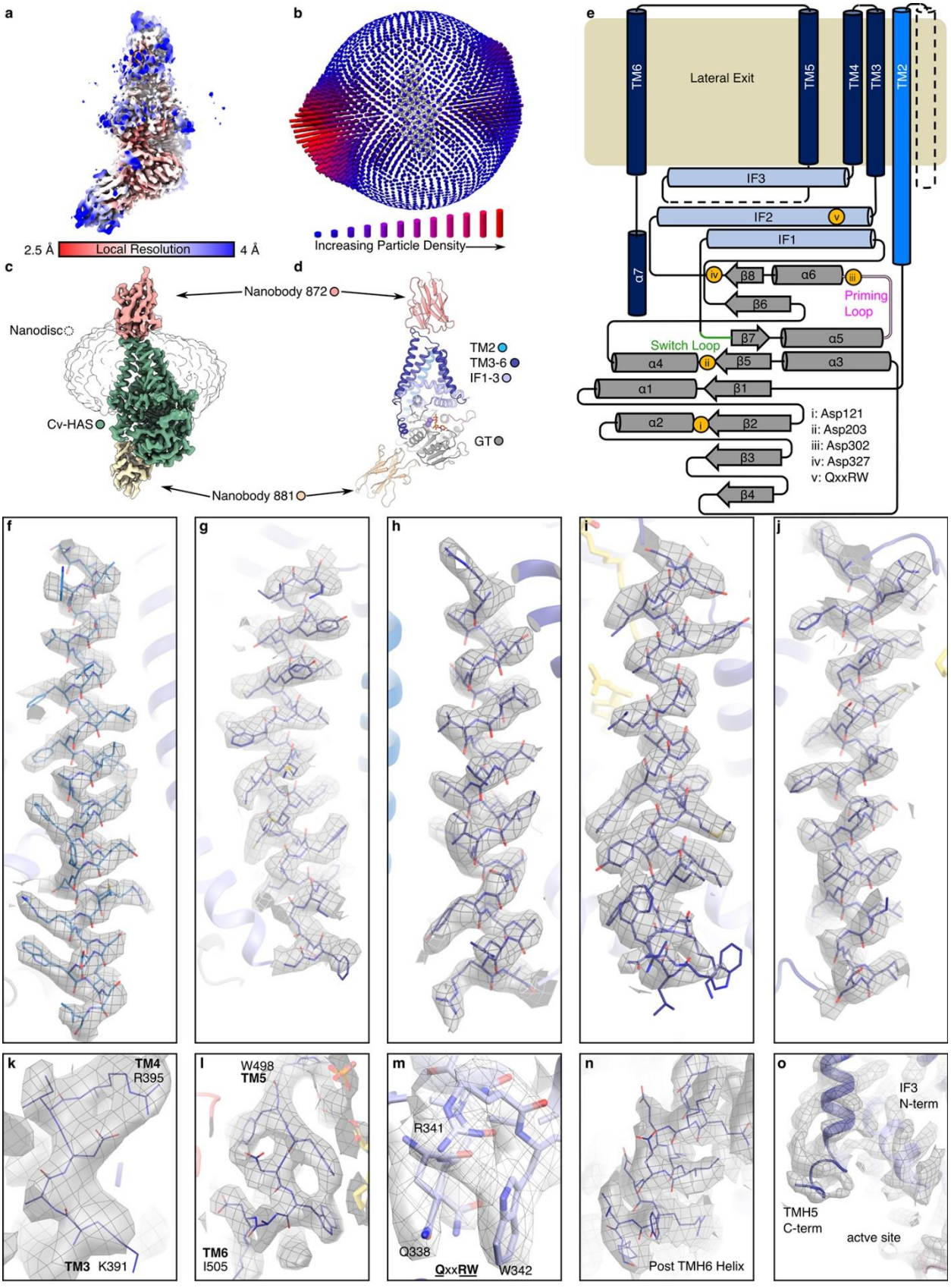
`

Fig. S2-3



Supplemental Figure 2-3 | Cryo-EM data collection and processing of Cv-HAS D302N in the presence of substrate. This workflow generated the apo, substrate-bound, primed, and primed with open channel Cv-HAS structures.

Fig. S2-4


a
2.5 Å
Local Resolution
4 Å
b
Increasing Particle Density
c
Nanobody 872
Nanodisc
Cv-HAS
Nanobody 881
d
TM2
TM3-6
IF1-3
GT
e
TM6
Lateral Exit
TM5
TM4
TM3
TM2
IF3
IF2
IF1
α7
β8
α6
Priming Loop
β6
Switch Loop
β7
α5
α4
β5
α3
α1
β1
α2
β2
β3
β4
i: Asp121
ii: Asp203
iii: Asp302
iv: Asp327
v: QxxRW
f
g
h
i
j
k
TM4
R395
TM3 K391
l
W498
TM5
TM6
I505
m
R341
Q338
QxxRW
W342
n
Post TMH6 Helix
o
IF3
N-term
TMH5
C-term
actve site

Supplemental Figure 2-4 | Map quality and model building of UDP-bound Cv-HAS.
(**a-d**) Map overview, estimated resolution based on FSC, and particle orientation distribution. (**e**) Secondary structure elements and topology of Cv-HAS. (**f-j**) TM helices 2 to 6 of Cv-HAS. (**k**) TMH3-4 extracellular loop. (**l**) The extracellular TMH5-6 loop. (**m**) The QxxRW motif. (**n**) The C-terminal cytosolic helix. (**o**) The unresolved TMH5-IF3 loop. All maps are contoured at 7.0σ.

Supplemental Figure 2-5 | Predicted location of TMH1.
(**a**) Relationship of evolutionarily coupled residues within Cv-HAS' TM and GT regions, generated in MapPred based on 65,535 sequences. TMH1 is shown at its predicted location as a violet cylinder. (**b**) RoseTTAfold models of full-length Cv-HAS. Cv-HAS is shown as a surface and its TMH 2 as a blue cylinder. TMH1 is shown as a cartoon at its predicted locations. (**c**) An AlphaFold2 predicted structure of human HAS2 (coloured blue to red from its N- to C-terminus) overlaid with the Cv-HAS structure shown as a grey cartoon and semi-transparent surface. (**d**) TMH1 remains disordered when two cytosolic nanobodies are used for cryo-EM analyses. (**e**) Catalytic activity of TMH1 truncated Cv-HAS. Left: Western blot of IMVs used for in vitro activity measurements. Right: Catalytic activity of the indicated Cv-HAS mutants expressed relative to the wild type enzyme. The assay quantifies ^{3}H-labelled HA by scintillation counting. Control reactions in the absence of UDP-GlcA served as background and are subtracted. Error bars represent deviations from the means with n = 3 independent experiments. On the next page.

`

Fig. S2-5

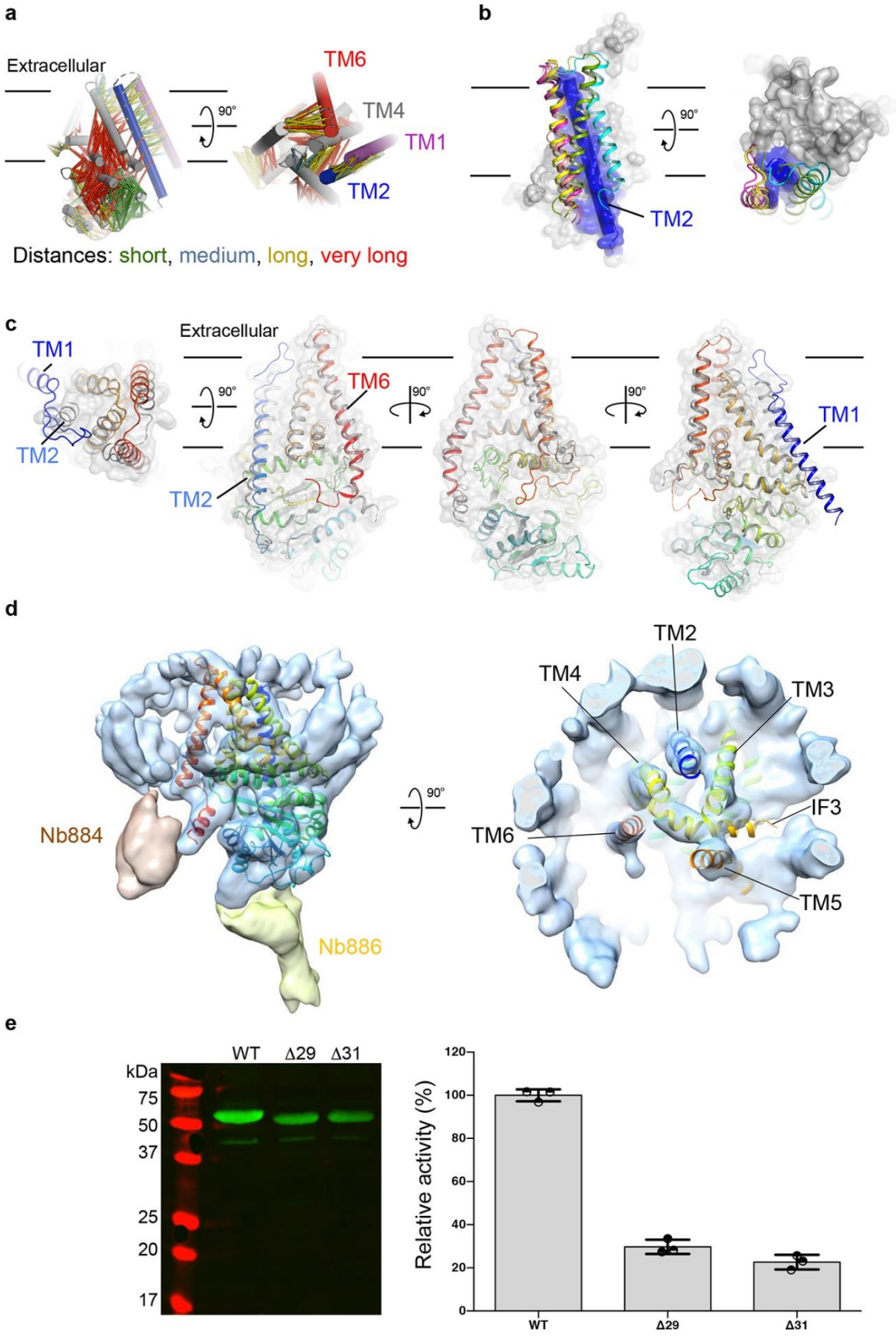
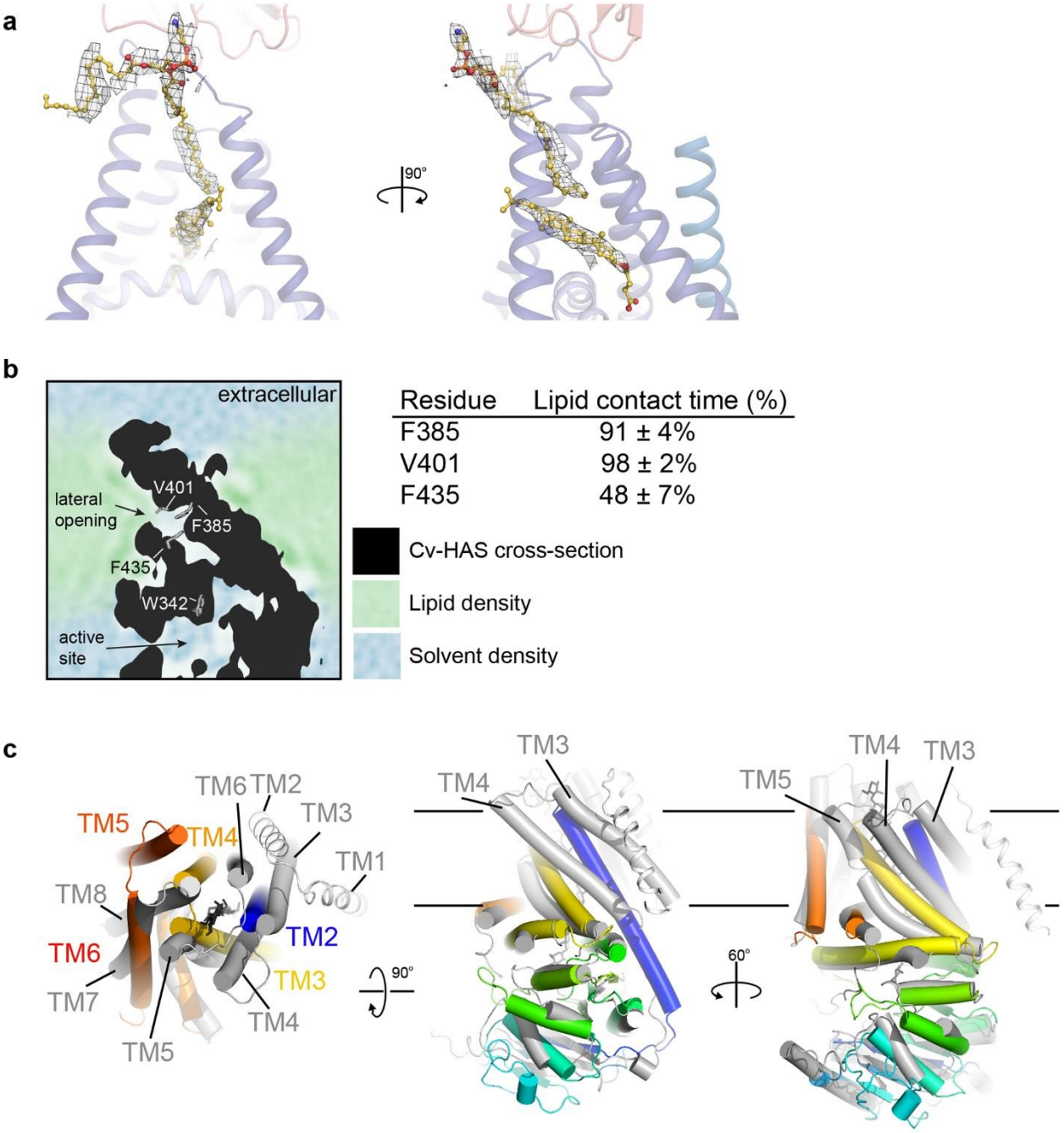

a
Extracellular
90°
TM6
TM4
TM1
TM2
Distances: short, medium, long, very long
b
90°
TM2
c
TM1
Extracellular
90°
TM6
90°
90°
TM2
TM1
d
Nb884
90°
TM2
TM4
TM3
IF3
TM6
TM5
Nb886
e
WT
Δ29
Δ31
kDa
75
50
37
25
20
17
Relative activity (%)
120
100
80
60
40
20
0
WT
Δ29
Δ31

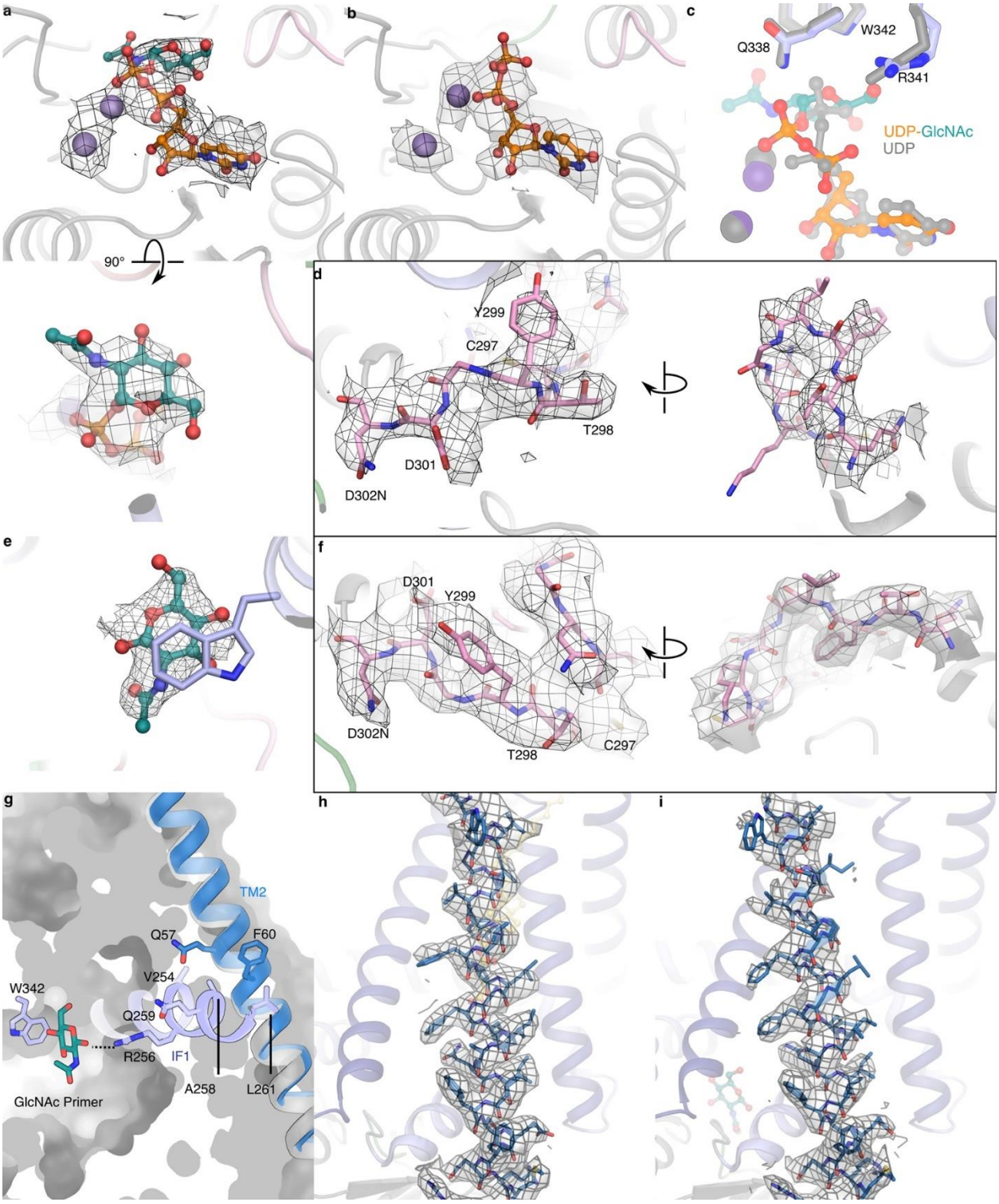
Fig. S2-6



Supplemental Figure 2-6 | Lipids plug the lateral channel opening.
a) Representative map regions for modelled lipids contoured at 7.0σ (from the UDP-GlcNAc bound set). (**b**) 2D slice from MD simulations of Cv-HAS (black area) within a POPE bilayer. Water and lipid densities are coloured blue and green, respectively. Right panel: Lipid contact times with selected channel residues. (**c**) Comparison of the Cv-HAS (rainbow coloured from the N- to C-terminus) and RsBcsA (grey, 4P00). Cellulose associated with BcsA is shown as black sticks. Helices are shown as cylinders except BcsA's N-terminal two TMHs, which are shown as coils.

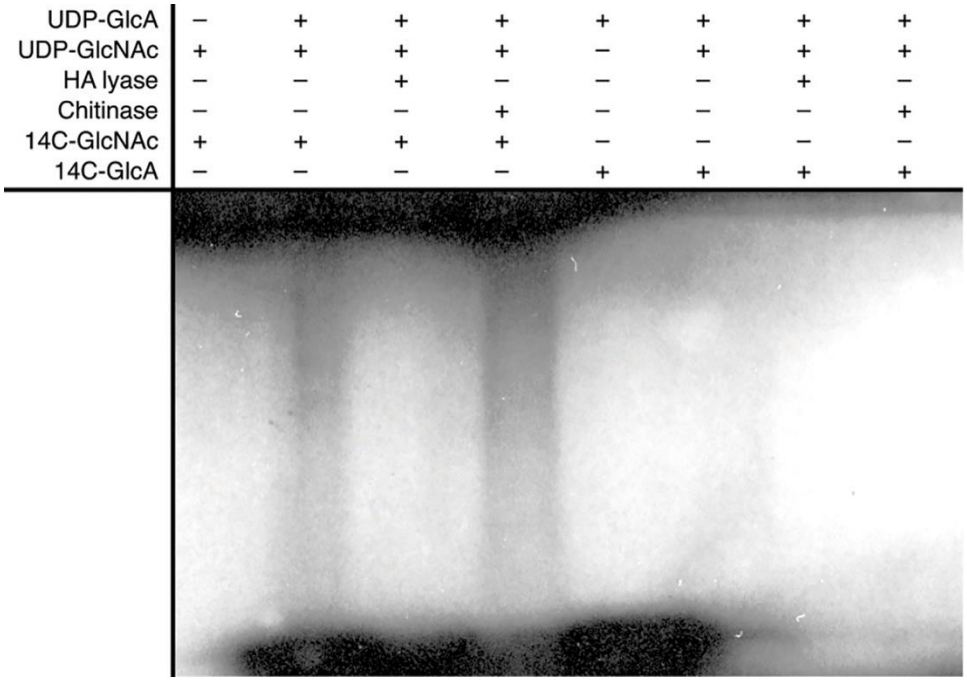
`

Fig. S2-7


a
b
c
W342
Q338
R341
UDP-GlcNAc
UDP
90°
d
Y299
C297
T298
D301
D302N
e
f
D301
Y299
D302N
T298
C297
g
h
i
TM2
Q57
F60
V254
W342
Q259
R256
IF1
A258
L261
GlcNAc Primer

Supplemental Figure 2-7 | Details of substrate-binding and of priming-induced conformational changes.
(**a** and **b**) Map quality for UDP-GlcNAc, UDP, and Mn^{2+} ligands. (**c**) Comparison of UDP and UDP-GlcNAc positions. (**d**) Map for the priming loop in nucleotide bound states. (**e**) Representative map for the GlcNAc primer. (**f**) Map for the priming loop in the primed states. (**g**) Contact point of TMH2 (open in blue, closed in grey) with IF1 in the primed state. (**h** and **i**) Map quality for TMH2 in a closed position (UDP-GlcNAc bound) and open position. All maps are contoured at 7.0σ.

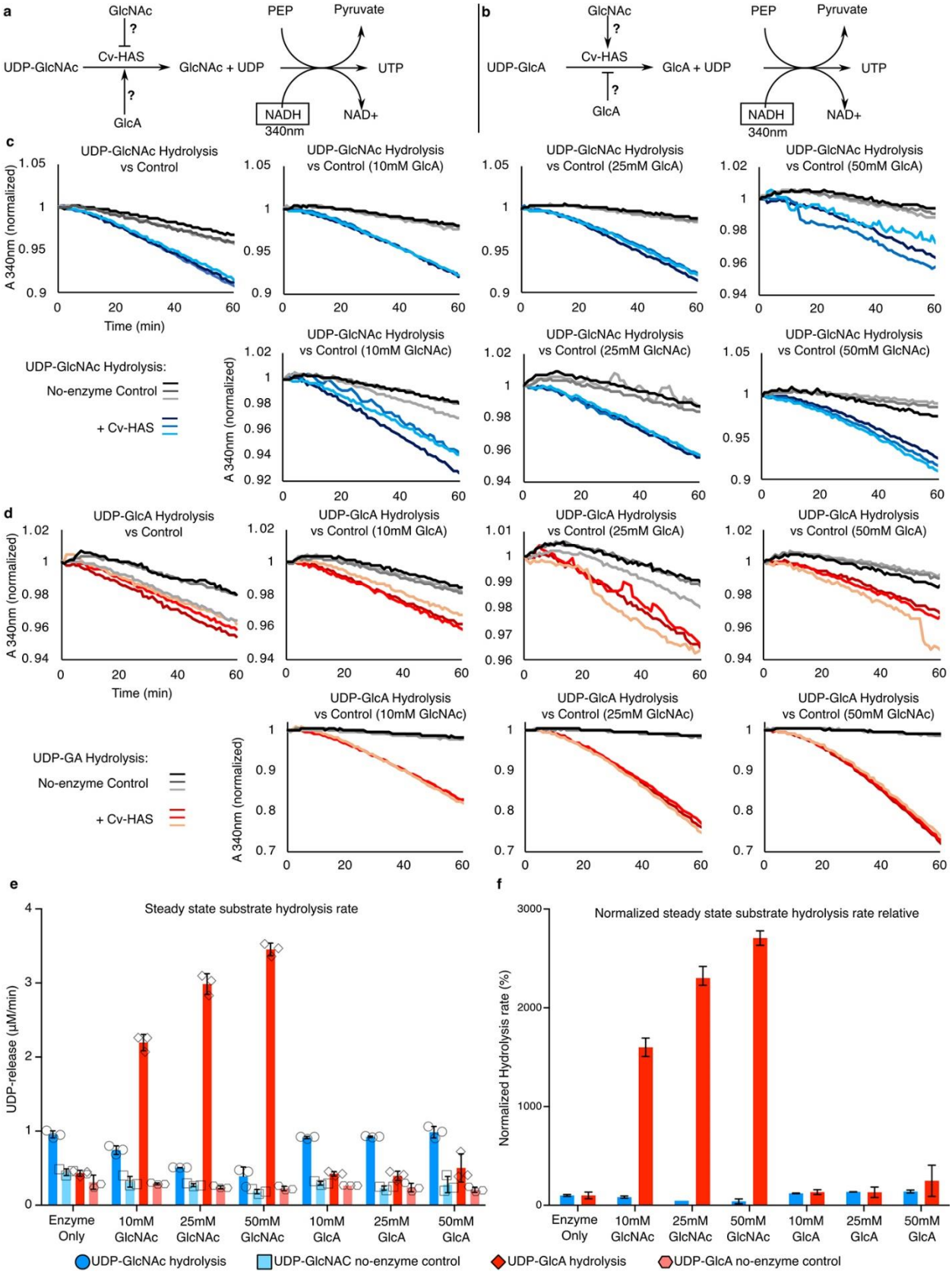
Fig. S2-8



Supplemental Figure 2-8 | GlcNAc priming of HA biosynthesis.
Shown is an autoradiogram of 14C-labelled HA after SDS-PAGE. The experiment has been repeated at least 3 times with essentially identical results.

`

Fig. S2-9

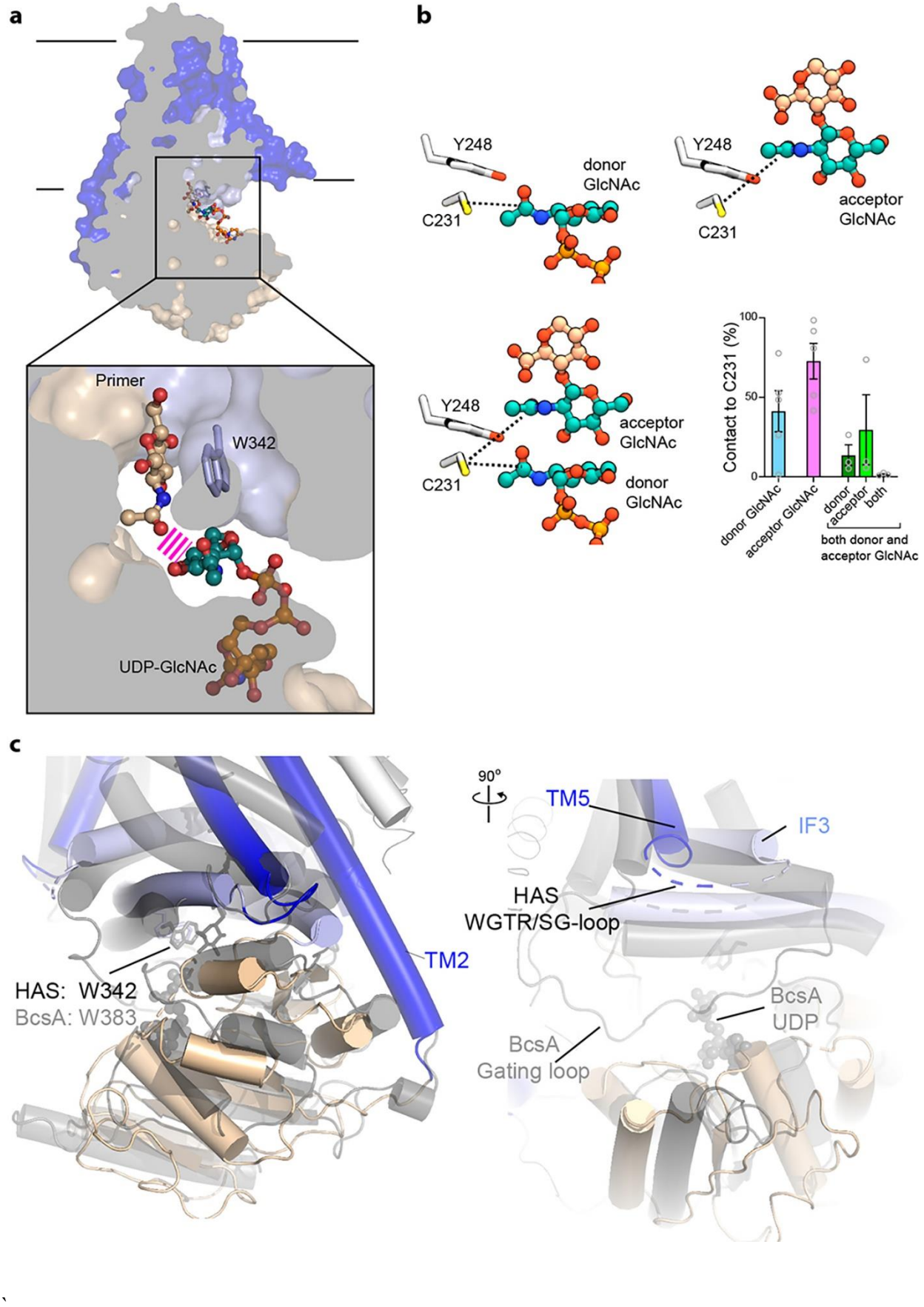

a
GlcNAc
?
Cv-HAS
UDP-GlcNAc
GlcNAc + UDP
?
GlcA
PEP
Pyruvate
UTP
NADH
340nm
NAD+
b
GlcNAc
?
Cv-HAS
UDP-GlcA
GlcA + UDP
?
GlcA
PEP
Pyruvate
UTP
NADH
340nm
NAD+
c
UDP-GlcNAc Hydrolysis vs Control
UDP-GlcNAc Hydrolysis vs Control (10mM GlcA)
UDP-GlcNAc Hydrolysis vs Control (25mM GlcA)
UDP-GlcNAc Hydrolysis vs Control (50mM GlcA)
A 340nm (normalized)
Time (min)
UDP-GlcNAc Hydrolysis:
No-enzyme Control
+ Cv-HAS
UDP-GlcNAc Hydrolysis vs Control (10mM GlcNAc)
UDP-GlcNAc Hydrolysis vs Control (25mM GlcNAc)
UDP-GlcNAc Hydrolysis vs Control (50mM GlcNAc)
d
UDP-GlcA Hydrolysis vs Control
UDP-GlcA Hydrolysis vs Control (10mM GlcA)
UDP-GlcA Hydrolysis vs Control (25mM GlcA)
UDP-GlcA Hydrolysis vs Control (50mM GlcA)
A 340nm (normalized)
Time (min)
UDP-GA Hydrolysis:
No-enzyme Control
+ Cv-HAS
UDP-GlcA Hydrolysis vs Control (10mM GlcNAc)
UDP-GlcA Hydrolysis vs Control (25mM GlcNAc)
UDP-GlcA Hydrolysis vs Control (50mM GlcNAc)
e
Steady state substrate hydrolysis rate
UDP-release (μM/min)
Enzyme Only
10mM GlcNAc
25mM GlcNAc
50mM GlcNAc
10mM GlcA
25mM GlcA
50mM GlcA
f
Normalized steady state substrate hydrolysis rate relative
Normalized Hydrolysis rate (%)
Enzyme Only
10mM GlcNAc
25mM GlcNAc
50mM GlcNAc
10mM GlcA
25mM GlcA
50mM GlcA
UDP-GlcNAc hydrolysis
UDP-GlcNAC no-enzyme control
UDP-GlcA hydrolysis
UDP-GlcA no-enzyme control

Supplemental Figure 2-9 | Effect of monosaccharides on substrate hydrolysis.
(**a** and **b**) Reaction schemes for UDP-GlcA and UDP-GlcNAc hydrolysis. (**c** and **d**) Raw absorbance measurements. (**e**) Quantification of hydrolysis rates in the presence of increasing monosaccharide concentrations. Blue and Red: Hydrolysis of UDP-GlcNAc and UDP-GlcA, respectively. Light and dark colours represent control reactions in the absence of enzyme. Right panel: Background subtracted hydrolysis rates. Error bars represent deviations from the means with n = 3 independent experiments.

