BIOCHEMISTRY
including biophysical chemistry & molecular biology

ARTICLE

pubs.acs.org/biochemistry

The Crystal Structure of *Escherichia coli* Group 4 Capsule Protein GfcC Reveals a Domain Organization Resembling That of Wza

Karthik Sathiyamoorthy,[†,⊥] Erez Mills,[‡,@] Titus M. Franzmann,[§] Ilan Rosenshine,[‡] and Mark A. Saper*[,†,∥]

[†]Program in Biophysics, University of Michigan, Ann Arbor, Michigan 48109-1055, United States

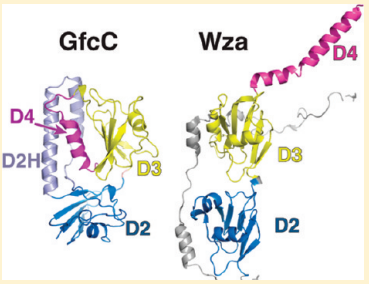
[‡]Department of Microbiology and Molecular Genetics, The Institute for Medical Research, Hebrew University of Jerusalem, 91120 Jerusalem, Israel

[§]Department of Molecular, Cell and Developmental Biology, University of Michigan, Ann Arbor, Michigan 48109-1048, United States

[∥]Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan 48109-1055, United States

Supporting Information

ABSTRACT: We report the 1.9 Å resolution crystal structure of enteropathogenic *Escherichia coli* GfcC, a periplasmic protein encoded by the *gfc* operon, which is essential for assembly of group 4 polysaccharide capsule (O-antigen capsule). Presumed gene orthologs of *gfcC* are present in capsule-encoding regions of at least 29 genera of Gram-negative bacteria. GfcC, a member of the DUF1017 family, is comprised of tandem β-grasp (ubiquitin-like) domains (D2 and D3) and a carboxyl-terminal amphipathic helix, a domain arrangement reminiscent of that of Wza that forms an exit pore for group 1 capsule export. Unlike the membrane-spanning C-terminal helix from Wza, the GfcC C-terminal helix packs against D3. Previously unobserved in a β-grasp domain structure is a 48-residue helical hairpin insert in D2 that binds to D3, constraining its position and sequestering the carboxyl-terminal amphipathic helix. A centrally located and invariant Arg115 not only is essential for proper localization but also forms one of two mostly conserved pockets. Finally, we draw analogies between a GfcC protein fused to an outer membrane β-barrel pore in some species and fusion proteins necessary for secreting biofilm-forming exopolysaccharides.



Many bacteria secrete and export high-molecular weight polysaccharides that can associate with the outer membrane as a capsule (CPS) or are secreted into the environment as exopolysaccharide (EPS).[1] CPS and EPS are polymers comprised of repeating oligosaccharides, each comprised of a strain-specific structure of one to seven sugar residues. Although lipids anchor some capsules to the cell surface, the attachment mechanism for most is unknown. Capsules function as virulence factors to inhibit host cell phagocytosis and block killing by antimicrobial peptides or complement[2−4] and are often essential for pathogenicity.[5,6] Capsules and EPS may also mediate the attachment of bacteria to host tissues or inorganic surfaces, such as medical implants.

Progress is being made toward understanding the different types of CPS and EPS, and the mechanisms of their synthesis and translocation to the cell surface.[1,7] For Gram-negative bacteria, the best-characterized system is the group 1 capsule, typified by the *Escherichia coli* K30 capsule and the colanic acid EPS. Here a capsule-specific operon encodes the enzymes that assemble the lipid-linked oligosaccharide repeat units in the cytoplasm. An inner membrane oligosaccharide-specific glycosyl transferase, Wzy, adds units to the growing polymer in the periplasm. A complex of the inner transmembrane protein, Wzc, an essential transmembrane tyrosine kinase,[8,9] and the outer membrane auxiliary (OMA) protein, Wza,[10,11] are necessary for the export of the polymer through the outer membrane. The crystal structure of a Wza octamer [Protein Data Bank (PDB) entry 2J58 (Figure 1A)] showed that each protomer is comprised of three domains and a carboxyl-terminal amphipathic helix.[12] The helices from each of the eight chains form a porelike channel (≈17 Å in diameter) thought to reside in the outer membrane, a fold previously not observed in bacterial outer membrane proteins. Three-dimensional reconstruction by electron microscopy of a complex between the Wza octamer and a Wzc tetramer showed a continuous conduit spanning the entire periplasm,[11] yet questions about how the polysaccharide enters this conduit and how the capsule is anchored to the cell surface remain.

Enteropathogenic (EPEC) and enterohemorrhagic (EHEC) strains of *E. coli* secrete a group 4 capsule (GFC) assembled from the same O-antigen repeat units that comprise the much shorter lipopolysaccharide (LPS).[13−15] The EHEC GFC facilitates adhesion to host cells and is important for host colonization in vivo.[15] Also, GFC contributes to virulence in *Vibrio* and *Salmonella* strains.[2,5,16]

In EPEC and EHEC strains, group 4 capsule expression requires the seven-gene *gfc* operon (Figure 1B).[14,15] The essential *gfc* genes, *gfcE*, *etp*, and *etk*, are paralogous to the *cps* genes encoding the group 1 capsule proteins Wza, Wzb, and Wzc, respectively.[10,14,17,18] The *gfcABCD* genes encode proteins of unknown function but are also

Received: November 23, 2010
Revised: March 30, 2011
Published: March 30, 2011

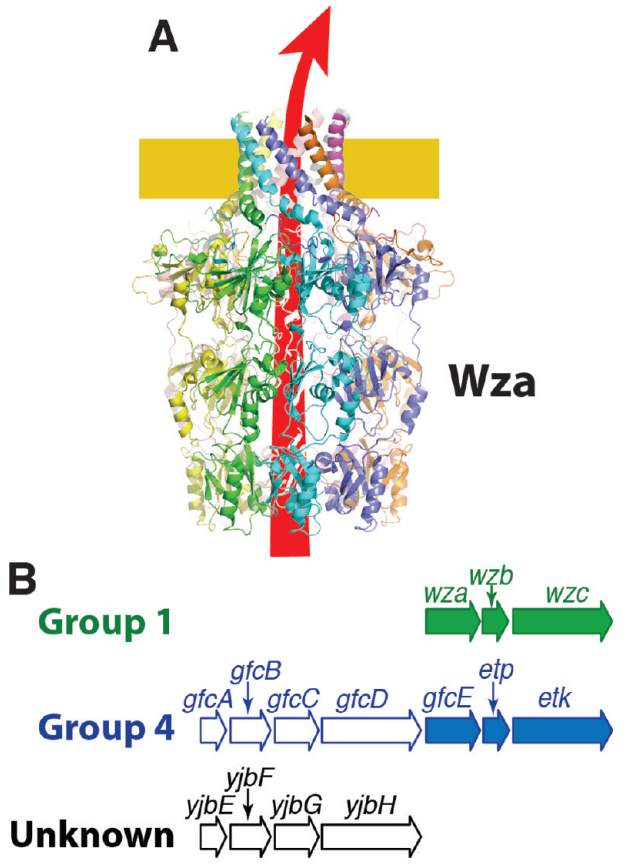


Figure 1. Wza and operons for capsule export. (A) Cartoon representation of *E. coli* Wza (PDB entry 2J58) as an octamer. It forms an octameric pore in the outer membrane.[12,50] (B) *E. coli* operons essential for polysaccharide secretion. The *yjbEFGH* operon regulates a yet uncharacterized polysaccharide.[20]

essential for GFC expression.[14] Homologues of these genes are found in at least 29 genera of Gram-negative bacteria. To date, characterization of the homologous genes in three species points to a role in polysaccharide secretion. The *yjbEFGH* operon, also on the *E. coli* chromosome, encodes paralogous proteins that are 30−65% identical to the *gfcABCD* proteins. Under control by the Rcs stress regulon, they facilitate secretion of an unidentified exopolysaccharide during osmotic stress.[19,20] Orthologs of *gfcBCD* in *Vibrio anguillarum*, *wbfDCB*, also affected secretion of a group 4-like exopolysaccharide important for virulence.[21] In *Vibrio parahemolyticus* O3:K6, genes homologous to *wza*, *wzb*, and *wzc* are not necessary for expression of K-antigen capsular polysaccharide, but genes VP0218−VP0215, homologous to *gfcABCD*, were important for regulating the amount and size of the expressed capsule.[22]

Little is known about the structure or function of the *gfcABCD*-encoded proteins. GfcA is a short inner membrane protein that is rich in Thr, Ala, and Gly residues. GfcB is an outer membrane lipoprotein with a β-barrel-like structure [PDB entry 2in5 (W. Zhao et al., 2006, unpublished observations)], and GfcD is predicted to be an outer membrane 22-strand β-barrel lipoprotein.[23,24] Such outer membrane β-barrels also serve as polysaccharide export pores in some exopolysaccharide systems.[25−27] Because GfcD is encoded upstream of the Wza paralog GfcE, the presumed export protein for GFC, we hypothesized that the *gfcABCD*-encoded proteins are also involved in the translocation of polysaccharide or an accessory molecule across the outer membrane.

