

Crystal Structure of the Major Periplasmic Domain of the Bacterial Membrane Protein Assembly Facilitator YidC*

Received for publication, October 30, 2007, and in revised form, December 12, 2007 Published, JBC Papers in Press, December 19, 2007, DOI 10.1074/jbc.M708936200

David C. Oliver and Mark Paetzel¹

From the Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

The essential bacterial membrane protein YidC facilitates insertion and assembly of proteins destined for integration into the inner membrane. It has homologues in both mitochondria and chloroplasts. Here we report the crystal structure of the *Escherichia coli* YidC major periplasmic domain (YidC_{EC}P1) at 2.5 Å resolution. This domain is present in YidC from Gram-negative bacteria and is more than half the size of the full-length protein. The structure reveals that YidC_{EC}P1 is made up of a large twisted β-sandwich protein fold with a C-terminal α-helix that packs against one face of the β-sandwich. Our structure and sequence analysis reveals that the C-terminal α-helix and the β-sheet that it lays against are the most conserved regions of the domain. The region corresponding to the C-terminal α-helix was previously shown to be important for the protein insertase function of YidC and is conserved in other YidC-like proteins. The structure reveals that a region of YidC that was previously shown to be involved in binding to SecF maps to one edge of the β-sandwich. Electrostatic analysis of the molecular surface for this region of YidC reveals a predominantly charged surface and suggests that the SecF-YidC interaction may be electrostatic in nature. Interestingly, YidC_{EC}P1 has significant structural similarity to galactose mutarotase from *Lactococcus lactis*, suggesting that this domain may have another function besides its role in membrane protein assembly.

The biogenesis of membrane proteins is a fundamental aspect of cell biology that involves proteinacious factors. In Gram-negative bacteria, most proteins destined for the cell envelope are targeted to the inner membrane in a post-translational manner via the cytosolic chaperone SecB or in a co-translational manner via the ribonucleoprotein SRP and its inner membrane-bound receptor FtsY (1). At the membrane, pro-

teins endowed with a Sec-dependent N-terminal signal peptide are exported across, or into, the inner membrane via the Sec system. The Sec system consists of the proteins SecY, SecE, and SecG, which form a heterotrimeric protein-conducting channel in the inner membrane (2); SecA, a cytosolic ATPase motor protein that unfolds (3) and pushes polypeptide substrates through the SecYEG channel (4); and the proteins SecD, SecF, and YajC, which form a heterotrimeric complex that interacts with SecYEG (5). The SecDFYajC complex has been proposed to (i) promote the release of substrate proteins from the SecYEG translocase following translocation (6) and/or (ii) enhance protein translocation by regulating SecA membrane cycling (7, 8). Proteins intended for integration into the inner membrane engage the essential protein YidC, which directly contacts transmembrane segments (9) and facilitates insertion (10), folding (11), and assembly (12, 13) of proteins into the inner membrane. Depending on the nature of the substrate, YidC can function in a Sec-dependent (SecYEG-YidC) (14, 15) or Sec-independent (“YidC only”) manner (16). It is thought that for Sec-dependent substrates, large hydrophilic domains are first exported across the membrane into the periplasm via the SecYEG channel, followed by movement of the transmembrane regions from the channel into the lipid bilayer; the latter step may be facilitated by YidC (9). How YidC promotes membrane protein insertion in a Sec-independent manner is unknown. YidC has been shown to co-purify with components of the Sec translocase (15), and a direct interaction between YidC and the SecDFYajC complex has been demonstrated, specifically with SecD and SecF (17, 18). YidC also plays a role in the biogenesis of lipoproteins (19), but its role in this process is not clear.

