Molecular Organization of the *E. coli* Cellulose Synthase Macrocomplex

Justin F. Acheson[1], Ruoya Ho[1], Nicolette F. Goularte[2], Lynette Cegelski[3], Jochen Zimmer*[1]

Affiliations:

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

[2] Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, 94305, USA

[3]Department of Chemistry, Stanford University, Stanford, CA, 94305, USA

*: Corresponding author: jz3x@virginia.edu

First Paragraph

Cellulose is frequently found in communities of sessile bacteria, called biofilms.[1] *E. coli* and other enterobacteriaceae modify cellulose with phosphoethanolamine (pEtN) to promote host tissue adhesion.[2,3] The *E. coli* pEtN cellulose biosynthesis machinery contains the catalytic BcsA-BcsB complex synthesizing and secreting cellulose, in addition to five other subunits.[1] The membrane-anchored periplasmic BcsG subunit catalyzes pEtN modification.[3,4] Here we present the cryo electron microscopy structure of the ~1 MDa *E. coli* Bcs supramolecular complex, consisting of one BcsA enzyme associated with six copies of BcsB. BcsB homo-oligomerizes primarily through interactions of its carbohydrate-binding domains as well as intermolecular beta-sheet formation. The BcsB hexamer creates a half-spiral whose open side accommodates two BcsG subunits directly adjacent to BcsA's periplasmic channel exit. The cytosolic BcsE and BcsQ subunits associate with BcsA's regulatory PilZ-domain. The macrocomplex is a fascinating example of cellulose synthase specification.

Main Text

Cellulose is an amphipathic linear glucose polymer produced primarily by vascular plants as a cell wall component but also by several bacteria, oomycetes, and tunicates.[1,5,6] The polymer is deposited on the cell surface where it performs its diverse biological functions.

Bacterial biofilms are 3-dimensional sessile communities consisting primarily of polysaccharides and proteinaceous fibers, but also include outer membrane (OM) vesicles as well as nucleic acids.[7] Due to reduced susceptibility to antimicrobial treatments, biofilms account for approximately 80% of nosocomial infections.[8] Accordingly, enzymatic degradation of biofilm polysaccharides has been shown to significantly increase their susceptibilities to conventional antibiotics.[9]

Cellulose is produced by several bacteria as a biofilm polysaccharide. Its physical form ranges from ordered microfibrils resembling plant cell wall fibers to

amorphous polymers embedded in complex composite materials.[7,10] Several enterobacteria, including *E. coli*, *Klebsiella pneumoniae*, and *Salmonella enterica*, modify cellulose in the periplasm with pEtN to stabilize its interaction with curli fimbriae on the cell surface, which in turn facilitates host tissue adhesion.[2]

The essential catalytically active cellulose synthase subunit is BcsA, a membrane-embedded processive glycosyltransferase (GT).[11] BcsA polymerizes UDP-activated glucose (UDP-Glc) via an intracellular GT domain and also secretes the nascent cellulose polymer through a channel formed by its own transmembrane (TM) segment.[12] BcsA associates with a periplasmic inner membrane-anchored protein, BcsB, containing a C-terminal TM helix preceded by two copies of carbohydrate-binding and flavodoxin-like domains (CBD and FD, respectively).[11] In addition, BcsA is allosterically activated by the bacterial signaling molecule cyclic-di-GMP (cd-GMP). The activator binds to BcsA's C-terminal PilZ domain tightly associated with the GT domain, thereby releasing an auto-inhibited state.[13]

Cellulose is translocated across the OM through a pore formed by BcsC, a ~130 kDa protein containing 19 predicted periplasmic tetratricopeptide repeats (TPR) as well as a C-terminal 16-stranded OM porin.[14,15] All cellulose biosynthesis systems identified to date in Gram-negative bacteria further express a non-essential periplasmic cellulase, BcsZ, of unknown biological function.[1,16,17]

While BcsA, BcsB and BcsC form the essential core of most Gram-negative cellulose biosynthesis systems, additional accessory subunits confer distinct physical properties to the cellulosic materials produced, such as the formation of cellulose microfibrils or pEtN-modification. Accordingly, the *E. coli* (Ec) cellulose synthase complex consists of 9 subunits, of which 7 are likely localized to the inner membrane, Extended Data Fig. 1.[1] Here, the BcsA-B complex associates with BcsE, F, G, R, and Q, of which BcsF and R are small 60-70 residues long polypeptides. BcsE is a cd-GMP binding protein of ~59 kDa that increases cellulose biosynthesis *in vivo.* The protein is organized into three regions of which the N-terminal domain likely

interacts with the membrane-integrated BcsF and the remainder binds cd-GMP as well as BcsQ.[18] In solution, BcsQ forms a complex with BcsR at a 2:2 molecular stoichiometry and localizes to the cell poles *in vivo* where cellulose accumulates in *E. coli* and *S. enterica*.[19,20] The BcsQR complex is likely recruited to the membrane due to its interactions with BcsE.[18]

BcsG is a periplasmic membrane-anchored pEtN transferase that transfers pEtN, most likely from phosphatidylethanolamine, to approximately 50% of cellulose's glucosyl units, thereby generating pEtN cellulose.[3,4,21] Bacterial two-hybrid assays confirmed protein-protein interactions between BcsE and BcsF, BcsF and BcsG, as well as BcsG and BcsA.[3, 22]

The inner membrane associated *Ec* BcsA-B-E-F-G-R-Q cellulose synthase macrocomplex has recently been characterized by negative stain electron microscopy analyses. These studies revealed a megadalton size Bcs complex organized into several layers.[22] Starting on the periplasmic side, the complex contains a crown-shaped hemi-circle formed from six BcsB copies, followed by a disc-shaped platform likely representing the TM segments, as well as two cytosolic layers assigned to the soluble regulatory subunits. Due to limited resolution, the complex stoichiometry and the locations of BcsG, BcsQR and BcsE were not determined. In particular, it also remained unresolved how many copies of BcsA are part of the complex, which correlates with the number of cellulose polymers formed.[22]

To gain mechanistic insights into pEtN cellulose biosynthesis, we determined the cryo electron microscopy (cryo-EM) structure of the recombinantly expressed and catalytically active *Ec* Bcs macrocomplex. Our structure reveals an unexpected complex stoichiometry in which only one BcsA enzyme associates with a crescent-shaped BcsB hexamer as well as two copies of BcsG. On the cytosolic side, BcsA interacts with two soluble species, representing BcsE and the BcsQR complex, via its C-terminal PilZ domain. Accordingly, we observe robust stimulation of *in vitro* cellulose biosynthesis by the BcsQR complex and only slight activation by BcsE. Our

114 analyses provide unique insights into the biosynthesis and secretion of biofilm
115 polysaccharides likely shared by other exopolysaccharide synthases.
116
117 **Results**
118 *Isolation of a catalytically active cellulose synthase supramolecular complex*
119 The catalytically active *Ec* Bcs complex was recombinantly expressed in *E. coli* C43.
120 Co-expression of all Bcs subunits, including the periplasmic BcsZ and OM BcsC
121 proteins as well as the cd-GMP-generating AdrA enzyme, resulted in significant
122 flocculation of the bacteria in liquid cultures, even in the presence of a catalytically
123 inactive BcsG, Extended Data Fig. 2a.[3] This aggregation was not observed in the
124 absence of the OM BcsC subunit, likely stalling cellulose biosynthesis in the
125 periplasm. The *in vivo* formation of authentic pEtN cellulose was confirmed by
126 Congo-red staining of macro-colonies grown on agar plates,[23] as well as solid state
127 NMR analysis of the cellulosic material isolated from liquid cultures of the
128 engineered *Ec* strain (Fig. 1a and Extended Data Fig. 2b and c). The ^{13}C CPMAS solid-
129 state NMR spectrum of the recombinantly expressed polymer is comparable to that
130 for natively produced pEtN cellulose.[3] The amino carbon (C8) appears uniquely at
131 41 ppm and the C4 chemical shift of 83 ppm indicates it is a non-crystalline
132 cellulosic polymer. The 62-66 ppm region corresponds to that for the C6 and C7 of
133 unmodified glucose and the C6' and C7' carbons of the pEtN glucose units.
134
135 To analyze the contributions of the individual Bcs components to BcsA's catalytic
136 activity, we co-expressed BcsA together with different membrane-integrated Bcs
137 subunits, prepared inverted membrane vesicles (IMVs), and quantified *in vitro*
138 cellulose production in a 30 min synthesis reaction upon incorporation of ^{3}H-
139 labeled glucose by scintillation counting, as described, Fig. 1b and Extended Data
140 Fig. 3a.[24]
141 As previously reported for components from *Rhodobacter sphaeroides* as well as *E.*
142 *coli*, BcsA is catalytically inactive in the absence of BcsB, Fig. 1b.[24] Co-expressing
143 BcsA and BcsG alone does not restore catalytic activity comparable to BcsA-B.

Similarly, the cellulose biosynthetic activity of the BcsA-B complex does not increase in the presence of BcsG, BcsF, or BcsFG, suggesting that these are non-regulatory accessory subunits. Of note, expressing BcsA in the presence of BcsB and BcsG significantly reduces its proteolysis in IMVs, perhaps reflecting increased stability of the BcsABG complex, Fig. 1b.

To evaluate the contributions of the cytosolic soluble Bcs subunits to cellulose biosynthesis, we incubated the purified BcsA-B-F-G complex with separately purified BcsE (N-terminally fused to maltose binding protein, MBP), the BcsQR complex, or the BcsE-Q-R trimeric complex. We observe approximately six-fold increased BcsA catalytic activity in the presence of BcsQR, while addition of BcsE alone accounts for only about a 3-fold increase, Fig. 1c. Addition of the BcsE, -Q, and -R components, which form a stable soluble heterotrimeric complex, Extended Data Fig. 3b, stimulates biosynthesis comparable to BcsQR alone, Fig. 1c.

To purify the inner membrane-associated BcsA-B-E-F-G-Q-R macrocomplex (IMC), the recombinantly expressed complex was extracted from the membrane and purified by two affinity chromatography steps using poly-His and Strep-tags added to the C- and N-termini of BcsA and BcsB, respectively. Size exclusion chromatography in the detergents lauryl- and decyl-maltose neopentyl glycol supplemented with cholesterylhemisuccinate (see Methods) purified the catalytically active IMC complex (Extended Data Fig. 2d and e). The presence of all IMC subunits was confirmed by Western blotting and mass spectrometry fingerprinting, Extended Data Fig. 2f and Extended Data Table 1. BcsF-derived peptides were not detected by mass spectrometry, yet its co-purification with the IMC was confirmed by Western blotting after inserting a C-terminal Myc-tag, Extended Data Fig. 2g.

The purified and detergent-solubilized IMC synthesizes a water-insoluble polymer *in vitro* that is degraded by a β-1,4-specific glucanase, Extended Data Fig. 2e, demonstrating cellulose formation in the absence of a pEtN donor, as previously

described.[24] The IMC is strongly activated by cd-GMP, leading to 5-fold stimulation of *in vitro* catalytic activity under saturating conditions, Extended Data Fig. 2h. Under these conditions, likely BcsA and BcsE interact with the activator. The complex binds cd-GMP and UDP-Glc with apparent affinities of ~2 and ~430 μM, respectively, in good agreement with similar studies on *Rhodobacter sphaeroides* BcsA (Rs BcsA), Extended Data Fig. 2h and i.[24]

Architecture of the IMC

The purified IMC used for *in vitro* activity assays was also used for structural analyses by single particle cryo-EM. Cryo as well as negative stain EM imaging readily reproduced particle shapes consistent with previous analyses of glutaraldehyde cross-linked IMC complexes.[22] However, our cryo-EM analyses provide unique insights into the detailed organization of the Bcs complex. The overall IMC structure has a resolution of approximately 7 Å. Local refinement of the extracellular crown resolved the BcsB hexamer at about 3.4 Å resolution, Fig. 2 and Extended Data Fig. 4.

Viewed from within the membrane plane, the IMC forms a triangular structure with a large periplasmic region approximately 172 Å wide, followed by a disc-shaped micelle-embedded region about 163 Å in diameter, and three cytosolic regions about 73, 59, and 66 Å long, Fig. 2c. The periplasmic region is a crescent-shaped array of six ring-like volumes representing BcsB, whereas the intracellular regions represent the cytosolic region of BcsA as well as two associated species, identified as BcsE and BcsQ-R (discussed below). The entire IMC spans ~210 Å along the membrane normal and its cytosolic domains protrude from the membrane into the cytosol by about 76 Å.