GfcC is a 248-residue periplasmic protein and member of the DUF1017 (PF06251) family of proteins of unknown function (Pfam database[28]). DUF1017 family members were annotated to contain a β-grasp-like domain. The β-grasp fold is a four- or five-strand mixed β-sheet with one helix. First identified in the eukaryotic protein ubiquitin, the fold is present in a diverse group of proteins, including numerous enzymes; iron−sulfur, RNA, and cofactor binding proteins; and those involved with post-translational modification.[29] Included in the superfamily of β-grasp folds is the SLBB (soluble ligand-binding β-grasp) domain, a Pfam sibling to DUF1017.[30] The SLBB family includes the two β-grasp domains in Wza and other polysaccharide export proteins,[31] as well as a variety of enzymes and ligand-binding proteins. In Wza, the SLBB domains of each subunit pack against those from adjacent subunits to form the cylindrical octamer. Some Wza-like "outer membrane auxiliary proteins" also have a DUF1017 fold at the C-terminus.[31]

In this paper, we begin to explore the role of the *gfc*-encoded proteins, by reporting the crystal structure of GfcC at 1.9 Å resolution. The structure reveals a domain architecture similar to that of Wza, but in a compact, soluble protein.

EXPERIMENTAL PROCEDURES

GfcC Sequence Alignment. To evaluate the location of residues conserved in homologues of EPEC GfcC (GI|215486102), we downloaded sequences from a precomputed Blast search of GfcC (BLink at NCBI, June 2010) with Blast scores greater than 100. Sequences that were ≥98% identical to any other sequence were deleted to avoid bias. Similarly, proteins longer than 290 residues were excluded. Thirty-five sequences were aligned with ClustalX[32] and evaluated with Jalview.[33] Sequence conservation was calculated on this alignment with Scorecons.[34]

Design of the GfcC Construct. The *gfcC* gene encoding amino acid residues 22−248 (i.e., without the signal peptide) was amplified by polymerase chain reaction (PCR) from genomic DNA derived from *E. coli* O127:H6 strain E2348/69 with forward primer 5′-CGACCCATGGCGCAAGGAATGGTGACT-3′ and reverse primer 5′-CGTCTCGAGCTCAGGAACACGTTGCGTTA-3′. After digestion with NcoI and XhoI, the purified oligonucleotide was cloned into a similarly cut pETBlue2 vector (EMD Biosciences), which encoded a C-terminal Leu-Glu-His_6 tag. Sequencing the resultant plasmid DNA (University of Michigan DNA Sequencing Facility) confirmed the predicted sequence.

Expression and Purification of GfcC. Tuner(DE3)pLacI cells (Novagen/EMD Biosciences) were transformed with the pETBlue2-GfcC plasmid. Typically, 1 L of Terrific broth was inoculated with 10 mL of an overnight culture and grown with shaking at 37 °C, until A_{600} reached ≈0.8. After the sample had cooled to room temperature, IPTG was added to a final concentration of 0.2 mM and protein expression proceeded for 18 h at room temperature. Following centrifugation for 20 min at 5000 rpm in a Beckman JLA-10.1 rotor, cells were suspended in 40 mL of 300 mM NaCl, 20 mM Tris-HCl buffer (pH 8.0) and one-half of an EDTA-free protease inhibitor cocktail tablet (Sigma). Cells on ice were disrupted by sonication (Branson) with the macro tip operating at 40% of the maximal power amplitude for 5 min, cycling 30 s on and 15 s off. After the lysate had been centrifuged at 48400*g* (20000 rpm in a JA-20 rotor) for 40 min, the supernatant was passed through a 0.22 μm syringe filter (Millipore) and applied to an 8 mL column of Talon Superflow

(Clontech) by the ÄKTAfplc system (GE Biosciences). After being washed, bound protein was eluted with a linear gradient from 0 to 300 mM imidazole in lysis buffer. GfcC eluted at about 60 mM imidazole. Fractions containing GfcC were dialyzed overnight in 4 L of lysis buffer, concentrated, and loaded onto a HiLoad 16/60 Superdex 75 (GE Healthcare) size exclusion column equilibrated with the same buffer. Purified protein for crystallization trials was concentrated to ≈23 mg/mL (as determined by A_{280}) with a 10 kDa cutoff centrifugal filter (Amicon).

The selenomethionine derivative of GfcC (SeMet-GfcC) was purified similarly. Before induction with IPTG, the cells were centrifuged and resuspended in 1 L of minimal medium with amino acids (Athena) and selenomethionine (0.8 μg/mL). Buffers during the subsequent purification included 2 mM TCEP (Soltec Ventures Inc.) to prevent selenomethionine oxidation.

Structure Determination. Protein drops were screened for crystallization by sitting drop vapor diffusion with several commercially available screens (Qiagen pHClear and PACT, Hampton Research Index). The best crystals for the native protein were grown from drops containing 1 μL of GfcC (23 mg/mL) and 1 μL of 1.6 M ammonium sulfate, 0.1 M Tris-HCl (pH 8.0) (pHClear E11), and 0.2 μL of 30% (w/v) 1,5-diaminopentane dihydrochloride (Hampton Research), equilibrated against 50 μL of a reservoir solution. Drops were set up at 4 °C and later heated to 22 °C to nucleate crystals. SeMet-GfcC crystallized with 1.5 M ammonium sulfate, 0.1 M NaCl, and 0.1 M Bis-Tris (pH 6.5) (Index C6) as the precipitant. Crystals were harvested and frozen in the original reservoir solution containing 15% glycerol as a cryoprotectant.

Anomalous diffraction data at three wavelengths were collected from the SeMet-containing crystals at the LS-CAT beamlines (Advanced Photon Source, Argonne, IL) and processed with HKL2000.[35] Only the data at the Se absorption peak wavelength (0.97930 Å) were phased by the single-wavelength anomalous diffraction method with the AutoSol scheme in Phenix.[36] Eighty-three percent of the residues were built in the initial maps. The remainder of the structure was built manually and refined against the peak anomalous Se data in the presence of NCS restraints using Phenix and Coot.[37] Final steps in the refinement included adding the His_6 tag of chain A and TLS parameters. Complete statistics are listed in Table 1. The final structure was deposited in the Protein Data Bank (RCSB) as entry 3P42.

Structures were superimposed with LSQMAN, first with a 6 Å cutoff and then with a 3.5 Å cutoff. All figures in this paper (except Figure 2) were prepared with MacPyMOL.[38]

RESULTS

GfcC Homologues Form a Subgroup of DUF1017 Proteins. Proteins representing 29 genera of proteobacteria, homologous in sequence to residues 76−248 of GfcC and consisting of fewer than 290 residues, form a subgroup of the Pfam DUF1017 family. The alignment of 35 homologues (available as Supporting Information) gives identities from 19 to 89%, with the majority being between 20 and 40%, and all have signal sequences consistent with secretion to the periplasm. Most are encoded in operons between *gfcB* and *gfcD* gene homologues. Figure 2 shows the alignment of eight representative sequences highlighting the identical residues. The region defined by DUF1017 begins at residue 76 of GfcC where homology is most prevalent. In addition to the GfcC homologues, the DUF1017 family also includes longer proteins with domains homologous to GfcC.

Table 1. Diffraction Data and Structure Refinement Summary[a]

	SeMet-GfcC(22−248)-His_6
beamline	21 ID-D, LS-CAT, APS
wavelength (Å)	0.97930
space group	$P2_1$
unit cell dimensions	a = 68.83 Å, b = 99.98 Å, c = 69.02 Å, β = 91.74°
no. of molecules per asymmetric unit	4
resolution (Å)	34.5−1.91 (1.98−1.91)
total no. of reflections	71575 (8700)
redundancy	7.4 (5.3)
R_{merge}[b]	0.091 (0.469)
I/σ_I	30.88 (2.98)
Refinement	
no. of working reflections	133527 (12580)
completeness (%)	98.5 (96)
R_{work}[c]	0.17 (0.26)
R_{free}[d]	0.22 (0.31)
no. of atoms (non-hydrogen)	7860
protein	7143
water	702
sulfate	15
mean B_{iso} (Å^{-1}, non-hydrogen)	37.1
protein	36.8
solvent	39.4
rmsd for bonds (Å)	0.008
rmsd for angles (deg)	1.110
Molprobity score	1.06[e]
Ramachandran favored (%)	98.68
Ramachandran outliers (%)	0.22
Molprobity clash score (no. of clashes)	0.98 (39)

[a] Values in parentheses are for highest-resolution shell. [b] $R_{merge} = [\sum_{hkl}\sum_{i}|I_i(hkl) - \bar{I}(hkl)|]/[\sum_{hkl}\sum_{i} I_i(hkl)]$. [c] $R_{work} = (\sum_{hkl}||F_{obs}| - |F_{calc}||)/(\sum_{hkl}|F_{obs}|)$ calculated over all reflections used in refinement. [d] R_{free} is similar to R_{work} but calculated from the 5% of the total number of reflections omitted from the refinement. [e] Calculated with hydrogen atoms positioned explicitly by Phenix.

Determination of the GfcC Structure. For structural studies, the mature form of GfcC (residues 22−248) was expressed and purified with a carboxyl-terminal hexahistidine (His_6) tag. The 236-residue (predicted M_r = 26051) soluble protein migrated at about 26 kDa by sodium dodecyl sulfate−polyacrylamide gel electrophoresis and at 28 kDa by gel filtration. Mass spectrometry of the purified native protein showed a major peak at 25919 Da that corresponded exactly to a monomer of GfcC without the mass of the N-terminal Met residue. The selenomethionine-containing protein (SeMet-GfcC) mass spectrum showed the primary species at 26060 Da, which is identical to native GfcC with three selenomethionine residues.