Supplemental Figure 2-10 | Likely mechanism of alternating substrate polymerization and comparison with cellulose synthase. (**a**) Superimposition of substrate-bound and primed Cv-HAS structures. The close distance between the primer and donor sugar is indicated by grey bars. (**b**) Contact likelihood between C231 and GlcNAc for the systems in (**a)** over the last 125 ns of each simulation. In the case of GlcNAc being in both donor and acceptor positions, both GlcNAc units are less likely to bind C231, and exhibit very high variance regarding binding poses. Of particular note, the chance of both GlcNAc being in the C231 pocket at the same time is very low (ca. 1.5 ± 0.8%). (**c**) Cv-HAS is superimposed with the *Rhodobacter sphaeroides* (Rs) BcsA-B complex (PDB: 4P00) based on secondary structure matching. Rs-BcsA-B is coloured grey and Cv-HAS is coloured blue and green for its TM and GT domains. The cellulose polymer associated with Rs-BcsA-B is shown as black sticks. On the next page.

`

Fig. S2-10


a
Primer
W342
UDP-GlcNAc
b
Y248
donor GlcNAc
C231
Y248
acceptor GlcNAc
C231
Y248
acceptor GlcNAc
C231
donor GlcNAc
Contact to C231 (%)
100
50
0
donor GlcNAc
acceptor GlcNAc
donor
acceptor
both
both donor and acceptor GlcNAc
c
HAS: W342
BcsA: W383
TM2
90°
TM5
IF3
HAS
WGTR/SG-loop
BcsA
UDP
BcsA
Gating loop

3. Chapter 3: Molecular Insights into Capsular Polysaccharide Secretion via ABC Transporters

Jeremi Kuklewicz[1] and Jochen Zimmer[1,2]

[1]Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia, USA

[2]Howard Hughes Medical Institute

*Corresponding author: jz3x@virginia.edu

In this chapter, I will present an unpublished article that forms the foundation of our understanding of how gram-negative CPS translocation is mediated by the ABC transporter KpsMT. I identified a CPS secretion system suitable for structural analyses and characterized the KpsMT-KpsE complex in six different states. The main advances include revealing a KpsE-encased ABC transporter, identifying conformational rearrangements of KpsMT during ATP hydrolysis, and delineating substrate recognition within a lipid-exposed electropositive canyon of the ABC transporter. My structural findings were validated by extensive *in vivo* encapsulation experiments in an engineered *E. coli* strain.

The wealth of data presented below has been accumulated since I joined the lab in June 2017 and will soon be submitted for peer review.

`

3.1 Introduction

Bacteria commonly display complex carbohydrates on their surfaces to fortify cell boundaries [1]. In gram-negative bacteria, capsular polysaccharides (CPS) are lipid-linked high molecular weight biopolymers that form protective capsules, serve as adhesives, or camouflage pathogens inside their hosts [2]. Therefore, CPS are potent virulence factors and targets for the development of novel antimicrobials.

CPS are predominantly linear, acidic polysaccharides of several hundred sugar units. Their biosynthesis follows two different pathways. In the Wzy-dependent pathway, short lipid-linked oligosaccharides are polymerized in the periplasm by Wzy with the help of the polysaccharide co-polymerase (PCP) Wzc [1]. During assembly, the CPS is exported across the outer membrane (OM) through the barrel shaped Wza octamer [2,3].

In contrast, the ABC transporter-dependent biosynthesis pathway relies on complete CPS biosynthesis on the cytosolic side of the inner membrane (IM), followed by export across the cell envelope [1,2]. CPS secretion is mediated by an ABC transporter containing KpsT and KpsM as its nucleotide-binding (NBD) and transmembrane (TM) domains, respectively [4]. The transporter partners with the periplasmic and IM anchored KpsE subunit, a class-3 PCP, as well as KpsD, assumed to form an OM pore.

CPS are synthesized on a phosphatidylglycerol (PG) lipid anchor that is usually extended by 5-9 Kdo (3-deoxy-D-manno-oct-2-ulosonic acid) sugars by the KpsS and KpsC enzymes (**Fig. 3-1 a-c)** [5-8]. The Kdo glycolipid then serves as an adaptor for CPS biosynthesis either by a single bifunctional glycosyltransferase (GT) or an ensemble of GTs, depending on the serotype.

`

Following or concomitant to biosynthesis, the CPS is recognized by KpsMT and translocated through a trans-envelope conduit consisting of KpsMT, KpsE and KpsD (**Fig. 3-1c**).

KpsMT belongs to class-5 of ABC transporters with structural homology to the O-antigen and teichoic acid transporters that recognize undecaprenyl-linked glyco-polymers [9-11]. *In vivo* substitution experiments indicated that KpsMT has limited specificity towards the CPS structure [12-14], suggesting that the conserved Kdo glycolipid moiety mediates substrate recognition.

We delineate how gram-negative CPS are recognized and secreted across the cell envelope by combining biochemical analyses with super-resolution CPS imaging and single particle cryogenic electron microscopy (cryoEM). First, genetic engineering of *E. coli* with CPS components from *Pasteurella multocida* reveals capsule formation with CPS secretion 'hotspots'. Second, cryoEM analyses of a complex of KpsMT and KpsE in apo and ATP bound states provide detailed insights into the association of the ABC transporter with a periplasmic octameric KpsE 'cage'. Third, a substrate-bound state of the KpsMT-KpsE complex reveals how KpsM recognizes the glycolipid in an electropositive canyon and identifies rigid body movements within the TM region to accommodate the substrate. Fourth, site directed mutagenesis and *in vivo* functional analyses support a model by which KpsMT binds and translocates its glycolipid substrate via conserved Arg residues at the interface of the KpsM subunits.

3.2 Results

3.2.1 The CPS secretion system exhibits broad substrate specificity

To analyze CPS biosynthesis structurally and functionally, we reconstituted CPS formation in an acapsular off-the-shelf *E. coli* laboratory strain. *E. coli* C43 (DE3) is acapsular because it lacks the

`

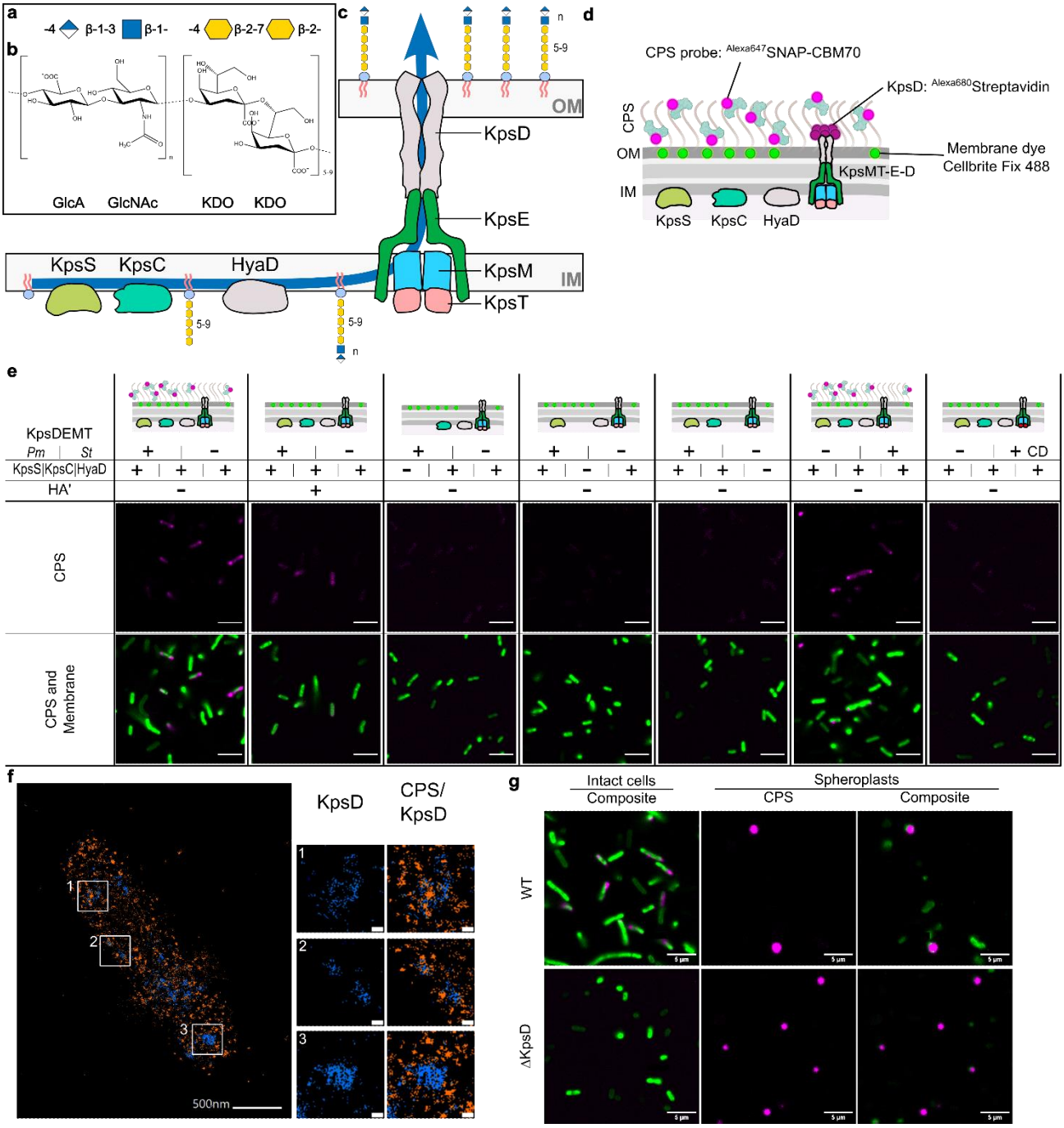


Fig. 3-1

functional CPS polymerase, all other CPS components are present [15]. Therefore, the endogenous group 1 CPS gene cluster was removed from this strain by CRISPR Cas9/Lambda Red mediated

Figure 3-1 | E. coli engineering for CPS encapsulation.
(a) and **(b)** Hyaluronan disaccharide and Kdo repeat units with their symbolic representations (blue square: GlcNAc, blue/white diamond: GlcA, yellow hexagon Kdo). **(c)** Schematic of the type-2 CPS biosynthesis pathway with its crucial components labeled. HyaD is the Pm hyaluronan synthase. **(d)** Diagram of the *in vivo* encapsulation assay and labeling strategy. **(e)** Confocal images of E. coli C43ΔCPS1 expressing different CPS components as indicated in the top panels. Membranes were stained with Cellbrite Fix 488 and CPS with ^{Alexa647}SNAP-CBM70. HA': hyaluronidase digestion, CD: 'catalytically dead' (Walker B) KpsMT$_{E151Q}$. *St* KpsE used in this assay contains the engineered disulfide bridge (L77C, S138C). **(f)** Representative dual color Minflux nanoscopy of an encapsulated cell. CPS (orange) is labeled with ^{Alexa647}SNAP-CBM70, KpsD (blue) is labeled with ^{Alexa680}streptavidin. Panels 1-3 are examples of CPS/KpsD hotspots (scalebar: 50 nm). **(g)** Confocal images of inner membrane linked CPS. Spheroplasts were prepared from encapsulated (WT, top row) and acapsular but CPS expressing cells in the absence of KpsD (ΔKpsD, bottom row). Membranes are stained as in panel e.

recombineering (see Methods) to avoid functional overlap with recombinantly expressed CPS genes, giving rise to the *E. coli* C43ΔCPS1 strain (**Fig. S3-1a, S3-2 and S3-3**). Next, we expressed the CPS biosynthetic and secretion components from *Pasteurella multocida* (*Pm*), a gram-negative pathogen [16], in the *E. coli* C43ΔCPS1 strain under the control of inducible T7 promoters, (**Fig. S3-1b**) and Methods. *P. multocida* produces a hyaluronan (HA) capsule [17], consisting of linear polysaccharides of alternating N-acetylglucosamine and glucuronic acid units, Fig. 1a and b. HA can be detected on the cell surface using a carbohydrate binding module (CBM70) [18] fused to a SNAP-tag charged with an Alexa^{647} fluorophore, (**Fig. 3-1 d-g and Fig. S3-1c**) and Methods.

In the presence of all *Pm* CPS components, the *E. coli* surface can be labeled with the HA-specific probe for confocal fluorescence microscopy (**Fig. 3-1e and Fig. S3-4**). Prior treatment with a hyaluronidase substantially reduces labeling, indicating that the CBM70 probe indeed detects surface exposed HA. Most cells exhibit distinct HA puncta frequently but not exclusively localized to the poles. A similar pattern of CPS distribution has been described previously for *E. coli* K1 and related capsules [19].

`

Control conditions in the absence of KpsS or KpsC, implicated in the synthesis of the Kdo-linked glycolipid anchor, do not detect extracellular HA above noise levels (**Fig. 1c and e**). Two additional *Pm* CPS subunits of unknown function, HyaB and HyaE, also influence encapsulation. No capsule was detected in absence of HyaB, while the capsule produced in the absence of HyaE has an aberrant morphology with a seemingly more diffuse CPS signal (**Fig. S3-1d and S3-4**). As expected, HA is not detected in the absence of HyaD, the bifunctional HA synthase (**Fig. 3-1c and e**) [17].

As described in detail below, the CPS secretion system from the gram-negative bacterium *Schlegelella thermodepolymerans* (*St*) is particularly suitable for structural analyses. Substituting the *Pm* CPS secretion machinery (consisting of KpsMT, KpsE and KpsD) with the corresponding subunits from *St* in the engineered *E. coli* C43ΔCPS1 strain enables HA secretion (**Fig. 3-1e**). Under the same conditions, rendering *St* KpsT catalytically inactive by introducing the Walker B mutation into its NBD abolishes HA detection, demonstrating that HA encapsulation depends on a functional *St* CPS secretion system (**Fig. 3-1e**).

Lastly, high resolution MINFLUX nanoscopy [20] provides insights into the capsule organization and distribution of secretion components (**Fig. 3-1f and Fig. S3-1 e-i**). The capsule produced from *Pm* components was imaged using the above described CBM70 probe. With the resolution provided by Minflux nanoscopy, we resolved CPS surface distributions ranging from clusters to uniform coats, (**Fig. 3-1f, Fig. S3-1h, and S3-8 - S3-11**), similar to those previously described [19], thereby further validating our heterologous encapsulation system.

`

To localize CPS together with the OM secretion pore KpsD, an engineered C-terminal Strep-tag of KpsD was detected using Alexa^{680}-conjugated streptavidin, (**Fig. S3-1 e-g**), enabling dual color Minflux nanoscopy [21]. An AlphaFold2 [22] predicted model of octameric KpsD, (**Fig. S3-1e, f and Fig. S3-5**), suggests that only 1 or 2 streptavidin tetramers can bind the KpsD complex simultaneously. Accordingly, a single KpsD octamer is expected to give rise to a fluorophore localization cloud of about 14 nm in diameter (**Fig. S3-1e and f**).

Combining the CPS and KpsD labeling strategies, we performed two-dimensional dual color Minflux localizations on the engineered *E. coli* cells (see Methods). KpsD is clustered into distinct 'hotspots' ranging from ~20 to 50 nm in diameter (**Fig. 3-1f, Fig. S3-1h, and S3-8 - S3-11**). No labeling was observed in the absence of KpsD or in the presence of untagged KpsD, suggesting that streptavidin specifically recognizes Strep-tagged KpsD (**Fig. S3-6**). The HA signal, in contrast, is more uniformly distributed across the cell surface (**Fig. 3-1f and Fig. S3-10,11**). Some KpsD clusters are proximal to CPS, suggesting that these are indeed HA secretion sites. Although the number of KpsD clusters cannot be determined with the current labeling strategy, we estimate fewer than 100 clusters per cell, based on their size and the observed KpsD localizations. KpsD clustering did not change in the absence of CPS or KpsE, suggesting that KpsD oligomers likely form prior to secretion system assembly and/or CPS engagement (**Fig. S3-12, 13**).

3.2.2 The inner membrane-integrated transporter components are sufficient for CPS secretion

As discussed in detail below, structural analyses of the KpsMT-KpsE complex bound to a putative substrate molecule suggest that KpsD is unnecessary for substrate translocation across the IM. We

`

tested this hypothesis by performing *in vivo* CPS translocation assays in the presence and the absence of KpsD. In the absence of an OM pore, any translocated CPS would accumulate in the periplasm where it can be detected after removing the OM.

Labeling the engineered *E. coli* cells with the HA-specific probe did not reveal any surface exposed HA in the absence of KpsD. However, after removing the OM, several spherically shaped cells exhibited strong HA labeling, suggesting periplasmic accumulation of HA under these conditions (**Fig. 3-1g and Fig. S3-14**). Similar results have been obtained previously for *E. coli* K5 CPS [23]. As also observed previously [24], cells expressing all CPS components and thus displaying CPS on their surfaces are elongated, sometimes appearing as tubes. This is not the case in the absence of KpsD. Accordingly, the obtained spheroplasts of the KpsD-lacking cells are smaller in diameter compared to the ones obtained from elongated cells (**Fig. 3-1g**). This suggests that, under the experimental conditions, CPS surface exposure impacts cell division.

3.2.3 The CPS ABC transporter is surrounded by a KpsE octamer

Structural analyses of the CPS secretion system were performed using *St* components due to complex stability in non-denaturing detergents. The inner membrane associated KpsM, KpsT, and KpsE subunits were expressed and purified for cryoEM analysis as described in the Methods. To stabilize the KpsE cage, an intermolecular disulfide bond was engineered by introducing cysteines at positions 77 and 138 (L77C, S138C) based on an AlphaFold2 predicted KpsE dimer model and 'Disulfide by Design 2.0' [25] (see Methods) (**Fig. S3-15a**). The obtained KpsMT-KpsE complex is catalytically active *in vivo* and *in vitro* and hydrolyzes ATP at a rate of about 40 nmol/(min mg) (**Fig. 3-1f and Fig. S3-15c**). As described in detail in the next section, we determined several

`

KpsMT-KpsE complex structures in different nucleotide bound states (**Fig. S3-16 - S3-18, and Table 3**). The most complete KpsE assembly described hereafter was obtained with an ATP-bound KpsMT transporter.

The KpsMT transporter is surrounded by a cage of eight KpsE subunits that extends about 80 Å into the periplasm (**Fig. 3-2 and Fig. S3-15f and g**). Each KpsM subunit interacts with four KpsE protomers, hence the KpsMT-KpsE complex is a dimer of a KpsMT half transporter associated with four KpsE subunits. Starting within the membrane-embedded region, the KpsE cage can be divided into TM, dome, and crown regions (**Fig. 3-2e and f**). Most intermolecular interactions occur within the dome that forms a tight periplasmic ring surrounding the ABC transporter.

KpsE is anchored to the IM via N- and C-terminal TM helices (**Fig. 3-2 b-e**). The helices interact through Leu, Ile, and Val rich surfaces, with the N-terminal helix forming the outer and the C-terminal helix the inner ring of a membrane-integrated 'carousel' (**Fig. 3-2d, e, and i**). The following dome region contains a 4-stranded twisted β-sheet (β1-β4), four short α-helices (α1-α4), as well as the N-terminal segment of helix α5, which continues into the crown (**Fig. 3-2e and Fig. S3-15f**). While most of the dome region is formed from residues directly following TM helix 1 (residues 34-172), the last β-strand (β4), residues 322-329, precedes the C-terminal TM helix.

`

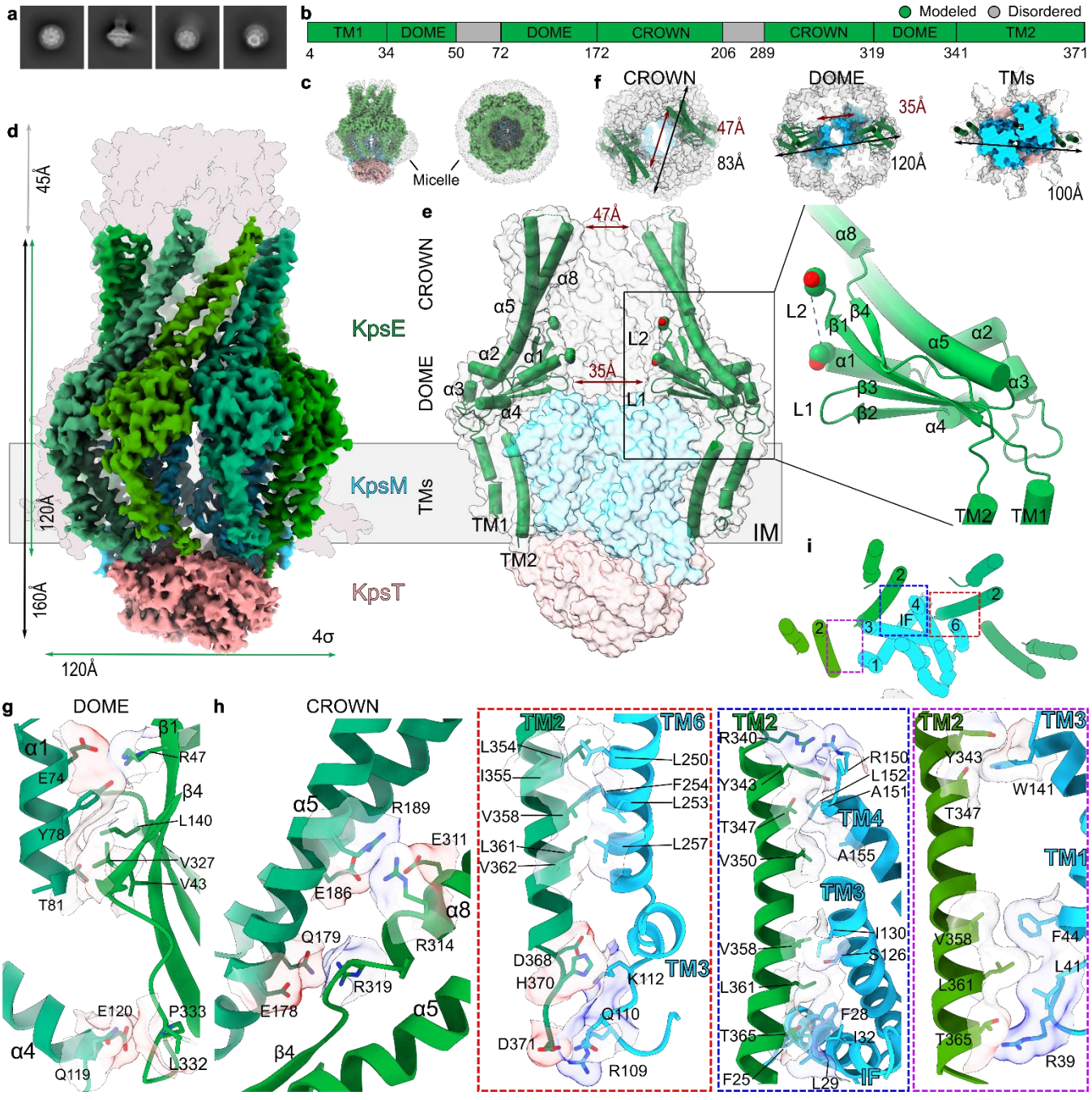


Figure 3-2 | The structure of the KpsMT-KpsE complex.
(a) Representative 2D class averages of KpsMT$_{E151Q}$-E. **(b)** Domain organization and model completeness of KpsE. **(c)** CryoEM map of KpsEMT$_{E151Q}$-E in an LMNG micelle shown. KpsE: green, KpsM: blue, KpsT: salmon. **(d)** Map of ATP-bound KpsMT$_{E151Q}$-E overlayed with a surface representation of an AlphaFold2 predicted octameric full length KpsE model (colored gray). Map contoured at 4σ level. **(e)** Cartoon representation of KpsE. **(f)** Section views of the KpsE octamer with inner and outer diameters. **(g)** and **(h)** Interactions between neighboring KpsE protomers. **(i)** Interactions between KpsE and KpsM.

Two loops extend from the β-sheet into the octamer lumen (L1 and L2), of which the longer L2 loop is disordered (residues 50-72), (**Fig. 3-2b, e and Fig. S3-15i**). Based on an AlphaFoldF2 model, L2 is predicted (with low confidence) to point away from the IM into the KpsE tunnel (**Fig. S3-15f, g and S3-5**). The crown is formed by a helical hairpin consisting of helix α5 extending from the dome and helix α8 (residues 291-319). The helices are connected by a disordered region (residues 205-290) that is predicted to form another helical hairpin, most likely to interact with KpsD, (**Fig. 3-2d, e, Fig. S3-15f and g**). The interior of the KpsE crown is hydrophilic (**Fig. S3-15h**). Starting at the cytosolic side of the membrane, the KpsE cage is about 120 Å long (**Fig. 3-2d**). Including the unresolved region extending the crown helices and based on an AlphaFold2 model, (**Fig. S3-15f and g**), the assembled cage would extend by about 120 Å from the IM's periplasmic surface into the periplasm (**Fig. 3-2d**). The dome region cylindrically extends the ABC transporter with outer and inner diameters of approximately 120 and 35 Å, respectively. The inner tunnel widens to about 47 Å within the crown, which has an outer diameter of about 83 Å (**Fig. 3-2f**).

3.2.4 Stabilization of the KpsE cage and interactions with KpsMT

Most of the cage-stabilizing interactions occur within the dome via its β-sheet (**Fig. 3-2g**). Helices α1 and α4 of a neighboring subunit pack against the twisted β-sheet surface. Within α1, Glu74, Tyr78, and Thr81 contact the sheet. Glu74 and Tyr78 are in close proximity to Arg47 of β1, and the methyl group of Thr81 packs into a hydrophobic pocket formed from Val43 of β1, Leu140 of β3, and Val327 of β4. The N-terminus of α4 contacts the loop connecting β4 with TM helix 2 in

`

the neighboring subunit via Gln119 and Glu120. These residues interact with the backbone carbonyl and amide groups of Leu332 and Pro333 in the β4-TM2 loop, respectively (**Fig. 3-2g**). The interprotomer interactions within the crown are less pronounced and confined to the dome-proximal region. Here, Glu178 and Gln179 of α5 are in close proximity to Arg319 at the C-terminal end of α8 in the neighboring subunit. Similar interprotomer interactions exist between Glu186 and Arg189 of α5 and Arg314 and Glu311 of α8, (**Fig. 3-2h**).

KpsMT contacts the KpsE cage within the membrane and the periplasm (**Fig. 3-2d, e, f and i**). The contacts are predominantly hydrophobic within the membrane segment and mediated by charged or polar residues within the membrane flanking regions (**Fig. 3-2i**). Each KpsM contacts three KpsE subunits. First, KpsM's TM helices 3 and 6 together with the N-terminal segment of its interface helix (IF, explained below) form a hydrophobic pocket that accommodates the end of KpsE's C-terminal TM helix. The C-terminal helix of the next KpsE protomer (counted counterclockwise viewed from the periplasm), fits into a groove formed by the C-terminal region of KpsM's IF helix, the central portion of its TM helix 3, as well as the N-terminus of TM helix 4 (**Fig. 3-2i**). TM helix 2 of the third KpsE subunit interacts with the C- and N-terminal segments of KpsM's TM helices 3 and 1 on the membrane's periplasmic and cytosolic sides, respectively (**Fig. 3-2i**).

3.2.5 KpsMT is a type 5 ABC transporter

We determined cryoEM structures of the KpsMT-KpsE complex in two nucleotide-free states (Apo 1 and Apo 2), bound to ADP:AlF_4^- to mimic an ATP post-hydrolysis state, as well as in an ATP-complexed conformation after introducing the Walker B mutation (E151Q) into KpsT's active site

`

(**Fig. 3-3 and Fig. S3-19**). Overall, the KpsMT transporter shows structural homology to the O-antigen (WzmWzt) and teichoic acid (TarGH) ABC transporters, as well as ABCG1 and PmtCD (**Fig. S3-19c**)[10,11,26,27]. However, in contrast to WzmWzt and TarGH, KpsT lacks the gate helix at

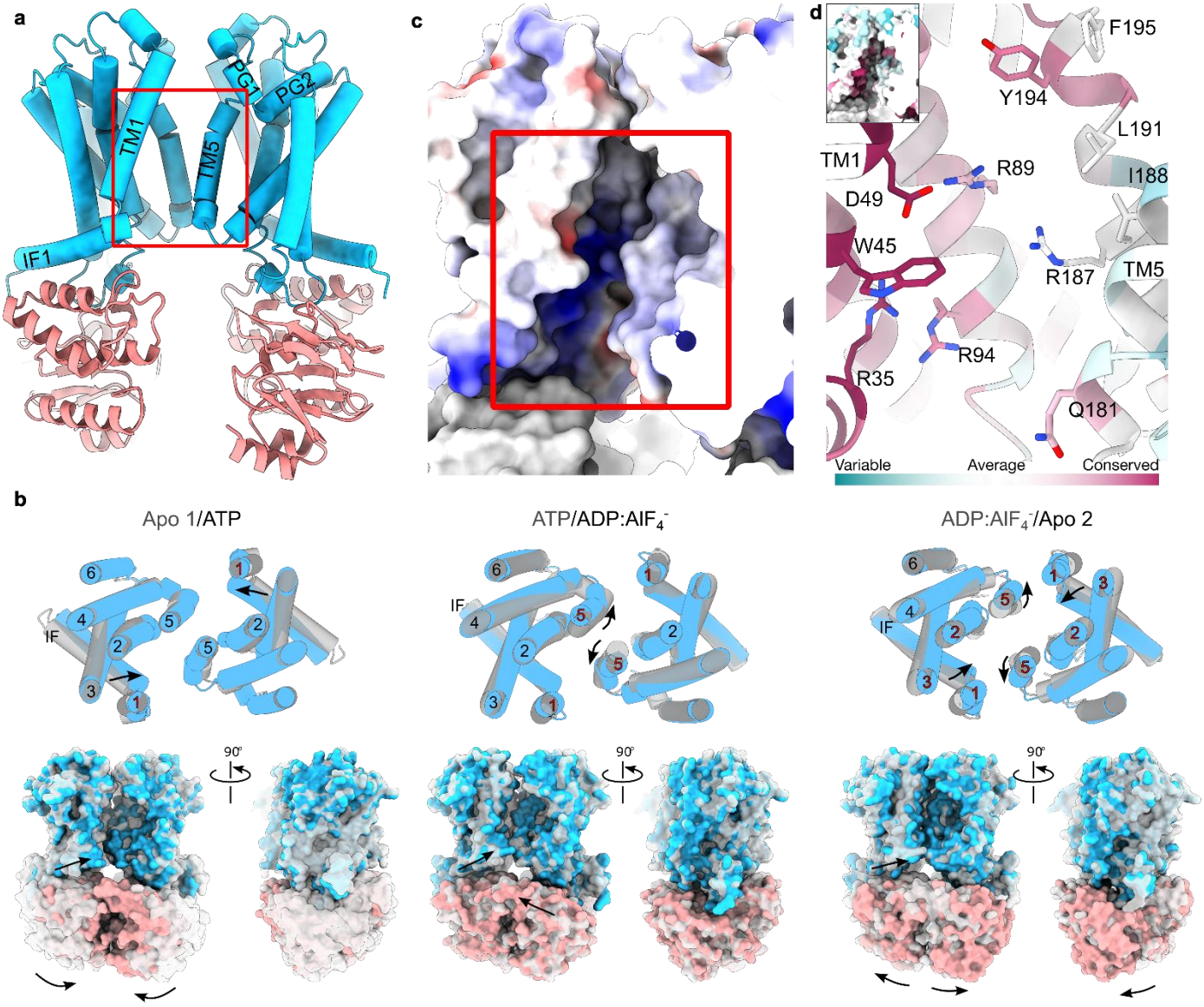


Figure 3-3 | Structure and conformational changes of the KpsMT ABC transporter.
(a) KpsMT Apo 1 structure. KpsM is shown as cylindrical helices in blue, KpsT as a cartoon in salmon. PG: periplasmic gate helix, IF: interface helix. **(b)** Conformational changes and rigid body movements of KpsMT when transitioning between nucleotide bound states (grey to colored). Helices indicated by red font color frame the TM canyon and modulate its opening. **(c)** Surface representation of coulombic electrostatic potential, red: -10, blue:10 kcal/(mol·*e*) of the polysaccharide canyon of KpsMT (red box). **(d)** Evolutionary conservation (ConSurf [48]), of the residues forming the polysaccharide canyon. Inset: surface representation of evolutionary conservation of the polysaccharide canyon of KpsMT.