BcsB forms a crescent-shaped scaffold

Subtraction of the micelle-embedded and cytosolic particle components, followed by 3D classification and non-uniform and local refinement of the periplasmic domain

(see Methods) generated a high quality density map of the BcsB hexamer enabling de-novo model building, Extended Data Fig. 4 and 6 and Table 2.

Similar to the *R. sphaeroides* homolog,[11] *Ec* BcsB consists of four distinguishable domains that are arranged in a triangular shape, Fig. 3a and Extended Data Fig. 7. The structural repeat unit is formed by a carbohydrate-binding domain (CBD) connected to an α/β-domain previously referred to as flavodoxin-like domain (FD). The CBD-1 – FD-1 pair is repeated and rotated by 180 degrees (CBD-2 – FD-2), such that the FD and CBD regions interact with their counterparts. The CBDs contact and are linked to each other via extended loops harboring a conserved disulfide bridge between Cys182 and Cys500.

BcsB's C-terminal TM helix following FD-2 is not resolved in this density map due to signal subtraction but is described below in the context of BcsA, Extended Data Fig. 4.

As a hexamer, the BcsB subunits interact via the faces of their FD and CBD regions, Fig. 3b and d. The protomers, arranged in a hemi-circle and labeled A-F in a clockwise fashion when viewed from the periplasm, point their disulfide bond-stabilized CBD loops toward the center. The FD and TM-anchor regions are located at the hexamer's periphery. Each subunit is rotated by approx. 40 degrees around the central pseudo-symmetry axis. Further, when viewed along the membrane plane, each BcsB subunit rotates by ~10 degrees away from the membrane surface around a hinge located near its TM-anchor. Thereby, BcsB domains near the center of the hemi-circle exhibit the greatest, while those near its periphery the least displacement from the membrane, Fig. 3c. The resulting half-spiral is 195 Å wide and 60 Å tall. With the exception of chain F, all subunits lack a short loop (residues 447 to 465) located on the membrane proximal side of the hexamer where interactions with lipid head-groups are possible.

Complex formation buries large, otherwise solvent exposed surfaces between the protomers. Chain B shares 7,418 and 7,613 Å^2 surface area with chains A and C, respectively.

The interface between BcsB subunits is stabilized by an 'Oligomerization Loop (OL)' (residues 617-659). This region belongs to FD-2 and is helical in chain A of the *Ec* BcsB hexamer, Fig. 3d and Extended Data Fig. 7 and 8 Yet, sandwiched between BcsB subunits, the OL forms a 3-stranded β-sheet that interacts with residues 345-352 of the counterclockwise neighboring subunit, resulting in a 4-stranded anti-parallel β-sheet at the interface, Fig. 3d.

Additionally, the loop connecting OL's β-strands 2 and 3 of each BcsB protomer occupies a large pocket created at the interface with the preceding subunit. In this pocket, the side chain of the loop's Arg633 contacts the backbone carbonyl oxygens of Pro361 and Trp363 of the neighboring subunit. Similarly, the carboxyl groups of the loop's Asp643 and Asp646 interact with Arg365 and Arg368 in the other subunit, Fig. 3d and Extended Data Fig. 8.

The OL region is either observed or predicted to be helical in the heterodimeric BcsA-B complexes of *R. sphaeroides* and *Komagataeibacter xylinus* (Kx), respectively, both sharing about 16% sequence identity with *Ec* BcsB, Extended Data Fig. 8.[11,25] In the homo-oligomerizing *Ec* BcsB, however, the OL evolved structural plasticity to adjust to the different environments within the oligomer.

At the center of the half-circle, the disulfide-tethered CBD loops stack on top of each other to form a deep groove, with the chain A and F loops being closest to and farthest away from the membrane interface, respectively, Fig. 3b and 4. This peculiar loop stacking is primarily stabilized by backbone interactions in which the peptide-linkage of Cys500 in one chain forms β-sheet interactions with the backbone of Gly495 in the following subunit. Due to the rotation and tilt of the hexamer's subunits, the disulfide loop of chain F is translated by approximately 40 Å farther away from the membrane compared to chain A. The disulfide-loops create a

deep electronegative canyon at the pseudo symmetry axis with the stacked disulfide bridges lining its base, Fig. 4.

The Bcs complex contains a single BcsA subunit

Our initial cryo-EM map identified several cytosolic and TM components associated with the BcsB hexamer, Fig. 2c Refinement of the IMC with a focus on the TM and cytosolic regions indicated the presence of TM proteins only underneath and in front of the first BcsB subunit, Extended Data Fig. 5. The remainder of the micelle-surrounded volume is essentially featureless, as expected for detergent molecules and disordered BcsB TM-anchors, Fig. 5.

Local refinement of a sub-volume comprising the first two BcsB subunits together with the corresponding TM and cytosolic volume significantly improved the map quality, allowing the rigid body docking of a poly-Ala BcsA backbone model (see Methods), Extended Data Fig. 5 and 9. While BcsA's TM helices could be individually positioned, some even revealing the locations of large side chains guiding register assignments, the cytosolic GT and PilZ domains were poorly resolved with the exception of several interface helices near the GT/membrane boundary. Yet, because the TM helices tightly interact with the cytosolic domains,[11] positioning the TM segments also places the GT and PilZ sub-domains into their corresponding volumes, Fig. 5a and Extended Data Fig. 9.

R. sphaeroides and *Ec* BcsA share a conserved core architecture including the *Ec* BcsA region from TM helix 2 to the C-terminal amphipathic helix following its PilZ domain. The *Ec* enzyme, however, contains 133 extra N-terminal residues as well as an additional C-terminal helix, Fig. 5a and Extended Data Fig. 10. The N-terminal region is predicted to form a TM helix between residues 28 to 50, thereby placing the N-terminus in the periplasm. The occurrence of this extension correlates with the presence of the *bcsG* gene in the operon, thus likely the formation of pEtN cellulose.

Only the first BcsB subunit interacts with BcsA. It's C-terminal TM-anchor packs into a groove formed by BcsA's TM helices 2 and 3, as also observed in *Rs* BcsA, Fig. 5b and Extended Data Fig. 9.[11] The TM-anchor is preceded by a short amphipathic helix that forms the hinge around which BcsB protomers rotate in the above-described hexameric arrangement, Fig. 5b. The overall *Ec* BcsA-B complex organization is similar to crystal structures of the hetero-dimeric *Rs* BcsA-B complex, Extended Data Fig. 10. Of note, also *Ec* BcsB's CBD-2 sits on top of BcsA's TM channel, thereby likely bending the secreted cellulose chain away from the BcsB hexamer (see Discussion), Fig. 5a and Extended Data Fig. 10b.[11]

Strong helical density 'in front' of BcsA indicated the location of its N-terminal TM helix. Although the connection with TM2 is only partially resolved in our EM map, its position at the interface and previously determined interaction with BcsG (discussed below) support its assignment as TM1, Extended Data Fig. 9.

The N-terminal TM helix of *Ec* BcsA is tilted by about 45 degrees relative to the membrane plane and interacts with TM3 as well as BcsB's TM-anchor, Fig. 5a. Its C-terminal end is in proximity to a long amphipathic interface helix following BcsA's PilZ domain at the cytosolic water-lipid interface. TM1 is connected to TM2 via an unresolved region followed by an amphipathic helix directly preceding TM2, in agreement with secondary structure predictions, Fig. 5a and Extended Data Fig. 1.

The C-terminal region of *Ec* BcsA following the PilZ domain, residues ~805-871, forms a helical pair at the water-lipid interface, Fig. 5a. One helix directly follows the PilZ β-barrel, as also observed in *Rs* BcsA, the other sits on top of the preceding helix running in the opposite direction. The helices are profoundly amphipathic, including a cluster of Arg residues (Arg853, 854 and 857) as well as several Trp, Val, Leu and Phe residues. Viewed from the periplasm, the two helices are arranged in a 'V' straddling the C-terminal end of BcsA's TM1, Extended Data Fig. 9 and 10a.

The Bcs complex contains two copies of BcsG

BcsG is a membrane-bound pEtN transferase containing five predicted N-terminal TM helices connected by short cytosolic loops as well as a periplasmic catalytic domain, Extended Data Fig. 1. The local refinement of a sub-volume comprising the micelle-embedded and cytosolic BcsA and BcsE volumes resolved a cluster of at least twelve TM helices directly in front of BcsA, Fig. 5c and Extended Data Fig. 5. One of the helices interacts with BcsA's TM1 and perhaps also with its extreme C-terminus. Further, at least 4 cytosolic interface helices are resolved in close proximity to the first ten TM helices. Strikingly, the helices can be grouped into two almost identical sets, each containing five TM and two interface helices, Fig. 5c and d. The remaining two TM helices farthest away from BcsA are less well resolved, yet could represent BcsF's TM segments (see below).

Next to BcsA, BcsG is the only other Bcs component containing multiple (i.e. 5) TM segments. We therefore assign the pair of five TM helices to two copies of BcsG. This assignment is supported by recent structures of the lipopolysaccharide receptor PbgA and the lipid-A pEtN transferase EptA that are distant homologs of BcsG.[26-28] The cytosolic interface helices may belong to BcsF, together with the two remaining poorly resolved TM segments. BcsF is 63 residues long and predicted to contain an N-terminal TM helix followed by a helical cytosolic region, Extended Data Fig. 1. In this position, BcsF is ideally positioned to interact with BcsG and BcsE, as experimentally confirmed.[3, 22]

Although present in the purified Bcs complex, Extended Data Fig. 2f, BcsG's periplasmic catalytic domain is not resolved in our cryo-EM maps. It is connected to the last TM helix via a ~40 residue long linker likely conferring significant flexibility. However, at low contour levels, we observe weak additional density right above the BcsG-assigned TM segments. The density sits in between the first and the last subunit of the BcsB hexamer and is sufficiently large to accommodate two copies of the crystal structure of *Ec* BcsG's periplasmic domain, Fig. 5e and Extended Data Fig. 11.[4]

BcsQ and BcsE interact with BcsA's PilZ domain

The Bcs complex contains the cytosolic subunits BcsE, Q and R. BcsE is a ~59 kDa protein with a recently determined crystal structure of its central and C-terminal domains.[18] BcsE significantly increases *in vivo* pEtN cellulose production in *S. enterica* and *E. coli*.[3, 19] It has been proposed to be part of a novel regulatory pathway involving BcsE, F and G for pEtN cellulose biosynthesis. BcsQ, a ~26 kDa protein, is not essential for cellulose biosynthesis *in vitro*,[24] yet it likely targets the Bcs complex to the cell poles for localized cellulose secretion.[20] It is a homolog of MinD, a bacterial cell division protein that dimerizes and associates with the cell membrane in an ATP-dependent manner.[29,30] Recent studies on isolated BcsQ demonstrated that it requires BcsR for structural stability in solution.[18] *In vitro*, the BcsQR complex is targeted to the membrane by BcsE, which interacts with BcsF.[18]

Our cryo-EM map of the entire Bcs complex reveals two additional densities associated with BcsA's PilZ domain, Fig. 2 and 6a. The larger density has no apparent association with the TM region and points away from the periplasmic BcsB hexamer, roughly underneath the BcsG dimer. The second bilobed density extends from BcsA's PilZ domain to the BcsG-associated interface helices, tentatively assigned to BcsF. These cytosolic components are weakly associated with the Bcs core complex as they were absent in a significant number of particles. Shape reconstructions of these smaller complexes resulted in volumes revealing only BcsA domains on the cytosolic membrane side, Extended Data Fig. 12a and b. Considering the size and shape of the densities as well as previously reported interactions of BcsQ and BcsR as well as BcsE and BcsF, [3, 22,18] we assign the large density adjacent to BcsA's PilZ domain to a $BcsQ_2R_2$ tetrameric complex and the bilobed micelle-associated density to a monomer of BcsE, Fig. 6. BcsE interacts with membrane-integrated BcsF most likely via its N-terminal region and with the BcsQR complex via its C-terminal cd-GMP-binding domain.[18] The proposed location of BcsE within the Bcs complex is consistent with these interactions. Further, the crystal structures of dimeric MinD (homologous to BcsQ) and BcsE fit the assigned densities well, Extended Data Fig. 12c and d.[3,29,18]

To confirm that the bilobed density indeed represents BcsE, we replaced wild type BcsE with a construct N-terminally fused to MBP. The fusion protein forms a stable complex with isolated BcsQR in solution, Extended Data Fig. 3b, and co-purifies with the Bcs components, suggesting that MBP does not abolish BcsE's interactions with other Bcs subunits. Although MBP is flexibly linked to BcsE and thus not completely resolved, the MBP-containing cryo EM map only reveals mass differences near the contact site with the membrane. At this position, BcsE's N-terminal domain likely interacts with BcsF, Extended Data Fig. 4 and 12d-f.