GfcC-His_6 crystallized in two different forms. The SeMet-GfcC crystal was in space group $P2_1$ with four molecules in the asymmetric unit. Diffraction data from this crystal were phased to 1.9 Å by single-wavelength anomalous diffraction (SAD), and the refined structure had an R_{work} of 0.173 and an R_{free} of 0.217 (see Table 1). The three predicted SeMet residues (but not the N-terminal SeMet) were resolved in each chain. The mean rmsd between noncrystallographic symmetry (NCS)-related chains is 0.66 Å (C_α), ranging from 0.58 between chains A and B to 0.73 Å between chains A and D. The structure of native GfcC crystals, in tetragonal space group $P4_32_12$ with two molecules per asymmetric

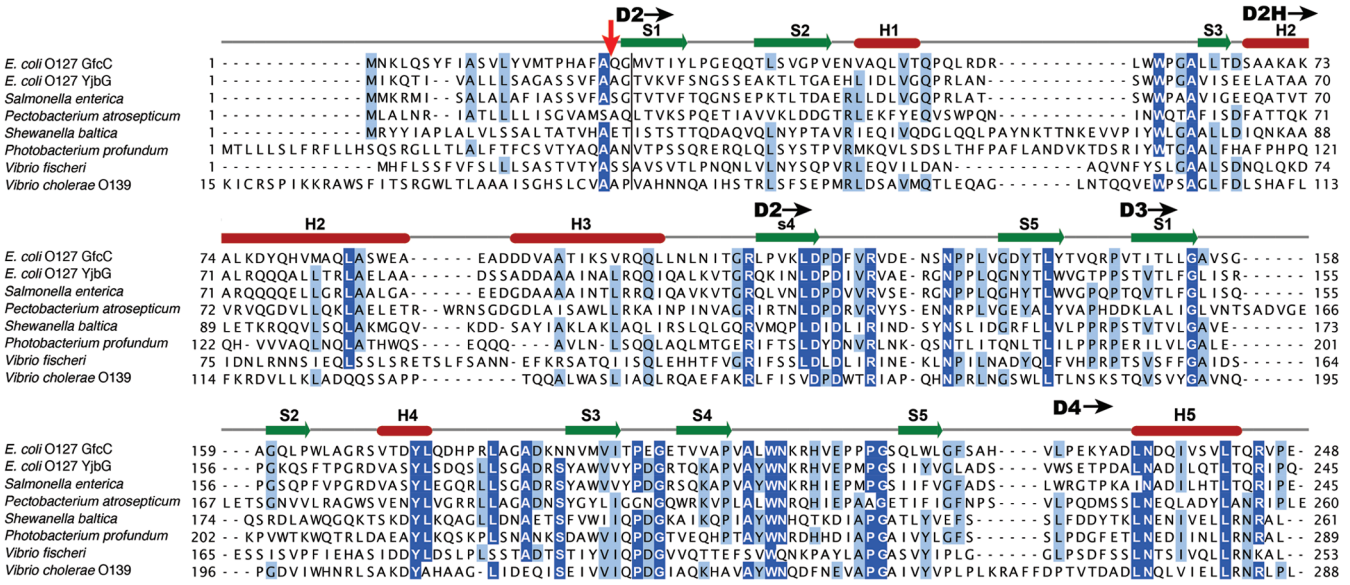


Figure 2. Alignment of selected sequences homologous to *E. coli* GfcC. Thirty-five sequences were selected from the precomputed Blast results of EPEC GfcC (NCBI, February 2010). All had alignment scores of >100, consisted of fewer than 290 residues, and did not have more than 98% identity with any other sequence in the group. Eight sequences from this alignment are shown here (complete alignment available as Supporting Information). Residues are shaded if >60% of the sequences in the alignment of 35 sequences are identical. Dark backgrounds with white letters represent residues identical in >80% of the sequences. The red arrow indicates the predicted signal peptidase 1 cleavage site in *E. coli* GfcC.

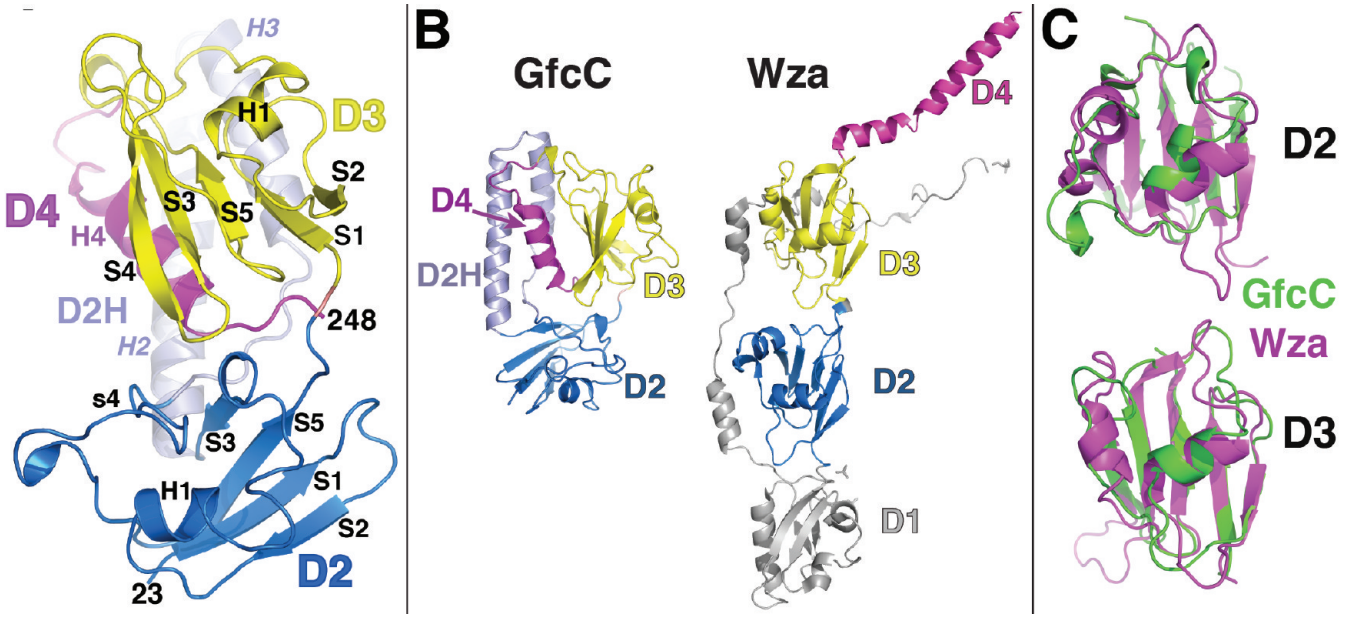


Figure 3. Structure of GfcC and comparison to Wza. (A) Cartoon representation of GfcC oriented with the two β-grasp domains in front. Coloring highlights the structural domains, as defined in the text: blue for D2, bluish-gray for D2H, yellow for D3, and magenta for D4. Most secondary structures are labeled in black. Note that names of the secondary structure elements in D2 and D3 correspond to a generic β-grasp domain.[30] (B) Comparison of GfcC and the Wza monomer (in the conformation found in the octamer, PDB entry 2J58[12]). Corresponding domains are colored similarly. GfcC does not have an equivalent of the D1 domain, while Wza lacks D2H. GfcC is rotated 90° in the vertical direction from panel A. (C) GfcC domains (green) superimposed on the corresponding domains of Wza (magenta). The top structure is D2 of GfcC on domain 2 (residues 169−252) of Wza. The bottom structure is D3 of GfcC superimposed on domain 3 (residues 254−344) of Wza. Structures were superposed with LSQMAN.[51]

unit, was determined by molecular replacement using chain A of the SeMet-GfcC structure as a search model (results not shown), but the final structure was of poorer quality than the SeMet-GfcC structure and will not be discussed further. The GfcC polypeptides in the asymmetric unit of both space groups pack as apparent dimers.

GfcC Contains Two β-Grasp Domains. For the sake of simplicity, we divide GfcC into three domains denoted here as D2