`

the interface with KpsM, which has been suggested to recognize undecaprenyldiphosphate-linked substrates. The KpsMT substrate is anchored to a PG phospholipid instead [6].

KpsM starts with an N-terminal amphipathic IF helix that runs roughly parallel to the cytosolic water-lipid interface, followed by six proper TM helices, (**Fig. 3-3a, b and Fig. S3-19a**). In the complete transporter, the two KpsM subunits are separated from each other with minimal contacts between TM helices 5. Similar arrangements have been observed for structurally related family-5 lipid ABC transporters, including ABCG1 and ABCG5/8 [26,28].

3.2.6 Rigid body movements of KpsMT

The conformational transitions relating the nucleotide-free, ATP, and ADP:AlF_4^-bound KpsMT states can be described as rigid body movements, (**Fig. 3-3b**). No major structural rearrangements were observed within the KpsE cage.

In a nucleotide-free state, our cryoEM analyses resolve two apo conformations of KpsMT (**Fig. 3-3a, b and Fig. S3-19a, b**). Apo 1 displays well separated NBDs and only minimal interactions between the KpsM subunits via TM helix 5. In this conformation, the gap between TM helices 1 and 5 of opposing KpsM protomers creates a deep electropositive canyon that extends from the NBD into the TM segment. The canyon ends near the periplasmic membrane surface at the interface between TM helix 1 and a short periplasmic reentrant helix (PG1) following TM helix 5 of the opposing subunit (**Fig. 3-3c, d and Fig. S3-19b**).

`

As expected, in the presence of ATP, the NBDs are closely associated, with an Mg:ATP complex bound at each nucleotide binding site (**Fig. 3-3b, Fig. S3-17 and S3-19**). The conformation of the KpsM pair in the ATP-bound state resembles the Apo 1 conformation with an open electropositive canyon (**Fig. S3-19a, b**). A notable difference, however, exists at the cytosolic entrance to the canyon. Here, the IF helix and the N-terminal segment of TM helix 1 shift by about 5 Å towards TM helix 5 of the opposing KpsM subunit, thereby narrowing the canyon (**Fig. 3-3b**).

Transitioning from the ATP to the ADP:AlF_4^--bound KpsMT conformation includes an approximately 5-degree rigid body clockwise rotation of the KpsM subunits around an axis running through the center of TM helices 2-4 of each subunit (**Fig. 3-3b and Fig. S3-19a**). This movement narrows the distance between the opposing TM helices 5 while also widening the interprotomer gap between TM helix 1 and 5.

In the Apo 2 state, the NBDs are closer together and slightly shifted against each other (**Fig. 3-3b and Fig. S3-19a**). A similar arrangement has been observed in the crystal structure of the isolated Wzt NBD of the O antigen ABC transporter [10]. Accordingly, the KpsM subunits move towards each other by about 12 Å as compared to the Apo 1 state, thereby decreasing the separation of the membrane-embedded subunits and narrowing the electropositive canyon (**Fig. S3-19b**).

3.2.7 The KpsM canyon binds a putative glycolipid substrate

The *in vivo* CPS secretion experiments described above, (**Fig. 3-1d**), demonstrate that the *St* ABC transporter recognizes and translocates a lipid-linked HA polymer formed by the *Pm* biosynthetic CPS components. We purified this CPS substrate in a detergent solubilized state from the

`

membrane fraction of the engineered *E. coli* using an affinity column of immobilized CBM70 (see Methods). The obtained glycolipid displays electrophoretic mobility similar to HA standards of about 500 kDa and is degraded by hyaluronidase, confirming the purification of a HA glycolipid of >1000 disaccharide units, (**Fig. S3-20a**). Using the same purification approach, only background levels of HA were obtained from cells lacking KpsS or KpsC. These enzymes produce the poly Kdo linker connecting the PG lipid with the HA polymer, (**Fig. 3-1 a-e and Fig. S3-20j**), suggesting that the isolated glycolipid indeed represents the KpsMT substrate [5].

To determine how KpsMT recognizes the glycolipid, the purified KpsMT-KpsE complex was preincubated with the isolated glycolipid prior to cryo grid preparation (see Methods). CryoEM analyses of this sample provided two additional cryoEM maps similar to the Apo 1 state with poorly defined KpsT subunits. Here, KpsT was placed after focused refinement and rigid body docking (see Methods) (**Fig. S3-16 and S3-18**).

In these structures, the above-described electropositive canyon at the interface between the KpsM subunits is open (**Fig. 3-4**). However, in contrast to all other structures, we identify two states of a lipid molecule bound inside the canyon (referred to as state-1 and state-2) (**Fig. 3-4 a-d and Fig. S3-18 and S3-20**). In state-2, the phosphate group of the lipid is shifted by about 7 Å towards the extracellular side, compared to state-1 (**Fig. 3-4e**). Both lipid densities are connected to a globular density that readily accommodates a Kdo sugar (**Fig. 3-4 a-e and Fig. S3-18 and S3-20**). From this putative first Kdo residue, additional weaker density extends past the interface of the KpsM subunits towards the KpsE carousel (**Fig. S3-20b, c**). While the density could accommodate 2-3 additional Kdo units, the map quality is too weak for assignment. Therefore, we modeled the

`

putative ligand as a PG-linked Kdo monosaccharide. Of note, similar ligand densities are observed at the opposing KpsM dimer interface of the pseudo 2-fold symmetric transporter (**Fig. S3-20b, c**). However, these densities are of lower quality and have not been interpreted.

In state-1 where the putative Kdo sugar is closer to the cytosolic water-lipid interface, the sugar moiety is sandwiched between TM helices 1 and 5 of opposing KpsM protomers (**Fig. 3-4b and Fig. S3-20d**). Although the exact position of the Kdo moiety is speculative, it is surrounded on one side by Trp45 of TM helix 1 and, via the opposite face of the sugar ring, Gln181, Ile185 and Ile188 of TM helix 5 of the other KpsM protomer. The following phosphate group is near Arg94 of TM helix 2 (**Fig. 3-4b**). All of these residues are conserved among CPS ABC transporters.

The lipid's diacylglycerol moiety extends through the canyon towards the extracellular water-lipid interface. One lipid acyl chain snakes through the hydrophobic segment of the canyon where it is surrounded by Phe85 and Phe86 of TM helix 2, as well as Leu191, Tyr194 and Phe195 of TM helix 5 of the same protomer (**Fig. 3-4b**). The second acyl chain points away from the KpsM subunits and does not mediate extensive protein interactions.

In state-2, the lipid is shifted by about 7 Å towards the periplasm, such that the carboxyl group of the putative Kdo sugar contacts Arg35 and Arg94 of the IF helix and TM helix 2, respectively (**Fig. 3-4d, e and Fig. S3-20e**). The carboxyl group likely remains in hydrogen bonding distance to Trp45 and the lipid's phosphate group contacts Arg187 of TM helix 5. Further, Gln181 near the C-terminal end of TM helix 5 in the opposing KpsM subunit is suitably positioned to contact the Kdo

`

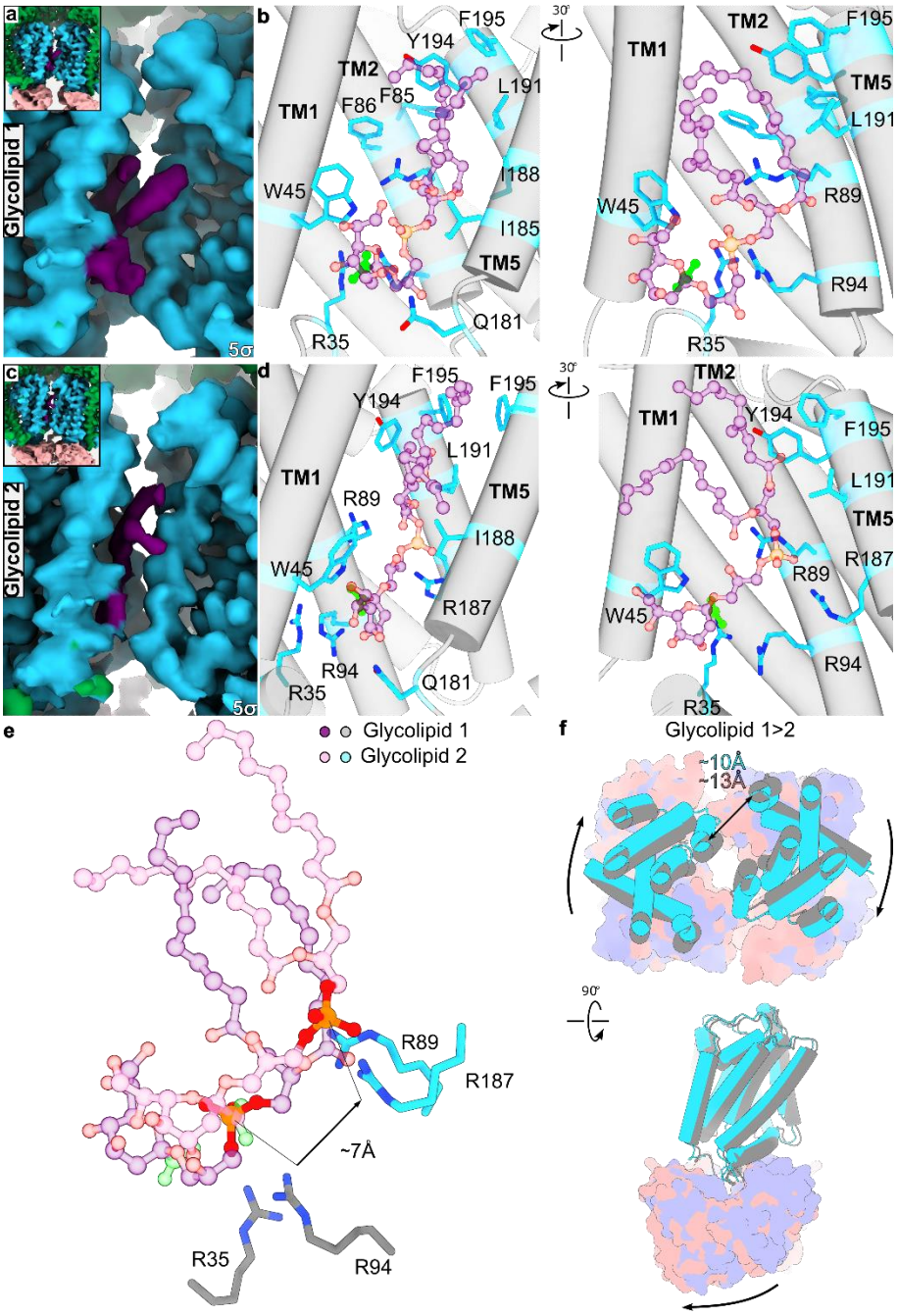


Figure 3-4 | Glycolipid loading into the KpsMT-E complex.
(a) CryoEM map of glycolipid 1 colored magenta. The protein is colored as in Fig 2. **(b)** Residues surrounding glycolipid 1, modeled as a Kdo-PG lipid molecule. Kdo's carboxyl group is highlighted in lime for orientation. **(c)** and **(d)** As panel a and b but for glycolipid 2. **(e)** Overlay of glycolipid 1 and 2 including two pairs of arginine residues likely involved in phosphate coordination. **(f)** Conformational differences between glycolipid 1 (grey/purple) and glycolipid 2 (blue/salmon) bound states of KpsMT.

`

moiety. Relative to the estimated cytosolic and periplasmic membrane boundaries, the glycolipid sits about halfway across the membrane, in agreement with it representing a translocation substrate (**Fig. 3-4c and d**).

The second lipid binding pose further correlates with an approximately 6 Å rigid body translation of the KpsM subunits against each other, thereby widening the space between the KpsM protomers, and narrowing the entry to the polysaccharide canyon by ~3 Å, (**Fig. 3-4f**). Extending this KpsM movement could create a channel forming state as observed for the O antigen WzmWzt transporter[29].

To test the functional relevance of the observed protein-glycolipid interactions, we altered canyon-lining residues in *St* KpsM (which are also present in *Pm* KpsM) and monitored the ability of the mutated transporter to secrete the HA glycolipid *in vivo* (**Fig. S3-20 f-h and S3-21**). None of the positively charged residues of the canyon (Arg35, Arg89, and Arg94) can be replaced with Lys or Ala without disrupting CPS export. Further, Trp45, juxtaposed with the putative Kdo moiety, cannot be replaced with Phe or Leu, underscoring its importance for CPS secretion. The generated mutants form stable KpsMT-KpsE complexes that can be co-purified, suggesting that the subunits indeed fold and assemble properly (**Fig. S3-20i**). The only exception is an Arg89 to Lys mutation, which impairs complex formation and was therefore excluded.

3.3 Discussion

Several ABC transporters are known to secrete high molecular weight biopolymers, including polypeptides and polysaccharides. In an extended conformation, the polymers exceed the length

`

of the transporters many times, likely requiring a stepwise translocation mechanism. Similar to unfolded proteins secreted by the bacterial Type-1 Secretion System (T1SS), CPS are transported in a single step across the gram-negative cell envelope. In contrast to the T1SS [30], however, a single CPS ABC transporter is surrounded by eight copies of the PCP KpsE that cylindrically extend into the periplasm. This architecture resembles the macrolide efflux system MacAB (**Fig. S3-22a**).

The KpsE structure provides an example of a class-3 PCP associated with an ABC transporter. Class 1 and 2 PCPs are associated with Wzy-dependent polysaccharide biosynthesis pathways and likely control the processivity of the polymerization reaction (**Fig. S3-22b**) [31]. It is conceivable that the unresolved KpsE helices α6 and α7 at the tip of the oligomer interact with an octamer of KpsD. This CPS subunit is predicted by AlphaFold2 to assemble into a barrel-shaped structure with its C-terminal domain likely traversing the OM (**Fig. S3-1 and S3-22a**). The ability to detect KpsD's C-terminal Strep tag on the cell surface supports this model. Thus, a translocating CPS is likely surrounded by the secretion system to avoid mislocalization to the periplasm.

The KpsMT-KpsE substrate is a lipid-linked polysaccharide. Other known ABC transporters involved in lipid transport across the bacterial inner membrane and/or the periplasm include the phospholipid trafficking MlaFEDB and the lipopolysaccharide extracting LptB2FGC complexes (**Fig. S3-22a**)[32,33]. None of these resembles the secretion system-like KpsMT-KpsE or MacAB-TolC systems. In addition, the CPS secretion system provides a unique example of a transporter recognizing a hydrophobic and charged substrate.

`

The KpsMT conformation in apo and nucleotide-bound states reveals neither a TM channel nor intracellular or extracellular funnels that could accommodate the substrate polymer. Instead, the membrane integrated KpsM subunits form a narrow lipid exposed canyon that may allow polysaccharide translocation. By analogy with other characterized CPS secretion systems, the translocation substrate contains a conserved core consisting of a PG lipid linked to a Kdo oligosaccharide, followed by a strain and/or serotype specific polysaccharide chain. Because CPS ABC transporters are interchangeable across CPS secretion systems, it is likely that the transporters recognize the substrate's conserved lipid and/or polyanionic oligosaccharide core to initiate secretion. The conserved electropositive canyon at the KpsM interface likely serves this purpose.

Indeed, the isolated HA glycolipid binds to this location. Although the entire substrate molecule is not resolved, the polysaccharide chain likely extends from the transporter towards the cytosolic segments of the KpsE octamer, right above the interface of the KpsT subunits (Fig. 3-5). As observed *in vitro*, CPS secretion could start with spontaneous substrate binding to the KpsM canyon, facilitated by electrostatic interactions and shape complementarity (**Fig. 3-5 and Fig. S3-23**). Migration of the substrate through the KpsM canyon, as indicated by the lipid bound states 1 and 2, could be facilitated by ATP hydrolysis or occur without additional energy input. It is possible that additional movements of the KpsM subunits against each other upon substrate binding close the canyon and create a TM channel surrounding the translocating CPS (**Fig. 3-5 and Fig. S3-23**). As the substrate further migrates through the transporter, the lipid headgroup could reorient in the membrane and exit the transporter laterally into the periplasmic leaflet of the inner membrane. A likely lateral gate exists between TM helix 1 and the PG1 reentrant helix following TM helix 5 in

`

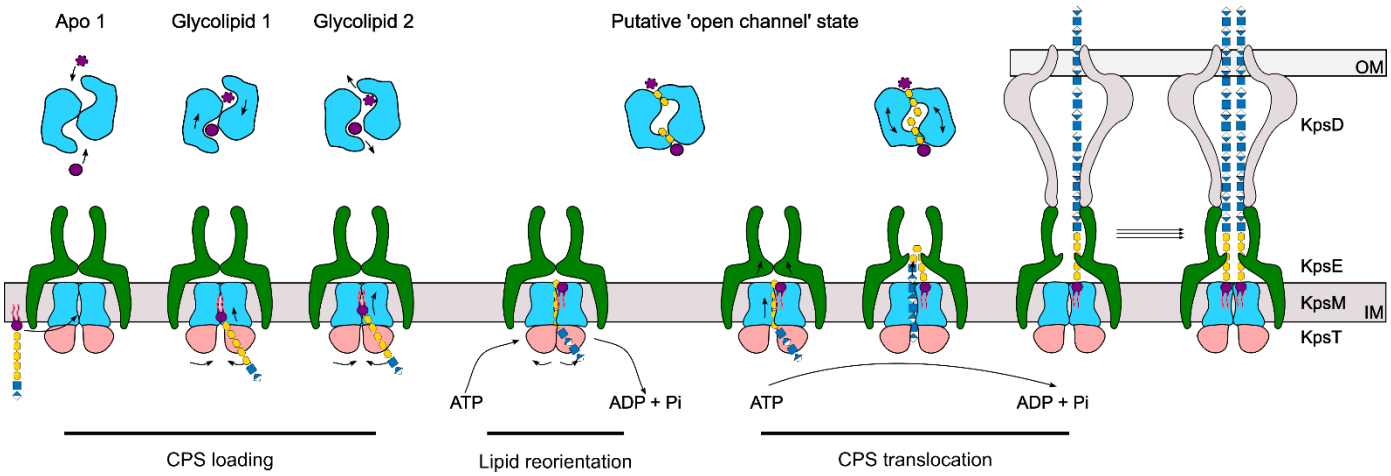


Figure 3-5 | Model of CPS translocation.
CPS secretion likely starts with glycolipid loading into the KpsM canyon halfway across the membrane. Structural flexibility of KpsMT may facilitate different binding poses. NBD closure upon ATP binding may reorient the glycolipid in the IM, thereby loading the CPS into the transporter. Sliding of the KpsM subunits against each other could create a TM channel around the translocating CPS. Consecutive rounds of ATP hydrolysis translocate the CPS across the IM and into the KpsE cage. CPS induced rearrangements within KpsE's crown may favor interactions with KpsD to form a trans-envelope conduit. The substrate's lipid moiety may exit laterally into the IM. Multiple CPS polymers could be secreted by a single secretion system.

the opposing KpsM subunit. In agreement with KpsMT's pseudo-2-fold symmetry, we observe putative substrate densities on either side of the KpsM dimer. It is thus conceivable that two substrate molecules can be translocated simultaneously.

The above-described model also suggests that CPS translocation starts with the lipid moiety and not with the polymer's opposite end, the terminal CPS glycosyl unit. It is currently unknown whether the lipid anchor is extracted from the IM and integrated into the OM, or whether it remains in the inner membrane while the CPS spans the cell envelope. Our structural data do not suggest a mechanism for lipid extraction from the IM's periplasmic leaflet. Instead, it is possible that the lipid dissociates from the transporter after reorientation in the IM, with the following CPS being pushed into the KpsE-KpsD trans-envelope tunnel (**Fig. 3-5**). The volume of the KpsE octamer

`

and the space surrounding the ABC transporter inside the KpsE cage would allow accumulating several CPS molecules. The disordered L2 loop of the KpsE dome extends into the KpsE tunnel, perhaps to form a semipermeable seal. This loop would have to reposition to accommodate CPS inside the KpsE tunnel.

The CPS translocation substrate isolated from the engineered *E. coli* suggest an HA length exceeding 1000 disaccharide repeat units. With an estimated distance of about 10 Å per repeat unit, this correlates with a polymer length of approximately one micron, about half the length of an *E. coli* cell. The patched distribution of KpsD in the OM and uniform CPS surface labeling revealed by dual color Minflux nanoscopy indicate that CPS coverage can be achieved from limited CPS secretion sites, suggesting the spreading of the polysaccharides on the cell surface. Future detailed insights into a complete CPS secretion system during translocation and its cellular dynamics will be necessary to derive a complete model of capsule biogenesis.

3.4 Materials and methods

3.4.1 DNA manipulation

3.4.1.1 Plasmid design – *Pm* CPS genes

All Pm CPS genes were codon optimized for *E. coli* expression and synthesized by Bio Basic Inc., and cloned into pETDuet and pACYCDuet plasmid backbones, giving rise to expression plasmids #5 and #6, pETDuet Hex Secretion System and pACYCDuet Biosynthesis components, respectively. Plasmid #6 encodes KpsC (with an N-terminal Strep-Tag II), KpsS (N-terminally 8xHis tagged), HyaE (N-terminally Myc tagged), HyaB (N-terminally FLAG tagged) and HyaD (N-terminally 8xHis tagged), each under a separate T7 promoter and lac operator. Plasmid #5

`

encodes KpsD (with a C-terminal Strep-Tag II), KpsE (N-terminal S-tag), KpsM, and KpsT (C-terminally 8xHis tagged), each under a separate T7 promoter and lac operator. In both plasmids, each CPS gene is flanked by a unique restriction site. Single gene deletions for in vivo encapsulation assays were achieved by restriction enzyme (NEB) digest and religation using T4 ligase (NEB). Restrictions enzymes used in single gene deletions from plasmid #6 are.: NcoI – KpsC, BamHI – KpsS, SacI – HyaE, EcoRI – HyaB, XhoI – HyaD; and for plasmid #5 are: NdeI – KpsD, BamHI – KpsE.

3.4.1.2 Two plasmid genome editing system

Plasmid design was based on the work of Huang, C., et al. [34], in which efficient *E. coli* genome editing was achieved using a two-plasmid approach: plasmid #*7*- *pCasJZ* and #*8*- *pUC57_region1_tet*. Plasmid #*7* encodes *Sp*Cas9 under an arabinose promoter, the red λ phage proteins Gam, Beta and Exo under T7 promoter, the SacB cassette, and the chloramphenicol resistance cassette. Plasmid #*8* encodes two 500 bp long homology regions (1 and 2), flanked by PAM (1 and 2) and target sequences (1 and 2), two guiding RNA sequences with target sequences (1 and 2) and the tetracycline resistance locus. PAM 1 and 2 and the target sequences 1 and 2 were chosen to be upstream and downstream, respectively, of CPS Region 1 in the *E. coli* C43 genome. Homology regions 1 and 2 are 500 bp sequences upstream and downstream of target sequences 1 and 2.

Plasmid #*7 was generated* using the Gam, Beta, Exo, λ tL3 terminator, SacB, SacB promoter, araBAD promoter and *Sp*Cas9 elements from plasmid pCasPA (addgene #113347) that were cloned by Gibson Assembly (NEB) into MCS-1 of the pACYCDuet plasmid. Next, *araBAD* and

`

araC were cloned in place of the deleted MCSII of that plasmid. The homology regions, guiding RNAs, PAM, and target sequences were synthesized by Gene Universal into the pUC57 vector. Then, using Gibson assembly, the ampicillin resistance cassette was switched for tetracycline resistance cassette from pBR322 plasmid, giving rise to plasmid #*8.*

3.4.1.3 Genome editing

The procedure is described in detail elsewhere [34]. Shortly, plasmids #*7* and #*8* were co-transformed into *E. coli* C43 (DE3). Working concentrations of isopropyl-β-D-thiogalactopyranoside (IPTG), arabinose, chloramphenicol, tetracycline, glucose, and sucrose were: 1 mM, 20 mM, 25 mg/L, 10 mg/L, 1% and 2%, respectively. Cells were plated on LB agar plates, supplemented with chloramphenicol, tetracycline and glucose. Single colonies were used to inoculate 2 mL of LB media supplemented with chloramphenicol, tetracycline and glucose. After 2 h of growth at 37 °C, IPTG was added, and expression of red λ phage proteins was induced. After an additional 1 hour, arabinose was added and the expression of *Sp*Cas9 and transcription of sgRNAs was induced. After 3 hours, cells were plated on LB agar plates containing chloramphenicol, tetracycline and arabinose and incubated overnight at 37 °C. Positive colonies were verified by colony PCR using primers flanking the 500 bp-long homology regions and sequencing (Fig. S3-2 and 3). Then, the positive clone was grown in LB containing sucrose and tetracycline to remove plasmid #*7.* No chloramphenicol-resistant colonies were detected. Genome-edited C43 cells lacking the CPS1 region, termed C43ΔCPS1, were made electrocompetent and used in *in vivo* encapsulation assay.

`

3.4.1.4 Plasmid design – CBM70 and SNAP-CBM70

The codon optimized gene encoding *Streptococcus pneumoniae* Carbohydrate-binding Module (CBM70) [18], was synthesized by Bio Basic Inc. and cloned into the pET30 vector with a C-terminal 10xHis tag, generating plasmid #*9*. Then, using Gibson Assembly, the SNAP-tag from the pSNAP-tag vector (addgene #101135) was N-terminally fused to CBM70 with a GSSMGS linker, creating plasmid #*10*.

3.4.1.5 Plasmid design – *St* KpsDEMT

The codon optimized KpsD, E, M, and T genes from *Schlegelella thermodepolymerans* were synthesized and cloned into expression vectors by Gene Universal, generating plasmids #*1a*, #*3a* and #*4*. Plasmid #*1a* contains KpsM in MCS-I and KpsT (C-terminal FLAG) in MCS-II in a pETDuet backbone. Plasmid #*3a* contains KpsE (C-terminal 10xHis tagged) in MCS-I in pCDFDuet backbone. Plasmid #*4* contains KpsD in a pET30 backbone. To increase the expression yields of KpsM and KpsE, we introduced 3 amino acids at the third through fifth position of the polypeptide chains (K_3-I_4-H_5) (using polymerase incomplete primer extension (PIPE) cloning) that were shown to increase translation initiation [35]. KpsE was further modified by introducing two cysteines at positions 77 and 138, giving rise to plasmids #*1b* and #*3b.* Next, using Gibson assembly, KpsD was inserted into MCS-II of plasmid #*3b*, creating plasmid #*3c,* used for *in vivo* encapsulation assay.

3.4.1.6 Mutagenesis

Mutagenesis was performed by PIPE cloning with overlapping primers on plasmid #*1b* and #*3b*, using Phusion HF DNA polymerase (NEB), resulting in plasmids #*17.-24.*

`

3.4.2 Protein and CPS expression

All bacterial cultures described in this work were grown at 37°C and shaking at 220 rpm, unless noted otherwise. Working concentrations: ampicillin 100 mg/L, kanamycin 50 mg/L, streptomycin 50 mg/L, chloramphenicol 25 mg/L. Appropriate plasmids were transformed into C43 cells (for protein purification) or C43ΔCPS1 cells (for *in vivo* encapsulation, Minflux, CPS purification), for overnight growth in the presence of suitable antibiotics. All harvested cell pellets in this study were flash-frozen in liquid nitrogen and stored at −80 °C for further use.

For Western blotting - For expression testing, an overnight starter culture of cells expressing all *Pm* CPS components was used to inoculate 1 L of LB media supplemented with ampicillin and chloramphenicol. At OD_{600} of 0.6, protein expression was induced with 100 mg/L of IPTG. Growth was continued for another 3-4 h, after which cells were harvested (4500 rpm for 20min) and flash-frozen in liquid N_2. This cell pellet was used to prepare inverted membrane vesicles (IMVs), as described previously by [36]. The IMVs were then run on a 12.5% polyacrylamide gel and analyzed using western blot techniques detecting the engineered affinity tags, as described previously [37].

For in vivo encapsulation assays and spheroplasting - 20 mL LB cultures were inoculated with a single stab of the appropriate transformants and grown in the presence of appropriate antibiotics and 100 mg/L of IPTG. Growth was carried out for 6-8 h, after which cells were harvested. Spheroplasts were prepared as described by Morgan et al (2017) [38]. After removal of the outer membrane, spheroplast were resuspended in PBS supplemented with 200 mM sucrose, and used for *in vivo* encapsulation assays.

For CPS purification – 8x 1 L of 2xLB supplemented with appropriate antibiotics were inoculated from an overnight starter culture. At OD of 0.6, the medium was cooled to 20°C, then cells were induced using 100 mg/L of IPTG, and grown overnight, after which they were harvested.

`

For CBM70, SNAP-CBM70, and KpsEMT protein purification – 6x 1 L of 2xLB supplemented with appropriate antibiotics were inoculated from an overnight starter culture. At OD_{600} of 0.6, protein expression was induced using 200 mg/L of IPTG, and cells were grown for additional 3-4h at 37 °C, after which they were harvested.

3.4.3 *In vivo* encapsulation assay

200 µL of cells at OD_{600}= 4 was washed with ice cold PBS 3 times and then incubated on ice with 10 µL of 2 mg/mL of ^{Alexa647}SNAP-CBM70 (see below) for a total of 2 h. For CPS digestion, samples were treated with 1 mg/mL of bovine testicular hyaluronidase (MP Biomedicals) for 2 h on ice. For 2 color Minflux nanoscopy, 12.5 µL of ^{Alexa680}Streptavidin (ThermoFisher Scientific) was added after 1 hour of incubation, followed by incubation on ice for 30 min. Then 2ul of 100x Cellbrite Fix 488 (Biotum) was added for additional 30 min, after which cells were washed 3-times in 1000 µL of ice-cold PBS, fixed with 4% PFA (Electron Microscopy Sciences) in a total of 1000 µL PBS for 20 min, blocked with 50 mM NH_4Cl in PBS for 30 min, resuspended in 200 µL ice cold PBS, and imaged.

3.4.3.1 Confocal microscopy

Imaging was performed on a Zeiss LSM980 confocal microscope with an Airyscan detector, at 40x using a 'water-immersion' objective. Membrane and CPS channels were recorded sequentially using 488 nm and 633 nm excitation lasers, respectively, and suitable filter sets. The pixel size was set to 52 nm. Images were processed in ImageJ/Fiji.