Discussion

Cellulose synthases are evolutionarily conserved enzymes that polymerize glucose molecules into linear polysaccharides. Cellulose is an incredibly versatile structural element with native biological applications as a cell wall structure in plants[5] and contributing to adhesion and protective community architectures in bacterial biofilms.[10,2] Reflecting this functional diversity, cellulose biosynthesis systems exhibit great variability in terms of complex organization,[31] activation in response to different stimuli,[32] as well as chemical modification.[33,23]

Work on *Rs* BcsA revealed that it forms a structural and functional entity with the BcsB subunit.[24] Therefore, previous low-resolution analyses of the *Ec* Bcs macrocomplex revealing six BcsB copies suggested the presence of six BcsA-B sub-complexes, with a corresponding maximum number of cellulose polymers synthesized.[22] In contrast, our cryo-EM analysis reveals an unexpected organization in which one BcsA protein associates with six BcsB copies that self-organize into a crescent-shaped crown. Therefore, a Bcs complex only synthesizes a single cellulose polymer at a time, which, in turn, likely only requires a single BcsC subunit enabling passage across the OM, Fig. 6b.[15]

Solid state NMR analysis of pEtN cellulose revealed the modification of about 50% of the polymer's glucosyl units with pEtN at the C6 hydroxyl groups.[3] Our cryo-EM analyses strongly suggest that the Bcs complex contains two BcsG copies. Recent

biochemical analyses of BcsG's periplasmic catalytic domain indicate that the protein functions as a monomer.[4,21] Because neighboring glucosyl units are rotated by about 180 degrees in the cellulose polymer, two BcsG copies may be necessary to modify its C6 hydroxyl groups located on opposite sides of the polymer.

In *Rs* BcsA-B, the bound nascent cellulose polymer kinks sharply at BcsA's periplasmic channel exit at the interface with BcsB.[11] This kink is induced by BcsB's CBD-2, which sits right on top of the channel exit. In the *Ec* Bcs complex, the first BcsB copy occupies a similar location, right above the BcsA channel exit, Fig. 5a and b and Extended Data Fig. 10b. Therefore, although not resolved in our EM map (but probably present), the cellulose polymer secreted by the *Ec* complex likely also bends away from the BcsA-B interface towards BcsG, thereby facilitating pEtN modification, Fig. 6b.

The exact functions of BcsF and BcsR are currently unknown. Zouhir et al recently demonstrated that BcsR is necessary for proper folding of BcsQ and forms a soluble complex with this subunit at a 2:2 molar stoichiometry.[18] The size of the BcsA-associated density observed in our EM map suggests that a $BcsQ_2R_2$ complex also interacts with BcsA.

Zouhir et al further showed that BcsE interacts with the $BcsQ_2R_2$ complex, likely also as a dimeric species. Dimerization of BcsE is further corroborated by the crystal structure of its core and C-terminal domains. Here, two BcsE protomers associate via their cd-GMP-binding C-terminal domains, stabilized by a bridging single cd-GMP molecule.[18] Our map is consistent with the presence of a single BcsE species, yet the proposed position of its C-terminal domain near the BcsQR complex would enable interactions with a second copy, perhaps stabilized by elevated cd-GMP concentrations. Accordingly, Krasteva et al observe a larger cytosolic volume associated with the glutaraldehyde cross-linked Bcs complex, compared to our cryo EM structure.[22] It is possible that, without cross-linking, the second BcsE copy dissociates during purification due to cd-GMP depletion.

The assigned positions of BcsE and BcsQR in the Bcs complex are further supported by bacterial two-hybrid analyses with BcsA's PilZ domain, demonstrating direct interactions.[22] The same study also suggested an interaction of BcsA and BcsG via BcsA's N-terminal domain, again in agreement with our assignment.

In vivo pEtN cellulose production significantly decreases in the absence of BcsE, while removal of BcsQ has a less dramatic effect.[3] *In vitro*, however, we observe a robust increase of BcsA's catalytic activity induced by BcsQR and only minor activity increases in the presence of BcsE, Fig. 1c. The molecular details of this apparent discrepancy are currently unknown. It is possible that the cytosolic Bcs subunits are important for proper assembly and positioning of the entire Bcs complex *in vivo* or interactions with additional factors, not captured *in vitro*.

The self-assembly of BcsB into a crown-shaped hexamer with only one BcsA subunit was unexpected and currently lacks structural homologs. Similar high molecular weight structures have not been observed for *Rs* or *Kx* BcsA-B, [11,25] most likely due to structural differences of BcsB, Extended Data Fig. 8. The function of this large periplasmic scaffold is unknown, although it may restrict cellulose flexibility by creating a confined reaction chamber for pEtN modification. BcsB self-association is likely terminated by BcsG, which occupies the space between the first and last BcsB subunits, Fig. 5e.

The BcsB hexamer creates an electronegative groove near its pseudo symmetry axis that could facilitate pEtN cellulose migration towards the OM. Similar scaffolds may be formed by periplasmic or membrane-bound subunits of related exopolysaccharide biosynthesis systems, including alginate, acetylated cellulose, and poly N-acetylglucosamine synthases.[34,35]

Cellulose is translocated across the OM through BcsC, a 16-stranded porin containing a large periplasmic domain.[15] It is currently unclear how BcsC interacts with the inner membrane complex. However, considering cellulose's flexibility and

475 tendency to self-aggregate, a direct BcsB-BcsC interaction shielding the polymer
476 seems necessary. The length of BcsC's periplasmic N-terminal domain and the
477 architecture of the IMC make a direct BcsC-BcsB interaction likely, Fig. 6b.[14]
478
479

Main References

1 Romling, U. & Galperin, M. Y. Bacterial cellulose biosynthesis: diversity of operons, subunits, products, and functions. *Trends Microbiol* **23**, 545-557, (2015).

2 Hollenbeck, E. C. *et al.* Phosphoethanolamine cellulose enhances curli-mediated adhesion of uropathogenic Escherichia coli to bladder epithelial cells. *Proc Natl Acad Sci U S A* **115**, 10106-10111, (2018).

3 Thongsomboon, W. *et al.* Phosphoethanolamine cellulose: A naturally produced chemically modified cellulose. *Science* **359**, 334-338, (2018).

4 Anderson, A. C., Burnett, A. J. N., Hiscock, L., Maly, K. E. & Weadge, J. T. The Escherichia coli cellulose synthase subunit G (BcsG) is a Zn(2+)-dependent phosphoethanolamine transferase. *J Biol Chem* **295**, 6225-6235, (2020).

5 Keegstra, K. Plant cell walls. *Plant Physiol* **154**, 483-486, (2010).

6 Melida, H., Sandoval-Sierra, J. V., Dieguez-Uribeondo, J. & Bulone, V. Analyses of extracellular carbohydrates in oomycetes unveil the existence of three different cell wall types. *Eukaryot Cell* **12**, 194-203, (2013).

7 McCrate, O. A., Zhou, X., Reichhardt, C. & Cegelski, L. Sum of the Parts: Composition and Architecture of the Bacterial Extracellular Matrix. *J. Mol. Biol.*, (2013).

8 Stewart, P. & Costerton, J. Antibiotic resistance of bacteria in biofilms. *Lancet* **358**, 135-138, (2001).

9 Snarr, B. D. *et al.* Microbial glycoside hydrolases as antibiofilm agents with cross-kingdom activity. *Proc Natl Acad Sci U S A* **114**, 7124-7129, (2017).

10 Iguchi, M., Yamanaka, S. & Budhiono, A. Bacterial cellulose; a masterpiece of nature's arts. *J Mater Sci* **35**, 261-270, (2000).

11 Morgan, J., Strumillo, J. & Zimmer, J. Crystallographic snapshot of cellulose synthesis and membrane translocation. *Nature* **493**, 181-186, (2013).

12 McNamara, J. T., Morgan, J. L. W. & Zimmer, J. A molecular description of cellulose biosynthesis. *Annu Rev Biochem* **84**, 17.11-17.27, (2015).

13 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).

14 Nojima, S. *et al.* Crystal structure of the flexible tandem repeat domain of bacterial cellulose synthesis subunit C. *Sci Rep* **7**, 13018, (2017).

15 Acheson, J. F., Derewenda, Z. S. & Zimmer, J. Architecture of the Cellulose Synthase Outer Membrane Channel and Its Association with the Periplasmic TPR Domain. *Structure* **27**, 1855-1861 e1853, (2019).

16 Mazur, O. & Zimmer, J. Apo- and Cellopentaose-bound Structures of the Bacterial Cellulose Synthase Subunit BcsZ. *J Biol Chem* **286**, 17601-17606, (2011).

17 Yasutake, Y. *et al.* Structural characterization of the Acetobacter xylinum endo-beta-1,4-glucanase CMCax required for cellulose biosynthesis. *Proteins* **64**, 1069-1077, (2006).

18 Zouhir, S., Abidi, W., Caleechurn, M. & Krasteva, P. V. Structure and Multitasking of the c-di-GMP-Sensing Cellulose Secretion Regulator BcsE. *mBio* **11**, (2020).

19 Fang, X. *et al.* GIL, a new c-di-GMP-binding protein domain involved in regulation of cellulose synthesis in enterobacteria. *Mol Microbiol* **93**, 439-452, (2014).

20 Le Quéré, B. & Ghigo, J.-M. BcsQ is an essential component of the Escherichia coli cellulose biosynthesis apparatus that localizes at the bacterial cell pole. *Mol Microbiol* **72**, 724-740, (2009).

21 Sun, L. *et al.* Structural and Functional Characterization of the BcsG Subunit of the Cellulose Synthase in Salmonella typhimurium. *J Mol Biol* **430**, 3170-3189, (2018).

22 Krasteva, P. V. *et al.* Insights into the structure and assembly of a bacterial cellulose secretion system. *Nat Commun* **8**, 2065, (2017).

23 Thongsomboon, W., Werby, S. H. & Cegelski, L. Evaluation of Phosphoethanolamine Cellulose Production among Bacterial Communities Using Congo Red Fluorescence. *J Bacteriol* **202**, (2020).

24 Omadjela, O. *et al.* BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for *in vitro* cellulose synthesis. *Proc Natl Acad Sci U S A* **110**, 17856-17861, (2013).

25 Du, J., Vepachedu, V., Cho, S. H., Kumar, M. & Nixon, B. T. Structure of the Cellulose Synthase Complex of Gluconacetobacter hansenii at 23.4 A Resolution. *PloS one* **11**, e0155886, (2016).

26 Clairfeuille, T. *et al.* Structure of the essential inner membrane lipopolysaccharide-PbgA complex. *Nature* **584**, 479-483, (2020).

27 Fan, J., Petersen, E. M., Hinds, T. R., Zheng, N. & Miller, S. I. Structure of an Inner Membrane Protein Required for PhoPQ-Regulated Increases in Outer Membrane Cardiolipin. *mBio* **11**, (2020).

28 Anandan, A. *et al.* Structure of a lipid A phosphoethanolamine transferase suggests how conformational changes govern substrate binding. *Proc Natl Acad Sci U SA* **114**, 2218-2223, (2017).

29 Ghosal, D., Trambaiolo, D., Amos, L. A. & Lowe, J. MinCD cell division proteins form alternating copolymeric cytomotive filaments. *Nat Commun* **5**, 5341, (2014).

30 Lackner, L. L., Raskin, D. M. & de Boer, P. A. ATP-dependent interactions between Escherichia coli Min proteins and the phospholipid membrane *in vitro*. *J Bacteriol* **185**, 735-749, (2003).

31 Nixon, B. T. *et al.* Comparative Structural and Computational Analysis Supports Eighteen Cellulose Synthases in the Plant Cellulose Synthesis Complex. *Sci Rep* **6**, 28696, (2016).

32 Ross, P. *et al.* Regulation of cellulose synthesis in Acetobacter xylinum by cyclic diguanylic acid. *Nature* **325**, 279-281, (1987).

33 Spiers, A. J., Bohannon, J., Gehrig, S. M. & Rainey, P. B. Biofilm formation at the air-liquid interface by the Pseudomonas fluorescens SBW25 wrinkly spreader requires an acetylated form of cellulose. *Mol Microbiol* **50**, 15-27, (2003).