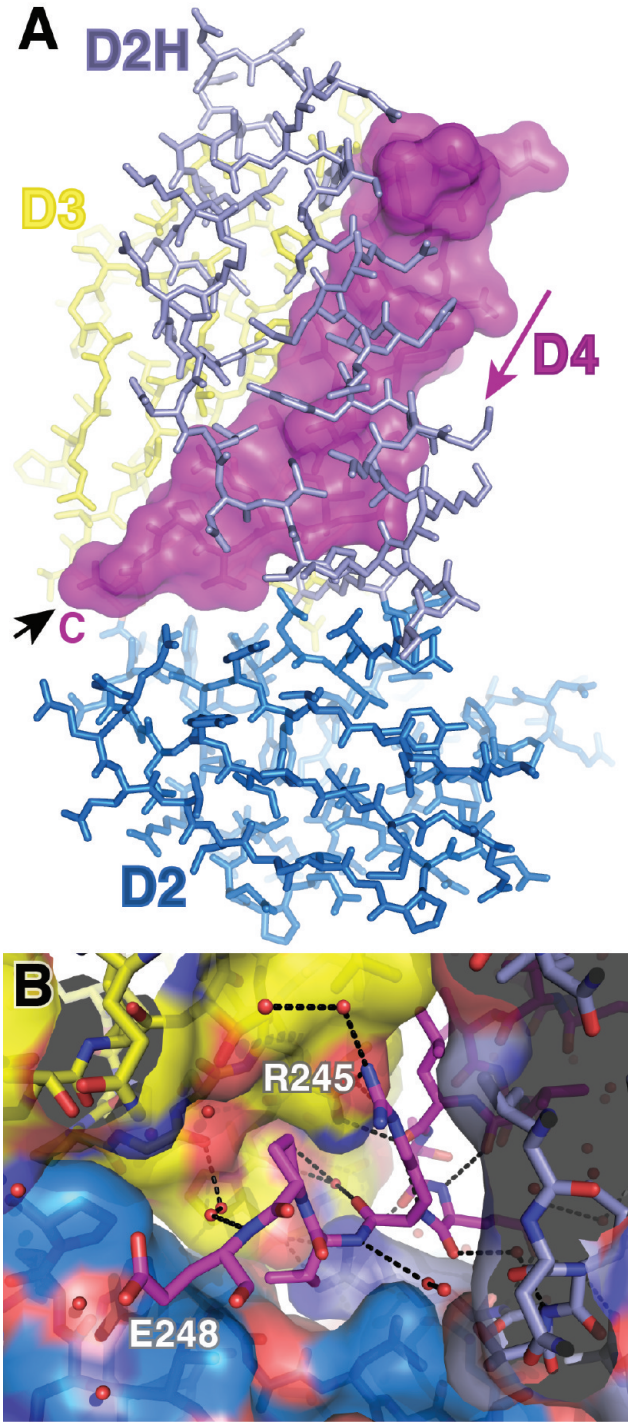


Figure 4. GfcC D4 helix packed between the D3 domain and D2H helices. (A) GfcC is rotated approximately 180° in the vertical direction from the view in Figure 1A. Domain colors are the same as those in Figure 1A. The long arrow designates the N to C polypeptide direction of D4, shown with a solvent contact surface. The short arrow points to C-terminal residue Glu248. (B) Close-up of the C-terminal residues of GfcC (245−248, magenta) passes through a tunnel formed by the other domains. The view is in the direction of the short arrow in panel A. The polar interactions with the other domains are mostly water-mediated hydrogen bonds. Note that the highly conserved Arg245 side chain does not form hydrogen bonds with the protein and is solvent accessible, suggesting that its role may be more than structural.

Table 2. Comparison of β-Grasp Domains[a]

	GfcC D2	GfcC D3	Wza D2	PDB entry 1vjk
GfcC D2		2.2 (49) 0.245 6%		
Wza D2	2.11 (53) 0.211 13%	1.65 (71) 0.152 18%		
Wza D3	1.83 (42) 0.219 10%	1.46 (61) 0.140 12%	1.19 (71) 0.110 20%	
PDB entry 2jx5[b]	1.83 (34) 0.224 not determined	1.71 (42) 0.210 not determined		1.62 (60) 0.167 not determined
PDB entry 1vjk[c]	1.72 (40) 0.203 not determined	1.37 (47) 0.160 not determined		

[a] For each comparison, the first row is the root-mean-square deviation (rmsd) for n α-carbon atoms (n in parentheses), the second row is a relative rmsd as calculated by LSQMAN, and the third row is the percent identity of the matched residues. [b] Ubiquitin domain adjacent to the S27a ribosomal subunit of *Giardia lamblia*. [c] Putative molybdopterin converting factor, subunit 1 from *Pyrococcus furiosus*.

(residues 22−147), D3 (residues 148−226), and D4 (residues 227−248) (see Figure 3A,B). D2 and D3 are β-grasp domains that typically have a central four- or five-strand mixed β-sheet with an α-helix on one face.[30] D3 has five β-strands, but the sheet in D2 has only four (S1−S3 and S5). Residues 116−121 (s4) form an extended strand, topologically equivalent to S4 in a typical β-grasp domain.[30] It is angled ≈45° from the rest of the D2 β-sheet and forms only one main chain hydrogen bond with adjacent strand S3.

In D2, the loop between strands S3 and S4 of the canonical β-grasp domain is replaced with a helical hairpin (H2 and H3, residues 68−115, bluish-gray in Figure 3A,B), here termed domain D2H. The hairpin is more than 40 Å long and consists of a seven-turn helix followed by a reverse turn, a shorter 4.5-turn helix, and a five-residue extended chain. The hairpin forms an ≈1100 Å^2 contact with the D4 helix and the edge of the D3 sheet near the S1−S2 loop. Most of the contacts are nonpolar, but D2H also forms eight hydrogen bonds with D3 and D4 (Table S1 of the Supporting Information).

Carboxyl-terminal α-helix D4 (magenta in Figure 3A,B) is amphipathic and folds onto the β-sheet of D3 (centered over S5) adjacent to the α-helical hairpin (D2H) on its other side (Figure 4A). The carboxyl terminus of the D4 helix passes through a large hole between the other domains to the opposite side of the protein and appears to be trapped between D3 and D2H (Figure 4B). This suggests that the D4 helix might fold onto D3 before D2H can pack next to it.

The structural similarity of D2 and D3 (rmsd = 2.2 Å for 49 residues) parallels that of either domain with β-grasp domains from other proteins (Table 2). The most noteworthy difference is the fact that the loop following S4 in D2 (residues 121−132) is wider and four residues longer than that in D3 (residues 202−209) and much longer than that in another β-grasp domain from a putative molybdopterin converting factor (PDB entry 1vjk). This loop in D2 contains several conserved residues that may be functionally

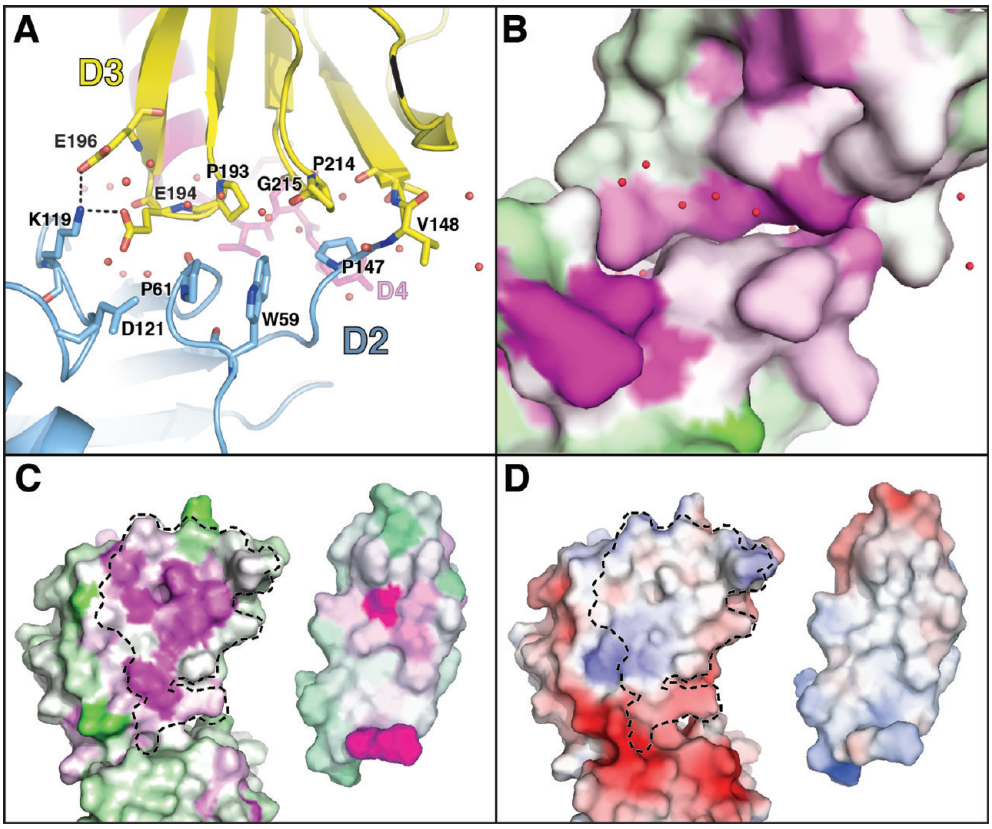


Figure 5. Interdomain interactions. (A) The interface between D2 and D3 is small and comprised of mostly van der Waals contacts and only two hydrogen bonds. Residues in the D2−D3 interface that are within 4 Å of the other domain are depicted as sticks. Also shown in sticks are the C-terminal residues of D4. Domain colors are light blue for D2, yellow for D3, and magenta for D4. (B) Conservation of the D2−D3 interface (same view as panel A). The surface representation is progressively colored by the degree of residue conservation from green (not conserved) to white (somewhat conserved) to magenta (highly conserved). (C) D2H interacts with a conserved surface on D3 and D4. Surface colored according to the degree of conservation as in panel B. On the left is the solvent contact surface of GfcC without D2H. The region buried in the presence of D2H, including residues 153−159, 180−182, and 192−195 from D3 and residues 219−245 from D4, is outlined. On the right is the D2H helical hairpin domain rotated 180° to show the surface that interacts with D3 and D4. (D) The interaction of D2H with GfcC is primarily nonpolar. Same views as in panel C. The electrostatic field at the solvent accessible surface is colored from blue (−5 kT/e) to white to red (5 kT/e). Electrostatics calculated with APBS[52] as implemented in MacPyMOL.[38]

significant, either as part of pocket P1 or in an observed dimer interface (see below).

Domain Interactions. The interface between the two β-grasp domains (excluding D2H) is small, comprising a surface of 420 Å^2 with only a couple of hydrogen bonds (Figure 5A and Table S1 of the Supporting Information). The D2 residues in the interface are residues 59, 61, and 62 from a 3_{10}-helix before S3, and Lys119 of S4, none of which are conserved in the alignment of homologues. The interacting surface of D3 is formed by two loops (S3 and S4) and the loop preceding S5, all relatively conserved in GfcC homologues. Figure 5B shows the interface colored by degree of conservation. Arg115 (at the end of D2H) also stabilizes the position of the two domains by interacting with both D2 and D3 (see below). Despite relatively few interactions between D2 and D3, residues in the D2−D3 linker (residues 146−148) and the interface have relatively low *B* factors, suggesting little movement between the two domains.

Helical hairpin D2H packs against a surface formed by portions of D3 and D4 (Figure 5C,D). The mostly nonpolar interface is approximately 1100 Å^2 but also includes 11 hydrogen bonds (Table S1 of the Supporting Information). While the contacting residues from D2H are not very conserved among GfcC homologues, the residues from D3 and D4 in the interface are highly conserved (Figure 5C). The interactions of hairpin D2H with D3 are likely responsible for fixing the position of D3 with respect to D2, and not the D2−D3 interface described above.