`

3.4.3.2 Minflux – sample mounting, imaging buffer and nanoscopy

Fixed cells were applied on glass slides precoated with poly-L-lysine. Samples were mounted in the imaging buffer as described by Carsten et al (2023). Shortly, gold nanoparticles (NanopartzInc., A11-200-CIT-DIH-1-10) were used as fiducials. GLOX buffer supplemented with 14 mM MEA (Cysteamine) was used as imaging buffer. Samples were sealed with EliteDouble22 (Zhermack). MINFLUX nanoscopy, and corresponding confocal microscopy were carried out using an in-house MINFLUX set up [39]. In 2 color Minflux, a single event originates from one of two different red fluorophores (Alexa 647 or Alexa 680), and is split into Cy5 near and Cy5 far detectors. The number of photons from this single event reaching both detectors is represented as a detector channel ratio (DCR) and is characteristic for each emitter. DCR values were acquired experimentally for each fluorophore separately, and then used to assign the color of fluorophores to localizations in 2 color Minflux experiments. Pixel size in rendered images was based on localization precisions of raw burst.

3.4.4 Protein purification

When possible, all purification steps were performed at 4°C.

CBM70, SNAP-CBM70 purification – Cell pellets were thawed and resuspended in 10 % glycerol, 100 mM NaCl, and 20 mM Tris pH 7.5, then incubated for 1 hour with 1 mg/mL lysozyme. 1 mM PMSF was added and cell suspensions were lyzed by 3 passes through a microfluidizer. Intact cells were removed by low-speed centrifugation for 25 min at 12,500 rpm, in a JA-20 rotor (Beckman). The supernatant was centrifuged for 1 hour at 200,000 g in Ti45 rotor (Beckman) and the insoluble material was discarded. The supernatant was spiked with 20 mM imidazole and incubated with Ni-NTA resin for 1 hour with agitation. Resin was washed with 1) 1 M NaCl, PBS (pH 7.4), 40

`

mM imidazole and 2) PBS, 60 mM imidazole. Protein was eluted by 30 min incubation in PBS containing 320 mM imidazole and concentrated using 10 kDa filter (Amicon) to 1mL. Then, in case of CBM70, the sample was run over an S200 16/60 gel filtration column equilibrated in PBS, the peak fractions were collected and used for αHA column preparation, or, in case of SNAP-CBM70, dialyzed against PBS overnight, aliquoted, flash-frozen in liquid N_2 and stored at −80 °C.

KpsEMT purification - Cell pellets were processed as described above. After low-speed centrifugation, membranes were isolated from the lysate by centrifugation for 2 hours at 200,000 g in a Ti45 rotor, then harvested and flash-frozen in liquid N_2 and stored at −80 °C. Membranes were thawed and resuspended in 300 mM NaCl, 20 mM Tris pH 7.5, 10 % glycerol, 40 mM imidazole, 1 % n-dodecyl-beta-maltoside (DDM), and 0.1 % cholesterol hemisuccinate (CHS), and incubated for 1 hour with agitation. Aggregated material was removed by centrifugation at 200,000 g for 30 min, and the supernatant was incubated with Ni-NTA resin for 1 hour. Resin was washed with 1) 1.5 M NaCl, 20 mM Tris pH 7.5, 10 % glycerol, 40 mM imidazole, 0.1 % Lauryl Maltose Neopentyl Glycol (LMNG); 2) 300 mM NaCl, 20 mM Tris pH 7.5, 10 % glycerol, 80 mM imidazole, and 0.1 % LMNG. Protein was eluted by a 30 min incubation in 300 mM NaCl, 20 mM Tris pH 7.5, 5 % glycerol, 400 mM imidazole, and 0.05 % LMNG and concentrated to 500µl using a 100 kDa filter (Amicon), followed by overnight incubation on ice. The next day, the sample was run over a S6-increase 10/300 gel filtration column equilibrated in 100 mM NaCl, 50 mM Tris pH 7.5, 0.025% LMNG. This buffer was supplemented with 5mM $MgCl_2$ for $KpsEMT_{E151Q}$ preparations. The peak fractions were collected and concentrated to 2-3 mg/mL using a 100 kDa filter and used for grid preparation or *in vitro* ATPase activity assays. KpsEMT in complex with ADP: AlF_4^- was purified similarly. The concentrated Ni-NTA elution sample was dialyzed

`

overnight against buffer containing 100 mM NaCl, 50 mM Tris pH 7.5, 0.05% LMNG, 5% glycerol, 10 mM NaF, 2 mM $AlCl_3$, 5 mM ADP, and 5 mM $MgCl_2$. The same buffer containing 0.025% LMNG and lacking glycerol was used to equilibrate the S6-increase gel filtration column. Peak fractions were collected based on elution times, concentrated to ~2-3 mg/mL and used for cryo grid preparation.

3.4.5 ATPase activity assay

KpsEMT's ATPase activity was quantified using an enzyme-coupled assay as previously described by Caffalette et al (2019) [40]. Peak fraction of the complex eluting from S6-increase column was concentrated to 0.5-1 mg/mL and used for activity assays. ATPase activity was initiated by adding ATP, and the depletion of NADH was monitored at 340 nm for 1 h at 27 °C in a SpectraMax plate reader. The rate of NADH depletion was converted to nmols of ATP hydrolyzed using an ADP standardized plot. Data was processed in Excel and GraphPad Prism. All experiments were performed at least in triplicate and error bars represent deviations from the means.

3.4.6 Anti-HA affinity column preparation

Purified CBM70 was coupled to NHS-activated Sepharose 4 Fast Flow beads (Cytiva) following the manufacturer's protocol. In short, the resin was washed with 1) MQ water, 2) 1 mM HCl, and 3) PBS. Then 20 mL of 5 mg/mL CBM70 in PBS was mixed with 25 mL of the washed resin, and left agitating for 2 days at 4°C. After that, liquid was drained from the beads, and the resin was washed with PBS, followed by blocking buffer (PBS containing 200 mM ethanolamine) for 24 h with agitation at 4°C. The beads were washed, and stored in 20% ethanol at 4°C.

`

3.4.7 CPS purification

Cell pellets and membranes were prepared as described above. Membranes were resuspended in PBS containing 1 % LMNG and incubated for 1 hour at room temperature with agitation. Aggregated material was removed by centrifugation at 200,000 g for 30 min, and the supernatant was incubated with αHA resin (see above) for 1 hour. Next, the beads were washed 3-times with PBS containing 0.1 % LMNG. The CPS was eluted from the column by 30 min incubation with 2 M NaCl, 100 mM sodium citrate pH 3.0, and 0.01 % LMNG, concentrated to 250ul using a 3 kDa filter (Amicon), and dialyzed overnight against PBS in a 3.5 kDa dialysis membrane. The next day, the sample was run on the 1.5 % agarose gel (Ultra-pure agarose, Invitrogen) or 4-20 % gradient polyacrylamide gel (Bio-Rad), and stained with Stains All dye (Sigma) as described [41]. The obtained CPS sample was also used for cryoEM analyses.

3.4.8 ^{Alexa647}SNAP-CBM70 preparation

All steps were carried out in the dark at 4°C. SNAP-CBM70 aliquots were thawed and mixed with DMSO-solubilized Alexa647 (NEB) in 1:1 molar ratio, in the presence of 1 mM DTT. The sample was incubated with agitation for 6-8 h and run over an S200 10/30 gel filtration column equilibrated in PBS to separate ^{Alexa647}SNAP-CBM70 from the free dye. Peak fractions with string absorbances at 280 and 671 nm were aliquoted at 2 mg/mL, flash-frozen in liquid N_2, and stored at −80 °C.

3.4.9 Grid preparation and data collection

To obtain the ATP and glycolipid bound states, $KpsEMT_{E151Q}$ and WT KpsEMT were supplemented with 2 mM ATP or 30 μL of lipid-linked HA, respectively, prior to grid preparation.

`

Quantifoil holey carbon grids (Cu 1.2/1.3, 300 mesh) were glow-discharged in the presence of 2 drops (about 200 μL) of amylamine. 4 μL of sample was applied, blotted for 4-10 seconds with a blot force of 4-7 at 4 °C and 100% humidity, then plunge frozen in liquid ethane using a Vitrobot Mark IV (FEI).

Cryo-EM data were collected at the University of Virginia Molecular Electron Microscopy Core (MEMC) on a Titan Krios (FEI) 300-kV electron microscope using a Gatan Imaging Filter (GIF) and a K3 direct electron detection camera. Movies were collected in EPU (Thermo Fisher Scientific) at a magnification of 81,000x with an energy filter width of 10 eV, using counting mode with a total dose of 51 e−/Å over 40 frames, and with a target defocus of −1.0 to −2.0 μm.

3.4.10 CryoEM data processing

All datasets were processed in cryoSPARC versions 3.3 and 4.0 [42]. Raw movies were subjected to patch motion correction and patch contrast transfer function (CTF) estimation. For all 4 datasets, particles were automatically selected by blob picker to generate initial temples, followed by template picker. After varying rounds of 2D classification, selected particles were used for ab initio reconstructions in C1, followed by heterogenous refinement.

For Dataset 1 yielding Apo 1 state, C2 symmetry was applied in both Non-uniform and Local refinements. To improve the KpsT density, focused 3D classification followed by Non-uniform and Local refinement was applied. Using the Phenix Combine Focused Maps job [43], a composite Map1 was created from map A and B, based on the model and half maps from the focused refinements.

For Dataset 2 yielding the ATP-bound state, C2 symmetry was applied during Non-uniform refinement, followed by Local refinement in C1. To improve the density of KpsE's crown region,

`

focused 3D classification followed by Non-uniform and Local refinement was applied. Using Phenix Combine Focused Maps job, a composite Map2 was created from map A and B, based on the model and half maps from the focused jobs.

For Dataset 3 yielding the ADP:AlF_4^--bound state, C2 symmetry was applied in Non-uniform refinement.

For Dataset 4 yielding the Glycolipid 1 and 2 and Apo 2 states, C1 symmetry was applied in both Non-uniform and Local refinement. Next, the particles were 3D classified, into 7 classes using 7 identical Apo 1 volumes as input. Three of the resulting classes (class 0, 1 and 4) revealed distinct states and subjected to Non-uniform refinement. Class 0 and 1 had noticeable extra density in the polysaccharide canyon. To improve the density of the putative substrate, 3D classification focused on KpsM was performed for both classes. To improve the KpsT density in classes 0 and 1, a separate 3D classification focused on KpsT was performed, resulting in improved KpsT density for the Glycolipid 1 and 2 state (respective Maps A), enabling rigid body docking of AF2 predicted KpsT. Particles from classes 0 and 1 were also subjected to 3D classification focused on the KpsM subunits, which resulted in improved glycolipids density maps (respective Maps B). Then maps focused on KpsT subunits and maps focused on the Glycolipid containing KpsM subunits (respective Maps A and B for both states) were combined, resulting in composite Maps 4 and 5.

Class 4 from the original 3D classification job revealed a novel arrangement of KpsM's TM helices. As for classes 0 and 1, the KpsT map quality was improved by 3D focused classification, followed by Non-uniform refinement. The improved map allowed rigid body docking of AF2 predicted KpsT. The focused maps A and B were combined using Phenix Combine Focused Maps job, resulting in Map6.

`

Maps were sharpened based on models using either Autosharpen (Maps 1, 2, 3, and 6) or Local Anisotropic Sharpening (Maps 4 and 5) jobs in Phenix:refine. Half maps were used to generate: global resolution estimates using EMBL's Fourier Shell Correlation (FSC) server, and local resolution estimates using cryoSPARC's Local resolution estimation job.

3.4.11 Model building and refinement

To generate the initial model of KpsEMT's Apo1 state, the Alphafold2 [22] models of the individual subunits were rigid body docked into the EM map using Chimera ref and the model was iteratively real-space refined in Coot and Phenix:refine [43,44]. The obtained structure was used to build all other states. Chain completeness for all states is reported in Table 4. For Apo 2, Glycolipid 1 and 2 states two sets of real-space refine jobs were run, with and without the KpsT subunits rigid body docked into the model, Table 1. For Glycolipid bound states 1 and 2, the putative substrate model was drawn using Coot Ligand Builder based on 1,2-dipalmitoyl-phospatidyl-glycerole (PG) and 4 Kdo sugars linked by β–2-7 and β–2-4 glycosidic linkages. The SMILES output from this job was used in Phenix eLBOW to generate coordinate and constraint files. To preserve the correct linkage geometry connecting the first two Kdo units during refinement, the generated poly-Kdo-PG lipid was trimmed to 2 Kdo sugar units, docked into the substrate density in Coot, and real space refined in Phenix.refine. Following this refinement, the second Kdo sugar was removed from the substrate and the whole model was refined again in Phenix.refine. After the Glycolipid 1 and 2 models were refined, the corresponding maps were locally anisosharpened in Phenix [45], resulting in Maps 4 and 5. All structural figures were prepared using ChimeraX and Inkscape [46,47].

`

3.4.11.1 AlphaFold2 predictions

The full length KpsE octamer was predicted at servers of the University of Virginia Molecular Electron Microscopy Core. *Pm* and *St* KpsD was predicted on AlphaFold2 Collab servers using truncated protein sequences (no outer membrane signal sequence, N-terminal truncation), limit the octameric protein sequence to <3300 amino acids. Omitted regions were then backfitted on the predicted octameric backbone from a monomer model.

3.5 Acknowledgements

We are indebted to Parastoo Azadi and Artur Muszynski for efforts on glycolipid characterization. We thank Michael Purdy and Kelly Dryden from UVA's MEMC cryo EM core facility as well as Ruoya Ho for help with cryo EM data collection. We are grateful to Chris Whitfield and Jessica Matthias for discussions. JK is a recipient of the UVA Wagner graduate student fellowship and JZ acknowledges funding from UVA's Pinn Scholarship program to initiate the project. The initial phase of the project was funded by NIH grant R21AI1642, followed by R35GM144130 award to JZ. JZ is an investigator of the Howard Hughes medical Institute.

`

3.6 References

1 Caffalette, C. A., Kuklewicz, J., Spellmon, N. & Zimmer, J. Biosynthesis and Export of Bacterial Glycolipids. *Annu Rev Biochem* **89**, 741-768, (2020).

2 Whitfield, C., Wear, S. S. & Sande, C. Assembly of Bacterial Capsular Polysaccharides and Exopolysaccharides. *Annu Rev Microbiol* **74**, 521-543, (2020).

3 Nickerson, N. N. *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).

4 Pavelka, M. S., Jr., Wright, L. F. & Silver, R. P. Identification of two genes, kpsM and kpsT, in region 3 of the polysialic acid gene cluster of Escherichia coli K1. *J Bacteriol* **173**, 4603-4610, (1991).

5 Willis, L. M. & Whitfield, C. KpsC and KpsS are retaining 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) transferases involved in synthesis of bacterial capsules. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 20753-20758, (2013).

6 Willis, L. M. *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).

7 Yan, L. F. *et al.* A Revised Structure for the Glycolipid Terminus of Escherichia coli K5 Heparosan Capsular Polysaccharide. *Biomolecules* **10**, (2020).

8 Doyle, L. *et al.* Biosynthesis of a conserved glycolipid anchor for Gram-negative bacterial capsules. *Nature Chemical Biology* **15**, 632-+, (2019).

9 Thomas, C. *et al.* Structural and functional diversity calls for a new classification of ABC transporters. *Febs Letters* **594**, 3767-3775, (2020).

10 Bi, Y., Mann, E., Whitfield, C. & Zimmer, J. Architecture of a channel-forming O-antigen polysaccharide ABC transporter. *Nature* **553**, 361-365, (2018).

11 Chen, L. *et al.* Cryo-electron Microscopy Structure and Transport Mechanism of a Wall Teichoic Acid ABC Transporter. *mBio* **11**, (2020).

12 Lo, R. Y., McKerral, L. J., Hills, T. L. & Kostrzynska, M. Analysis of the capsule biosynthetic locus of Mannheimia (Pasteurella) haemolytica A1 and proposal of a nomenclature system. *Infect Immun* **69**, 4458-4464, (2001).

13 Ward, C. K. & Inzana, T. J. Identification and characterization of a DNA region involved in the export of capsular polysaccharide by Actinobacillus pleuropneumoniae serotype 5a. *Infect Immun* **65**, 2491-2496, (1997).

14 Buckles, E. L. *et al.* Role of the K2 capsule in Escherichia coli urinary tract infection and serum resistance. *J Infect Dis* **199**, 1689-1697, (2009).

15 Andreishcheva, E. N. & Vann, W. F. Escherichia coli BL21(DE3) chromosome contains a group II capsular gene cluster. *Gene* **384**, 113-119, (2006).

16 Chung, J. Y., Zhang, Y. & Adler, B. The capsule biosynthetic locus of Pasteurella multocida A:1. *FEMS microbiology letters* **166**, 289-296, (1998).

17 DeAngelis, P. L., Jing, W., Drake, R. R. & Achyuthan, A. M. Identification and molecular cloning of a unique hyaluronan synthase from Pasteurella multocida. *J Biol Chem* **273**, 8454-8458, (1998).

18 Suits, M. D. L. *et al.* Conformational analysis of the Streptococcus pneumoniae hyaluronate lyase and characterization of its hyaluronan-specific carbohydrate-binding module. *The Journal of biological chemistry* **289**, 27264-27277, (2014).

`

19 Phanphak, S. *et al.* Super-Resolution Fluorescence Microscopy Study of the Production of K1 Capsules by Escherichia coil: Evidence for the Differential Distribution of the Capsule at the Poles and the Equator of the Cell. *Langmuir* **35**, 5635-5646, (2019).
20 Balzarotti, F. *et al.* Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. *Science* **355**, 606-612, (2017).
21 Pape, J. K. *et al.* Multicolor 3D MINFLUX nanoscopy of mitochondrial MICOS proteins. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 20607-20614, (2020).
22 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, (2021).
23 Bronner, D. *et al.* Expression of the capsular K5 polysaccharide of Escherichia coli: biochemical and electron microscopic analyses of mutants with defects in region 1 of the K5 gene cluster. *J Bacteriol* **175**, 5984-5992, (1993).
24 McNulty, C. *et al.* The cell surface expression of group 2 capsular polysaccharides in Escherichia coli: the role of KpsD, RhsA and a multi-protein complex at the pole of the cell. *Molecular Microbiology* **59**, 907-922, (2006).
25 Craig, D. B. & Dombkowski, A. A. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics* **14**, 346, (2013).
26 Xu, D. *et al.* Structure and transport mechanism of the human cholesterol transporter ABCG1. *Cell Rep* **38**, 110298, (2022).
27 Zeytuni, N. *et al.* Structural insight into the Staphylococcus aureus ATP-driven exporter of virulent peptide toxins. *Sci Adv* **6**, (2020).
28 Lee, J. Y. *et al.* Crystal structure of the human sterol transporter ABCG5/ABCG8. *Nature* **533**, 561-564, (2016).
29 Spellmon, N. *et al.* Molecular basis for polysaccharide recognition and modulated ATP hydrolysis by the O antigen ABC transporter. *Nature Commun.* **13**, (2022).
30 Zhao, H., Lee, J. & Chen, J. The hemolysin A secretion system is a multi-engine pump containing three ABC transporters. *Cell* **185**, 3329-3340 e3313, (2022).
31 Wiseman, B., Nitharwal, R. G., Widmalm, G. & Hogbom, M. Structure of a full-length bacterial polysaccharide co-polymerase. *Nat Commun* **12**, 369, (2021).
32 Coudray, N. *et al.* Structure of bacterial phospholipid transporter MlaFEDB with substrate bound. *Elife* **9**, (2020).
33 Li, Y., Orlando, B. J. & Liao, M. Structural basis of lipopolysaccharide extraction by the LptB(2)FGC complex. *Nature* **567**, 486-490, (2019).
34 Huang, C., Guo, L., Wang, J., Wang, N. & Huo, Y.-X. Efficient long fragment editing technique enables large-scale and scarless bacterial genome engineering. *Applied microbiology and biotechnology* **104**, 7943-7956, (2020).
35 Verma, M. *et al.* A short translational ramp determines the efficiency of protein synthesis. *Nat Commun* **10**, 5774, (2019).
36 Blackburn, M. R. *et al.* Distinct reaction mechanisms for hyaluronan biosynthesis in different kingdoms of life. *Glycobiol* **28**, 108-121, (2018).
37 Maloney, F. P. *et al.* Structure, substrate recognition and initiation of hyaluronan synthase. *Nature* **604**, 195-201, (2022).
38 Morgan, J. L., Acheson, J. F. & Zimmer, J. Structure of a Type-1 Secretion System ABC Transporter. *Structure* **25**, 522-529, (2017).

`

39 Carsten, A. *et al.* MINFLUX imaging of a bacterial molecular machine at nanometer resolution. *Methods Appl Fluoresc* **11**, (2022).
40 Caffalette, C., Corey, R., Sansom, M.S.P., Stansfeld, P.J., Zimmer, J. A lipid gating mechanism for the O antigen ABC transporter. *Nature Commun.* **10**, (2019).
41 Gorniak, I., Stephens, Z., Zimmer, J. https://doi.org/10.1101/2023.05.11.540447. *bioRxiv*, (2022).
42 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, (2017).
43 Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-367, (2012).
44 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, (2004).
45 Adams, P. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, (2010).
46 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82, (2021).
47 Project, I. Available from: https://inkscape.org. (2020).
48 Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* **38**, W529-533, (2010).

`

3.7 Supplementary information

Table 3

State	Apo 1	ATP-bound	ADP:AlF_4^- bound	Apo 2	Glycolipid 1	Glycolipid 2
Refined Subunits	KpsEMT	KpsEMT	KpsEMT	KpsEM	KpsEM	KpsEM
Refined Chains	A,B,C,D,E,F,G,H,I,J,K,L	A,B,C,D,E,F,G,H,I,J,K,L	A,B,C,D,E,F,G,H,I,J,K,L	C,D,E,F,G,H,I,J,K,L	C,D,E,F,G,H,I,J,K,L	C,D,E,F,G,H,I,J,K,L
Rigid Body Fitted Subunits	-	-	-	KpsT	KpsT	KpsT
Rigid Body Fitted Chains	-	-	-	A,B	A,B	A,B
Data collection						
Microscope	Titan Krios					
Camera	K3, GIF, 10eV slit					
Magnification	81,000x					
Voltage (kV)	300					
Dose ($e^-/Å^2$)	50					
Defocus range (μm)	-1.8 to -1.0					
Pixel size (Å)	1.08					
Data processing						
Symmetry	C2	C1	C2	C2	C1	C1
Particles extracted #	6 899 265	7 300 079	3 649 073	11 520 747		
Final Particles	105 655 (Map A)	269 617 (Map A)	58 165	141 857 (Map A)	79 301 (Map A)	78 036 (Map A)
	43 811 (Map B)	36 731 (Map B)		46 064 (Map B)	86 755 (Map B)	85 882 (Map B)
Map resolution (Å)	3.1 (Map A)	3.1 (Map A)	4.4	3.3 (Map A)	3.4 (Map A)	3.3 (Map A)
	3.3 (Map B)	3.3 (Map B)		3.5 (Map B)	3.4 (Map B)	3.3 (Map B)
	3.1 (Composite)	3.1 (Composite)		3.4 (Composite)	3.4 (Composite)	3.3 (Composite)
FSC threshold	0.143	0.143	0.143	0.5	0.143	0.143
Refinement						
Initial model used	AF2	Apo 1				
Model resolution (Å)	3.4	3.3	4.0	3.4 (3.4)	3.7 (3.7)	3.5 (3.6)
Non-hydrogen atoms	20 318	24 235	20 242	17 840 (21 382)	18 495 (22 037)	19 199 (22 741)
Protein residues	2506	3001	2488	2194 (2646)	2274 (2726)	2368 (2820)
Ligands	0	2: ATP 2: Mg^{2+}	2: ADP:AlF_4^-	0	1: PGK	1: PGK
B Factors ($Å^2$)						
Protein	84.72	86.77	39.14	68.32 (104.21)	127.69 (144.56)	130.04 (137.93)
Ligands	-	71.34	76.52	-	137.33 (137.36)	152.39 (118.76)
RMS deviations						
Bond lenghts	0.003	0.003	0.003	0.003 (0.003)	0.003 (0.003)	0.003 (0.005)
Bond angles	0.601	0.572	0.671	0.624 (0.632)	0.682 (0.696)	0.627 (0.714)
Validation						
MolProbity score	1.62	1.50	1.76	1.67 (1.63)	1.67 (1.72)	1.59 (1.63)
Clashscore	8.10	7.14	9.67	7.39 (7.99)	8.28 (11.81)	8.64 (8.78)
Rotamer outliers (%)	0.23	0.16	0.55	0.46 (0.09)	0.2 (0.25)	0.29 (0.66)
Ramachandran plot						
Favored	96.93	97.45	96.3	96.17 (96.83)	96.58 (97.27)	97.37 (97.14)
Allowed	3.07	2.55	3.7	3.83 (3.17)	3.42 (2.73)	2.63 (2.86)
Outliers	0.0	0.0	0.0	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)

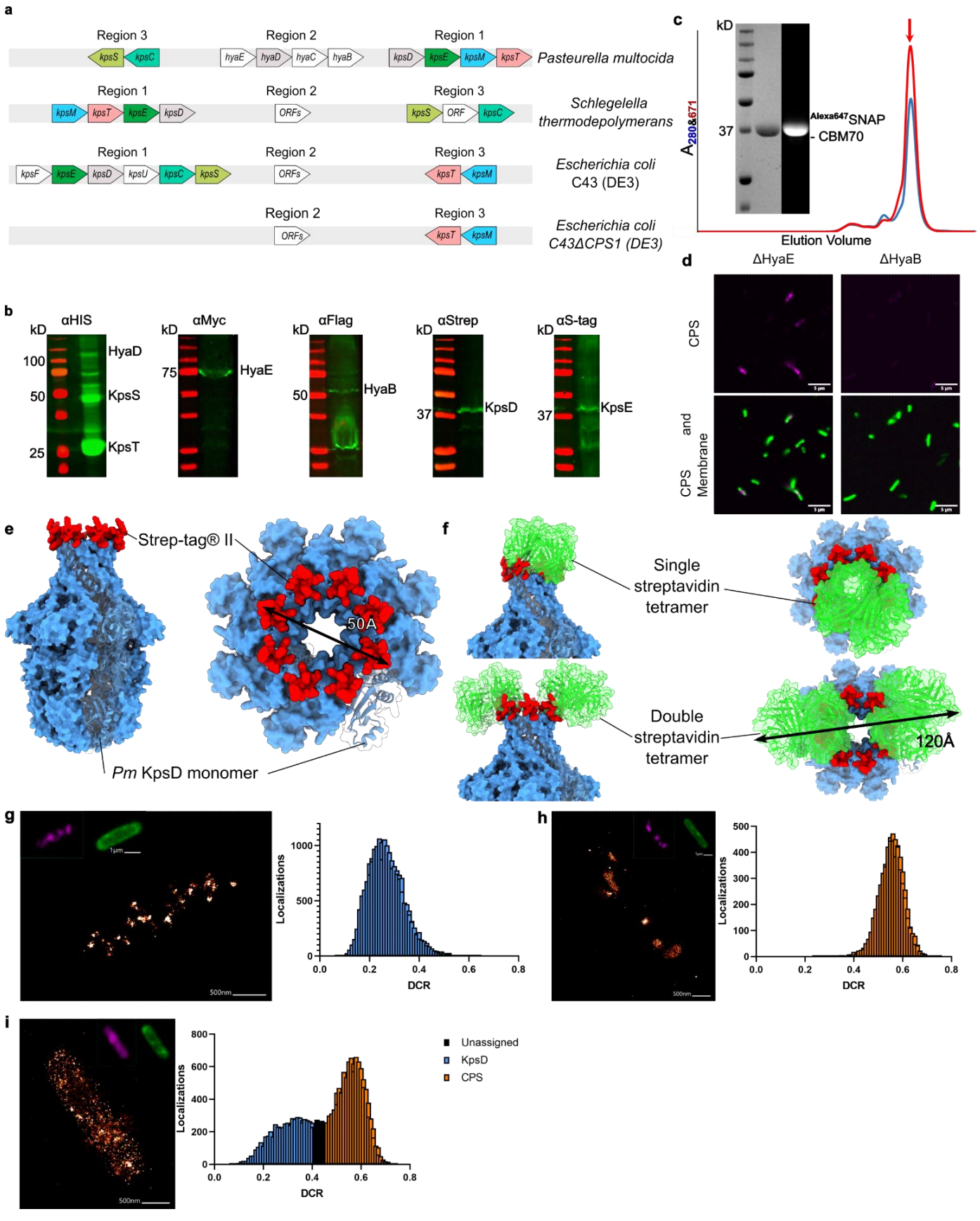
Table 3 | Cryo-EM data collection, refinement and validation statistics. Values in parentheses refer to statistics including the rigid body docked KpsT subunits.

Table 4

Protein	Chain ID	Apo 1	ATP-bound	ADP:AlF_4^- bound	Apo 2	Glycolipid 1	Glycolipid 2
KpsT	A	1-11,17-219	1-216	2-216	Rigid body fit (1-216)	Rigid body fit (1-216)	Rigid body fit (1-216)
	B	1-11,17-219	1-216	2-216	Rigid body fit (1-216)	Rigid body fit (1-216)	Rigid body fit (1-216)
KpsM	C	14-268	12-270	14-267	13-269	12-269	14-269
	D	14-268	14-267	14-267	13-269	14-269	14-269
KpsE	E	6-50,71-179,320-363	10-50, 71-204,289-368	12-48,73-180,321-361	10-48, 72-176,319-367	10-50,72-178,320-367	7-50, 72-185, 301-370
	F	4-50,72-176,321-371	4-50, 72-206, 289-371	7-48,72-175,322-370	6-48,72-172,321-370	3-52,72-179,319-370	6-50, 72-180, 299-371
	G	5-50,71-180,319-371	5-50, 70-204, 289-371	11-49,71,-175,320-371	10-49,72-177,320-370	9-50,72-181,305-371	6-50, 71-184, 299-370
	H	19-50,71-180,320-351	11-50, 71-204, 290-369	10-48,71-178,320-366	13-48,71-177,315-369	10-50,71-189,299-369	4-50, 71- 192, 299-369
	I	6-50,71-179,320-363	10-50, 70-204, 291-365	12-48,73-180,321-361	10-48,72-176,319-367	10-50,71-188,300-367	9-51, 71-187, 298-369
	J	4-50,72-176,321-371	9-50, 71-204, 289-371	7-48,72-175,322-370	6-48,72-172,321-370	3-50,72-183,308-370	3-50, 71-182, 299-370
	K	5-50,71-180,319-371	13-50, 71-202, 291-371	11-49,71,-175,320-371	10-49,72-177,320-370	6-50,72-189,301-371	6-50, 71-187,300-370
	L	16-50,71-180,320-351	10-50, 71-204, 290-367	10-48,71-178,320-366	13-48,71-177,315-369	10-50,72-183,303-369	4-50, 71-190, 302-371

Table 4 | Chain completeness for all of the models in this study.