34 Marmont, L. S. *et al.* Oligomeric lipoprotein PelC guides Pel polysaccharide export across the outer membrane of Pseudomonas aeruginosa. *Proc Natl Acad Sci U SA* **114**, 2892-2897, (2017).

35 Low, K. E. & Howell, P. L. Gram-negative synthase-dependent exopolysaccharide biosynthetic machines. *Curr Opin Struct Biol* **53**, 32-44, (2018).
36 *Schroedinger, LLC,* The PYMOL Molecular Graphics System, Version 2.0
37 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, (2004).

Figures and Figure Legends

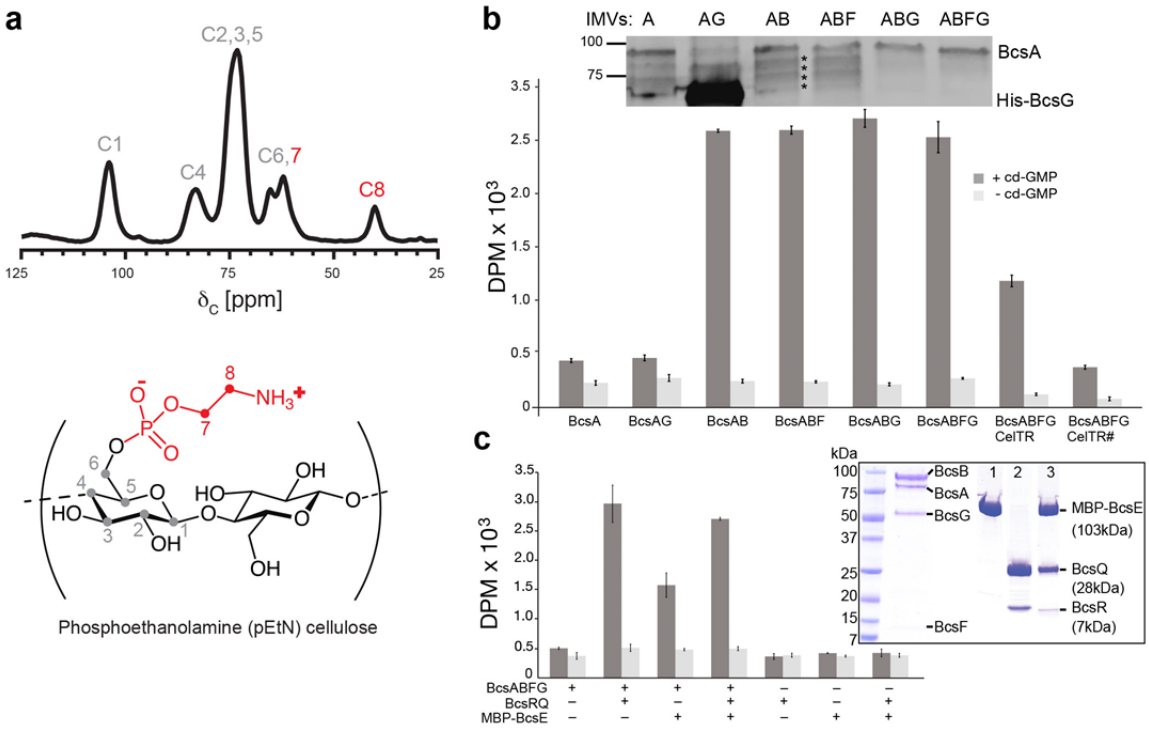

Figure 1| Functional characterization of the E. coli Bcs complex. (**a**) Solid state NMR analysis and one representation of pEtN cellulose produced *in vivo* by the recombinantly expressed Bcs complex. (**b**) Catalytic activity of BcsA co-expressed with the indicated subunits. Experiments were performed in IMVs and quantify the incorporation if ^{3}H-labeled glucose into cellulose. CelTR cellulase; (#) denotes cellulase digestion after synthesis and vesicle solubilization at pH 4.5. Inset: Western blot detecting His-tagged BcsA. (*) indicates BcsA proteolytic fragments. AG sample: Here, BcsG carried an N-terminal His-tag for co-expression with BcsA and partially overlaps with the BcsA band. DPM: disintegrations per minute. (**c**) Activity of the purified BcsABFG complex in the presence of the indicated cytosolic soluble components. Inset: Coomassie-stained SDS-PAGE of the purified components; 1: MBP-BcsE, 2: BcsQR, 3: MBP-BcsEQR.

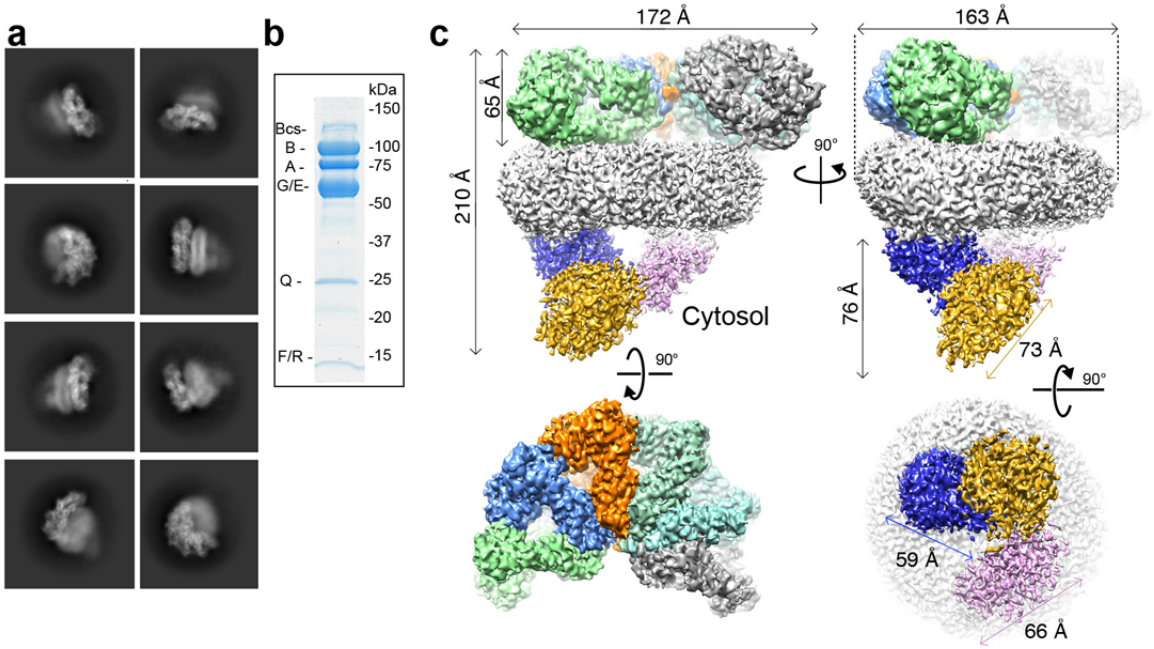

Figure 2| Structure of the Bcs complex. (**a**) Examples of 2D class averages of the Bcs complex. Each class contains between 8,000 and 20,000 particles (from cryoSPARC). (**b**) Coomassie-stained SDS-PAGE of the purified Bcs complex. (**c**) Overall structure of the Bcs complex. The periplasmic ring-shaped densities representing BcsB are colored individually, the micelle density is colored grey, and cytosolic domains are colored blue, orange and magenta. The map is Gaussian filtered at a width of 1.2 and contoured at 2.6σ .

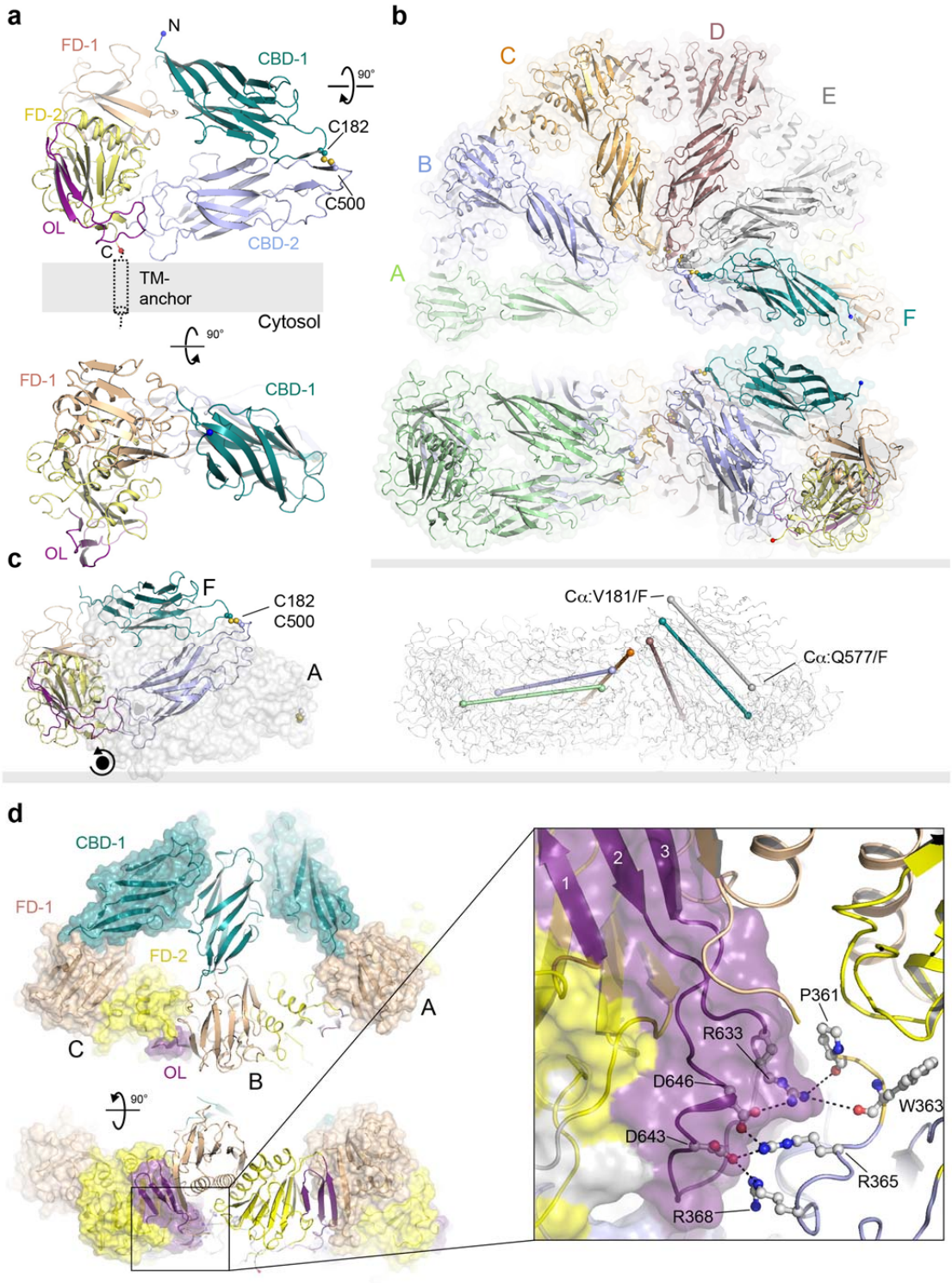
a
N
FD-1
CBD-1
90°
FD-2
C182
C500
OL
CBD-2
C
TM-
anchor
Cytosol
90°
FD-1
CBD-1
OL
b
C
D
E
B
A
F
c
F
C182
C500
A
Cα:V181/F
Cα:Q577/F
d
CBD-1
FD-1
FD-2
A
C
OL
B
90°
1
2
3
P361
R633
D646
W363
D643
R365
R368

Figure 3| BcsB forms a crescent-shaped periplasmic scaffold. (**a**) Domain organization of *Ec* BcsB. The individual domains are colored and labeled: FD/CBD; flavodoxin-like and carbohydrate-binding domains; OL: oligomerization loop. The location of the C-terminal TM anchor is indicated as a cartoon. (**b**) Six BcsB subunits assemble into a hemi-circle with a deep groove at its center containing a ladder of stacked disulfide bridges. (**c**) Superimposition of the first and last BcsB subunit of the hemi-circle. The rotation axis is indicated by a black circle. Right panel: Vectors indicating the distance between V181 and Q577 in each subunit reveal the increasing tilt of subunits within the hemi-circle. (**d**) Intermolecular interactions between BcsB protomers. Shown is a BcsB trimer with the flanking subunits shown as surfaces, the middle copy is shown as a cartoon. Interactions are predominantly between FD-2 and the OL.