More than 1200 Å^2 (58%) of the solvent accessible surface of the carboxyl-terminal helix D4 is buried between D3 and D2H (Figure 4A), with more than 50% of this surface contributed by highly conserved residues. Most of the interactions are nonpolar, van der Waals contacts, but there are also five hydrogen bonds and two salt bridges (not shown).

The relative orientation of the GfcC domains within one molecule varies slightly in all four molecules in the asymmetric unit, with the largest difference existing between chains A and C. When the D2 domains of chains A and C are superimposed, D2H of chain C is rotated almost 6° compared to D2H of domain A, resulting in a maximal difference of 3.1 Å at the distal Asp93 (Figure S1 of the Supporting Information). The axis of the D2H rotation passes between D2 and D3. This hints at inherent flexibility in the molecule that may play a role in its function.

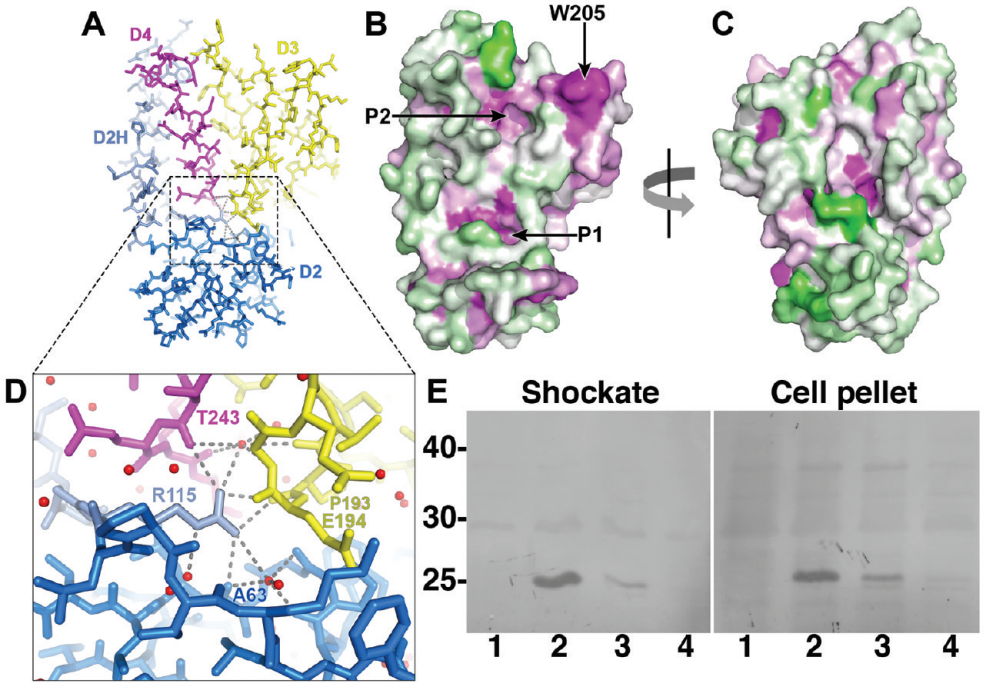


Figure 6. Conserved residues of GfcC and invariant residue Arg115. (A) Overall view of GfcC with domains colored as in previous figures. The orientation is the same as in Figure 1B. (B) Solvent contact surface colored by degree of conservation as in Figure 5B. The orientation is the same as in panel A. P1 and P2 point to pockets discussed in the text. W205 is totally exposed and highly conserved. (C) Same as panel B but with the molecule rotated 180° around the vertical axis. (D) The invariant Arg115, from the end of D2H, forms hydrogen bonds to waters and to main chain atoms of each of the other domains. The Arg115 side chain also forms part of the surface of pocket P1. (E) Arg115 mutants are poorly expressed or degraded in the periplasm as shown by Western blots of the supernatant (left) and cell pellet (right) after osmotic shock. The primary antibody is the mouse anti-His antibody: lane 1, EM4462 (EPEC *gfcC::kan*); lane 2, EM4462/pAP1720 expressing full-length GfcC-His_6; lane 3, EM4462/pAP1720-R115K; lane 4, EM4462/pAP1720-R115A. GfcC migrates at ~25 kDa. The presence of significant protein in the cell pellet likely represents intact cells with periplasmic GfcC. A similar experiment with sonicated whole cells showed no GfcC in the insoluble fraction (not shown). Methods for this experiment are given in the Supporting Information.

GfcC Domain Structures Are Similar to That of Wza. The domain structure of GfcC bears a striking resemblance to that of Wza, the octameric protein that forms the conduit for the export of group 1 polysaccharide through the outer membrane (Figure 3B).[12] Like GfcC, both the D2 and D3 domains of Wza have β-grasp folds. The respective domains were superimposed with LSQMAN, and the results are listed in Table 2 and shown in Figure 3C. From the structural alignment, the sequence identity is less than 12% between the two proteins. While the β-grasp domains of the two proteins have similar topologies, there are differences in the twist of the β-sheets and the number of α-helices crossing the β-sheet. D3 of Wza has a sixth β-strand contributed by the lipid-anchored N-terminal region (not shown in Figure 3C).

Search of known structures with the Dali server[39] showed that D3 of GfcC is significantly more similar to the Wza β-grasp domain (Z = 9.1−9.5) than it is to any other non-Wza β-grasp domain [Z = 6 for the next best fit (Figure 3C)]. On the basis of structural alignments summarized in Table 2, GfcC D3 was 18 and 12% identical in sequence to Wza D2 and D3, respectively.

The differences between the D4 domains in GfcC and Wza are striking and likely reflect the different functions of the two proteins (Figure 3B). In Wza, the carboxyl-terminal amphipathic helix (nine turns, 38 residues) extends away from D3 and has a slight kink. It oligomerizes with the D4 helices from seven other Wza protomers to form an ≈17 Å inner diameter pore that crosses the bacterial outer membrane (Figure 1A).[11,12] In the octamer, most of the nonpolar side chains from Wza D4 are exposed to lipid acyl chains. The Wza D4 domain does not interact with any of the other domains of Wza. Helix H5 of GfcC D4 is also an amphipathic helix, but only 13 residues long. The nonpolar face interacts with both D3 and the helical hairpin domain, D2H. While Wza has aromatic residues characteristic of transmembrane proteins at either end of the helix, the GfcC H5 helix begins with only a Tyr residue. Both D4 domains begin and end with Arg or Lys residues, which likely interact with phospholipid headgroups on either side of the membrane in the case of Wza.

Conserved Residues and Pockets. When we scored the conserved residues with SCORECONS[34] and colored a molecular contact surface (Figure 6B,C), we found that there were relatively few patches of highly conserved residues exposed on the surface. One prominent region was the loop following S4 in D3, where an exposed and invariant Trp205 (Figure 6B) clusters with conserved residues Asn206, Ala203, and Pro201. This may be a region that interacts with another protein or membrane. On the bottom surface of D2 (as viewed in Figure 6A but not shown) is another patch of exposed conserved residues, including Asn132, Arg128, Asp123, and Asp121. These residues form interactions across the dimer interface (see below).

All GfcC homologues contain the invariant Arg115 at the end of the D2H helical hairpin (Figure 6D), which forms hydrogen bond interactions with main chain atoms from D2 (Ala63), D3 (Pro193), and D4 (Thr243), but no salt bridges. To address the structural role of this conserved Arg, we expressed full-length

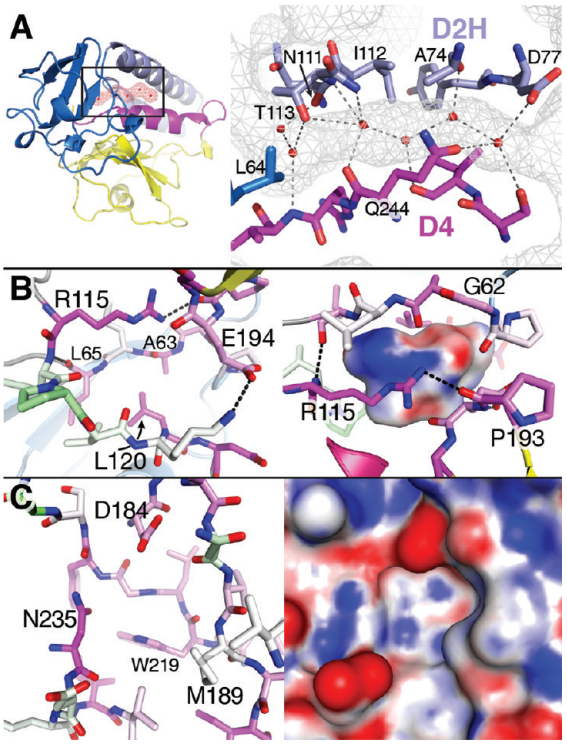


Figure 7. Cavities and pockets in GfcC. (A) A water-filled tunnel lies between D2H and D4 in molecules A and C of the asymmetric unit. The box in the left image shows the location of the image on the right. The mesh represents the solvent contact surface. The stick model is colored by domain. Depicted is the tunnel from molecule A in the asymmetric unit. In molecules B and D, the tunnel is not present because residues 111 and 112 (from D2H) move closer to D4 and displace the waters (not shown). (B) Pocket P1 is surrounded by mostly conserved residues, including Arg115, and main chain atoms. The left panel shows the pocket viewed from the outside. Carbon atoms are colored on the basis of the degree of conservation (from green to white to magenta). The right panel shows a view rotated by 90° from that on the left. Sticks are colored similarly. The surface is the solvent contact surface within the pocket and is colored by electrostatic potential (red for negative, white for neutral, and blue for positive) as calculated by APBS. (C) The left panel shows pocket P2 colored by degree of conservation. The right panel shows the solvent accessible surface colored by electrostatic potential.

wild-type GfcC or two mutants, Arg115→Lys and Arg115→Ala, from pAP1720 in EPEC *gfcC::kan* without IPTG induction (see the Supporting Information). Periplasmic proteins were released by osmotic shock. Supernatant and remaining whole cells were separated by sodium dodecyl sulfate gel electrophoresis and immunoblotted with anti-His antibody (Figure 6E). The amount of the Arg115 → Lys mutant was <30% of that of the wild type, while the Arg115 → Ala mutant was barely detected. Whole cell lysates showed bands smaller than GfcC in the mutants, but not in the wild type. These results suggest that replacing Arg115 with Lys or Ala and the concomitant loss of the polar interactions with the other domains produce a destabilized form of GfcC that is susceptible to degradation. Its importance for structural integrity does not preclude other functions for Arg115.