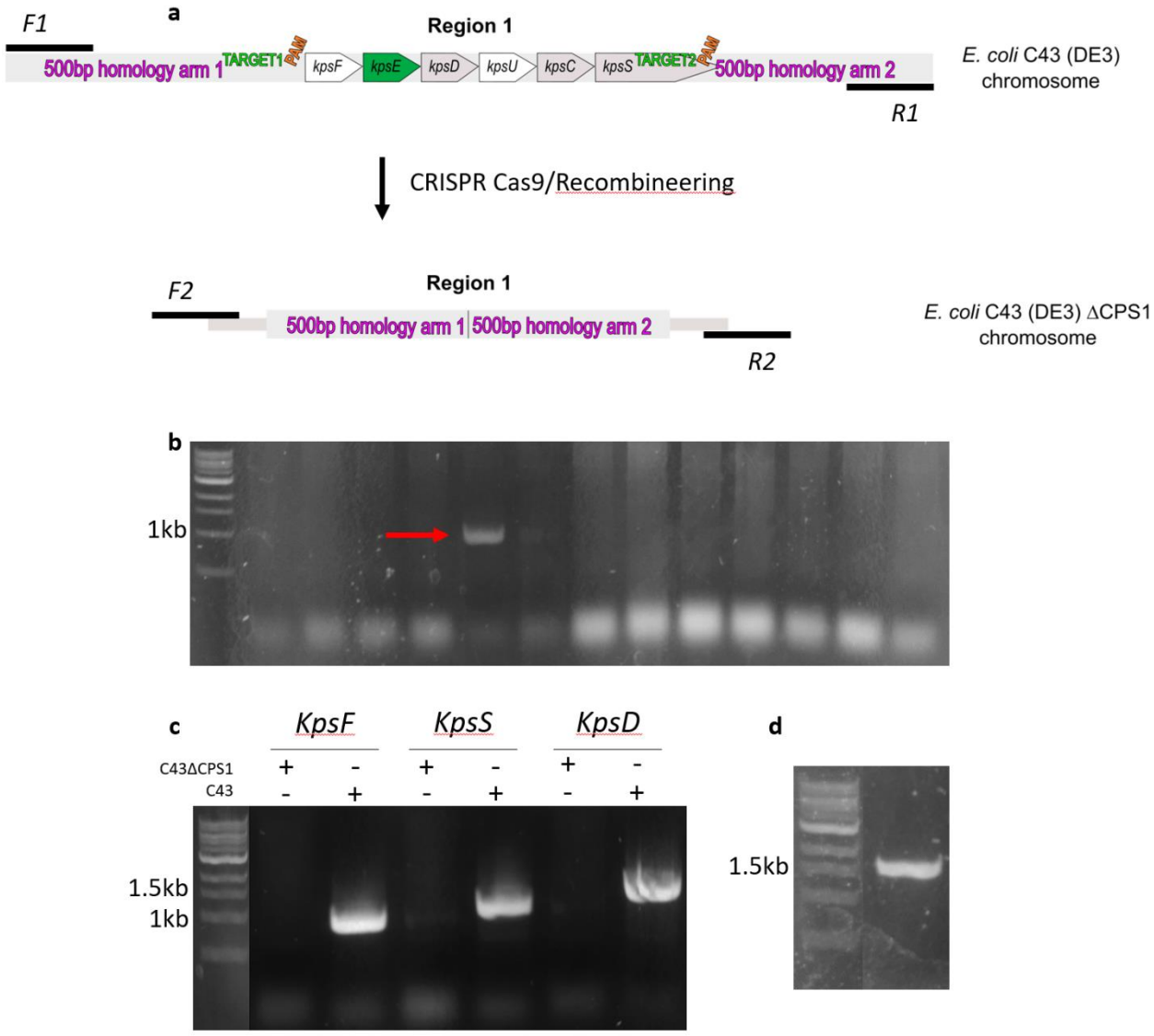
Fig. S3-1


a
Region 3
Region 2
Region 1
kpsS
kpsC
hyaE
hyaD
hyaC
hyaB
kpsD
kpsE
kpsM
kpsT
Pasteurella multocida
Region 1
Region 2
Region 3
kpsM
kpsT
kpsE
kpsD
ORFs
kpsS
ORF
kpsC
Schlegelella thermodepolymerans
Region 1
Region 2
Region 3
kpsF
kpsE
kpsD
kpsU
kpsC
kpsS
ORFs
kpsT
kpsM
Escherichia coli C43 (DE3)
Region 2
Region 3
ORFs
kpsT
kpsM
Escherichia coli C43ΔCPS1 (DE3)
b
αHIS
αMyc
αFlag
αStrep
αS-tag
kD
100
50
25
75
37
HyaD
KpsS
KpsT
HyaE
HyaB
KpsD
KpsE
c
kD
37
A280&671
Alexa647 SNAP - CBM70
Elution Volume
d
ΔHyaE
ΔHyaB
CPS
CPS and Membrane
e
Strep-tag® II
50Å
Pm KpsD monomer
f
Single streptavidin tetramer
Double streptavidin tetramer
120Å
g
1µm
500nm
Localizations
DCR
h
1µm
500nm
Localizations
DCR
i
500nm
Localizations
DCR
Unassigned
KpsD
CPS

Supplemental Figure 3-1 | Genetic organization and recombinant expression of CPS components. (a) Genomic organization of CPS loci in different organisms. *E. coli* C43ΔCPS1 was created in this study using CRISPR Cas9/recombineering method. ORFs: open reading frames without assigned function. **(b)** Recombinant expressions of *Pm* CPS components in E. coli. Shown are Western blots of inverted membrane vesicles containing the individually tagged, membrane bound capsular proteins. **(c)** Alexa^{647} loading of the in-house purified SNAP-CBM70 probe. Size exclusion chromatography profile for protein absorbance (280nm - blue) and fluorophore absorbance (671nm - red). The inset shows a Coomassie stained gel (left) and in gel fluorescence at 671nm emission (right). **(d)** Confocal images of E. coli C43ΔCPS1 cells expressing the Pm CPS components in the absence of HyaE or HyaB. CPS was stained as in Fig. 1 using the ^{Alexa647}SNAP-CBM70 probe, membranes were stained with Cellbrite Fix 488. **(e)** Surface/cartoon representation of the AlphaFold2 predicted octameric Pm KpsD structure (blue) with localization of the engineered C-terminal Strep-tag peptide (red). **(f)** Possible configurations of tetramers of ^{Alexa680}streptavidin (PDB:6j6j) bound to a KpsD octamer. A complex of a KpsD octamer bound to two ^{Alexa680}streptavidin tetramers would create a fluorophore cloud of about 14 nm diameter. **(g)** Minflux nanoscopy of KpsD in an encapsulated cell via ^{Alexa680}streptavidin labeling (left panel). Most localizations were detected via the Cy5-far detector (DCR: detector channel ratio <0.5) (right panel). **(h)** Minflux nanoscopy of CPS of an encapsulated cell using ^{Alexa647}SNAP-CBM70 (left panel). Most localizations were detected via the Cy5-near detector (right panel). For panels g and h, the experiments were repeated at least 3 times with similar results**. (i)** Left: Dual color Minflux nanoscopy localizations of the encapsulated cell presented in Fig. 1f with ^{Alexa647}SNAP-CBM70 labeled CPS and ^{Alexa680}streptavidin labeled KpsD without fluorophore assignment (left). Right: Histogram DCR values used for color assignment in Fig. 1f. Unassigned: localizations of overlapping DCR values that cannot be clearly assigned to either Alexa^{647} or Alexa^{680}.

`

Fig. S3-2



Supplemental Figure 3-2 | CRISPR Cas9/Recombineering of *E. coli* C43 genome.
(a) Deletion of Region 1 of CPS gene cluster. **(b)** Colony screening - genomic DNA amplified using primers directly flanking two homology arms (F1, R1) by colony PCR. Red arrow indicates a positive clone. **(c)** Comparison of genes amplified by colony PCR from within the CPS1 gene cluster in C43 and C43ΔCPS1 cells. **(d)** DNA fragment amplified from the genome of C43ΔCPS1 cells using primers annealing upstream and downstream of the deleted region (F2, R2). This fragment was then purified and sent for sequencing.

`

Fig. S3-3

Supplemental Figure 3-3 | Sequencing results of the fragment amplified from the genome of C43ΔCPS1 cells. Homology arms marked in magenta and pink and primers used in Fig. S3 -2 in black. Genomic sequence: top, sequencing results: bottom. Alignment presented in SnapGene by Dotmatics.

`

Fig. S3-4a

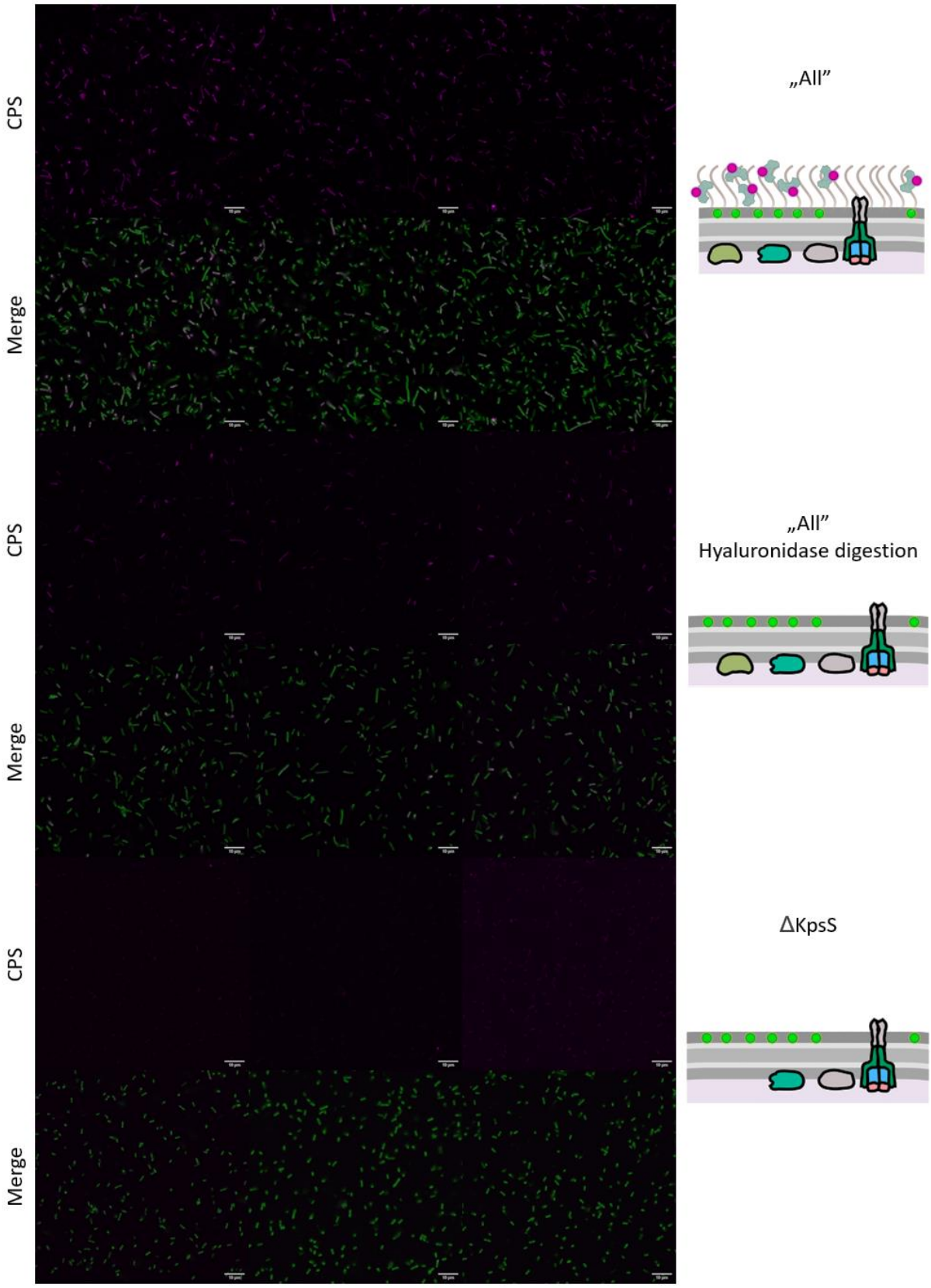

CPS
Merge
„All"
CPS
Merge
„All"
Hyaluronidase digestion
CPS
Merge
ΔKpsS

Fig. S3-4b

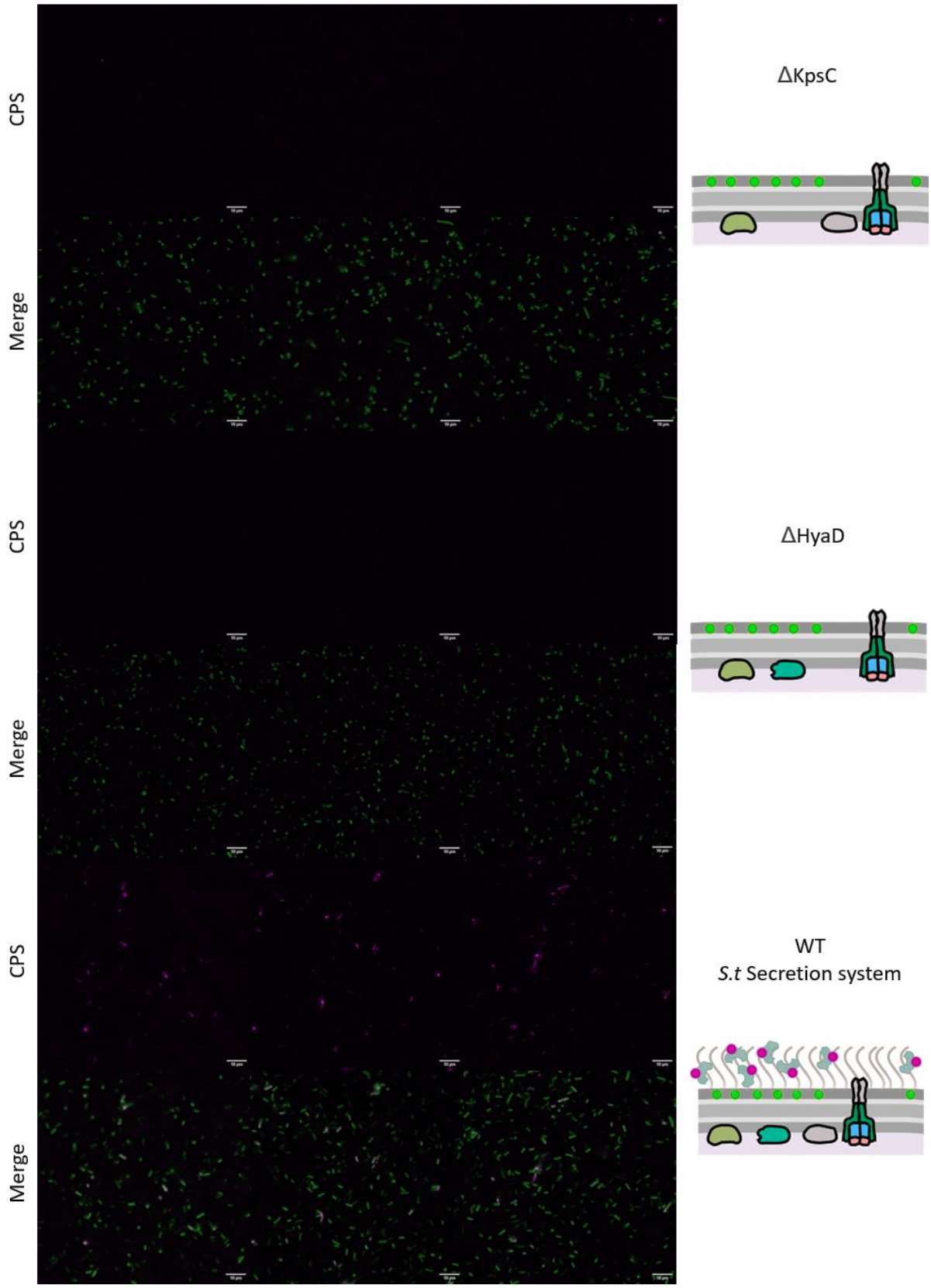
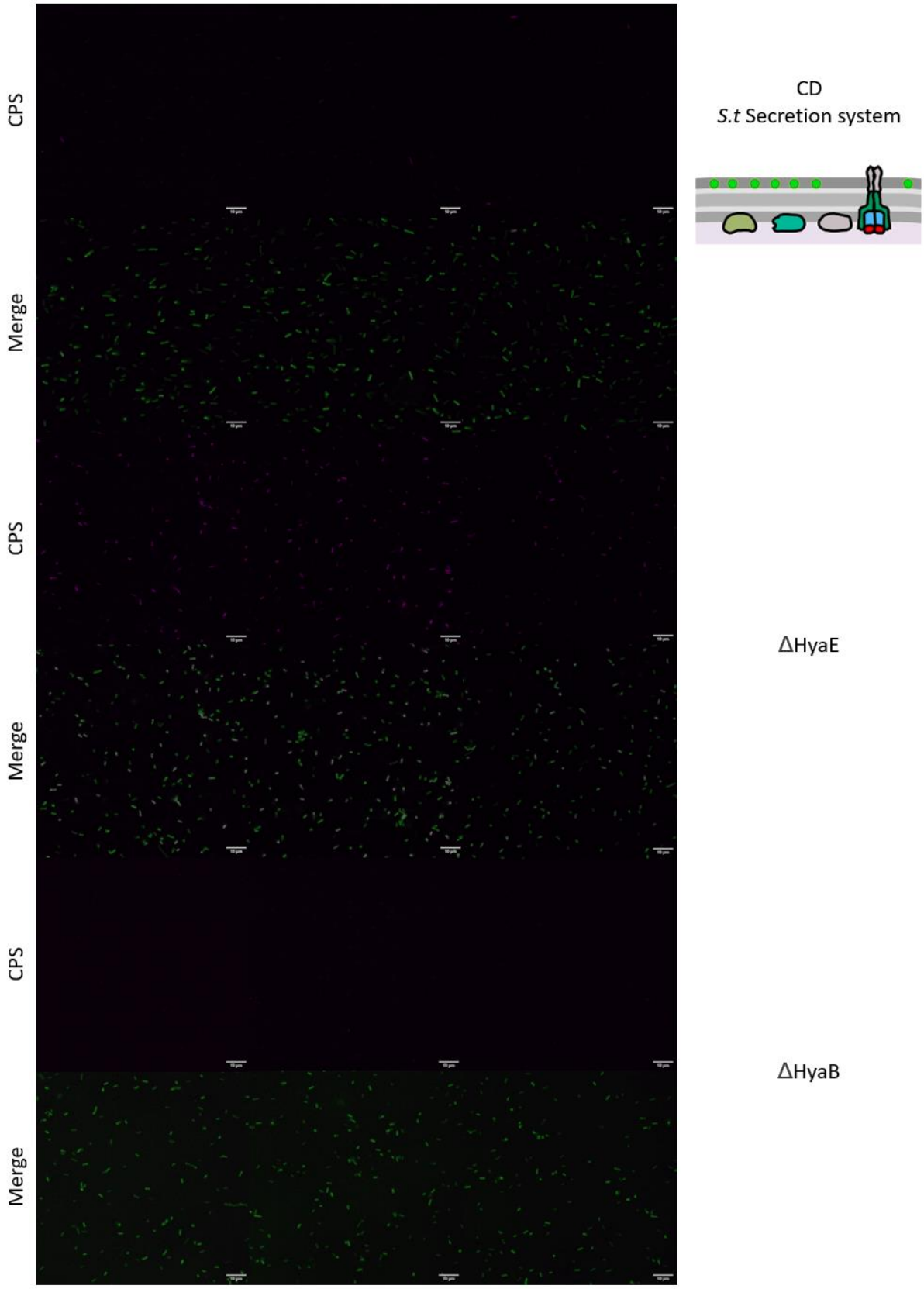

CPS
Merge
CPS
Merge
CPS
Merge
ΔKpsC
ΔHyaD
WT
S.t Secretion system

Fig. S3–4c


CPS
Merge
CPS
Merge
CPS
Merge
CD
S.t Secretion system
ΔHyaE
ΔHyaB

Supplemental Figure 3-4 (a-c) | Full fields of view and additional examples of confocal imaging presented in Figure 3 -1 and Figure S3 -1. Scalebar: 10μm.

Supplemental Figure 3-5 | pLDDT (predicted local distance difference test) values for all AlphaFold2 predicted models in the order of appearance in the main text. On the next page.

`

Fig. S3-5

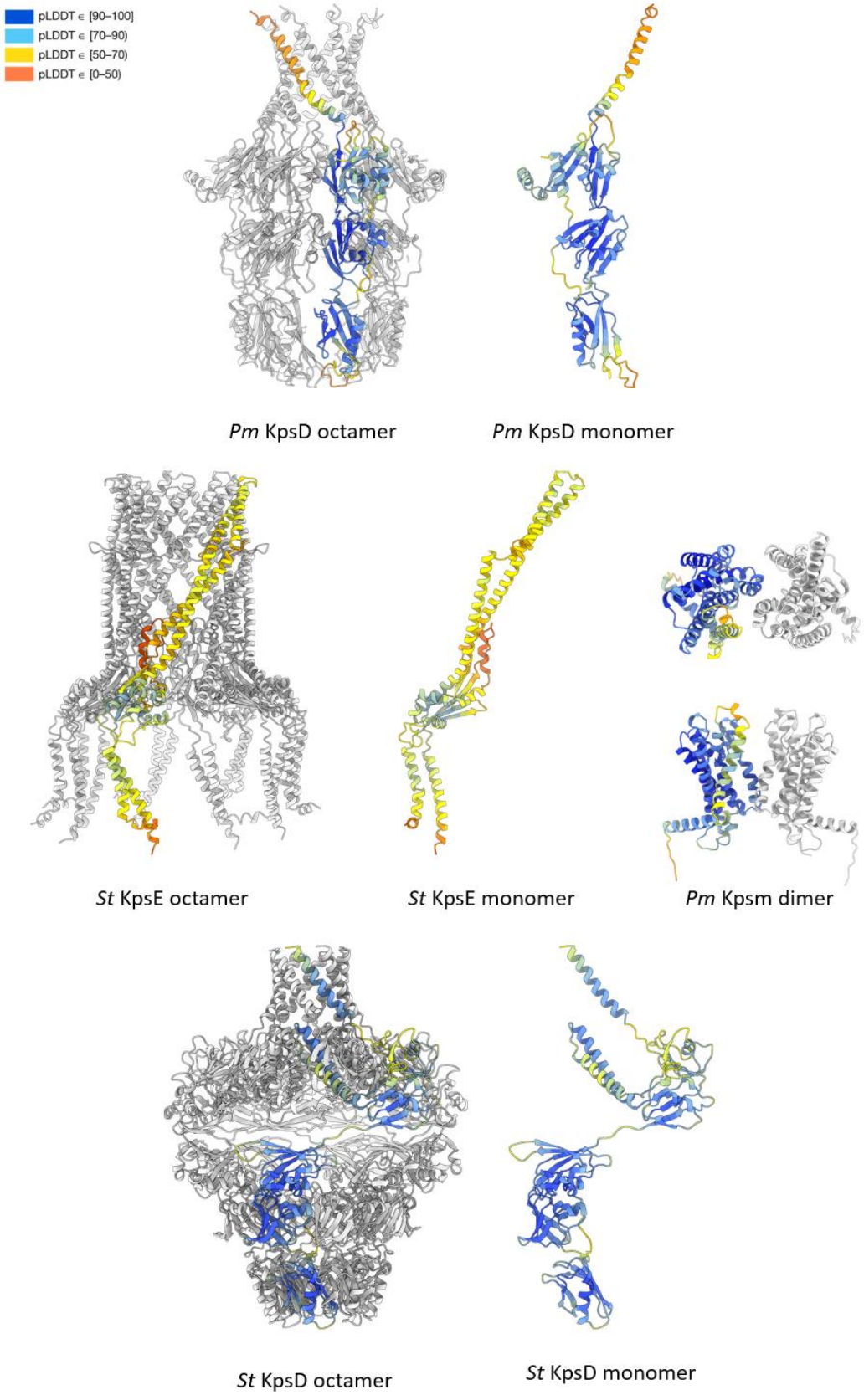
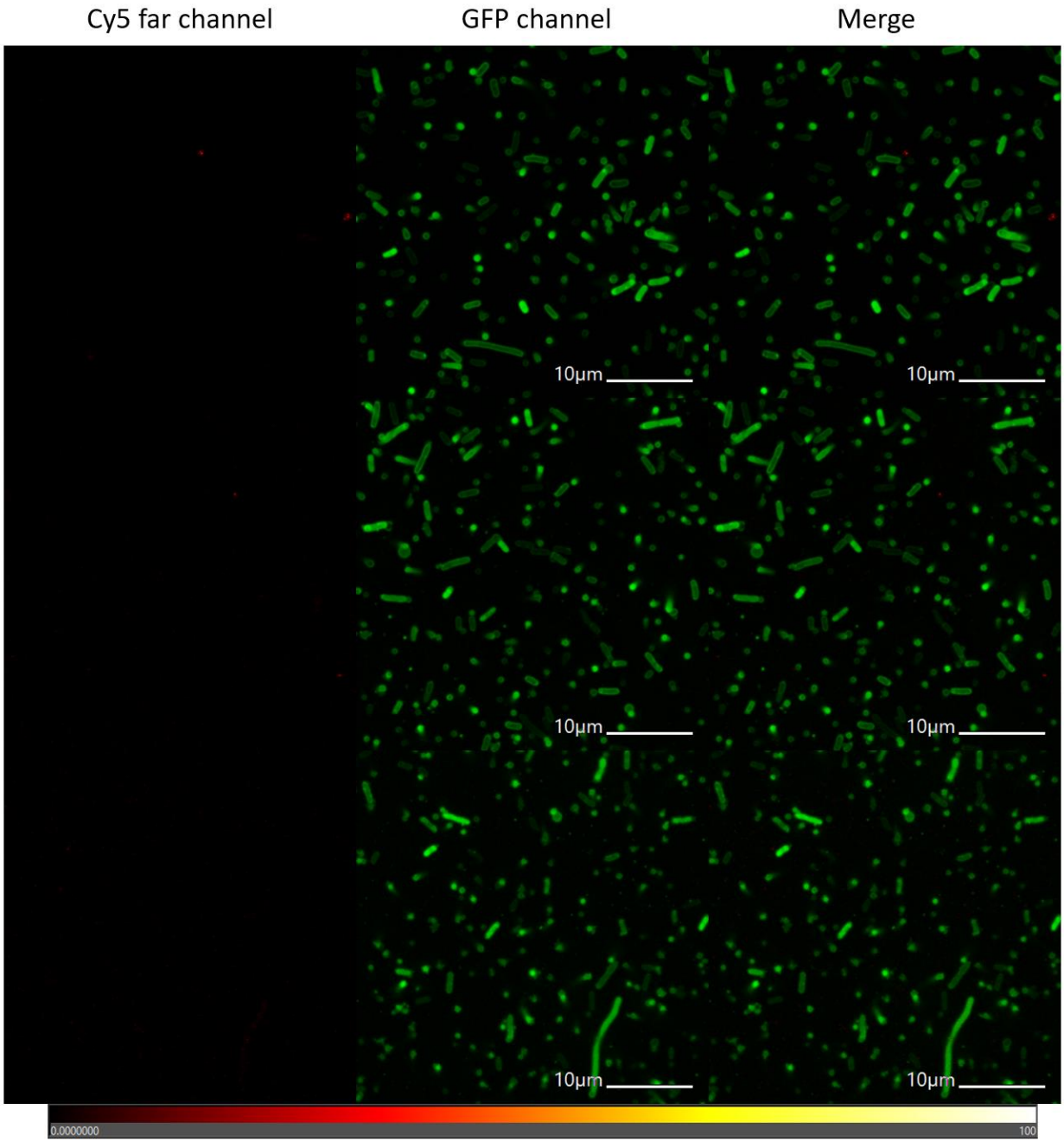

pLDDT ∈ [90–100]
pLDDT ∈ [70–90)
pLDDT ∈ [50–70)
pLDDT ∈ [0–50)
Pm KpsD octamer
Pm KpsD monomer
St KpsE octamer
St KpsE monomer
Pm Kpsm dimer
St KpsD octamer
St KpsD monomer

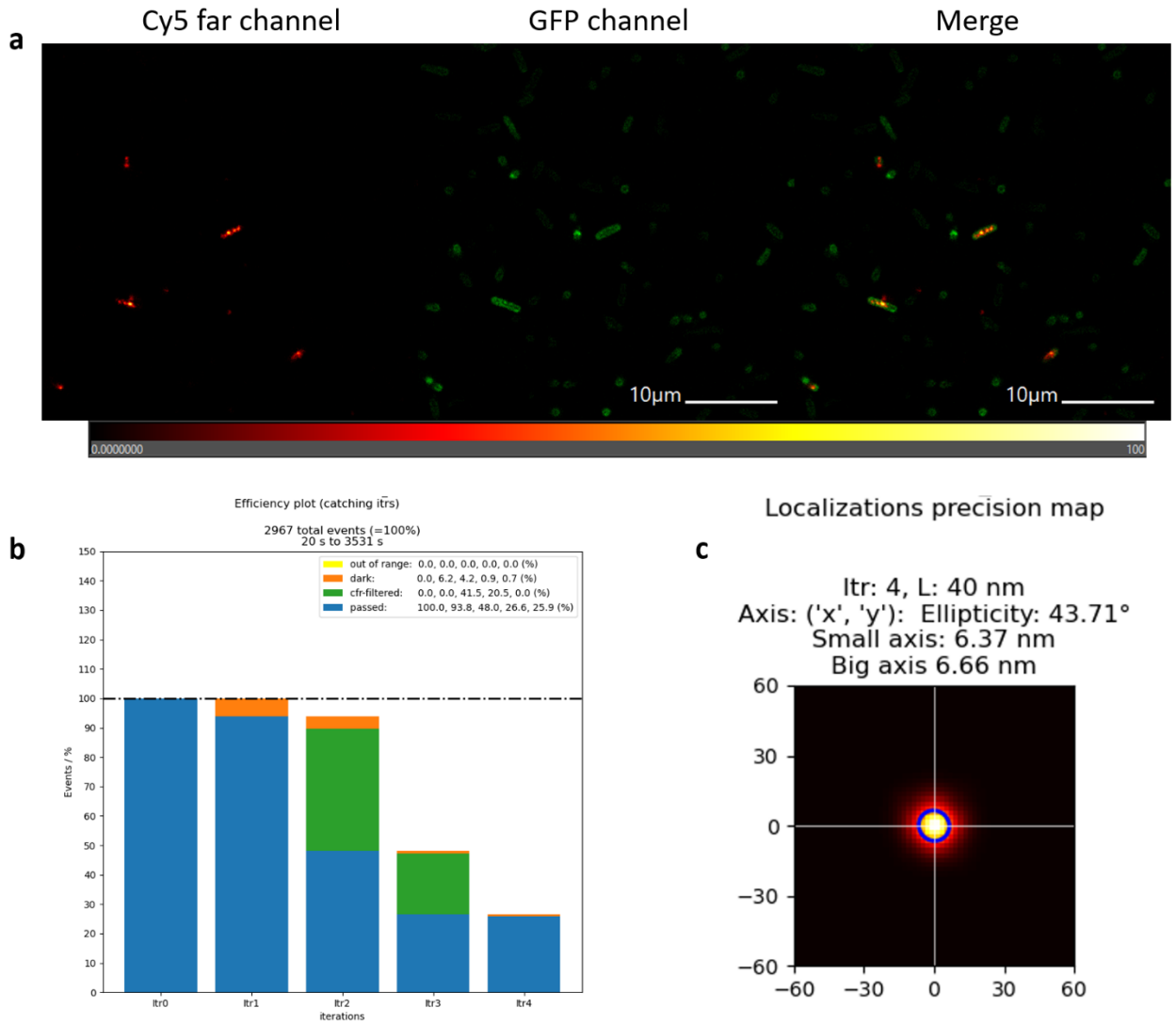
Fig. S3-6



Supplemental Figure 3-6 | ^{Alexa680}Streptavidin is not binding to cell not expressing Strep-tagged *Pm* KpsD. Here cells expressing *St* KpsMT-E-D were treated as cells in Fig 3-1 and S3-1. Little to no signal in Cy5 far channel was detected in three consecutive trials presented above. The intensity of signal detected on Cy5 channel is scaled to 100 for all of the presented examples.