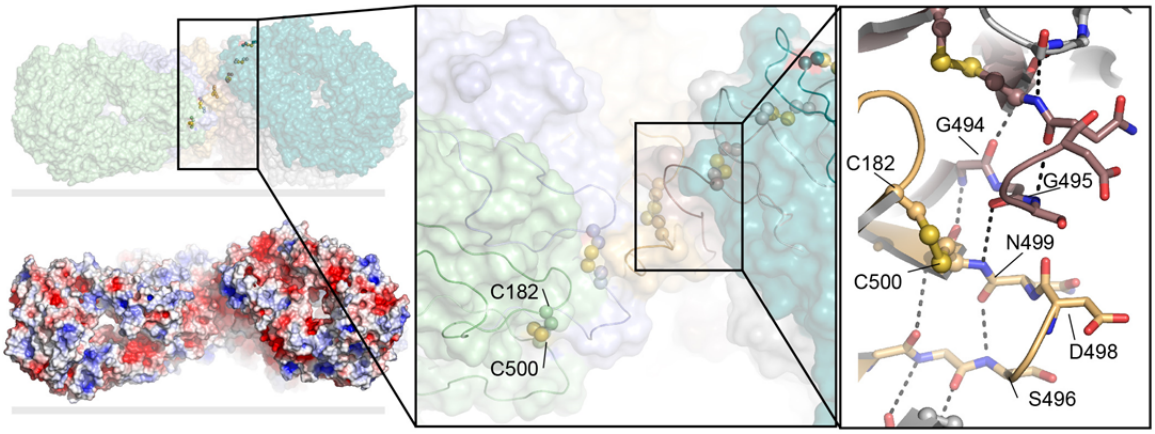

Figure 4| The BcsB hexamer forms a deep groove. Top left: Stacking of the disulfide bridges (shown as spheres) along the hexamer's groove. Bottom left: Surface electrostatic potential of the BcsB hexamer calculated in Pymol using the APBS plugin.[36] Red to blue: -5 to +5kT. Gray bars indicate the periplasmic membrane interface.

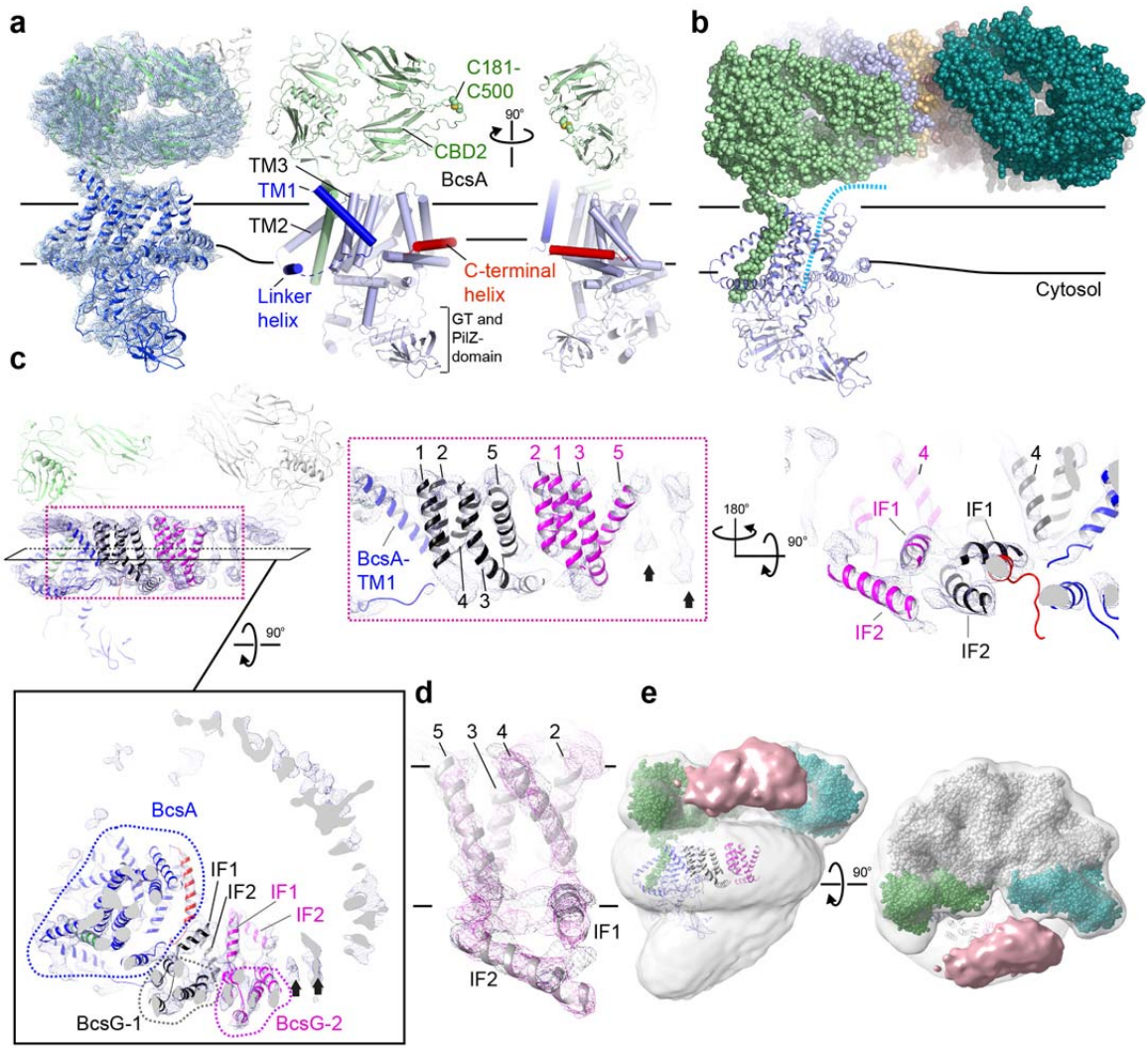

Figure 5| The Bcs complex contains one BcsA and two BcsG subunits. (**a**) Local refinement and poly-alanine model of *Ec* BcsA. Left: Locally refined map contoured at 6σ. Right: Backbone model of *Ec* BcsA together with the high-resolution structure of BcsB (colored green). BcsA is shown with cylindrical helices colored pale and dark blue as well as red for its C-terminal extension. (**b**) BcsA associates with the first BcsB subunit. Horizontal lines indicate membrane boundaries as indicated by the micelle shape. The nascent cellulose polymer present in pdb 4P00 is indicated by a dashed cyan line. (**c**) A cluster of TM helices adjacent to BcsA suggests the presence of two BcsG subunits. IF: Interface helices. Likely additional poorly resolved TM helices are indicated with black arrows. Contour level: 8σ. (**d**) Superimposition of densities assigned to helices shown in black and magenta in panel (c) in Chimera.[37] (**e**) Additional weak periplasmic density observed at low contour levels above the TM helix cluster shown as a solid light-brown surface. Contoured at 0.9σ.

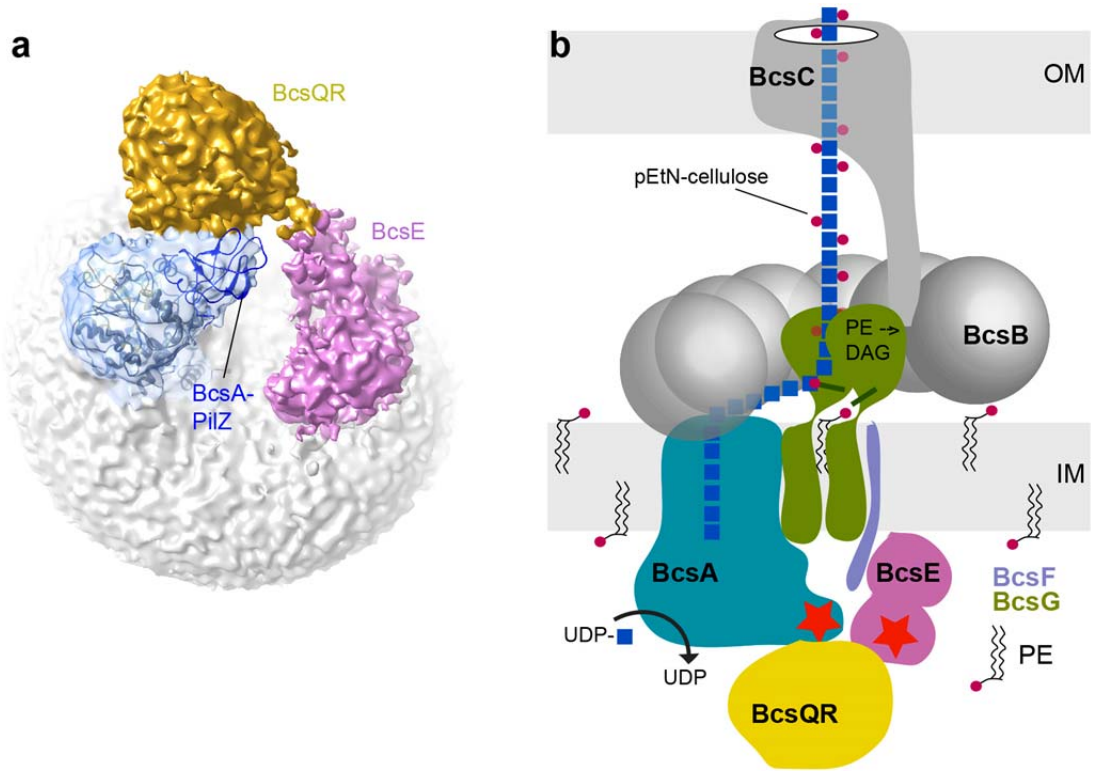

Figure 6| BcsA associates with cytosolic soluble subunits and model of pEtN cellulose biosynthesis. (**a**) Overall map of the Bcs complex viewed from the cytosol, contoured at 2.8σ. Density representing BcsA is shown as a blue semi-transparent surface and BcsA is shown as a cartoon colored blue. Additional volumes assigned to BcsE and the BcsQ complex are colored magenta and orange, respectively. (**b**) Model of pEtN cellulose synthesis and secretion. One BcsA enzyme synthesizes and secretes a cellulose polymer that is pEtN-modified by BcsG in the periplasm. A single BcsC subunit in the OM likely facilitates transport across the periplasm and the outer membrane (OM). BcsF facilitates interactions with BcsE. BcsA and BcsE bind cyclic-di-GMP (red star). PE: Phosphatidylethanolamine, DAG: Diacylglycerol.

Methods

Plasmid Design

To co-express the *Ec* Bcs complex including AdrA, three plasmids were developed. First, pETDuetEcBcs_A-12His_nSS-Strep-B_AdrA-6His was created by inserting BcsA in to MCS1 including a C-terminal dodeca histidine tag. BcsB, including the native secretion signal (nSS) peptide, was inserted into MCS2. A STREP II tag was inserted into BcsB between the nSS and the beginning of the mature polypeptide coding sequence using primer extension.

To express a third gene in pETDuet, EcAdrA was inserted between the BcsB stop codon and the T7 terminator region using PIPE cloning. AdrA was amplified from *Ec* K12 genomic DNA and inserted into pACYCDuet adding a C-terminal hexa-histidine tag. Using PIPE primers, the construct was amplified from pACYC_Ec_AdrA-6His beginning ~200 bp upstream of the start codon to include a spacer, T7 promoter, LacO and RBS site, while the C-terminus retained the hexa-histadine tag.

pCDF_EcBcs_R_Q_HA-E was produced by first using PIPE to amplify the coding sequence for both bcsR and bcsQ genes from *Ec* Ar3110 genomic DNA. This is a variant of W3110 that has had the premature stop codon in BcsQ repaired to the native coding sequence. The cassette was inserted into MCS, followed by PIPE to insert N-terminally HA-tagged BcsE into MCS2.

The final plasmid pACYCDuet_EcBcs_PelB-8His-C_Z_F_G-FLAG was constructed with genes from either plasmid or genomic DNA, and all plasmid sequences were originally derived from *Ec* K12 genomic DNA. Using PIPE, pACYCDuet_EcBcs_PelB-8His-C_PelB-Z-6His was created by amplifying PelB-8His-BcsC from pET20_EcBcs_PelB-8His-C and inserting into MCS1 and PelB-BcsZ-6His was amplified from pET20_EcBcs_PelB-Z-6His and inserted into MCS2. A second pACYCDuet construct, pACYC_EcBcs_F_G-FLAG was created by amplifying BcsG including a C-terminal FLAG tag and inserting it into MCS2. BcsF was then amplified from genomic DNA and inserted into MCS1. Finally, pACYCDuet_EcBcs_PelB-8His-C_Z_F_G-FLAG was assembled by amplifying BcsF_G-FLAG including ~200 bp upstream of BcsF to include the T7/Lac regulatory components and inserted into pACYCDuet_EcBcs_PelB-8His-C_PelB-Z-6His between the BcsZ C-terminus and T7

terminator regions using PIPE. The PIPE primers restored the BcsZ native C-terminus by adding a stop codon and omitting the hexa-His coding region.