We were struck by the numerous cavities and pockets in GfcC identified by the CASTp server.[40] The largest solvent accessible cavity was a tunnel of ≈400 Å^3 that passes between the helical hairpin domain and the D4 carboxyl-terminal helix in molecules A and C of the asymmetric unit (Figure 7A). The tunnel is exposed at each end to bulk solvent and contains seven water molecules that form hydrogen bonds to each other. Most interactions of water are to protein main chain atoms and three polar side chains (Figure 7A), highlighting the fact that a portion of the D2H hairpin is not closely packed with the other domains. Surprisingly, the tunnel is not present in chain B or D, where at least four of the bound waters are absent (not shown). In chains B and D, residues 111 and 112 move closer to the D4 helix compared to chain A. In chain B, the end of the D4 helix (residues 244 and 245) also adapts a different conformation. This reflects the plasticity of the structure in the context of crystal packing constraints. It also raises the possibility that GfcC can assume other conformations besides those observed in the crystal.

The largest pocket is located in D2 near the interface with D3 (Figure 7B, P1 in Figure 6B). Glu194 from the D3 β-grasp domain, Arg115 from the end of the D2H insertion, and residues 116−121 from the s4 strand surround the opening to the pocket. Residues 60−65 from a 3_{10}-helix and S3 form the back of the pocket. All residues with side chains exposed to the pocket are conserved except for Val118 and Lys119. The interior surface of the pocket is >60% polar, with 10 main chain amides or carbonyls available for hydrogen bonding to a potential ligand. The pocket in chain D is occupied by six water molecules and is exposed to bulk solvent in two places. The side chains of Arg115, Lys119, and Asp121 are also accessible in the pocket for interacting with a potential ligand. The pocket appears to be approximately the size of a sugar molecule, but such a molecule clashed with the protein when docked with Vina (results not shown).[41]

Another significant pocket is located between D4 and the β-sheet of D3 (Figure 7C, P2 in Figure 6B). It is surrounded by Asp236, Asn235, Phe222 (main chain), Asp184, Asn187, Val188 (main chain), and Met189. Trp219, including the ring imide, the main chain of residue 220, Gly221, and Val239 form the bottom surface of the pocket. Of these residues, Asp184 and Asn235 are mostly conserved. Six ordered water molecules are resolved in the pocket. In ubiquitin, which also has a β-grasp fold, the region homologous to this pocket is hydrophobic and binds interacting molecules.[42]

GfcC Apparent Dimer Interactions and a Resolved His_6 Tag. The four SeMet-GfcC molecules in the monoclinic asymmetric unit are arranged as two apparent dimers (A−B and C−D) stacked on top of each other. The protomers in each dimer are related by a noncrystallographic 2-fold axis (Figure 8A), while the dimers are related by an approximate 4_1-screw axis parallel to the *c* direction. The buried interface area within the dimer is 1215 Å^2, as calculated with PISA.[43] The mostly polar interface includes ≈17 hydrogen bonds and 14 salt bridges, with several water-mediated hydrogen bonds (Figure 8B). Each "half" interface consists primarily of residues from D2 interacting with D3 of the other protomer. Residues 122−129 of D2 adopt an extended β-conformation to pair with a similarly structured S4−S5 loop of D3 in the other protomer. Although there are several conserved residues in the interface [Asp123, Arg126, Asn132, Glu211, and Pro214 (Figure 8C)], analysis of the dimer interface with PISA suggests that the interface may not have enough buried nonpolar surface to stabilize the dimer in solution. To further characterize the oligomeric structure of GfcC, we measured its hydrodynamic properties by sedimentation velocity at several concentrations (Figure S2 and Table S2 of the Supporting Information) and showed that the protein is a monomer with an M_r of 25277 and a frictional coefficient (f/f_0) of 1.33.

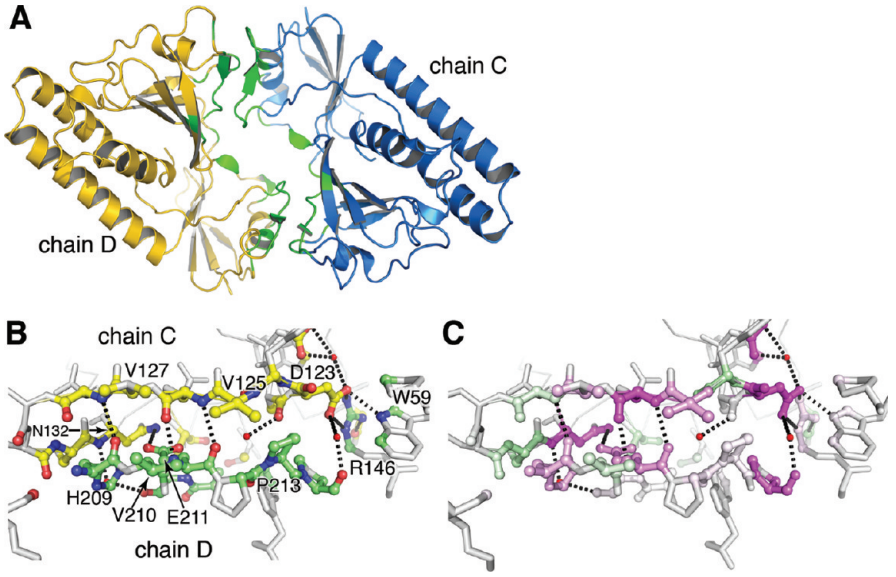


Figure 8. GfcC forms apparent dimers in the crystal. (A) Cartoon representation of two of the four chains of the asymmetric unit related by a noncrystallographic 2-fold axis. The interface regions are colored green and are formed by D2 and D3 of each chain. Also see Figure 9B. (B) Close-up of the hydrogen bonds at the interface showing antiparallel β-type hydrogen bonding between the two chains. Atoms represented as distinct spheres take part in the interaction. Yellow carbon atoms are from D2 of chain C and green atoms from D3 of chain D. (C) Same as panel B but with interface atoms colored by degree of conservation with magenta representing the most conserved residues.

Following refinement of the structure, we noticed difference electron density near the noncrystallographic 2-fold axis at the apparent dimer interface of molecules (chains) C and D (Figure 9). After noting that the density was consistent with a His residue and that the current C-terminal residue 249 of molecule A was ~15 Å distant, we suspected that the density might be a residue of the C-terminal His_6 tag of molecule A. We placed water molecules with zero occupancy in nearby solvent regions. After refinement, the entire His tag was resolved when the electron density was contoured at 0.7 standard deviation above the mean density. Figure 9B shows that the C-terminus (spheres) of A is positioned over the C−D interface. The final four residues take on a β-turn-like conformation to interact with molecules C and D (Figure 9C). His254, the residue found in the original difference density, forms hydrogen bonds with Asp121 (molecule D) and water 147, which in turn forms hydrogen bonds with Asp121 and Glu194 of molecule C. Besides hydrogen bonding interactions, Phe124 of molecule C adopts a conformation different from molecule D to make a $\pi-\pi$ stacking interaction with His243. Of the other molecules, only C-terminal residue 248 of molecule C is near the interface of molecules A and B, but the tag has no electron density. The C-termini of molecules B and D are not near another chain, but His251 of chain D is resolved in the electron density map.

DISCUSSION

The high-resolution structure of GfcC represents the first structure determined of the DUF1017 fold family. Although not recognized by Pfam, GfcC actually has two β-grasp domains rather than the one predicted by the DUF1017 classification. This is likely due to the presence of inserted helical hairpin D2H within the unrecognized β-grasp domain (D2), a feature that has not been described for other β-grasp domains, to the best of our knowledge. The GfcC D2H domain binds to a conserved surface formed by the D3 and D4 domains, implying that the C-terminal portion of the molecule may fold first. In the absence of D2H, the small contact interface between D2 and D3 may be insufficient to maintain the relative orientations of the two domains. We speculate that this would allow the connecting strand between the D2 and D3 domains to be more flexible and allow D2 and D3 to become separated as they appear in Wza. Such a state would also allow D2H to interact with other molecules.

Arg115 is invariant in all GfcC homologues and highly conserved in other DUF1017 proteins. This residue would be a key contact residue should some molecule bind in the P1 pocket (Figure 6B). Alternatively, its interactions with D2−D4 may stabilize the overall tertiary fold, a hypothesis that was supported by our observation that less GfcC was released upon osmotic shock when the R115K or R115A mutant (Figure 6E) was expressed in place of the wild-type full-length protein.