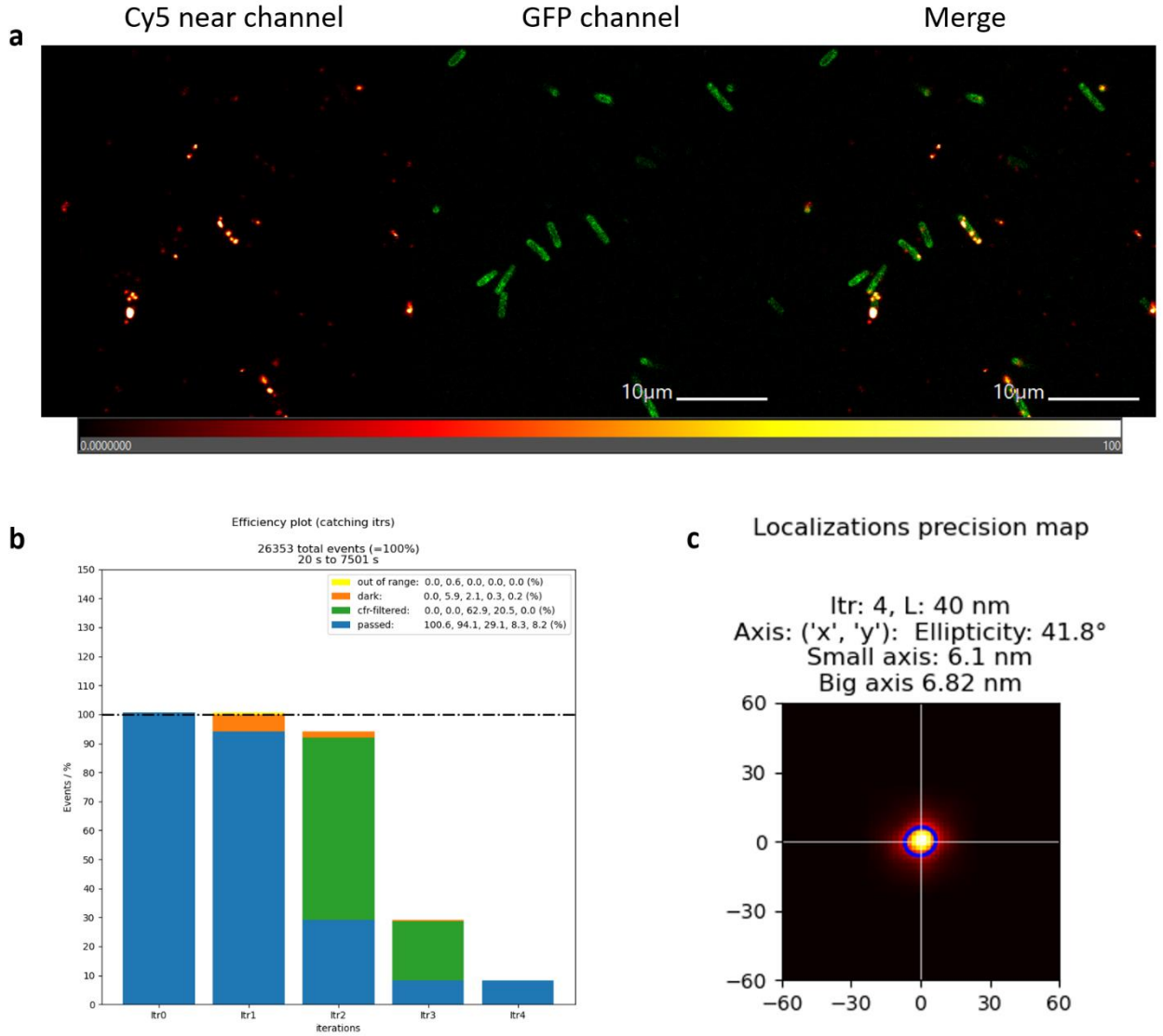
`

Fig. S3-7



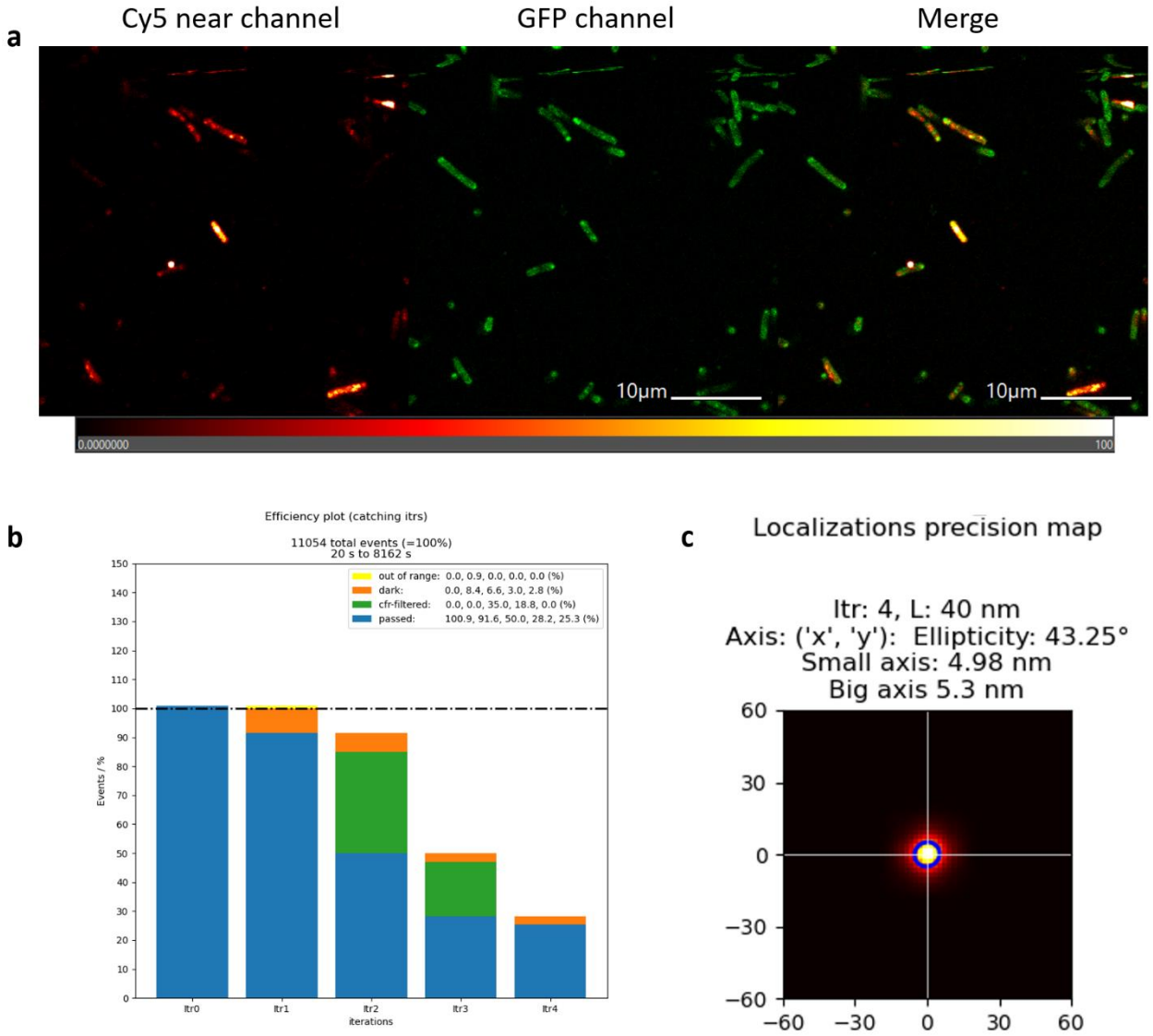
Supplemental Figure 3-7 | Minflux labeling of *Pm* KpsD.
(a) Confocal fields of view for single fluorophore experiments presented in Figure S3-1g. The intensity of signal detected on Cy5 channel is scaled to 100. **(b) (c)** Minflux dataset collection parameters for experiments presented in Figure S3-1g. **(b)** Efficiency plot of caught fluorophores. **(c)** Raw burst precision estimates.

Fig. S3-8



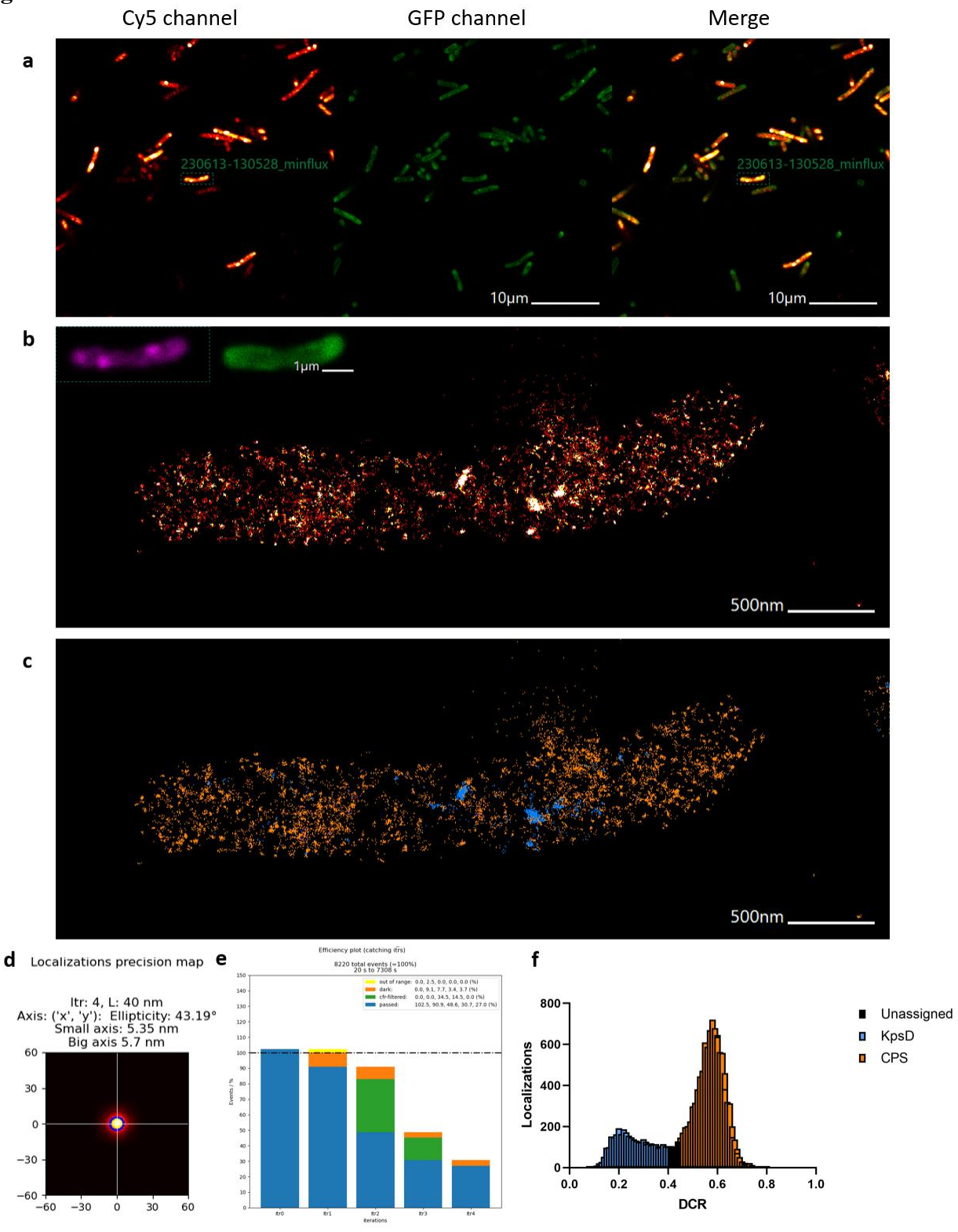
Supplemental Figure 3-8 | Minflux labeling of CPS.
(a) Confocal fields of view for single fluorophore experiments presented in Figure S3-1h. The intensity of signal detected on Cy5 channel is scaled to 100. **(b) (c)** Minflux dataset collection parameters for experiments presented in Figure S3-1h. **(b)** Efficiency plot of caught fluorophores. **(c)** Raw burst precision estimates.

Fig. S3-9



Supplemental Figure 3-9 | Minflux labeling of CPS and *Pm* KpsD.
(a) Confocal fields of view for single fluorophore experiments presented in Figure S3-1i. The intensity of signal detected on Cy5 channel is scaled to 100. **(b) (c)** Minflux dataset collection parameters for experiments presented in Figure S3-1i. **((b)** Efficiency plot of caught fluorophores. **(c)** Raw burst precision estimates.

Fig. S3-10


Cy5 channel
GFP channel
Merge
a
230613-130528_minflux
230613-130528_minflux
10µm
10µm
b
1µm
500nm
c
500nm
d
Localizations precision map
Itr: 4, L: 40 nm
Axis: ('x', 'y'): Ellipticity: 43.19°
Small axis: 5.35 nm
Big axis 5.7 nm
e
Efficiency plot (catching itrs)
8220 total events (=100%)
20 s to 7308 s
out of range: 0.0, 2.5, 0.0, 0.0, 0.0 (%)
dark: 0.0, 9.1, 7.7, 3.4, 3.7 (%)
cfr-filtered: 0.0, 0.0, 34.5, 14.5, 0.0 (%)
passed: 102.5, 90.9, 48.6, 30.7, 27.0 (%)
Events / %
Iterations
f
Localizations
DCR
Unassigned
KpsD
CPS

Supplemental Figure 3-10 | 2nd Example of 2 color Minflux dataset of a cell expressing „All" components and labeled as in Figure 3-1f. (a) Confocal field of view for this eperiment. **(b)** Minflux localizations of the combination of fluorophores, inlet: zoomed in confocal image of the cell used for Minflux. **(c)** Minflux localizations colored according to the DCR values. Localizations rendered with the pixel size based on the localization precision map **(d)**. **(e)** Efficiency plot of caught fluorophores. **(f)** DCR histogram.

Supplemental Figure 3-11 | 3rd example of 2 color Minflux dataset of a cell expressing „All" components and labeled as in Figure 3-1f. (a) Confocal field of view for this eperiment. **(b)** Minflux localizations of the combination of fluorophores, inlet: zoomed in confocal image of the cell used for Minflux. **(c)** Minflux localizations colored according to the DCR values. Localizations rendered with the pixel size based on the localization precision map **(d)**. **(e)** Efficiency plot of caught fluorophores. **(f)** DCR histogram. On the next page.

`

Fig. S3-11

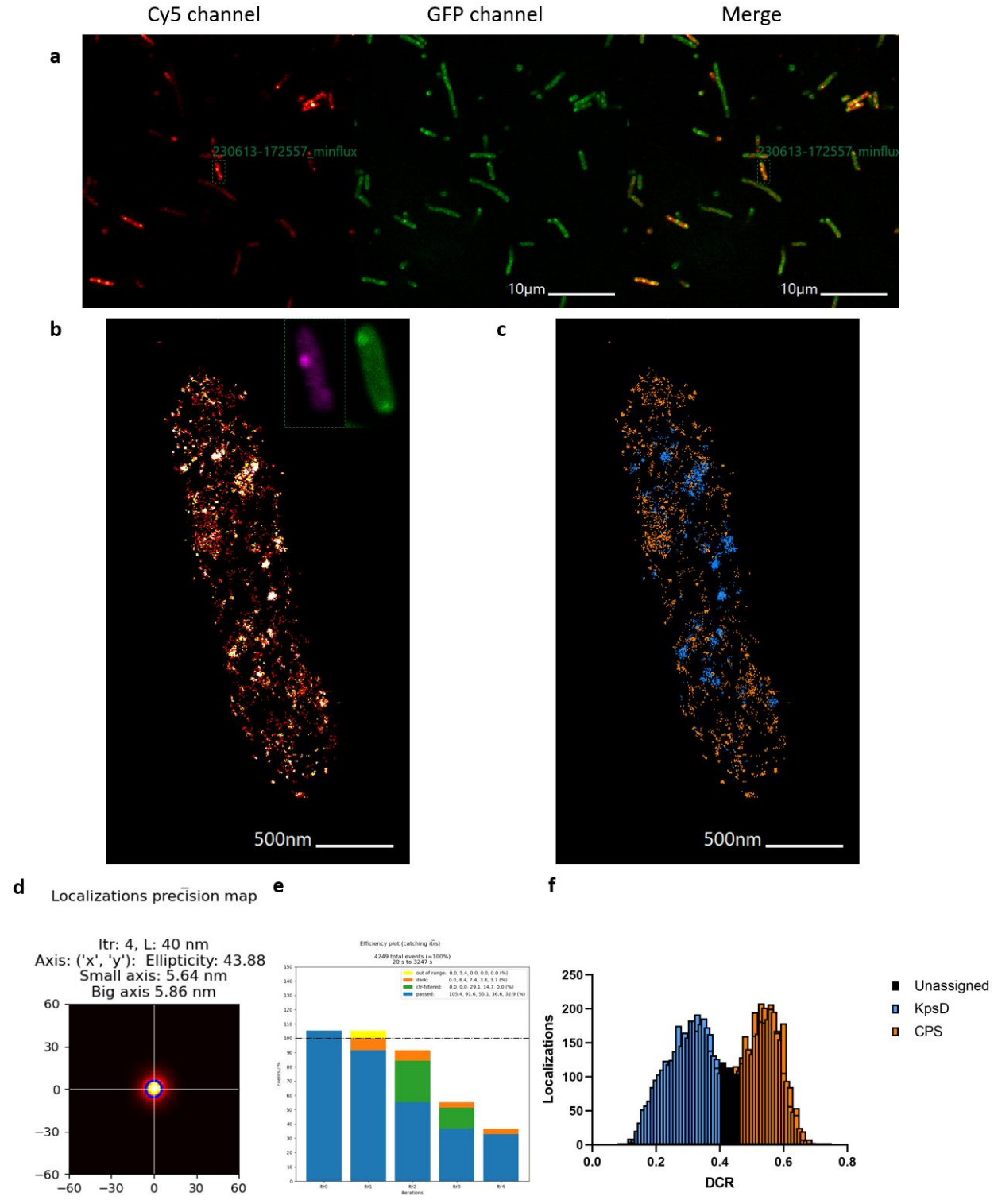
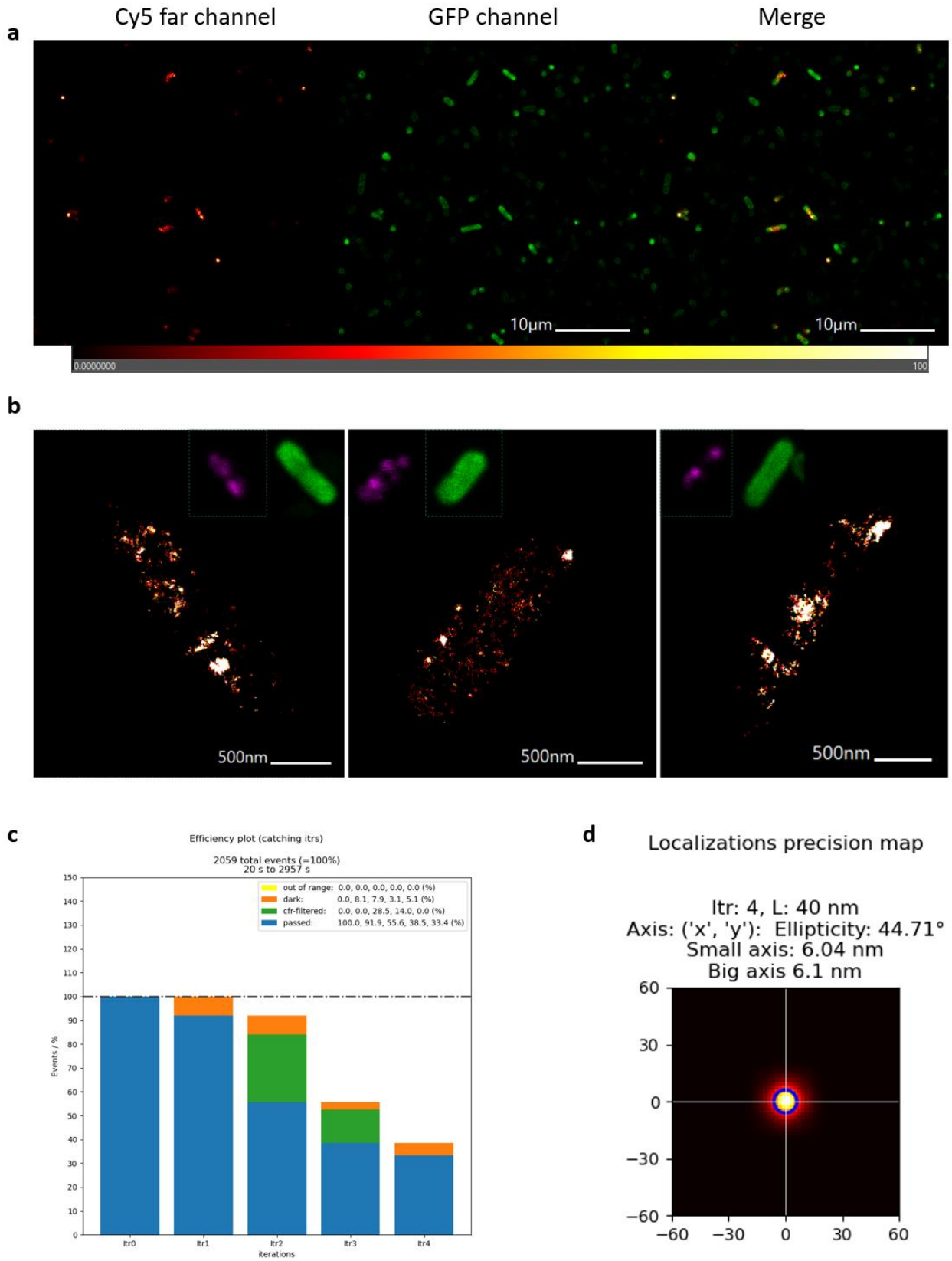

Cy5 channel
GFP channel
Merge
a
230613-172557_minflux
230613-172557_minflux
10µm
10µm
b
c
500nm
500nm
d
Localizations precision map
e
f
Itr: 4, L: 40 nm
Axis: ('x', 'y'): Ellipticity: 43.88
Small axis: 5.64 nm
Big axis 5.86 nm
Efficiency plot (catching itrs)
4249 total events (=100%)
20 s to 3247 s
Iterations
Localizations
DCR
Unassigned
KpsD
CPS

Fig. S3-12

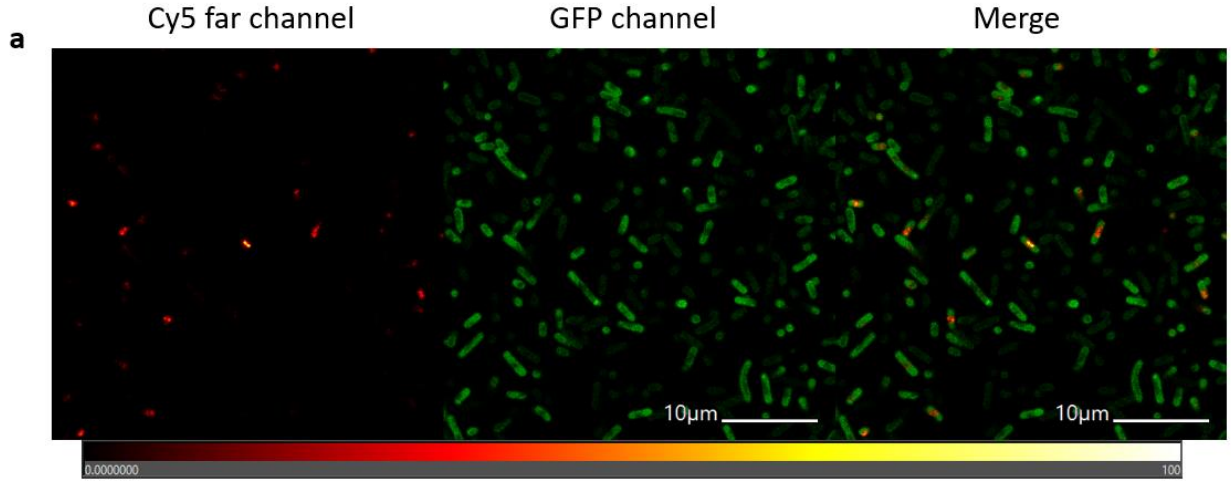

a
Cy5 far channel
GFP channel
Merge
10µm
10µm
0.0000000
100
b
500nm
500nm
500nm
c
Efficiency plot (catching itrs)
2059 total events (≈100%)
20 s to 2957 s
out of range: 0.0, 0.0, 0.0, 0.0, 0.0 (%)
dark: 0.0, 8.1, 7.9, 3.1, 5.1 (%)
cfr-filtered: 0.0, 0.0, 28.5, 14.0, 0.0 (%)
passed: 100.0, 91.9, 55.6, 38.5, 33.4 (%)
Events / %
itr0
itr1
itr2
itr3
itr4
iterations
d
Localizations precision map
Itr: 4, L: 40 nm
Axis: ('x', 'y'): Ellipticity: 44.71°
Small axis: 6.04 nm
Big axis 6.1 nm

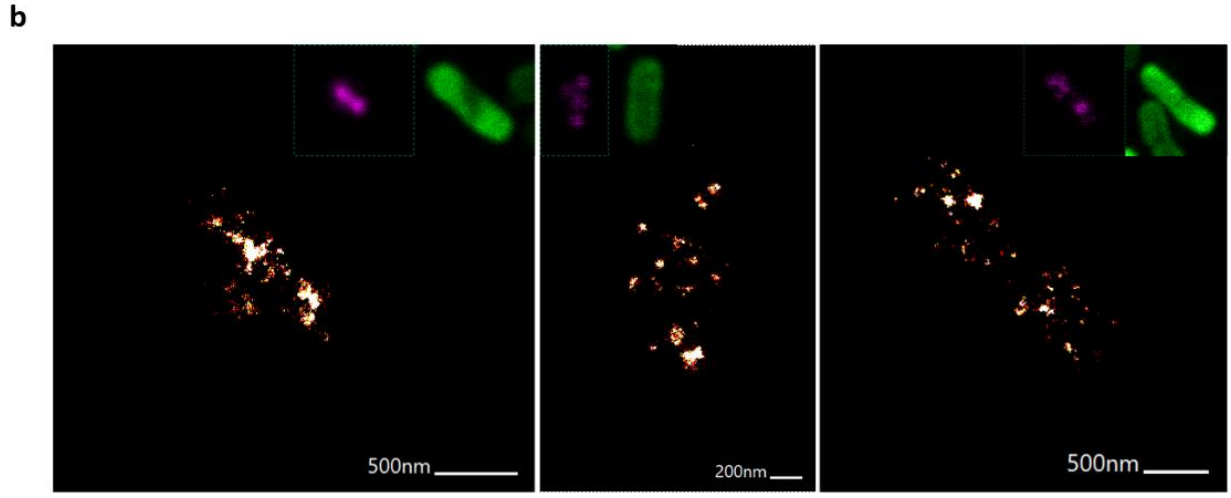
Supplemental Figure 3-12 | Minflux nanoscopy of ΔHyaD cells labeled with ^{Alexa680}Streptavidin. (a) Confocal fields of view for single fluorophore experiments done on ΔHyaD cells. The intensity of signal detected on Cy5 channel is scaled to 100. **(b)** Three examples of Minflux datasets collected on ΔHyaD cells labeled with ^{Alexa680}Streptavidin. **(c) (d)** Examples of Minflux dataset collection parameters for ΔHyaD datasets. **(c)** Raw burst precision estimates. **(d)** Efficiency plot of caught fluorophores.

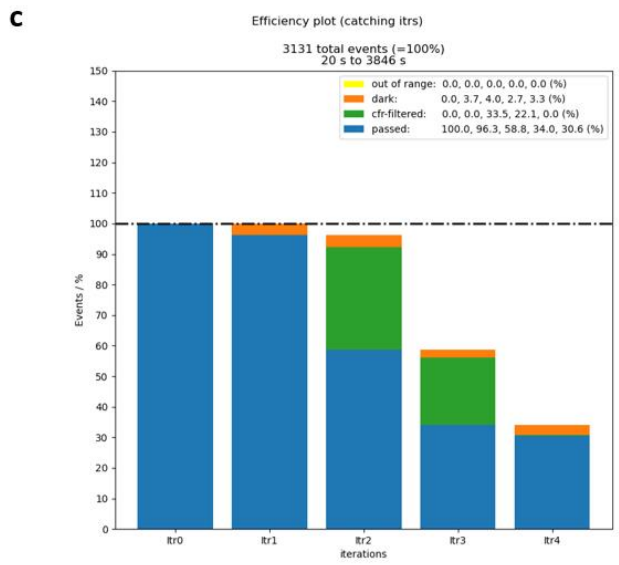
Supplemental Figure 3-13 | Minflux nanoscopy of ΔKpsE cells labeled with ^{Alexa680}Streptavidin. (a) Confocal fields of view for single fluorophore experiments done on ΔKpsE cells. The intensity of signal detected on Cy5 channel is scaled to 100. **(b)** Three examples of Minflux datasets collected on ΔKpsE cells labeled with ^{Alexa680}Streptavidin. **(c) (d)** Examples of Minflux dataset collection parameters for ΔKpsE datasets. **(c)** Raw burst precision estimates. **(d)** Efficiency plot of caught fluorophores. On the next page.