To express cytoplasmic components of the Bcs system several constructs were produced. A BcsE fusion to maltose binding protein was created using pVP68K (Center for Eukaryotic Structural Genomics) that contains a N-terminal octahis tag-MBP lacking the periplasmic signal sequence. BcsE was amplified from the pCDF_EcBcs_R_Q_HA-E plasmid without the HA-tag, while adding a TEV protease cleavage site between the MBP linker and BcsE. BcsQ was inserted into pACYCDuet MCS2 from K12 genomic DNA, the forward primer included a point mutation (stop to L) to repair the premature stop codon that inactivates cellulose production in K12 strains. An N-terminal octahis tag followed by a TEV cleavage site was then added using primer extension. pCDF_EcBcs_R_Q was produced during the first step to make pCDF_EcBcs_R_Q_HA-E as described above.

Protein Expression

To express the *Ec* Bcs complex, *Ec* Bl21 C43 cells were co-transformed by electroporation with pETDuetEcBcs_A-12His_nSS-Strep-B_Adra-6His, pACYCDuet_EcBcs_PelB-8His-C_Z_F_G-FLAG, and pCDF_EcBcs_R_Q_HA-E, and plated on LB-ampicillin/chloramphenical/streptomycin plates at 37°C overnight. The next day, colonies were selected for 5 mL LB with antibiotics starter cultures, and placed in a 37°C shaker incubator for 6 hrs. The starter cultures produced clumping cells likely due to leaky protein expression. One mL of starter culture was added to four 2.8 L plastic baffled flasks containing 1000 mL of TB-M-80155 using the following recipe: TB (24 g/L yeast extract, 12 g/L tryptone), 50 mL 20xM (1 M NH_4Cl, 0.5 M $Na_2H(PO_4)$, 0.5 M $KH_2(PO_4)$, 0.1 M $Na_2(SO_4)$) per liter, autoclaved and supplemented with 25 mL of sterile filtered 40x80155 autoinduction solution (32% (w/v) glycerol, 0.6% (w/v) glucose, 2% (w/v) lactose) per liter of media along with 1 mL of each 1000X antibiotic (Amp 200 mg/mL, Cam 34 mg/mL, Strep 100 mg/mL). Cells were grown overnight at 30°C with shaking at 220 rpm. After about 12 hrs, the temperature was dropped to 25°C and 400 µL of 1 M IPTG was added. Growth was continued for 4 – 5 hours. Cells were collected by centrifugation at 5000 rpm for 20

minutes. The collected cells (~75 g) were resuspended and homogenized using a glass dounce to a final volume of 250 mL in the Buffer A as follows: 50 mM HEPES 8.0, 300 mM NaCl, 5% glycerol, 0.5 M cellobiose, 5 mM $MgCl_2$, and 45 mM imidazole. Cells were incubated for a minimum of 30 minutes with 1 mg/mL lysozyme and *A. niger* cellulase (Millipore-Sigma C1184), as well as 2.5 mL of 100 mM PMSF to aid in cell lysis and prevent proteolysis. Cells were disrupted using a gas powered microfluidizer at 25,000 psi by four passes. The lysate was spun at 12,500 rpm for 20 minutes to remove non-lysed cells. The supernatant was collected and membranes were isolated by centrifugation at 42,000 rpm (138,000 g) for 2 hours and 4°C. Collected membranes were flash frozen in liquid N_2 and stored at -80°C until used.

To express BcsHisQ with BcsRQ or to express MBP-BcsE, growth was carried out in TB-M-80155 with appropriate plasmids and antibiotics as described above with the following differences. Precultures were grown to OD ~3 to 5 and 2.5 mL of preculture was added to four 2.8 L flasks containing 1 L of media. The cells were grown for 25 hours at 25°C, at which time the cells were centrifuged at 5000 rpm for 20 minutes and stored at -80°C until use.

Inner Membrane Vesicle Preparation

IMVs of the inner membrane BCS proteins focusing on BcsA were produced in a series of additive expressions caried out in 1 L cultures as follows; BcsA, BcsAG, BcsAB, BcsABF, BcsABG, and BcsABFG all additionally co-expressing AdrA. Cells were lysed in a C3 microfluidizer as described using Buffer B lacking detergents. The lysate was spun at 12,500 rpm for 20 minutes, then 25 mL was carefully added to a 2 M sucrose cushion made with the same buffer, and centrifuged for 2 h at 42,000 rpm (138,000 g). The dark ring at the cushion interface was carefully collected (~8 mL) and diluted to 65 mL with Buffer B, and further centrifuged for 90 min at 42,000 rpm. The pellet was gently rinsed with Buffer B followed by addition of 1 mL of buffer. The pellet was gently homogenized using a No. 6 round paintbrush followed by a 2 mL glass dounce. The IMVs were aliquoted into microfuge tubes and flash frozen in liquid N2 until needed. To normalize the concentration of BcsA to

later be used in activity assays, western blotting against the his-tag was performed. In order to remove the large amount of contaminating lipids and other proteins, prior to western blotting the IMVs were solubilized with 2% SDS, centrifuged for 20 minutes at 65,000 rpm in a bench top ultracentrifuge, incubated with 25 uL of nickel-NTA beads for 30 minutes, centrifuged at 700 g for 2 minutes, washed and centrifuged twice with buffer B containing 2% SDS. The beads were suspended in 40 uL of Buffer B containing 10 mM EDTA, 250 mM imidazole and 2% SDS followed by addition of 20 uL of 4X loading dye. 15 uL were loaded into each well, and represent the ratio of IMVs used in activity assays.

Protein Purification

Membranes were solubilized in Buffer A (12 mL/mL membrane suspension) containing Detergents A (0.5% LMNG, 0.01% DMNG, 0.01% CHS), 2 EDTA-free protease inhibitor tablets (Roche), and 125 µL 100mM PMSF using a glass homogenizer. After gentle rocking for 90 minutes at 4°C, non-solubilized material was removed by centrifugation at 138,000 g for 30 min. During this time ~5 mL of Ni-NTA resin (Pierce-Fisher) were equilibrated with Buffer A for batch binding. Equilibrated beads were added to the membrane extract and incubated at 4°C for 1 hour with gentle rocking. Resin was collected in a gravity flow column and washed with ~30 column volumes of Buffer A with 55 mM imidazole and Detergents B (0.01% LMNG, 0.002% DMNG, 0.002% CHS). The Bcs complex was eluted with Buffer A containing 400 mM imidazole and Detergents B. The eluent was immediately diluted in Buffer B (50 mM HEPES 8.0, 300 mM NaCl, 5% glycerol, 0.5 mM cellobiose, 5 mM $MgCl_2$, and Detergents B) to ~40% in the final volume to dilute the imidazole. During the Ni-NTA step 4 mL of streptactin resin (IBA) was washed and equilibrated with Buffer B at room temperature. Diluted eluent was passed over the streptactin resin twice at room temperature, washed with 5 column volumes of Buffer B, and eluted with Buffer B containing 2.5 mM desthiobiotin into a 100 kDa spin concentrator on ice. The streptactin eluent (~10 mL) was concentrated to ~600 µL and loaded onto a 10/30 superose 6 increase gel filtration column equilibrated with Buffer C (50 mM HEPES 8.0, 150 mM NaCl, 5 mM $MgCl_2$, 0.5 mM

cellobiose) with Detergents C (0.003% LMNG, 0.0006% DMNG, 0.0006% CHS). The Bcs complex eluted in a sharp peak at ~13 mL elution volume. Sample quality was assessed based on peak shape, SDS-PAGE, and negative stain imagining on a FEI F20 electron microscope.

Purification of the Bcs complex with MBP-BcsE was carried out by expressing all components with the exception of BcsE using pCDF_EcBcs_R_Q in place of pCDF_EcBcs_R_Q_HA-E. Before cell lysis a cell pellet from a 1 L growth of the MBP-BcsE fusion was mixed in, and all subsequent steps were as described for the full complex.

To purify BcsRQ cells were lysed as described above in 50 mM HEPES 8.0, 300 mM NaCl, 30 mM imidazole, 5% glycerol, and 0.3 mM TCEP. the lysate was spun at 138,000 g for 30 minutes. The lysate was applied to two 5 mL his-trap columns in tandem and washed with 15 column volumes of lysis buffer. A gradient to 350 mM imidazole in lysis buffer was performed and the major peak collected and concentrated using a 30 MWCO Amicon spin concentrator then applied to a superdex S200 size exclusion column equilibrated with 50 mM HEPES pH 8.0, 150 mM NaCl and 0.3 mM TCEP. Protein was concentrated to 100 μM based on absorbance at 280 with and extinction coefficient of 110 mM^{-1} for a 2:2 stoichiometry and molecular weight of 77 kDa.

To purify MBP-BcsE steps were as described for BcsRQ with the following addition. After histrap column elution the protein was applied to a XK 16 adjustable column with amylose resin (NEB) attached to a GE Purifier equilibrated in lysis buffer lacking imidazole. After washing to baseline the protein was eluted by addition of 10 mM maltose to the buffer. The protein was concentrated using a 100 MWCO Amicon spin concentrator prior to gel filtration. The final product was concentrated to 100 μM based on an extinction coefficient of 166 mM^{-1} and molecular weight of 103 kDa.

To purify the BcsRQ-MBP-BcsE complex, purified two-fold molar excess of MBP-BcsE was combined with BcsRQ and applied to an S200 size exclusion column. The protein was concentrated using a 100 MWCO Amicon spin concentrator prior to

gel filtration. The final product was concentrated to 100 μM based on an extinction coefficient of 414 mM^{-1} and molecular weight of 275 kDa assuming a 2:2:2 complex stoichiometry.

Western Blotting

Blotting procedures were performed after transferring SDS-PAGE gels to nitrocellulose membranes and incubated with specific primary antibodies for the following epitopes: pentahis (BcsA), STREPII (BcsB), FLAG (BcsG), HA (BcsE), MYC (BcsF), and maltose binding protein (MBP-BcsE). Nitrocellulose membranes were blocked with 5% BSA for 30 minutes briefly rinsed with tris buffered saline with tween (TBST), and incubated with primary antibody in 5% BSA at 4°C overnight. The following morning the membranes were washed three times by incubating with fresh TBST for at least five minutes. Membranes were then incubated with anti-mouse IgG with Dylight 800 fluorescent marker (Rockland) for 1 hour at ambient temperature followed by three washes. Blots were examined using an Odyssey Licor scanner at wavelengths 700 nm and 800 nm,

Mass Spectrometry

Purified Bcs from the high molecular weight front of the gel filtration elution peak was loaded into precast 4 – 20% TGX polyacrylyamide gel (Bio-Rad), and ran at 100 V until all the sample had migrated into the gel (~10 minutes). The top ~1.0 cm of the gel lane was excised and submitted to the Biomolecular Analysis Facility at the University of Virginia. The gel piece was digested with 20 ng/μL trypsin in 50 mM ammonium bicarbonate on ice for 30 min. Any excess enzyme solution was removed and 20 μL 50 mM ammonium bicarbonate added. The sample was digested overnight at 37°C and the peptides extracted from the polyacrylamide in a 100 μL aliquot of 50% acetonitrile/5% formic acid. This extract was evaporated to 10 μL for MS analysis. The LC-MS system consisted of a Thermo Electron Q Exactive HF mass spectrometer system with an Easy Spray ion source connected to a Thermo 75 μm x 15 cm C18 Easy Spray column (through pre-column). The nanospray ion

source was operated at 1.9 kV. The digest was analyzed using the rapid switching capability of the instrument acquiring a full scan mass spectrum to determine peptide molecular weights followed by product ion spectra (Top10 - 10 HCD) to determine amino acid sequence in sequential scans. This mode of analysis produces approximately 25000 MS/MS spectra of ions ranging in abundance over several orders of magnitude. The data were analyzed by database searching using the Sequest search algorithm against Uniprot *Ec* and specifically *Ec* Bcs proteins.