The physiological relevance of the apparent dimer (Figure 8A) observed in the crystal structure of GfcC cannot be resolved at this time. Apparent oligomers in crystal structures often occur because of crystal growth in the presence of high concentrations of protein. In GfcC, the intermolecular interactions are mostly polar, occurring between loops of each β-grasp domain, and do not involve the exposed surface of the domain's β-sheet as observed in the Wza oligomer. PISA server analysis calculates a ΔG^S (solvation energy gain upon formation of the complex) of −2.3 kcal/M, comparable to the energy gain if the interface were comprised of random surface atoms with the same surface area.[43] The analytical ultracentrifugation (Figure S2 of the Supporting Information) and gel filtration results are consistent with a monomer in solution, but because there are a number of conserved residues at the observed interface (Figure 8C) and β-grasps often function as protein interaction domains,[44] we cannot exclude the possibility that GfcC may form complexes with itself or another protein in the periplasm.

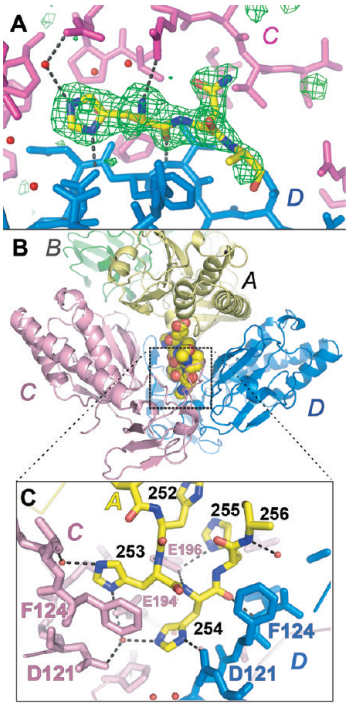


Figure 9. His_6 tag of GfcC resolved in one of the molecules (chains) in the asymmetric unit. In all panels, carbon atoms and ribbons are colored by chain: yellow for chain A, magenta for chain B, pink for chain C, and blue for chain D. (A) Simulated annealing, difference electron density map contoured at 3σ (green) before including the His_6 tag in refinement. The initial model of a tripeptide (yellow carbon atoms) is modeled near the 2-fold axis of the dimer formed by chains C and D. The modeled His is equivalent to His254 in panel C. (B) Overall view of the four monomers in the asymmetric unit. The His_6 tag (residues 251−256) of chain A (spheres) extends downward to bind near the 2-fold axis of the C−D dimer. (C) Three of the chain A His residues (253−255) form most of the interactions, including stacking of the C chain Phe124 ring with His253. Note that the dimer 2-fold noncrystallographic symmetry is broken near the binding site (compare Phe124 positions between the two chains).

Ongoing research in our laboratories with EPEC deletion mutants suggests that GfcC (and GfcD) is affecting some aspect of capsule expression on the cell surface, but a more precise function is not known. Although GfcC has domains that resemble Wza, it is not likely to be the export channel for the group 4 capsule because GfcE, the Wza paralog, probably fulfills that role. Also, GfcC lacks the polysaccharide export sequence domain common to outer membrane auxiliary proteins.[31] Although the GfcC C-terminal helix is amphipathic, it is significantly shorter than the one in Wza and may not be able to span the outer membrane.

These explanations do not preclude the possibility that other proteins containing a DUF1017 fold might function like Wza. Interestingly, several of the outer membrane auxiliary proteins have SLBB domains but are terminated by a DUF1017 domain.[31] An example is the 911-residue WbfF (formerly OtnA) protein that is essential for capsule secretion in *Vibrio cholera* O139[45] and in *Vibrio parahemolyticus* O3:K6.[22] Secondary structure prediction and sequence alignments suggest that the DUF1017 domain in WbfF has structural features similar to those of GfcC. It contains a D2H-like helical region (residues 715−771), an Arg774 that is equivalent to the invariant Arg115 of GfcC, a D3 domain, and a 22-residue C-terminal amphipathic helix (M. A. Saper, unpublished observations). Because the C-terminal amphipathic helix of WbfF is predicted to be similar in length to Wza, it might oligomerize if exposed and span the outer membrane. One could speculate that the DUF1017 domain acts as a temporary chaperone, allowing the C-terminal helix to pack between D3 and D2H as in GfcC, effectively hiding its nonpolar surfaces. Thus, it would be protected from premature oligomerization during transport of WbfF to the outer membrane. Of course, upon interactionn with the outer membrane, a major conformational change would be required to release the helix so it could oligomerize and form a pore. The D2 and D3 domains, like the Wza β-grasp domains, could then interact with neighboring subunits to maintain the channel architecture necessary for capsule translocation.

Another DUF1017 family member also provides clues about a possible function of GfcC. OtnG is a 995-residue protein encoded by *Burkholderia* sp. and several other gammaproteobacteria. Sequence annotation of OtnG, supported by sequence alignment with GfcC, shows that it is a fusion of GfcC (≈15% identical) and GfcD (≈40% identical) homologues. In fact, in *E. coli*, the *gfcC* and *gfcD* reading frames actually overlap by one base. As with *gfcC* in *E. coli*, the OtnG protein is encoded downstream of a gene homologous to *gfcB* and in a chromosomal region attributed to capsule and exopolysaccharide synthesis and export.[46] Often two adjacent genes on the same operon encode proteins that physically interact.[47] The fusion of adjacent genes into one polypeptide further supports the probability that the individual proteins directly interact.[48] Therefore, we propose that GfcC interacts with GfcD, a predicted β-barrel outer membrane lipoprotein, perhaps by binding its amino-terminal region or one of the GfcD periplasmic loops.

An analogous situation also occurs in the export system of the exopolysaccharide alginate from *Pseudomonas aeruginosa*. AlgK is a helical outer membrane-associated lipoprotein with a likely polypeptide binding site,[25] while AlgE, encoded immediately downstream of *algK*, is a β-barrel outer membrane lipoprotein and alginate export pore.[25,26,49] AlgK is proposed to interact with AlgE. Like OtnG described above, the exopolysaccharide systems for poly-β-1,6-*N*-acetyl-D-glucosamine and cellulose express proteins in which orthologs of AlgK and AlgE proteins are fused as one protein (BcsC or PgaA).[25]

We envision that GfcC will function at one of these steps in capsule expression: polymerization, polymer modification or termination, export from the periplasm, or attachment to the bacterial surface. Ongoing experiments aimed at identifying this step and identifying binding partners for GfcC and GfcD will contribute to understanding GfcC's structure and function.

■ ASSOCIATED CONTENT

Supporting Information. Text file in Clustal W format of an alignment of GfcC homologues, and a PDF file containing additional methods, figures and tables. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Program in Biophysics, University of Michigan, 930 N. University Ave., Ann Arbor, MI 48104-1055. Phone: (734) 764-3353. E-mail: saper@umich.edu.

Present Addresses
[⊥]Department of Structural Biology, Stanford University, Palo Alto, CA 94305-5126.
[@]Department of Immunology, University of Washington, Seattle, WA 98195.

Funding Sources
Funding provided by the Lady Davis Foundation (M.A.S.), the United States-Israel Binational Science Foundation (I.R. and M.A.S.), and the University of Michigan Rackham Graduate School (K.S.).

■ ACKNOWLEDGMENT

We thank Stefan Walter for assistance with the sedimentation velocity analyses and the Life Sciences Institute, University of Michigan, for use of the analytical ultracentrifuge. We appreciate the efforts of many who have developed and shared the crystallographic and protein analysis software packages cited in the text. We also thank Dervla Mellerick (Science Word Doctor, LLC) for editorial assistance. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corp. and the Michigan Technology Tri-Corridor for the support of this research program (Grant 085P1000817).

■ ABBREVIATIONS

CPS, capsule polysaccharide; EPS, exopolysaccharide; EPEC, enteropathogenic *E. coli*; EHEC, enterohemorrhagic *E. coli*; NCS, noncrystallographic symmetry; rmsd, root-mean-square deviation; SeMet, selenomethionine.

■ REFERENCES

(1) Whitfield, C. (2006) Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem. 75*, 39–68.

(2) Attridge, S. R., and Holmgren, J. (2009) *Vibrio cholerae* O139 capsular polysaccharide confers complement resistance in the absence or presence of antibody yet presents a productive target for cell lysis: Implications for detection of bactericidal antibodies. *Microb. Pathog. 47*, 314–320.

(3) Schneider, M. C., Exley, R. M., Ram, S., Sim, R. B., and Tang, C. M. (2007) Interactions between *Neisseria meningitidis* and the complement system. *Trends Microbiol. 15*, 233–240.

(4) Campos, M. A., Vargas, M. A., Regueiro, V., Llompart, C. M., Albertí, S., and Bengoechea, J. A. (2004) Capsule polysaccharide mediates bacterial resistance to antimicrobial peptides. *Infect. Immun. 72*, 7107–7114.

(5) Nakhamchik, A., Wilde, C., and Rowe-Magnus, D. A. (2007) Identification of a Wzy polymerase required for group IV capsular polysaccharide and lipopolysaccharide biosynthesis in *Vibrio vulnificus*. *Infect. Immun. 75*, 5550–5558.

(6) Guerry, P., and Szymanski, C. M. (2008) Campylobacter sugars sticking out. *Trends Microbiol. 16*, 428–435.

(7) Grangeasse, C., Cozzone, A. J., Deutscher, J., and Mijakovic, I. (2007) Tyrosine phosphorylation: An emerging regulatory device of bacterial physiology. *Trends Biochem. Sci. 32*, 86–94.

(8) Bechet, E., Gruszczyk, J., Terreux, R., Gueguen-Chaignon, V., Vigouroux, A., Obadia, B., Cozzone, A. J., Nessler, S., and Grangeasse, C. (2010) Identification of structural and molecular determinants of the tyrosine-kinase Wzc and implications in capsular polysaccharide export. *Mol. Microbiol. 77*, 1315–1325.