`

Fig. S3 -13


a
Cy5 far channel
GFP channel
Merge
10µm
10µm
0.0000000
100


b
500nm
200nm
500nm


c
Efficiency plot (catching itrs)
3131 total events (=100%)
20 s to 3846 s
out of range: 0.0, 0.0, 0.0, 0.0, 0.0 (%)
dark: 0.0, 3.7, 4.0, 2.7, 3.3 (%)
cfr-filtered: 0.0, 0.0, 33.5, 22.1, 0.0 (%)
passed: 100.0, 96.3, 58.8, 34.0, 30.6 (%)
Events / %
itr0
itr1
itr2
itr3
itr4
iterations

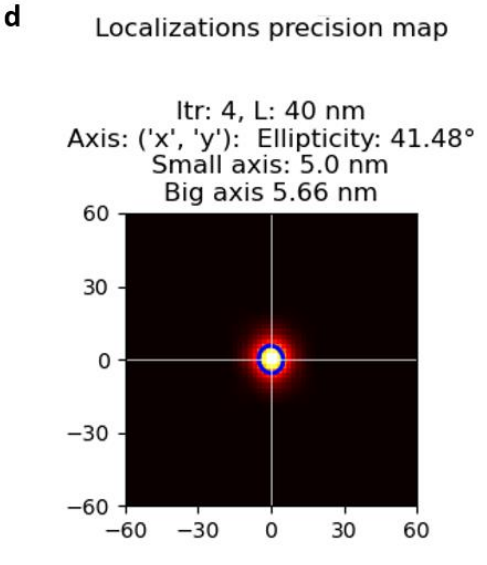
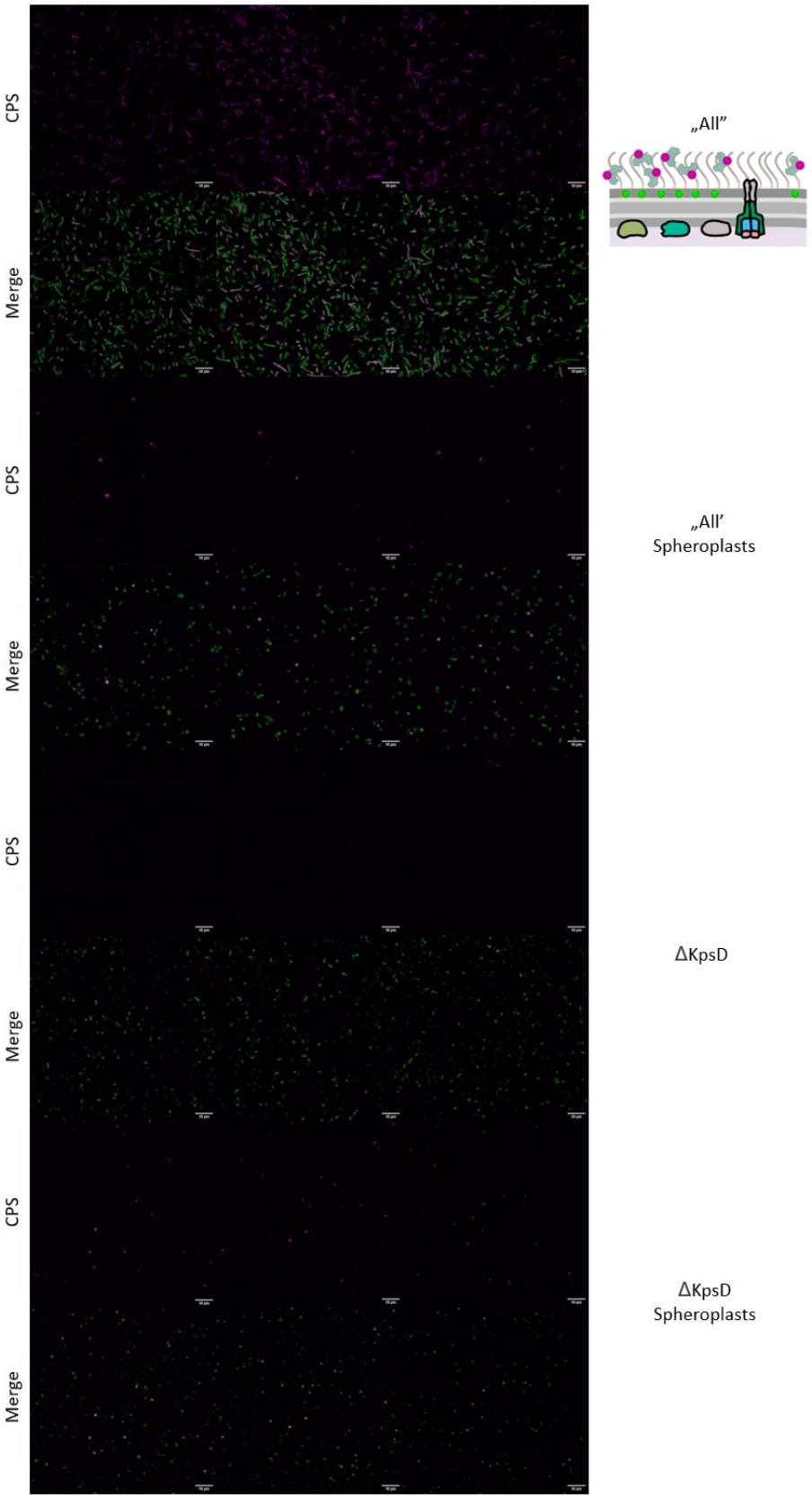

d
Localizations precision map
Itr: 4, L: 40 nm
Axis: ('x', 'y'): Ellipticity: 41.48°
Small axis: 5.0 nm
Big axis 5.66 nm

Fig. S3-14

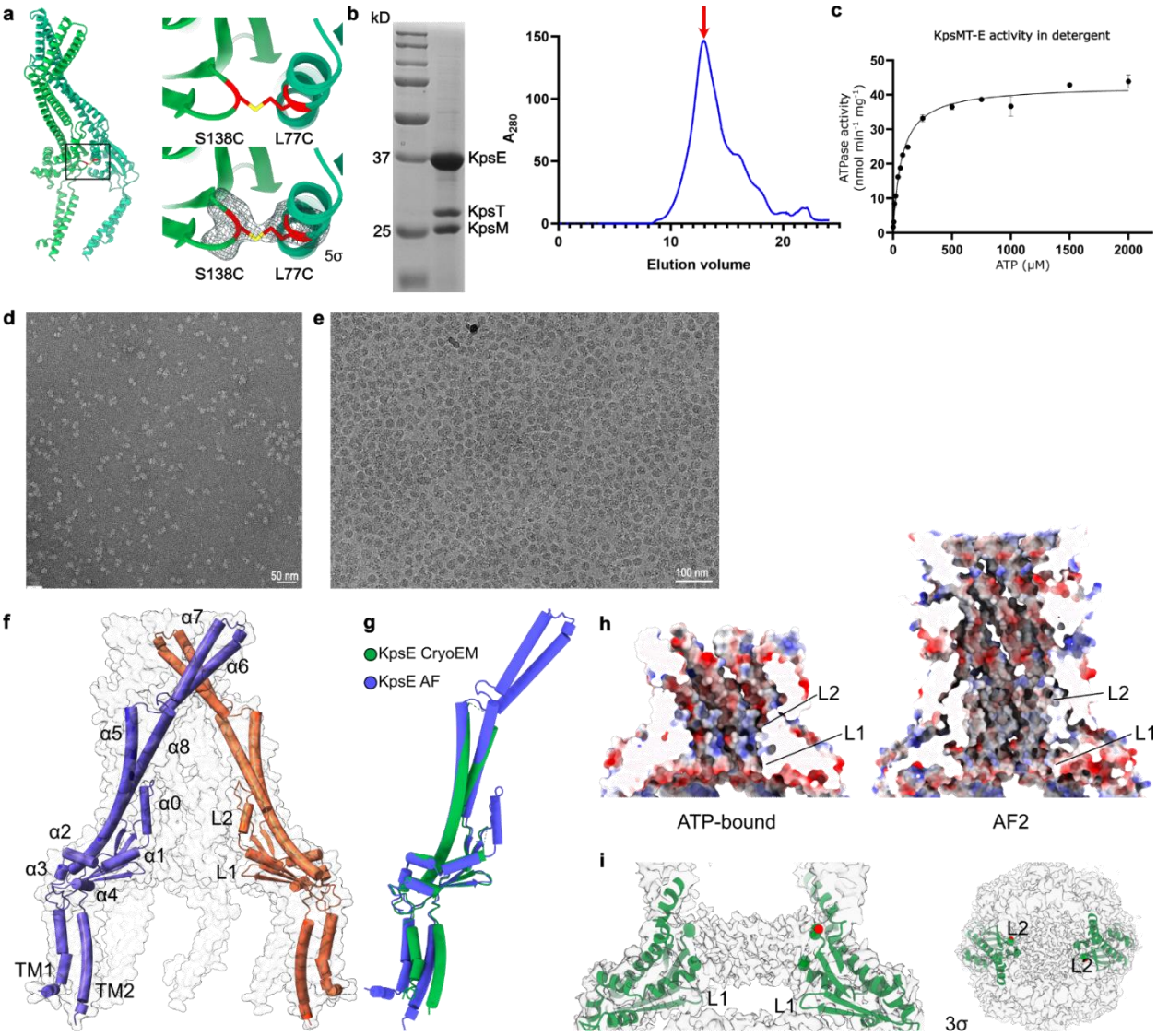

CPS
Merge
CPS
Merge
CPS
Merge
CPS
Merge
„All"
„All'
Spheroplasts
ΔKpsD
ΔKpsD
Spheroplasts

`

Supplemental Figure 3-14 | Full fields of view and additional examples of confocal imaging presented in Figure 3-1g. Scalebar: 10μm.

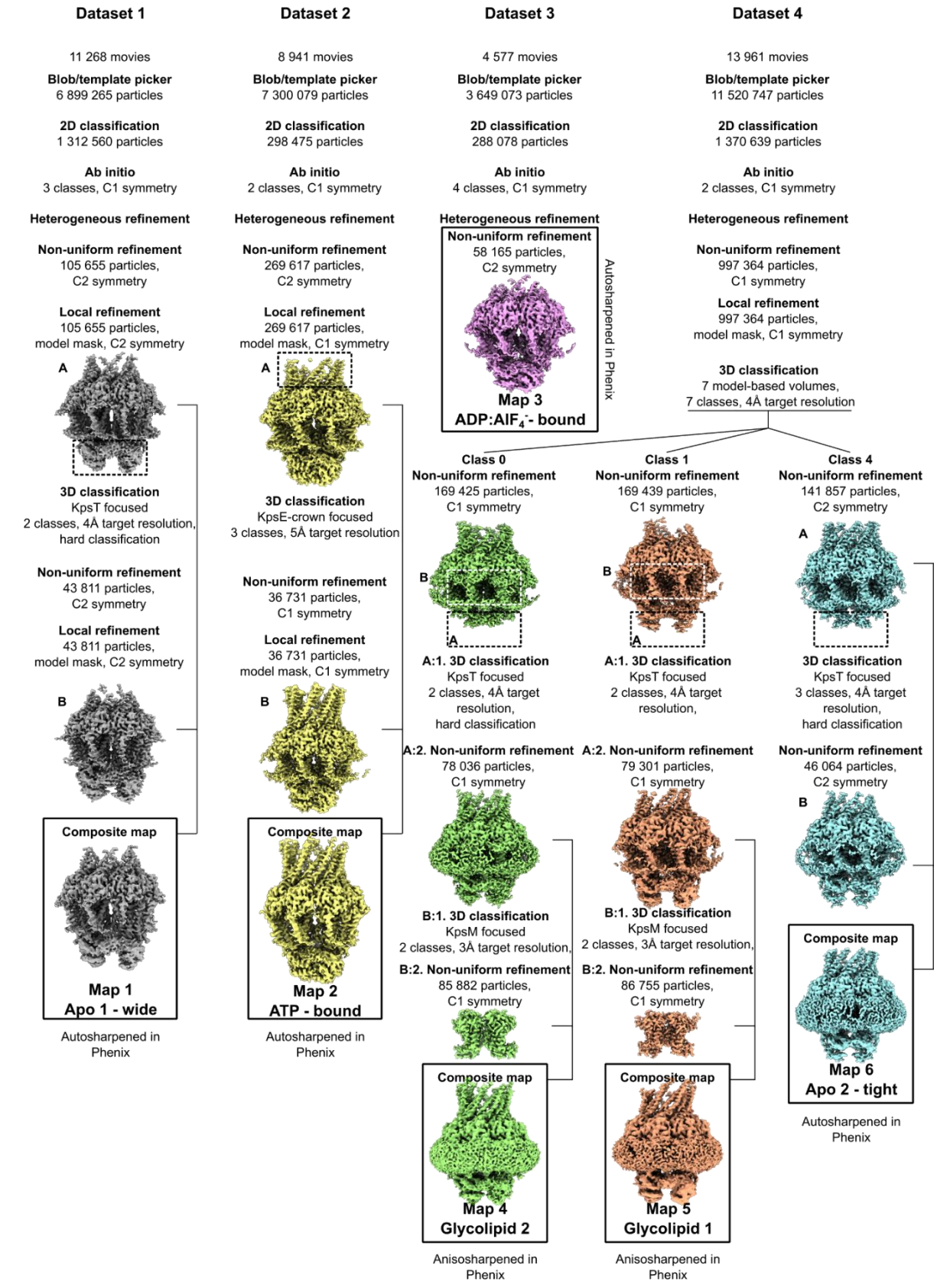
`

Fig. S3-15



Supplemental Figure 3-15 | KpsE stabilization, purification, cryo grid preparation and KpsE AlphaFoldF2 prediction. (a) Cartoon representation of an AlphaFold2 KpsE dimer model used to engineer an intermolecular disulfide bridge. Electron density of the engineered disulfide bond observed in the cryoEM map of the ATP-bound state. **(b)** Size exclusion chromatography profile of the KpsMT-KpsE complex (right) and peak fraction run on a reducing SDS PAGE Coomasie stained gel (left). **(c)** ATPase activity of the purified KpsEMT complex determined by monitoring the release of ADP via an enzyme coupled assay. **(d)** and **(e)** Negative stain and cryoEM micrographs of the KpsMT-KpsE complex. **(f)** An AlphaFold2 predicted full length St KpsE model indicating the localizations of three missing helices: α0 within the L2 loop and α6 and α7 of the crown. **(g)** Overlay of cartoon representations of the AlphaFold2 predicted and experimental KpsE structures. **(h)** Coulomb surface potential of the octameric KpsE complex calculated in ChimeraX [46] (red: -10, blue:10 kcal/(mol·e)). **(i)** Disordered density of the L2 loop visible at low contour levels.

Fig. S3-16


Dataset 1
11 268 movies
Blob/template picker
6 899 265 particles
2D classification
1 312 560 particles
Ab initio
3 classes, C1 symmetry
Heterogeneous refinement
Non-uniform refinement
105 655 particles,
C2 symmetry
Local refinement
105 655 particles,
model mask, C2 symmetry
A
3D classification
KpsT focused
2 classes, 4Å target resolution,
hard classification
Non-uniform refinement
43 811 particles,
C2 symmetry
Local refinement
43 811 particles,
model mask, C2 symmetry
B
Composite map
Map 1
Apo 1 - wide
Autosharpened in
Phenix
Dataset 2
8 941 movies
Blob/template picker
7 300 079 particles
2D classification
298 475 particles
Ab initio
2 classes, C1 symmetry
Heterogeneous refinement
Non-uniform refinement
269 617 particles,
C2 symmetry
Local refinement
269 617 particles,
model mask, C1 symmetry
A
3D classification
KpsE-crown focused
3 classes, 5Å target resolution
Non-uniform refinement
36 731 particles,
C1 symmetry
Local refinement
36 731 particles,
model mask, C1 symmetry
B
Composite map
Map 2
ATP - bound
Autosharpened in
Phenix
Dataset 3
4 577 movies
Blob/template picker
3 649 073 particles
2D classification
288 078 particles
Ab initio
4 classes, C1 symmetry
Heterogeneous refinement
Non-uniform refinement
58 165 particles,
C2 symmetry
Map 3
ADP:AlF4--bound
Autosharpened in Phenix
Class 0
Non-uniform refinement
169 425 particles,
C1 symmetry
B
A
A:1. 3D classification
KpsT focused
2 classes, 4Å target
resolution,
hard classification
A:2. Non-uniform refinement
78 036 particles,
C1 symmetry
B:1. 3D classification
KpsM focused
2 classes, 3Å target resolution,
B:2. Non-uniform refinement
85 882 particles,
C1 symmetry
Composite map
Map 4
Glycolipid 2
Anisosharpened in
Phenix
Dataset 4
13 961 movies
Blob/template picker
11 520 747 particles
2D classification
1 370 639 particles
Ab initio
2 classes, C1 symmetry
Heterogeneous refinement
Non-uniform refinement
997 364 particles,
C1 symmetry
Local refinement
997 364 particles,
model mask, C1 symmetry
3D classification
7 model-based volumes,
7 classes, 4Å target resolution
Class 1
Non-uniform refinement
169 439 particles,
C1 symmetry
B
A
A:1. 3D classification
KpsT focused
2 classes, 4Å target
resolution,
A:2. Non-uniform refinement
79 301 particles,
C1 symmetry
B:1. 3D classification
KpsM focused
2 classes, 3Å target resolution,
B:2. Non-uniform refinement
86 755 particles,
C1 symmetry
Composite map
Map 5
Glycolipid 1
Anisosharpened in
Phenix
Class 4
Non-uniform refinement
141 857 particles,
C2 symmetry
A
3D classification
KpsT focused
3 classes, 4Å target
resolution,
hard classification
Non-uniform refinement
46 064 particles,
C2 symmetry
B
Composite map
Map 6
Apo 2 - tight
Autosharpened in
Phenix

Supplemental Figure 3-16 | CryoEM data processing workflows.
Datasets 1,2 and 3 were processed in cryoSPARC version 3.3. Dataset 4 was processed in cryoSPARC version 4.0.

Supplemental Figure 3-17 | Local resolution estimates and map quality examples for the Apo 1 and ATP-bound KpsMT-KpsE complexes. (a) Local resolution estimate of Map 1 (Apo 1 state). Two copies of KpsE were removed to show the KpsM volume. **(b)** Gold-standard Fourier shell correlation (FSC) indicating the overall Map 1 resolution of 3.1 Å at FSC=0.143. **(c)** Examples of Map 1 for the indicated segments contoured to 5σ. **(d)** Local resolution estimate of Map 2 representing the ATP-bound state of $KpsMT_{E151Q}$-E. Two copies of KpsE were removed to show the KpsM volume. **(e)** Gold-standard Fourier shell correlation (FSC) indicating an overall Map 2 resolution of 3.1 Å at FSC=0.143. (**f**) Cryo-EM maps of the ATP:Mg^{2+} molecules. (**g**) Examples of Map 2 for the indicated segments contoured to 5σ, unless specified otherwise. On the next page.

`

Fig. S3-17

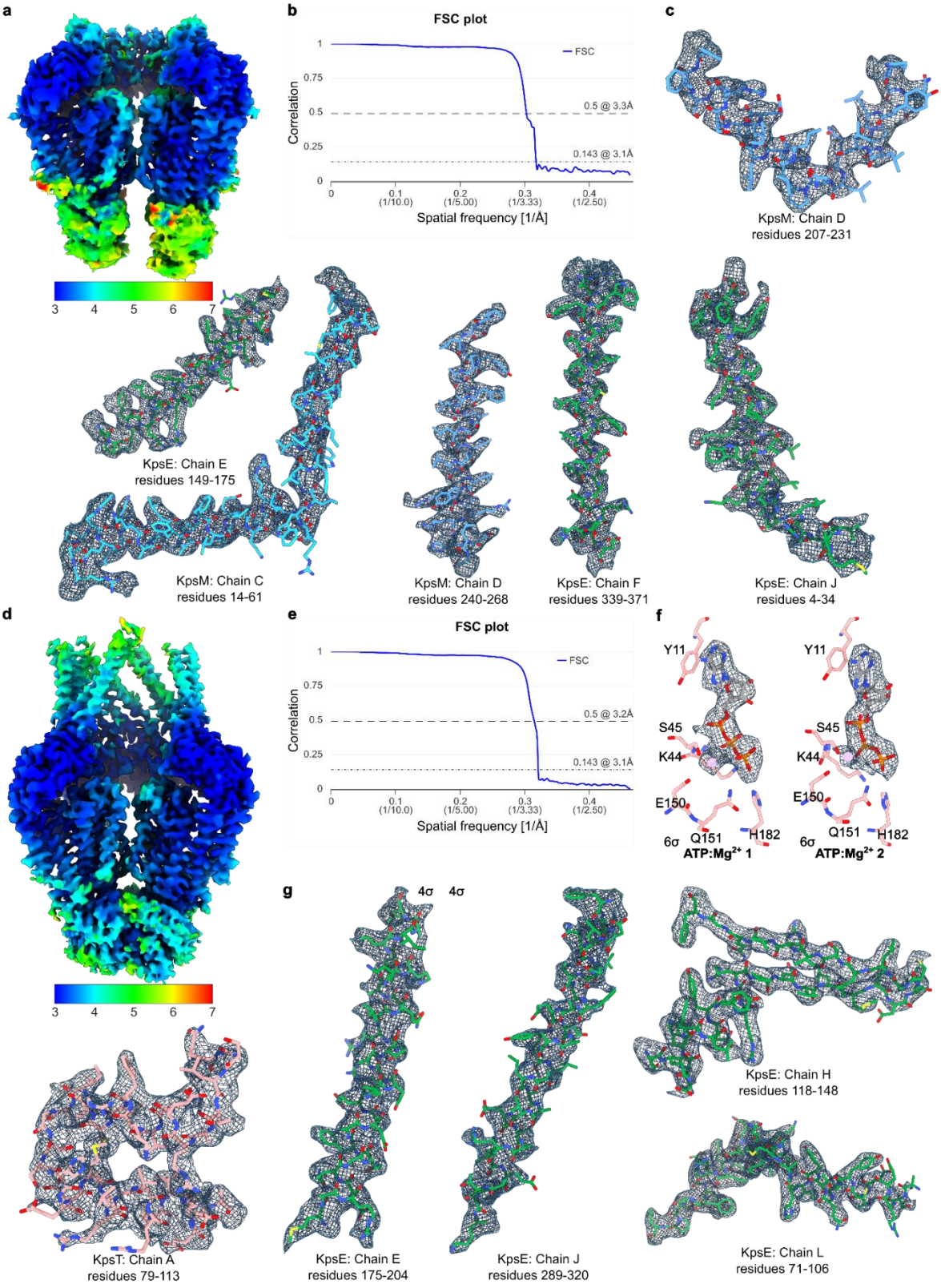
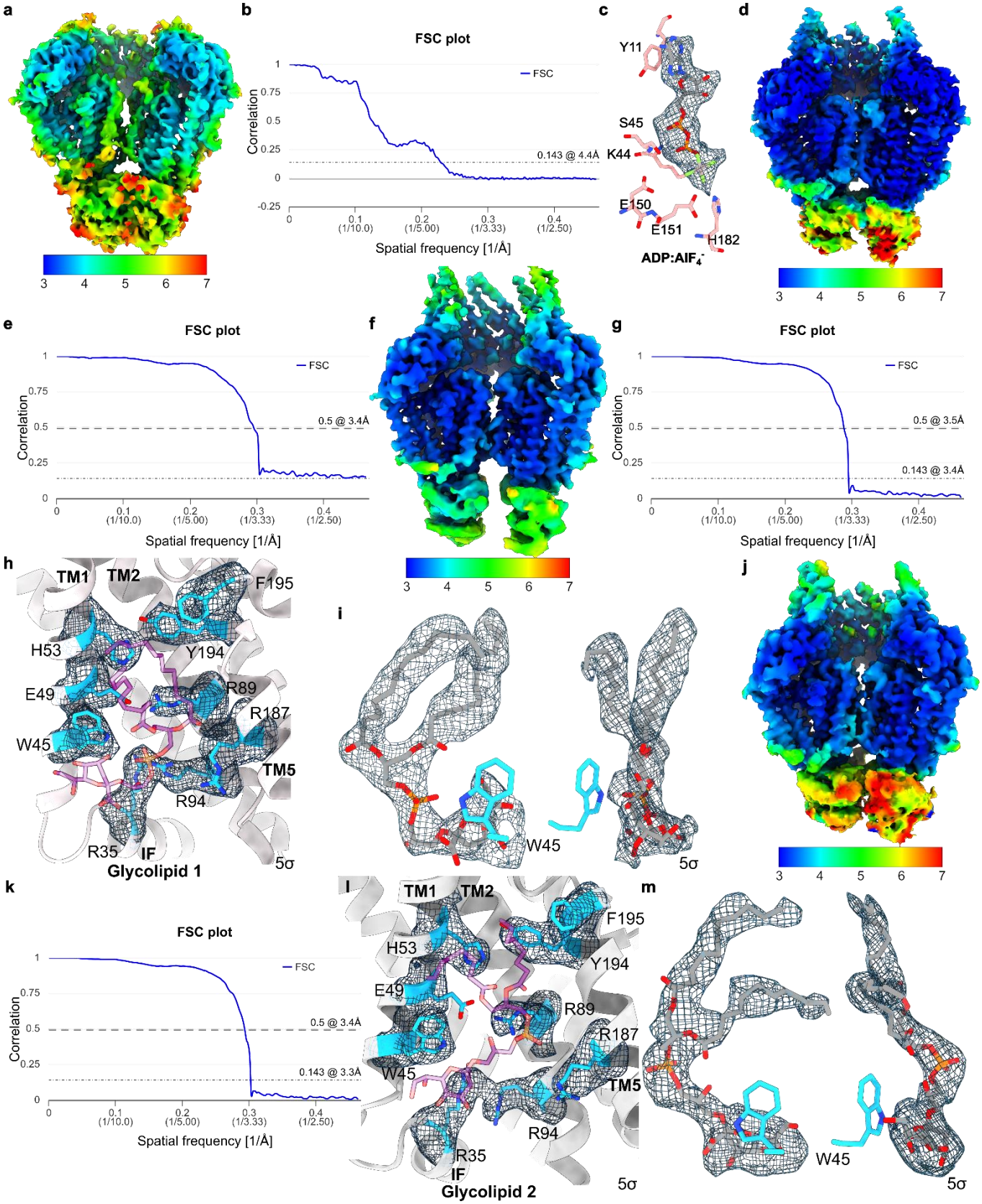

a
3 4 5 6 7
b
FSC plot
FSC
0.5 @ 3.3Å
0.143 @ 3.1Å
Correlation
Spatial frequency [1/Å]
c
KpsM: Chain D
residues 207-231
KpsE: Chain E
residues 149-175
KpsM: Chain C
residues 14-61
KpsM: Chain D
residues 240-268
KpsE: Chain F
residues 339-371
KpsE: Chain J
residues 4-34
d
3 4 5 6 7
e
FSC plot
FSC
0.5 @ 3.2Å
0.143 @ 3.1Å
Correlation
Spatial frequency [1/Å]
f
Y11
S45
K44
E150
Q151
H182
6σ
ATP:Mg2+ 1
ATP:Mg2+ 2
g
4σ
4σ
KpsT: Chain A
residues 79-113
KpsE: Chain E
residues 175-204
KpsE: Chain J
residues 289-320
KpsE: Chain H
residues 118-148
KpsE: Chain L
residues 71-106

Fig. S3-18


a
b
FSC plot
FSC
Correlation
0.143 @ 4.4Å
Spatial frequency [1/Å]
c
Y11
S45
K44
E150
E151
H182
ADP:AlF4-
d
3 4 5 6 7
e
FSC plot
0.5 @ 3.4Å
f
g
FSC plot
0.5 @ 3.5Å
0.143 @ 3.4Å
h
TM1
TM2
F195
H53
Y194
E49
R89
R187
W45
TM5
R94
R35
IF
Glycolipid 1
5σ
i
W45
5σ
j
k
FSC plot
0.5 @ 3.4Å
0.143 @ 3.3Å
l
TM1
TM2
F195
H53
Y194
E49
R89
R187
W45
TM5
R94
R35
IF
Glycolipid 2
5σ
m
W45
5σ

Supplemental Figure 3-18 | Local resolution estimates and map quality examples for the ADP:AlF4- and glycolipid bound states. **(a)** Local resolution estimate for Map 3 representing KpsMT-KpsE bound to ADP:AlF_4^-. Two copies of KpsE were removed to show the KpsM volume. **(b)** Gold-standard Fourier shell correlation (FSC), indicating an overall Map 3 resolution of about 4.4 Å. **(c)** CryoEM maps of the ADP:AlF_4^- complexes in Map 3. ALF_4^- is colored silver/green. **(d)** Local resolution estimation of Map 6 representing KpsMT-KpsE in the Apo 2 state. Two copies of KpsE were removed to show the KpsM volume. **(e)** Fourier shell correlation (FSC), indicating an overall Map 6 resolution of 3.4 Å at FSC=0.5. **(f)** Local resolution estimate of Map 5, KpsMT-KpsE complex bound to Glycolipid 1. Two copies of KpsE were removed to show the KpsM volume. **(g)** Gold-standard Fourier shell correlation (FSC), indicating an overall Map 5 resolution of 3.4 Å at FSC=0.143. **(h)** CryoEM maps of residues proximal to Glycolipid 1. **(i)** CryoEM map of Glycolipid 1. **(j)** Local resolution estimate of Map 4, KpsMT-KpsE complex bound to Glycolipid 2. Two copies of KpsE were removed to show the KpsM volume. (**k)** Gold-standard Fourier shell correlation (FSC), indicating an overall Map 4 resolution of 3.3 Å at FSC=0.143. **(l)** CryoEM density maps of residues proximal to Glycolipid 2. **(m)** CryoEM map of Glycolipid 2.

Supplemental Figure 3-19 | Comparison of the KpsMT architecture with other ABC transporters. **(a)** and **(b)** KpsMT architecture in ATP-bound, ADP:AlF_4^--bound, and Apo 2 states **(a)** and corresponding coulomb surface potentials (red: -10, blue:10 kcal/(mol·*e*)) **(b)**. The polysaccharide canyon region is highlighted by a red box. TM: transmembrane helix, PG: periplasmic gate helix, IF: interface helix. **(c)** Comparison of the KpsMT TM domain organization of KpsMT and other type-5 ABC transporters of known structure. GH: gate helix. On the next page.