Cryo-EM

Grid preparation

The *Ec* Bcs complex samples were collected as they eluted from the gel filtration column and centrifuged at 100,000 g for 10 minutes. For cryo grid preparation, several optimization experiments were performed using a Vitribot Mark IV plunge freezing robot. The best freezing conditions resulted from 300 mesh C-Flat 1.2/1.3 grids glow discharged with amylamine. The optimal concentration of Bcs ranged from 0.25 to 0.45 mg/mL based on absorbance at 280 nm using a general extinction coefficient of 1 au per mg of protein. 3.5 µL of sample was applied to the grid and incubated for 30 s at 100% humidity. For data collected with activator cd-GMP, protein was incubated with 10 µM cd-GMP for 15 minutes prior to grid application. Grids were blotted for 12 – 14 s with a blot force of 6 and plunged into liquid ethane.

Collection

Data was collected on a 300 keV FEI Titan Krios (Thermo Fisher Scientific) equipped with a Gatan K3 camera and Gatan Quantum energy filter using Latitude S. Movies were collected at 81,000X at 40 frames/movie in counting mode with a total dose of 51 $e^-/Å^2$ and pixel size 1.08 $Å^2$. A total of six grids were used to allow for image collection in optimal ice, under the same blotting and data collection conditions resulting in a combined 13,883 movies. Data collected for *Ec* Bcs incubated with cd-GMP from a single grid with minor differences. Thermo Fisher EPU software was used in place of Latitude S, and the total dose of 50 $e^-/Å^2$, and a total of 3306 movies

collected. Data collection of *Ec* Bcs with MBP-BcsE was also collected from a single grid using EPU with a total dose of 51 $e^-/Å^2$, and a total of 841 movies collected.

Data processing

All data was processed using cryoSPARC.[38] The movies were imported into cryoSPARC and gain reference corrected as individual sets for each grid. Motion correction was performed using Patch Motion correction excluding the first two frames. CTF parameters were estimated with Patch CTF, and manual curation was used to remove low quality movies. The movies were combined from each grid to give a final 11,007 movies after curation based on CTF estimated parameters. ~1000 particles with a box size of 400 pixels (432 Å) were manually selected and 2D classified to produce templates from one grid (755 movies), which were then used to pick a larger particle set from the combined movies. After several rounds of 2D classification high quality classes were used to re-pick 2,675,278 particles and 2D classified to a final set of 863,273 particles. These particles were used to generate two *ab initio* 3D volumes. The best model was further refined into six classes using heterogenous refinement in cryoSPARC. 3D classification showed high heterogeneity. Particles from the best two 3D volumes were combined and put through a second round of heterogenous refinement into four classes, and further processed as described below. A new particle set was generated using the Topaz Wrapper in cryoSPARC.[39] A training model was performed using particles from one 3D class showing the best density of all subunits (~10,000 particles) against 1,384 movies from one grid collection. The model was applied to each individual grid's movie set separately to process smaller data sets. The particles were extracted using a box size of 336 pixels (362 Å), down-sampled to 4.32 Å, 2D class averaged and recombined. The particles from the final class averages were re-extracted without down-sampling, put through another round of 2D classification giving 735,904 particles, and processed as described below.

In an attempt to improve complex stability a data set with the addition of 10 μM cd-GMP was collected. Movies were motion corrected and CTF parameters were estimated as described for the previous data. Particles were picked using the TOPAZ

wrapper after training a model against a random subset of 500 micrographs using particles generated from one round of manual picking and template picker. 409,611 particles were extracted from the curated 2964 micrographs using a box size of 360 pixel down-sampled to 90 pix (4.32 Å). After 2D classification, 282,918 particles were re-extracted without down-sampling, put through a round of 2D classification giving 253,491 particles. Further processing is described below.

804 movies collected from *Ec* Bcs with MBP tagged BcsE was motion corrected and CTF estimated as described above. The TOPAZ model used in particle picking for *Ec* Bcs with cdGMP was applied to the curated data set resulting in 163,853 particles, and further processing is described below.

BcsB hexamer refinement

A single 3D class with six copies of a periplasmic subunit was selected for local refinement. The signal of the membrane and cytosolic domains was removed as follows. Subtraction and local refinement masks were generated in Chimera[37] using a refined full volume mask after Homogenous Refinement of the selected 3D class. The micelle and cytosolic domains were removed with volume eraser and the map was subtracted from the full mask to generate an inverse mask to be subtracted. Both masks were imported into cryoSPARC. Particle subtraction was performed on 190,876 particles using the micelle/cytosolic mask and the refined full volume. The subtracted particles were 2D classified to remove poorly subtracted, or junk particles giving a final 138,286 particles. Using the generated periplasmic mask, local refinement produced maps of sufficient quality to allow *de novo* building of BcsB with a resolution range of ~3.4 – 4.9 Å.

Refinement of BcsB region from incubating with cd-GMP was carried out beginning with 3D heterogenous refinement generating three models by inputting Class I from a second round of 3D heterogenous refinement with where particles from the intial round's Class I and IV were combined. The best class contained 115,289 particles, and was further refined using homogenous followed by NU-refinement. Masks were generated as described above and the signal from the membrane and cytoplasmic domains was subtracted. Local refinement of remaining

BcsB volume was performed and resulted in a map with a resolution range of 3.2 to 5 Å.

BcsAB refinement

Performing particle subtraction and local refinement using the opposite masks generated in BcsB's refinement led to a modest improvement of BcsA's membrane and cytosolic domains. To further resolve BcsA density, a mask was generated for local refinement to cover two copies of BcsB and BcsA using Chimera and following the same methods as above. The resulting map revealed eight transmembrane helices of the core BcsA that aligned well with the known structure of *Rs* BcsA (PDB: 4P00) as well as additional helices identified as *Ec* BcsA's N-terminal and C-terminal helices.

Full Bcs complex refinement

Using particles generated through the Topaz wrapper, particles were classified through three rounds of heterogenous refinement starting with the original 3D volume used to generate the particle set. The final round of refinement gave four volumes containing a complete BcsB hexamer with only minor differences in the cytosolic domains. The volume with the best cytosolic densities was further refined through rounds of homogenous and NU-refinement giving a final volume consisting of 58,576 particles with a resolution range of 3.5 – 9 Å.

BcsAG refinement

Using the above full *Ec* Bcs volume, masks were generated in Chimera to remove the BcsB and BcsE signals using the refined mask as a template. The BcsBQ mask was used to subtract the corresponding signal from the 58,576 particles and the above 3D volume was locally refined with a mask covering the micelle, BcsA, and BcsQ regions.

Bcs with MBP-BcsE

2D classification resulted in 63, 657 particles that were put through a round of 3D heterogenous refinement using the Class 1 as described for BcsB refined from cd-GMP incubated complex. One class with 29,297 particles contained six well defined BcsB subunits with large cytoplasmic densities, and was further refined using NU-refinement.

Model building and refinement

Using locally refined maps of the BcsB hexamer, a single BcsB subunit (chain F) was build de-novo in Coot[40] using secondary structure predictions, similarities with RsBcsB (PDB: 4P00), and side chain densities as guides. The excellent map quality allowed building of the entire subunit. This subunit was then docked as a rigid body into the positions of the other five subunits, followed by manual adjustments and refinements in Coot. The BcsB hexamer was refined in Phenix.refine[41] including secondary structure and non-crystallographic symmetry restraints. The final model misses BcsB's transmembrane anchor (residues 731 to 778) for all subunits, chain A lacks residues 632 – 659, and chains A-E miss residues 446 – 466 due to flexibility and disorder. The final model has 94.3, 5.7, and 0% of residues in the preferred, allowed and disallowed regions of the Ramachandran diagram and has a MolProbity score of 2.0.

Due to significant heterogeneity, BcsA was modeled as a poly-alanine model only. The initial backbone model of *Ec* BcsA was generated using the *Rs* BcsA structure (PDB entry 4P00) as a template. Secondary structure and transmembrane topology predictions combined with sequence alignments were used to inform the extension or truncation of the *Rs* BcsA model in accordance with the *Ec* BcsA sequence. The model was placed into the BcsA density as a rigid body and manually adjusted by rigid body fitting of individual transmembrane helices. Resolved bulky side chains and the helical pitch guided placements. The initial rigid body docking of the entire BcsA poly-Ala model also placed the cytosolic GT and PilZ domains into their corresponding volumes. These domains were not individually adjusted due to insufficient resolution.

The final model contains residues 18 through 871 as a poly-Ala chain, missing residues include 46-112 that form the connection between TM helix 1 and 2. Only side chains of residues within TM helices used as register guides are included. The assignment of TM helix 1 is speculative due to the missing linker between TM helix 1 and 2. However, the assignment is based on predicted secondary structure of *Ec* BcsA's N-terminal region, TM topology predictions, and the recently demonstrated interaction of this region with BcsG.[22]

The final *Ec* BcsA model was combined with chain A of the refined BcsB hexamer (including side chains), which was also docked as a rigid body into the corresponding density. The model was refined in Phenix.refine with BcsA and BcsB as rigid bodies.

All model building was done in Coot. Figures were prepared using Pymol and Chimera.[36,37]

Cellulose enzyme assays

Purified Bcs was assayed for cellulose production as previously described.[24] Briefly, Buffer C with Detergent C was mixed with 5 mM UDP-Glc supplemented with 1.25 μCi of UDP-Glc[6-^{3}H] to create a reaction master mix. 20 μL reactions were performed by adding 5 μL of purified Bcs (0.5 mg/mL) to 12 μL of master-mix. 3 μL of a 200 μM stock solution of cyclic di-GMP (30 μM final), was added to initiate the reaction, and incubated at 30°C for 1 hour. Controls were performed replacing cd-GMP with ddH_2O, or addition of 1 mg/mL cellulase BcsZ. The reactions were terminated with 5 μL of 10% SDS and 25 μL were applied to Whatman 2MM blotting paper. Non-cellulosic material was removed by paper chromatography using 60% ethanol, and the origins were counted on a Beckman S Ls liquid scintillation counter. Each condition was performed in triplicate and error bars represent the deviations from the means.

Activity assays for *Ec* BcsA in IMVs were performed as described for 30 minutes in reaction master mix lacking detergents. To ensure the incubation time was in the linear phase, a time course with IMVs expressing the full complex was performed. To confirm the signal produced from the reaction was cellulose a

specific 1,4 betaglucanase, CelTR (Megazyme), was added to 1 mg/mL prior to initiation with cd-GMP. One set of reactions was stopped at 30 minutes by addition of 10 mM EDTA, followed by addition of 2% triton X-100 and 1 M sodium acetate pH 4.5 (final concentration 200 mM) to solubilize the IMVs and lower the pH to the cellulase's optimum range. These samples were incubated for 45 min at 40°C to allow the cellulase access to reaction product protected by stable vesicles and the total reaction was loaded on to the blotting paper. All activity assays were performed in triplicate, excluding the time course where duplicates were performed.

Purified BcsABFG activity assays

To assess whether cytoplasmic *Ec* Bcs proteins had an effect on BcsA activity pETDuetEcBcs_A-12His_nSS-Strep-B_Adra-6His, pACYCDuet_EcBcs_PelB-8His-C_Z_F_G-FLAG, while omitting pCDF_EcBcs_R_Q_HA-E, was expressed and purified as described for the full inner-membrane complex. The resulting detergent solubilized BcsABFG was assayed as described above with the following alteration. Protein concentration was estimated to be 670 nM based on an extinction coefficient of 906 mM-1 and molecular weight of 740 kDa for hexamer BcsB complexes. Cytoplasmic proteins BcsRQ, MBP-BcsE, and BcsRQ-MBP-BcsE were added in 10-fold molar excess. Controls for all individual cytoplasmic proteins without BcsABFG were also performed at 6.7 μM each. All assays were performed in triplicate.

Purified BcsABFG activity assays

To assess whether cytoplasmic Bcs proteins affect BcsA activity we co-expressed the following constructs in the absence of BcsQ, -R and -E: pETDuetEcBcs_A-12His_nSS-Strep-B_Adra-6His and pACYCDuet_EcBcs_PelB-8His-C_Z_F_G-FLAG. The BcsABFG complex was purified as described for the full IMC. The resulting detergent solubilized BcsABFG was assayed as described above with the following alterations. The assay buffer contained 0.003% LMNG, 0.0006% DMNG, and 0.0006% CHS and protein concentration was estimated to be 670 nM based on an extinction coefficient of 906 mM-1 and molecular weight of 742 for hexamer BcsB complexes.