(9) Obadia, B., Lacour, S., Doublet, P., Baubichon-Cortay, H., Cozzone, A. J., and Grangeasse, C. (2006) Influence of tyrosine-kinase Wzc activity on colanic acid production in *Escherichia coli* K12 cells. *J. Mol. Biol. 367*, 42–53.

(10) Drummelsmith, J., and Whitfield, C. (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.

(11) Collins, R. F., Beis, K., Dong, C., Botting, C. H., McDonnell, C., Ford, R. C., Clarke, B. R., Whitfield, C., and Naismith, J. H. (2007) The 3D structure of a periplasm-spanning platform required for assembly of group 1 capsular polysaccharides in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A. 104*, 2390–2395.

(12) Dong, C., Beis, K., Nesper, J., Brunkan-Lamontagne, A. L., Clarke, B. R., Whitfield, C., and Naismith, J. H. (2006) Wza the translocon for *E. coli* capsular polysaccharides defines a new class of membrane protein. *Nature 444*, 226–229.

(13) Goldman, R. C., White, D., Orskov, F., Orskov, I., Rick, P. D., Lewis, M. S., Bhattacharjee, A. K., and Leive, L. (1982) A surface polysaccharide of *Escherichia coli* O111 contains O-antigen and inhibits agglutination of cells by O-antiserum. *J. Bacteriol. 151*, 1210–1221.

(14) Peleg, A., Shifrin, Y., Ilan, O., Nadler-Yona, C., Nov, S., Koby, S., Baruch, K., Altuvia, S., Elgrably-Weiss, M., Abe, C. M., Knutton, S., Saper, M. A., and Rosenshine, I. (2005) Identification of an *Escherichia coli* operon required for formation of the O-antigen capsule. *J. Bacteriol. 187*, 5259–5266.

(15) Shifrin, Y., Peleg, A., Ilan, O., Nadler, C., Kobi, S., Baruch, K., Yerushalmi, G., Berdichevsky, T., Altuvia, S., Elgrably-Weiss, M., Abe, C., Knutton, S., Sasakawa, C., Ritchie, J. M., Waldor, M. K., and Rosenshine, I. (2008) Transient shielding of intimin and the type III secretion system of enterohemorrhagic and enteropathogenic *Escherichia coli* by a group 4 capsule. *J. Bacteriol. 190*, 5063–5074.

(16) Gibson, D. L., White, A. P., Snyder, S. D., Martin, S., Heiss, C., Azadi, P., Surette, M., and Kay, W. W. (2006)*Salmonella* produces an O-antigen capsule regulated by AgfD and important for environmental persistence. *J. Bacteriol. 188*, 7722–7730.

(17) Vincent, C., Duclos, B., Grangeasse, C., Vaganay, E., Riberty, M., Cozzone, A. J., and Doublet, P. (2000) Relationship between exopolysaccharide production and protein-tyrosine phosphorylation in gram-negative bacteria. *J. Mol. Biol. 304*, 311–321.

(18) Wugeditsch, T., Paiment, A., Hocking, J., Drummelsmith, J., Forrester, C., and Whitfield, C. (2001) Phosphorylation of Wzc, a tyrosine autokinase, is essential for assembly of group 1 capsular polysaccharides in *Escherichia coli*. *J. Biol. Chem. 276*, 2361–2371.

(19) Ferrières, L., Aslam, S. N., Cooper, R. M., and Clarke, D. J. (2007) The *yjbEFGH* locus in *Escherichia coli* K-12 is an operon encoding proteins involved in exopolysaccharide production. *Microbiology 153*, 1070–1080.

(20) Ionescu, M., and Belkin, S. (2009) Overproduction of exopolysaccharides by an *Escherichia coli* K-12 *rpoS* mutant in response to osmotic stress. *Appl. Environ. Microbiol. 75*, 483–492.

(21) Croxatto, A., Lauritz, J., Chen, C., and Milton, D. L. (2007) *Vibrio anguillarum* colonization of rainbow trout integument requires a DNA locus involved in exopolysaccharide transport and biosynthesis. *Environ. Microbiol. 9*, 370–382.

(22) Chen, Y., Dai, J., Morris, J. G., and Johnson, J. A. (2010) Genetic analysis of the capsule polysaccharide (K antigen) and exopolysaccharide genes in pandemic *Vibrio parahaemolyticus* O3:K6. *BMC Microbiol. 10*, 274.

(23) Zhai, Y., and Saier, M. H. (2002) The β-barrel finder (BBF) program, allowing identification of outer membrane β-barrel proteins encoded within prokaryotic genomes. *Protein Sci. 11*, 2196–2207.

(24) Remmert, M., Linke, D., Lupas, A. N., and Söding, J. (2009) HHomp: Prediction and classification of outer membrane proteins. *Nucleic Acids Res. 37*, W446–W451.

(25) Keiski, C.-L., Harwich, M., Jain, S., Neculai, A. M., Yip, P., Robinson, H., Whitney, J. C., Riley, L., Burrows, L. L., Ohman, D. E., and Howell, P. L. (2010) AlgK is a TPR-containing protein and the periplasmic component of a novel exopolysaccharide secretin. *Structure 18*, 265–273.

(26) Hay, I. D., Rehman, Z. U., and Rehm, B. H. (2010) Membrane topology of outer membrane protein AlgE, which is required for alginate production in *Pseudomonas aeruginosa*. *Appl. Environ. Microbiol. 76*, 1806–1812.

(27) Itoh, Y., Rice, J. D., Goller, C., Pannuri, A., Taylor, J., Meisner, J., Beveridge, T. J., Preston, J. F., and Romeo, T. (2008) Roles of *pgaABCD* genes in synthesis, modification, and export of the *Escherichia coli* biofilm adhesin poly-β-1,6-*N*-acetyl-D-glucosamine. *J. Bacteriol. 190*, 3670–3680.

(28) Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R., and Bateman, A. (2010) The Pfam protein families database. *Nucleic Acids Res. 38*, D211–D222.

(29) Burroughs, A. M., Balaji, S., Iyer, L. M., and Aravind, L. (2007) Small but versatile: The extraordinary functional and structural diversity of the β-grasp fold. *Biol. Direct 2*, 18.

(30) Burroughs, A. M., Balaji, S., Iyer, L. M., and Aravind, L. (2007) A novel superfamily containing the β-grasp fold involved in binding diverse soluble ligands. *Biol. Direct 2*, 4.

(31) Cuthbertson, L., Mainprize, I. L., Naismith, J. H., and 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*, 155–177.

(32) Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res. 25*, 4876–4882.

(33) Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009)Jalview Version 2: A multiple sequence alignment editor and analysis workbench. *Bioinformatics 25*, 1189–1191.

(34) Valdar, W. S. (2002) Scoring residue conservation. *Proteins 48*, 227–241.

(35) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. In *Macromolecular Crystallography Part A* (Carter, C. W., Ed.) pp 307−326, Academic Press, San Diego.

(36) Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D66*, 213–221.

(37) Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126–2132.

(38) *The PyMOL Molecular Graphics System*, version 1.3 (2010) Schrödinger, LLC.

(39) Holm, L., and Rosenström, P. (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res. 38*, W545–W549.

(40) Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. (2006) CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res. 34*, W116–W118.

(41) Trott, O., and Olson, A. J. (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem. 31*, 455–461.

(42) Winget, J. M., and Mayor, T. (2010) The diversity of ubiquitin recognition: Hot spots and varied specificity. *Mol. Cell 38*, 627–635.

(43) Krissinel, E., and Henrick, K. (2007)Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol. 372*, 774–797.

(44) Lytle, B. L., Peterson, F. C., Qiu, S. H., Luo, M., Zhao, Q., Markley, J. L., and Volkman, B. F. (2004) Solution structure of a ubiquitin-like domain from tubulin-binding cofactor B. *J. Biol. Chem. 279*, 46787–46793.

(45) Nesper, J., Schild, S., Lauriano, C. M., Kraiss, A., Klose, K. E., and Reidl, J. (2002) Role of *Vibrio cholerae* O139 surface polysaccharides in intestinal colonization. *Infect. Immun. 70*, 5990–5996.

(46) Reckseidler-Zenteno, S. L., Moore, R., and Woods, D. E. (2009) Genetics and function of the capsules of *Burkholderia pseudomallei* and their potential as therapeutic targets. *Mini Rev. Med. Chem. 9*, 265–271.

(47) Dandekar, T., Snel, B., Huynen, M., and Bork, P. (1998) Conservation of gene order: A fingerprint of proteins that physically interact. *Trends Biochem. Sci. 23*, 324–328.

(48) Enright, A. J., Iliopoulos, I., Kyrpides, N. C., and Ouzounis, C. A. (1999) Protein interaction maps for complete genomes based on gene fusion events. *Nature 402*, 86–90.

(49) Rehm, B. H., Boheim, G., Tommassen, J., and Winkler, U. K. (1994)Overexpression of *algE* in *Escherichia coli*: Subcellular localization, purification, and ion channel properties. *J. Bacteriol. 176*, 5639–5647.

(50) Beis, K., Collins, R. F., Ford, R. C., Kamis, A. B., Whitfield, C., and Naismith, J. H. (2004) Three-dimensional structure of Wza, the protein required for translocation of group 1 capsular polysaccharide across the outer membrane of *Escherichia coli*. *J. Biol. Chem. 279*, 28227–28232.

(51) Kleywegt, G. J., and Jones, T. A. (1997) Detecting folding motifs and similarities in protein structures. *Methods Enzymol. 277*, 525–545.

(52) Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001)Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A. 98*, 10037–10041.