`

Fig. S3-19

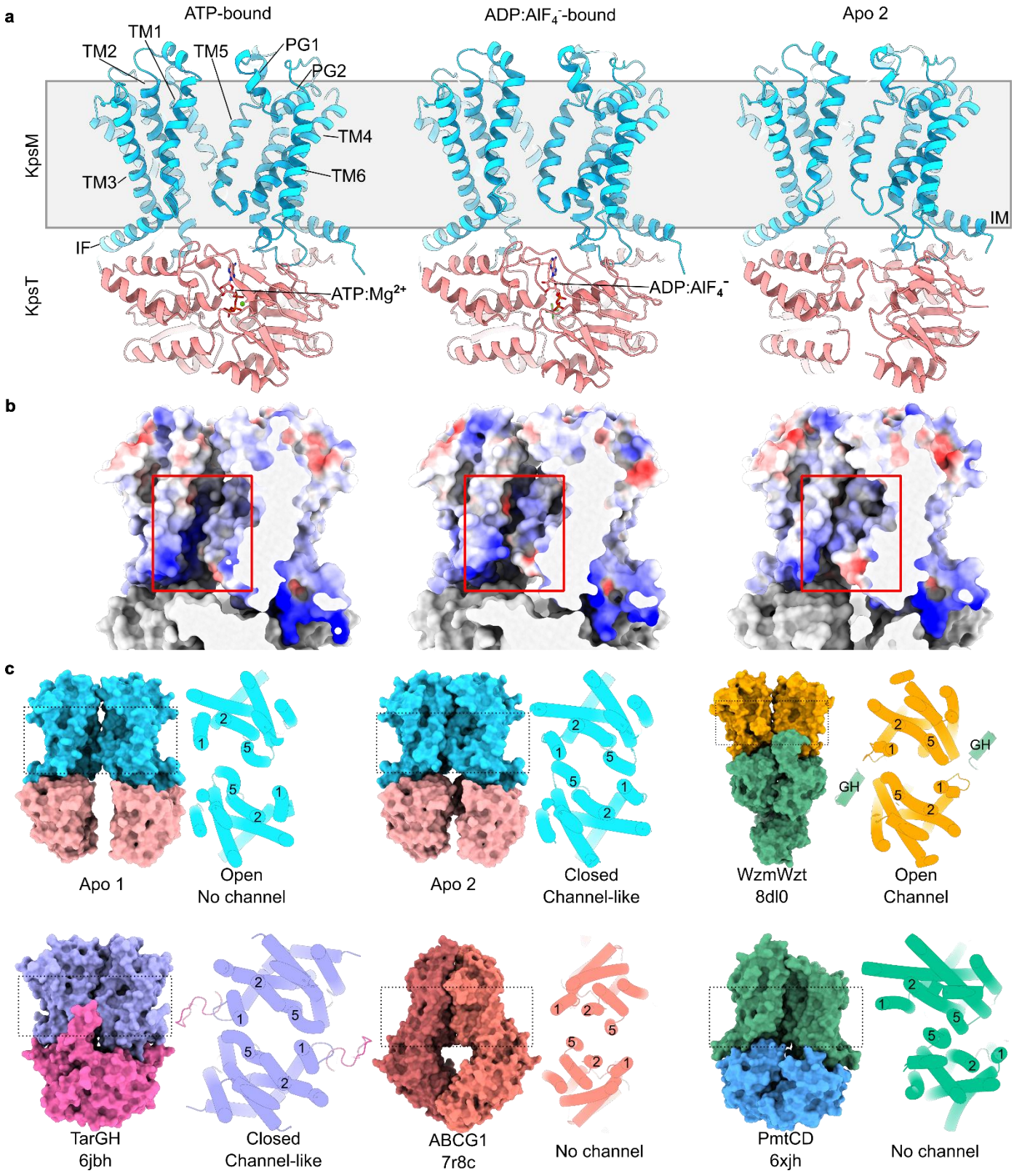
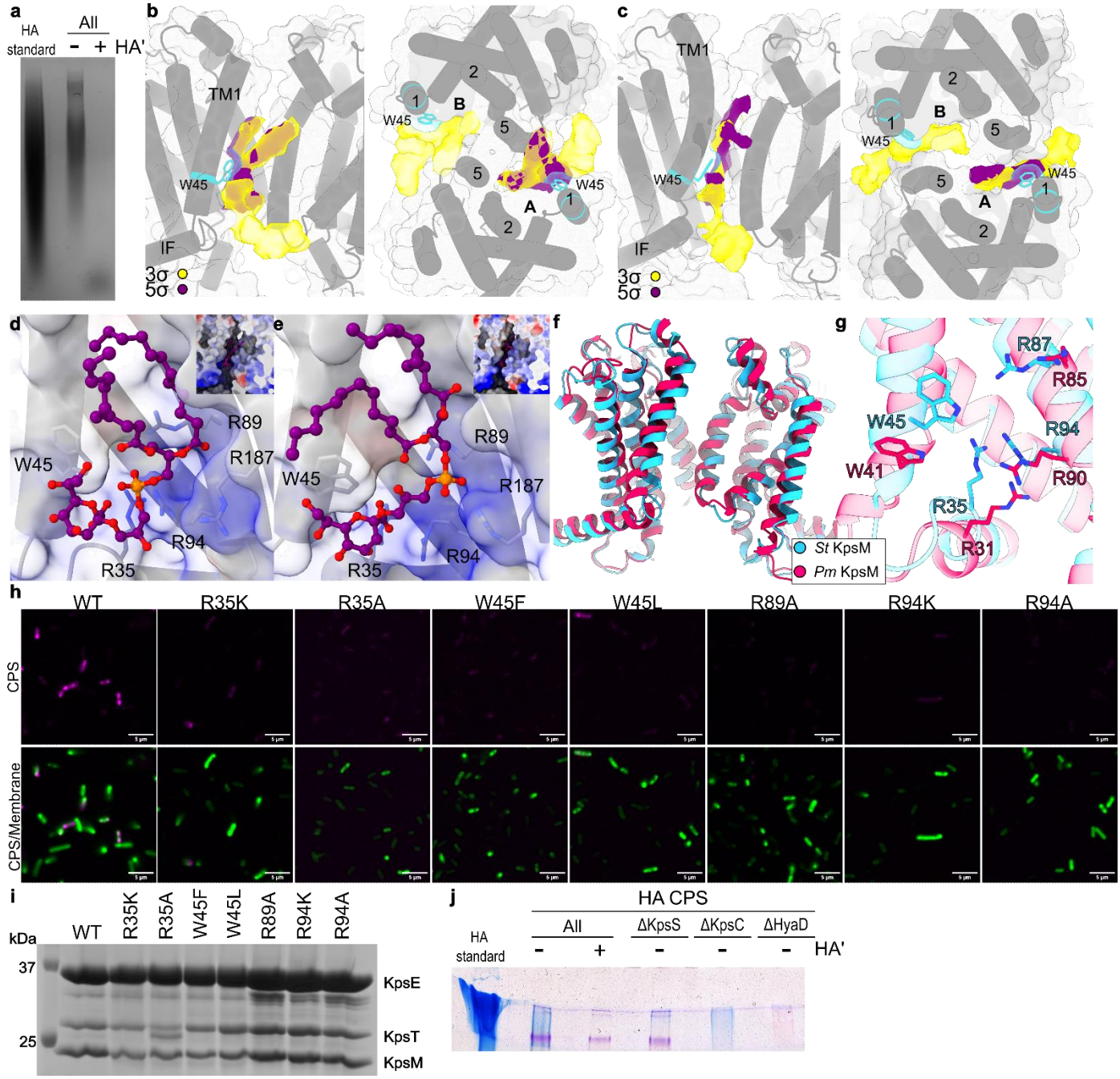

a
ATP-bound
ADP:AlF4-bound
Apo 2
TM1
TM2
TM5
PG1
PG2
TM4
TM6
TM3
KpsM
IM
IF
KpsT
ATP:Mg2+
ADP:AlF4-
b
c
Apo 1
Open
No channel
Apo 2
Closed
Channel-like
WzmWzt
8dl0
GH
Open
Channel
TarGH
6jbh
Closed
Channel-like
ABCG1
7r8c
No channel
PmtCD
6xjh
No channel

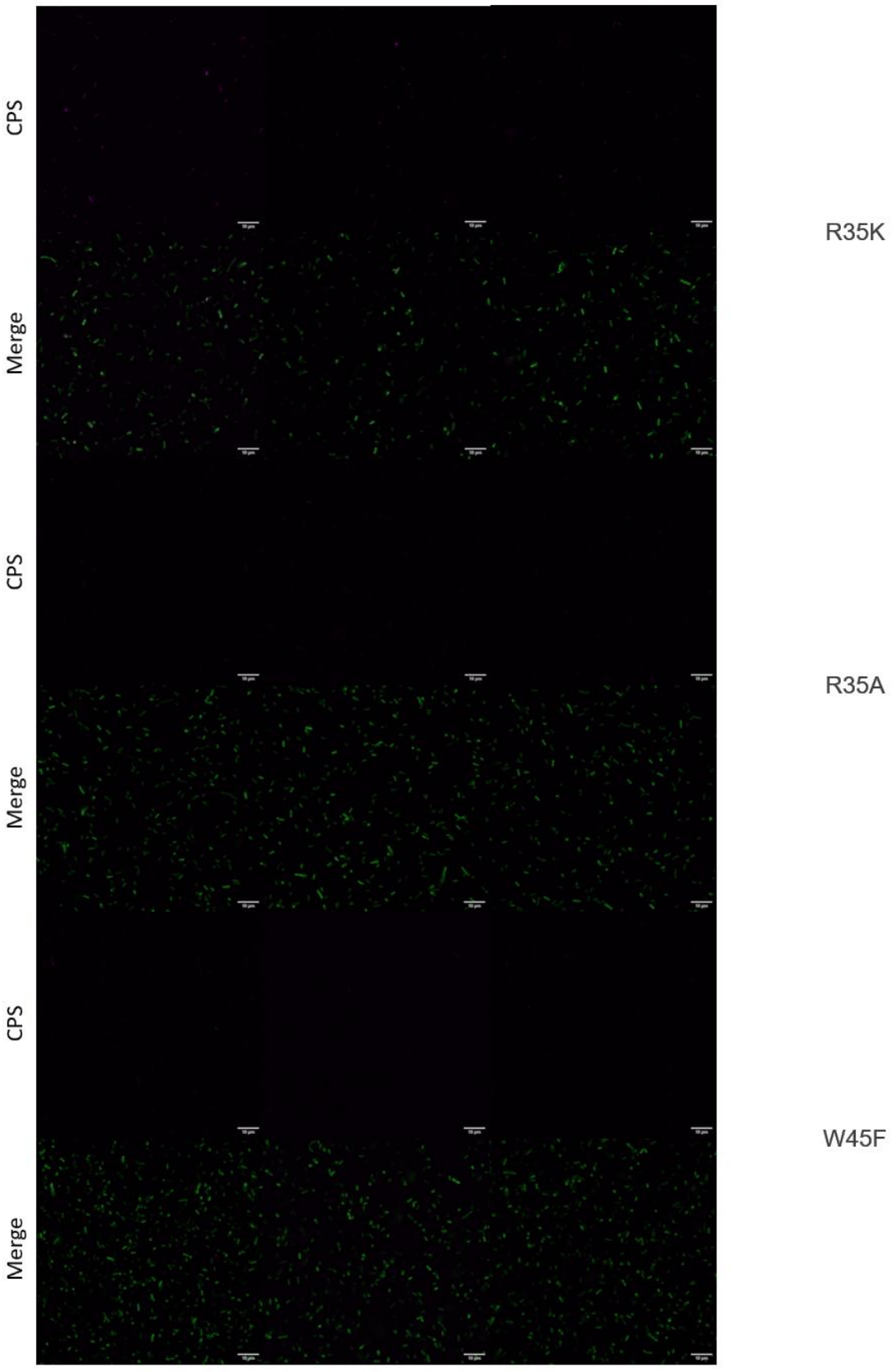
Fig. S3-20


a
HA standard
All
- + HA'
b
TM1
W45
IF
3σ
5σ
B
A
c
d
e
R89
R187
R94
R35
f
St KpsM
Pm KpsM
g
R87
R85
R94
R90
W45
W41
R35
R31
h
WT
R35K
R35A
W45F
W45L
R89A
R94K
R94A
CPS
CPS/Membrane
i
kDa
37
25
KpsE
KpsT
KpsM
j
HA CPS
All
ΔKpsS
ΔKpsC
ΔHyaD
HA standard
HA'

Supplemental Figure 3-20 | Glycolipid preparation and canyon mutagenesis. (a) Agarose gel stained with Stains-All dye of recombinantly expressed and purified *Pm* CPS, run against a 300-500 kDa HA standard. HA': hyaluronidase treatment, 'All': expression of all *Pm CPS* components in *E. coli* C43ΔCPS1. **(b)** and **(c)** Low contour level cryoEM volumes (yellow) extending past the glycolipid's putative first Kdo moiety (shown at high contour level in magenta). The low contour map could accommodate 2-3 additional Kdo units. **(d)** and **(e)** Additional views of the glycolipid in states 1 and 2 inside the positively charged canyon. **(f)** Overlay of the experimentally determined structure of *St* KpsM – Apo 1 with an AlphaFoldF2 predicted *Pm* KpsM model. **(g)** Overlay of residues contributing to the polysaccharide canyon, underscoring the conservation of crucial residues. **(h)** *In vivo* encapsulation assay of KpsM mutants analyzed by confocal fluorescence microscopy. CPS is labeled with the ^{Alexa647}SNAP-CBM70 probe, membranes are labeled with Cellbrite Fix 488. **(i)** SDS PAGE and Coomassie staining of purified KpsMT-KpsE complex mutants used in panel i. **(j)** Polyacrylamide gel stained with Stains-All dye of recombinantly expressed *Pm* CPS, run against 300-500 kDa HA standard. Components omitted during CPS expression are indicated above the lanes.

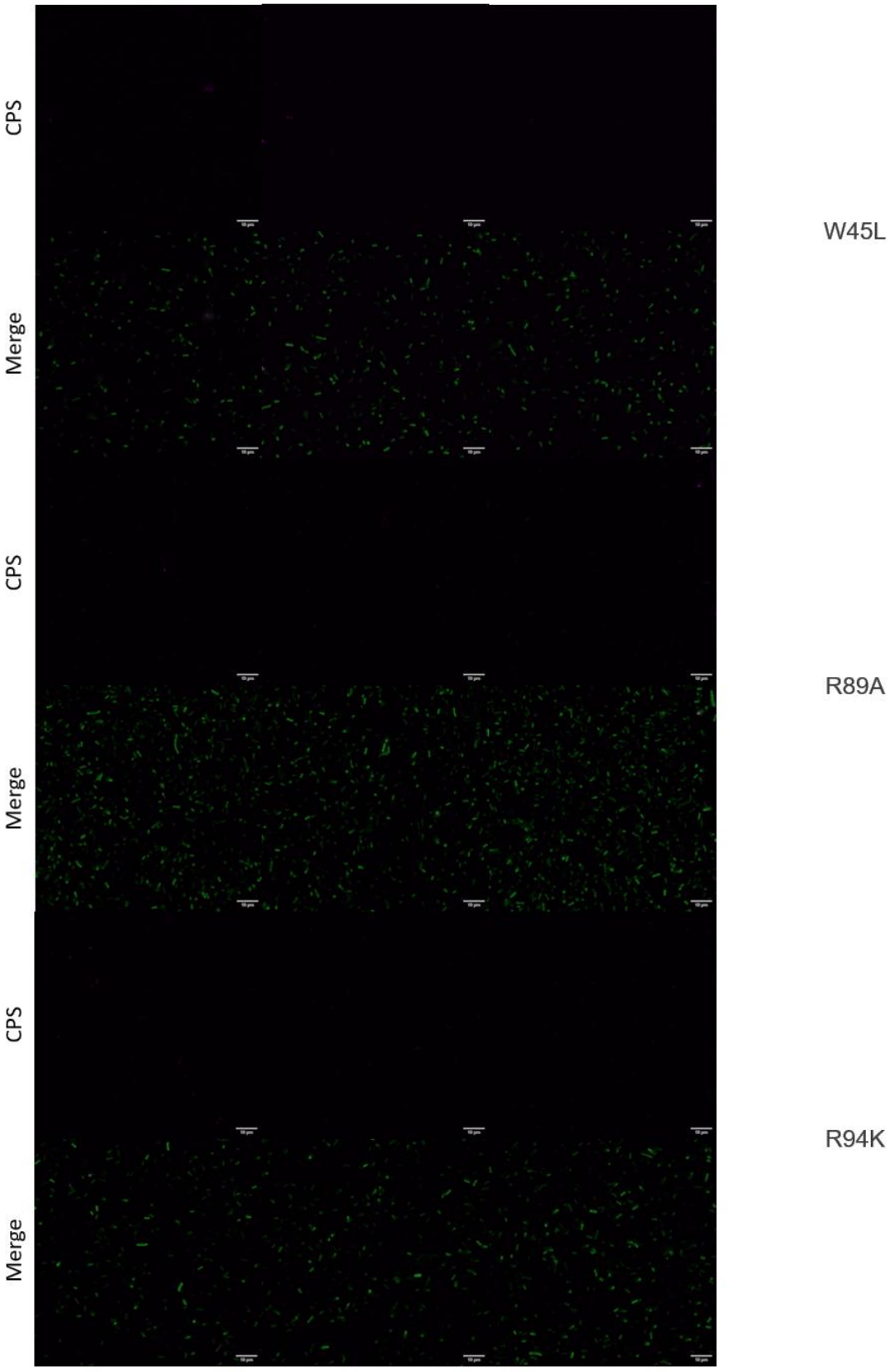
`

Fig. S3-21a



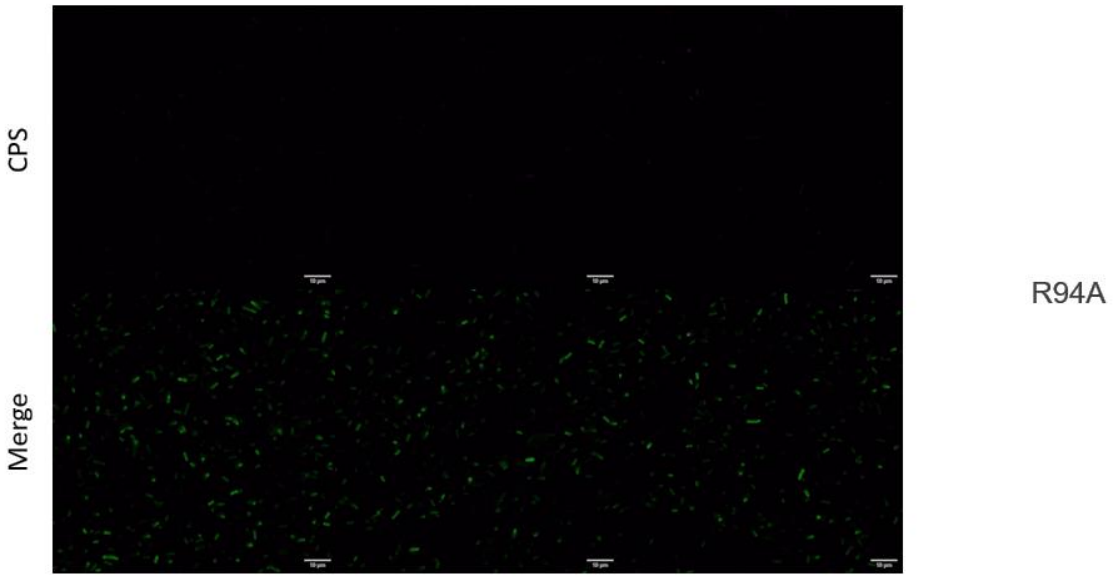
`

Fig. S3-21b



`

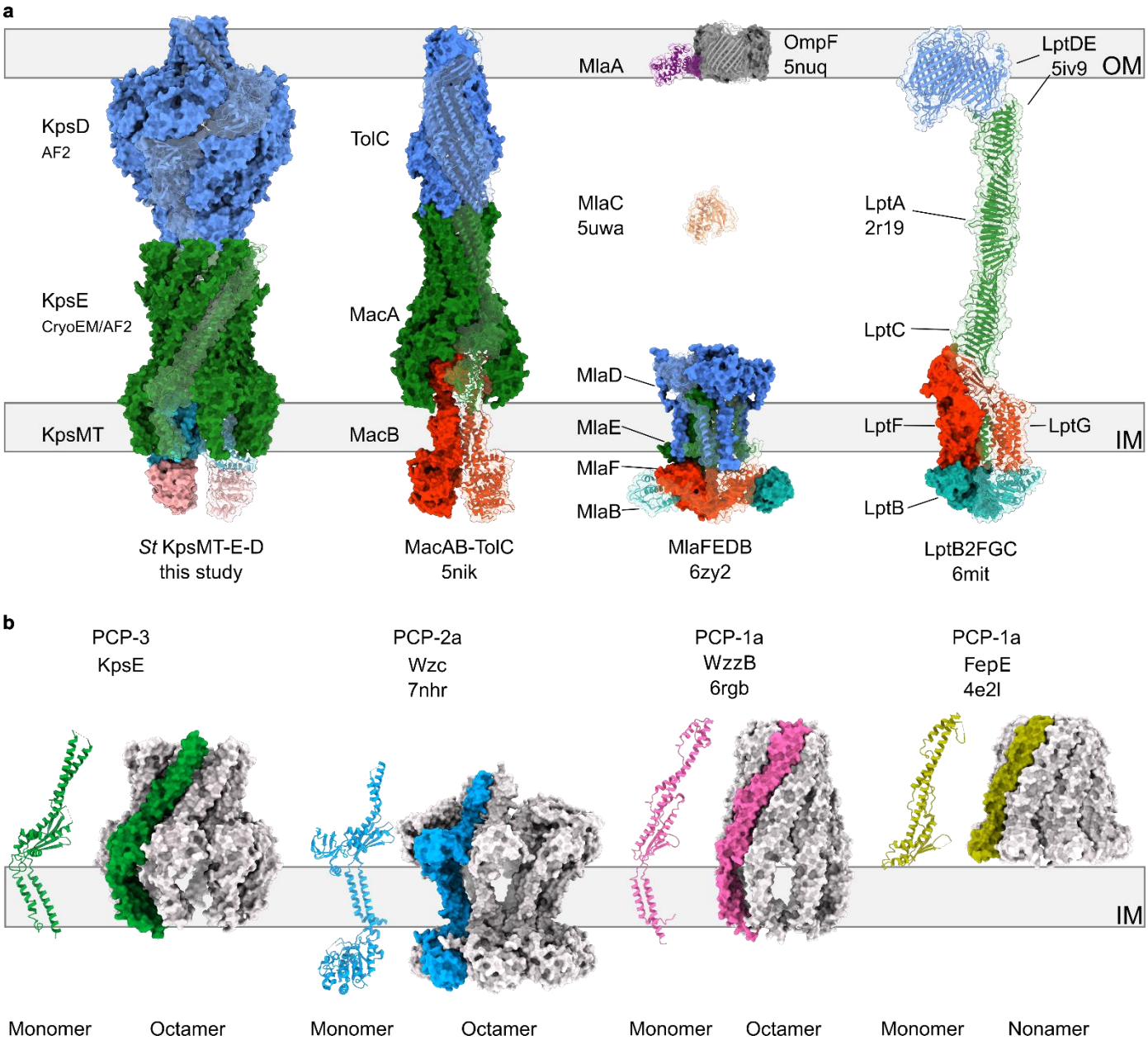
Fig. S3-21c



Supplemental Figure 3-21 (a-c) | Full fields of view and additional examples of confocal imaging presented in Figure S3 -20h. Scalebar: 10μm.

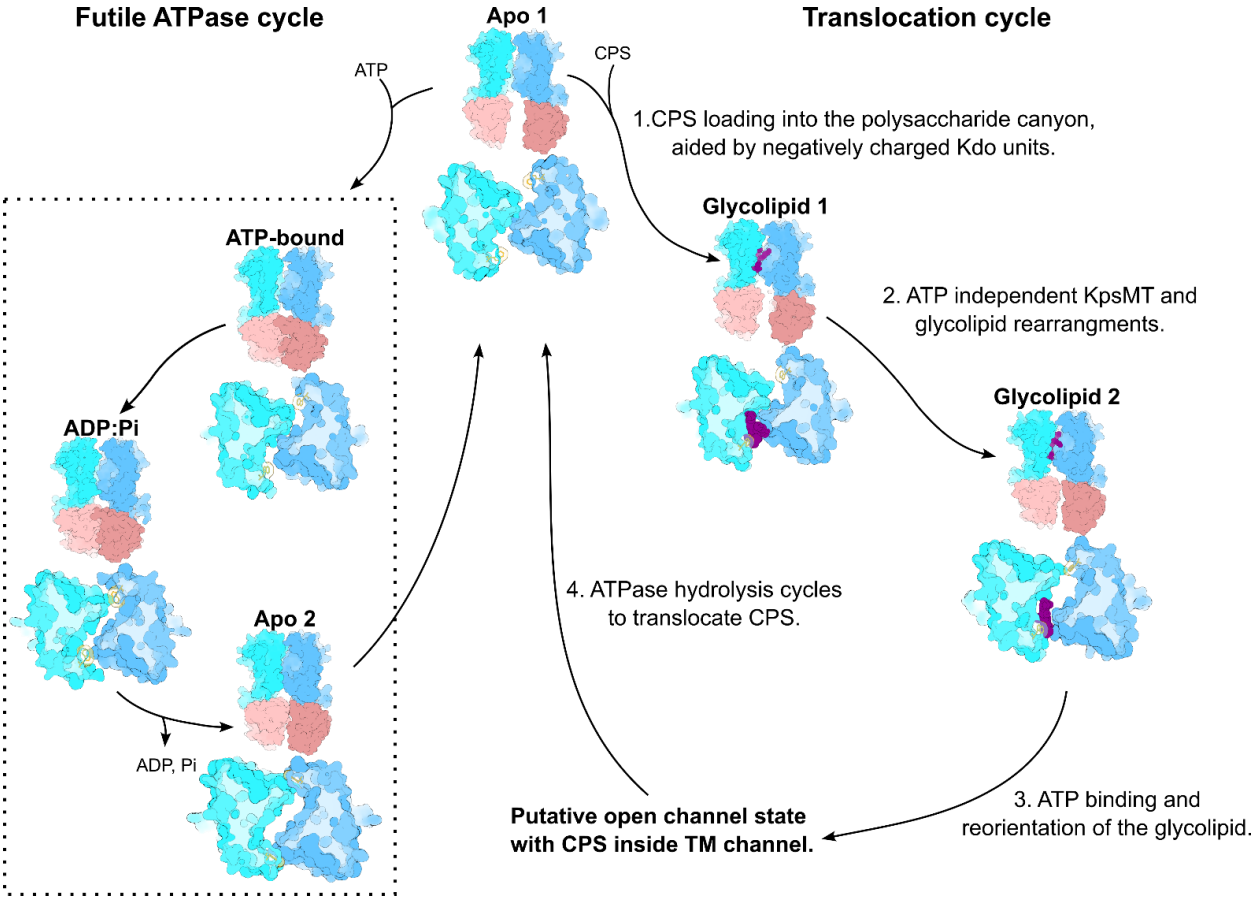
`

Fig. S3-22



Supplemental Figure 3-22 | ABC transporter based trans-envelope secretion systems and known types of polysaccharide co-polymerases. (a) Organization of trans-envelope ABC transporter based secretion systems. **(b)** Comparison of all three families of polysaccharide co-polymerases.

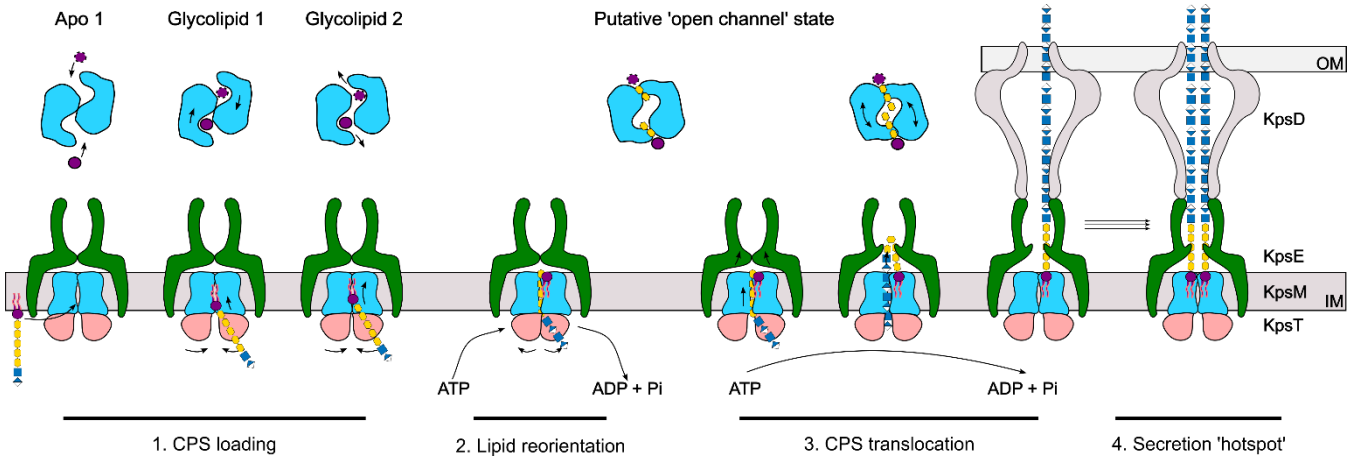
Fig. S3-23



Supplemental Figure 3-23 | Possible conformational transitions during futile and substrate-loaded ATP hydrolysis cycles.

3.8 Future directions

In the upcoming section, I will discuss potential future research directions that have the potential to provide further validation for the model proposed in Fig. 3-5 (repurposed and displayed below). The proposed experiments will directly address the suggested translocation steps outlined in the model, specifically: 1. CPS loading, 2. Lipid reorientation, 3. CPS translocation and 4. Secretion 'hotspot'.



1. CPS loading

The transition from Apo1 to Glycolipid 1 and 2 remains elusive. This can be addressed through:

1.1 Molecular dynamics simulations of the KpsMT ABC transporter in a lipid bilayer and, in the presence of small quantities of PG and Kdo-linked PG, could yield valuable insights into the process by which the substrate diffuses into the polysaccharide canyon. Additionally, these simulations could help determine whether the transition from Glycolipid 1 to Glycolipid 2 requires additional energy input.

1.2 Functional studies of the N-terminal part of KpsE. There is a possibility that the extreme N-terminus of KpsE plays a role in substrate engagement, potentially regulating the accessibility

`

of the polysaccharide canyon for substrate entry. This hypothesis can be examined through mutagenesis and deletion experiments, with the developed *in vivo* encapsulation assay as a readout.

2. Lipid reorientation and 3. CPS translocation

What are the initial steps of CPS translocation? How is the lipid reorientation energized? Is the CPS playing a role in stabilizing the putative open channel conformation? These open question can be addressed through CryoEM studies of the KpsMT-KpsE actively translocating the CPS. I intend to:

2.1 Incubate a purified complex of $KpsMT_{E151Q}$-KpsE with purified CPS, supplying ATP and magnesium prior to preparing the CryoEM grid.

2.2 Incubate a purified complex of KpsMT-KpsE with purified CPS, supplying ATP and magnesium prior to preparing the CryoEM grid.

In theory, either of these approaches should yield valuable insights and help confirm the presence of the postulated open-channel state. Furthermore, they should enable us to formulate more confident hypotheses about the driving forces behind CPS translocation initiation - whether it primarily hinges on ATP binding (2.1) or if ATP binding and subsequent hydrolysis are both necessary (2.2).

4. Secretion 'hotspot'

The existence of the fully formed secretion system that functions as a continuous conduit across the inner and outer membrane can be confirmed by the following approaches:

4.1 Cryo electron tomography on encapsulated cells, targeting *P. multocida* KpsMT-KpsE-KpsD. In this method I intend to enzymatically digest the capsule and use Streptavidin with nanogold beads as a fiducial marker to label KpsD.

`

4.2 CryoEM studies of *S. thermodepolymerans* KpsMT-KpsE-KpsD. I plan to purify the fully assembled secretion machinery by introducing disulfide bridges and employing chemical crosslinking. I propose to co-express the proteins of the secretion system alongside the biosynthetic components of the CPS operon, motivated by the strong likelihood that CPS plays a pivotal role in stabilizing the fully formed secretion system complex.

4.3 Minflux nanoscopy of CPS and the *P. multocida* secretion system. My plan involves engineering SNAP domain fusion constructs for either KpsT or KpsM. If functional, SNAP fused ABC transporter will allow for 2 color minflux experiments of: CPS with the secretion system as well as outer-membrane porin with the inner membrane KpsMT-KpsE.

`

4. Chapter 4: Concluding remarks

Complex carbohydrates are essential for countless biological processes across all kingdoms of life. Gram-negative bacteria employ cell surface polysaccharides to survive in hostile environments, thrive as pathogens hidden from the host immune system, or provide protection against antimicrobial agents. The gram-negative outer membrane is decorated with complex carbohydrates, which often originate from distinct pathways within the single organism. Comprised of Lipid A, short core oligosaccharides, and the long terminal O antigen polysaccharide, LPS is one of the best studied examples of such a complex carbohydrate. LPS not only contributes to the structure of the cellular envelope but also rigidifies the outer membrane. CPS, another crucial component in the carbohydrate mesh of the bacterial glycocalyx, is less studied yet equally important, compared to LPS. CPS exceed the length of LPS many times. With a length reaching 1000 sugar units, CPS can extend far from the cell surface. Understanding how CPS is synthesized and deposited on the cell surface could be pivotal in perturbing bacterial defense mechanisms and creating vulnerabilities.

Contrasting the two model systems I have studied during my PhD, HA synthesis and secretion by a synthase-dependent process (Cv-HAS) and spatio-temporally uncoupled synthesis and secretion (KpsMT-KpsE), reveals intriguing differences and commonalities.

Cv-HAS (and closely related vertebrate HASes) is a remarkable example of an enzyme that's fully optimized to play one specific function, HA synthesis, in the most efficient way imaginable. HAS utilizes a single catalytic domain to transfer two different donor sugars, form substrate-specific glycosidic linkages between them and, simultaneously, translocate the nascent HA polymer. As a

`

result, it functions as a glycosyltransferase and HA translocase with both processes being energized by the binding and turn-over of UDP-activated donor sugars.

HAS mediated synthesis does not require an external primer to initiate the reaction. The architecture of the acceptor binding site and biochemical data revealed that a GlcNAc monosaccharide can be positioned at the acceptor site to function as a primer for HA synthesis. Then, after the attachment of a GlcA-GlcNAc disaccharide unit, it is the binding of the following UDP-GlcNAc substrate that likely pushes the newly formed disaccharide into the channel. The growing polymer is initially stabilized and coordinated inside the channel by a set of positively charged lysine and arginine residues. This coordination changes to hydrophobic residues further into the channel, plausibly facilitating polymer release.

In comparison, CPS biosynthesis and translocation are decoupled. First, a group of enzymes synthesizes the lipid-linked poly-Kdo primer that is common for encapsulated gram-negative bacteria. Then, dedicated glycosyltransferases produce a specific and usually charged polysaccharide, e.g. HA. During or after this process, the CPS secretory machine, KpsMT-E, recognizes the lipid-linked poly-Kdo primer through a conserved and positively charged lipid-exposed cavity at the interface between KpsM subunits. By utilizing an external energy source – ATP, KpsMT is able to translocate its substrate, regardless of the chemical composition of the serotype-specific CPS polysaccharide chain. Due to this modular design and the existence of the universal CPS anchor, genes encoding glycosyltransferases can be horizontally transferred from species to species, leading to the formation of new capsular serotypes.

`

Given the different translocation mechanisms employed by Cv-HAS and KpsMT-E, it is quite notable that in both cases, the polysaccharide primer is coordinated by conserved positively charged residues that line the entrance to a TM channel or canyon. As the polymer is translocated, the interactions between the polymer and residues lining the secretion path become weaker and less conserved, thereby likely preventing translocation energy barriers.

The multi-enzymatic CPS biosynthetic apparatus is then a stark opposite of the elegant multifunctional HAS synthase. It serves as a universal platform for the export of various charged sugar polymers, favoring adaptability and interspecies compatibility over efficiency and specialization. My work on ABC transporter-mediated CPS translocation sheds light on the initial steps of this unique and poorly understood process. It should serve as an intriguing starting point for next experiments aiming to provide a full atomistic model of CPS translocation.

`