Cytoplasmic proteins BcsQR, MBP-BcsE, and BcsQR-MBP-BcsE were added in 10-fold molar excess.

Enzyme Coupled Assay

The formation of UDP during cellulose synthesis was quantified using an enzyme coupled assay, as previously described.[24] 100 µL reactions were performed in a corning 96 well clear flat bottom plate with the following conditions. A master-mix containing 0.1 mg/mL Bcs in Buffer C with Detergent C, 8 U of pyruvate kinase and 8 U of lactate dehydrogenase, 0.5 mM NADH and increasing concentrations of UDP-Glc from 0 to 5 mM. NADH absorbance at 340 nm was monitored for 2 hours at 30°C using a Molecular Devices SpectraMax.

To determine the Km of cd-GMP binding, UDP-Glc was kept constant at 5 mM and cd-GMP was titrated from 0 – 30 µM. Analysis was performed on the linear portion of the data, usually representing the 20 minutes after initiation of the reaction using GraphPad. All assays were performed in triplicate and error bars represent the deviations from the means.

Purification of *in vivo* synthesized cellulose

pEtN cellulose was extracted from transformed *Ec* C43 according to published procedures with slight variations.[3] Four 1 L cell cultures were grown in TB-M-80155 autoinduction media as described above. The cultures were grown with slow shaking for 24 hours at 28°C, and the cells were pelleted by centrifugation at 5,000 g for 15 minutes. Cells were resuspended in 10 mM Tris pH 7.4 and sheared with five rounds of the OmniMixer homogenizer (1 min on, 2 min rest) on ice. Cells were pelleted and removed by three rounds of centrifugation 10,000 g for 10 minutes. The supernatant was dialyzed against ddH_20 for 24 hours. The resulting solution was frozen, thawed, and centrifuged at 13,000 g pelleting the insoluble material. The pellet was treated with 4% SDS in Tris buffer overnight with subsequent Tris buffer washes and centrifugation steps to remove SDS. A final wash was performed in ddH_20. The resulting pEtN cellulose was lyophilized and analyzed using solid state NMR.

Solid-state NMR analysis

The ^{13}C cross-polarization magic-angle spinning (CPMAS) solid-state NMR measurement was performed at ambient temperature in an 89 mm bore 11.7 T magnet (Agilent Technologies, Danbury, CT) using an HCN Agilent probe with a DD2 console (Agilent Technologies) operating at 499.12 MHz for ^{1}H and 125.52 MHz for ^{13}C. Samples were spun at 7143 Hz in 36 µL capacity 3.2 mm rotors. ^{1}H-^{13}C matched cross-polarization transfer was performed at 50 kHz for 1.5 ms and proton decoupling was performed at 83 kHz. 100,000 scans were obtained for the spectrum with a 2s recycle delay. The ^{13}C chemical shift scale is referenced to TMS as 0.0 ppm using a solid adamantine sample at 38.5 ppm.

Congo-red fluorescence measurements

Small cultures expressing the Bcs complex either in its entirety, with a catalytically inactive BcsG subunit, or without the outer-membrane porin BcsC were tested for Congo-red binding and fluorescence using a UV light box and Molecular Devices SpectraMax plate reader exciting at 525 nm and reading emission at 610 nm as previously described.[23]

Methods References

38 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).

39 Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* **16**, 1153-1160, (2019).

40 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, (2004).

41 Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-367, (2012).

42 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl Acid Res* **32**, 1792-1797, (2004).

43 Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinform* **25**, 1189-1191, (2009).

44 Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucl Acid Res* **43**, W389-394, (2015).

Acknowledgements

We thank Kelly Dryden and Michael Purdy from the University of Virginia Molecular Electron Microscopy Core (MEMC) facility for help with data collection. MEMC is funded by NIH Recovery grant 1G20RR31199. The MEMC Titan Krios and Gatan K3/GIF detector were funded through NIH grants SIG S10-RR025067 and U24-GM116790, respectively. We thank Nicholas Sherman and the University of Virginia Biomolecular Analysis Center for mass spectrometry analysis and Georgios Skiniotis for discussions. JFA was supported in part by NIH grant 1F32GM126647-01. RH was supported by the Center for Lignocellulose Structure and Formation, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Basic Energy Sciences (Award Number DESC0001090). JZ is supported by NIH grant 5R01GM101001. LC acknowledges support from the National Science Foundation, NSF Award 2001189. N.G. is a recipient of the NSF Predoctoral Fellowship.

Author contributions

JFA and JZ designed the experiments. JFA cloned, expressed and purified the Bcs complex. RH and JFA screened and optimized cryo grid preparation. JFA processed all data. JZ and JFA built the atomic models. NG and LC performed the ssNMR analysis. All authors edited the manuscript.

Competing interests

The authors declare no competing interests.

Materials and Correspondence

Correspondence and requests for materials should be addressed to Jochen Zimmer (jz3x@virginia.edu).

Data Availability

Coordinates and EM maps have been deposited at the PDB/EMDB under accession codes XXX and YYY.

Extended Data Figure and Table Legends

Extended Data Table 1| Mass spectrometry peptide analysis. Identified peptides are shown in red.

Extended Data Table 2| Cryo-EM data collection, refinement and validation statistics.

Extended Data Figure 1| Predicted TM topologies and secondary structures of the *E. coli* cellulose synthase subunits.

Extended Data Figure 2| Expression, purification and characterization of the *Ec* Bcs complex. (**a**) Clumping cells upon expression of the recombinant Bcs genes with AdrA in liquid cultures. Cells without BcsC remain planktonic. (**b**) Cultures in (a) were grown on Congo red agar plates and photographed on a Fisher Scientific UV lightbox. (**c**) Overlay of the ssNMR spectrum shown in Fig. 1 with the pEtN standard. (**d**) Gel filtration purification of *Ec* Bcs, 256 and 280nm absorbances are shown in red and blue, respectively. Inset: SDS-PAGE of the peak fractions. (**e**) Catalytic activity determined by UDP-Glc[^{3}H] incorporation as described.[24] The synthesized polymer is sensitive to cellulase (BcsZ) degradation. DPM: Disintegrations per minute. (**f**) Western blot analysis of the purified *Ec* Bcs complex. Right panel: Coomassie-stained SDS-PAGE of the purified complex. Percentages indicate peptide covered in MS-sequencing based on the identified unique peptides (also shown). R/NR: Reducing and non-reducing conditions; * denotes a BcsG proteolytic fragment. (**g**) Identification of Myc-tagged BcsF by Western blotting in the purified complex (left Coomassie-stained SDS-PAGE; right: anti-Myc Western blot). (**h, i**) Kinetic analysis of the purified detergent-solubilized samples. UDP generation was quantified in real time using an enzyme-coupled assay that monitors the oxidation of NADH, as described.[24]

Extended Data Figure 3| Bcs catalytic activity and complex formation of MBP-BcsE and BcsQR. (**a**) Time course of cellulose biosynthesis from IMVs containing BcsABFG. (**b**) The BcsQR complex, MBP-BcsE, and the BcsQR-MBP-BcsE complex were analyzed by S200 size exclusion chromatography. The samples elute at the volumes indicated above the main peaks. Inset: Coomassie-stained SDS-PAGE of all samples (1: BcsQR-MBP-BcsE, 2: MBP-BcsE, 3: BcsQR).

Extended Data Figure 4| Outline of EM data processing and refinement of BcsB in cryoSPARC. (a) Representative micrograph and particle, 2D, and 3D refinement workflow representing processing route 1 of the initial combined data sets. Red boxes indicate combined classes used to generate volumes described in Extended Data Fig. 5. **(b)** Representative micrograph and particle, 2D, and 3D refinement workflow from the cd-GMP incubated data set. **(c)** BcsB local refinement masks, volume, resolution data, and orientation plot generating the map used to build the initial model. **(d)** BcsB local

refinement masks, volume, and resolution data, and orientation plot generating the final map used to build the final model.

Extended Data Figure 5| Outline of EM data processing in cryoSPARC. (a) Representative micrograph and particle, 2D, and 3D refinement workflow representing further processing of route 1 of the initial combined data. **(b)** Workflow of Route 2 re-picking particles with TOPAZ. **(c)** Representative micrograph and particle, 2D, and 3D refinement workflow from the IMC MBP-BcsE data set. **(d)** BcsAB local refinement masks and volume. **(e)** Full volume local resolution. **(f)** BcsAG local refinement subtraction masks and volume. **(g)** Full volume refinement of IMC MBP-BcsE volume. Red circles indicate areas of additional densities.

Extended Data Figure 6| Examples of BcsB density. (a-d) Map quality is shown for each of the BcsB domains of Chain B including β-strands, loops, and α-helices. Density is contoured to σ=2.0, and carve=2.0. **(e)** Disulfide bond between CBD-1 and CBD-2, σ=1.5 and carve=2.0. **(f)** Oligomerization β-strands and loops between FD domains of chain A and chain B, σ=1.5 and carve=2.0. Maps were converted to CCP4 format with Phenix, and figures were created in Pymol.

Extended Data Figure 7| Comparison of BcsB from *R. sphaeroides* and *E. coli*. The OL region forms an amphipathic interface helix at the periplasmic water/lipid interface in *R. sphaeroides* BcsB (PDB: 4P00). In *E. coli* BcsB, the OL is helical (residues 622-629) followed by a poorly order stretch (residues 631-658) in chain A. This region forms the interface with neighboring subunits in all other chains. Conserved cysteines covalently linking the CBDs are shown as spheres. The BcsB TM anchor is indicated as a cylinder.

Extended Data Figure 8| Comparison of BcsB sequences. BcsB sequences from *E. coli*, *K. xylinus,* and *R. sphaeroides* were aligned in MUSCLE[42] and displayed in Jalview[43] colored based on sequence conservation. Secondary structure elements observed in the *Ec* and *Rs* BcsB structures are indicated as bars and arrows for helices and β-strands, respectively. The *Kx* BcsB secondary structure of the OL was predicted in Jpred[44] (colored magenta). Residues mediating intermolecular contacts in the *Ec* BcsB hexamer are indicated by yellow circles.

Extended Data Figure 9| The *Ec* BcsA poly-alanine model. The *Ec* BcsA poly-Ala model was generated based on *Rs* BcsA (PDB: 4P00) and manually docked into the cryo-EM density. TM helices were adjusted individually by rigid body docking in Coot. The GT and PilZ domains were only corrected for missing or additional residues and not adjusted due to insufficient resolution. Side chains of residues used to guide the register assignment are included in the model. The map is contoured at 7.5σ.

Extended Data Figure 10| Comparison of *Rs* and *Ec* BcsA. (**a**) *Rs* BcsA (PDB: 4P00) was superimposed with the *Ec* BcsA model by secondary structure matching. Regions not present in *Rs* BcsA are shown in dark blue. (**b**) Comparison of the *Rs* and *Ec* BcsAB complexes. BcsB is colored green.

Extended Data Figure 11| Putative BcsG density. Overall Bcs cryo-EM map contoured at the indicated levels. Additional density is visible for BcsG at low contour levels. Bottom right: The map contoured at 0.8σ was Gaussian filtered at a width of 2.6 in Chimera. Two copies of *Ec* BcsG (PDB: 6PCZ) were manually docked into the putative BcsG density and colored blue to red from the N- to the C-terminus. Helices representing the helical cluster next to BcsA are shown as cartoons in black and magenta. The first and last BcsB subunits are shown as spheres in green and steel-blue, respectively.

Extended Data Figure 12| The cytosolic Bcs components. (**a**) 2D class averages of Bcs particles lacking BcsE and BcsQ/R and (**b**) corresponding 3D volume (blue surface) superimposed with the BcsE/QR-containing macrocomplex (mesh) revealing only cytosolic density corresponding to BcsA. (**c**) Localization of MBP-fused BcsE in the Bcs complex. Volumes are contoured at 0.7σ and Gaussian filtered at a width of 5.2 in Chimera. Densities assigned to BcsQR and BcsE are colored orange and magenta, respectively, for the MBP-BcsE containing complex. (**d**) Available structures of Bcs components manually docked into their corresponding densities. (**e**) Domain organization of BcsE with corresponding functions. Adapted from reference ([18]). (**f**) SDS-PAGE and Western blot analyses of the MBP-BcsE containing Bcs complex. The gels were run under reducing conditions in which BcsB co-migrates with BcsA. (1: Bcs complex, 2: MBP-BcsE control; a, b c: MBP-BcsE, reduced BcsA/BcsB, BcsG, respectively). (*) indicates non-specific antibody binding to the strong BcsA/B band.