Structurally, *Escherichia coli* YidC is a 548-amino acid polypeptide with a molecular mass of 61,526 Da and a predicted isoelectric point of 7.7. Saaf *et al.* (20) have experimentally mapped the topology of YidC and shown that it consists of 6 transmembrane regions (TM) with a large ~35-kDa periplasmic domain (residues 24–342) located between transmembrane regions 1 and 2 (Fig. 1A). Deletion analyses have revealed that YidC insertase function is located mainly in the C-terminal five transmembrane regions (18, 21). Remarkably, up to 90% of the 35-kDa YidC periplasmic domain (residues 25–323) can be deleted without affecting inner membrane protein biogenesis or cell viability (18, 21). YidC is conserved in all three domains of life and is homologous to the well characterized proteins Oxa1 and ALB3 that are found in the inner membrane of mitochondria and the thylakoid membrane of chloroplasts, respectively (22). Consistent with functional mapping studies of inser-

* This work was supported in part by a Canadian Institute of Health Research operating grant, a National Science and Engineering Research Council of Canada discovery grant, a Michael Smith Foundation for Health Research Senior Scholar award, a Canadian Foundation of Innovation grant (to M. P.), and a Michael Smith Foundation for Health Research postdoctoral fellow award (to D. C. O.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 3BLC) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

¹ To whom correspondence should be addressed: Simon Fraser University, Dept. of Molecular Biology and Biochemistry, South Science Bldg., 8888 University Dr., Burnaby, British Columbia V5A 1S6, Canada. Tel.: 604-291-4320; Fax: 604-291-5583; E-mail: mpaetzel@sfu.ca.

tase activity, amino acid sequence alignments reveal that the C-terminal ~200 residues of YidC, corresponding to transmembrane regions 2–5, are conserved in prokaryotic and eukaryotic versions of the protein (22). Further, Oxa1 has been shown to complement YidC insertase activity when expressed in *E. coli* (23), underscoring the functional significance of this region in catalyzing protein insertion. By comparison, the function of the YidC periplasmic domain remains largely unknown; nevertheless, the conservation of this domain in Gram-negative bacteria suggests that it performs a significant, but as yet unrecognized, role in the cell.

To gain further insight into the structure and function of YidC, we sought to determine the structure of the YidC periplasmic domain. We report here the crystal structure of residues 57–346 of *E. coli* YidC to 2.5 Å resolution.

EXPERIMENTAL PROCEDURES

Cloning and Mutagenesis—A 942-base pair DNA fragment, coding for residues 26–340 of *E. coli* YidC, was amplified from *E. coli* K-12 genomic DNA using the forward primer 5'-ATGCAAGCATATGGATAAAAACCCGCAACCTCAGG and the reverse primer 5'-ATGCCTACTCGAGGCTGCCGCGCGG-CACCAGCAGCGGCTGAGAGATGAACC that contain the restriction sites NdeI and XhoI, respectively. The resulting PCR product was ligated into vector pET20b (Novagen). The YidC_{EC}(26–340)His construct includes an N-terminal methionine and a C-terminal thrombin/hexahistidine affinity tag bearing the sequence LVPRGSLEHHHHH. DNA sequencing (Macrogen) confirmed that the YidC insert matched the sequence reported in the Swiss-Prot data base (P25714). To facilitate crystallization, several residues within this construct were targeted for mutagenesis using the QuikChange method (Stratagene). The primer pair 5'-GTACTCCACGCCTGACGCGGCGTATGCGGCATACGCGTTCGATACCATTGCCG and 5'-CGGCAATGGTATCGAACGCGTATGCCGCATACGCCGCTCAGGCGTGGAGTAC was used to construct a version of YidC_{EC}(26–340)His that bears the mutations E228A, K229A, E231A, K232A, and K234A. This construct (referred to as pYidC_{EC}P1) yielded crystals suitable for structure determination. The expressed pYidC_{EC}P1 encodes 330 residues has a molecular mass of 35,702 Da and a theoretical pI of 5.3.

Protein Expression and Purification—The expression plasmid pYidC_{EC}P1 was transformed into *E. coli* expression strain BL21(DE3) and used to inoculate (1:100 back dilution) 3 liters of Luria Bertani medium containing ampicillin (100 µg/ml). Cultures were grown at 37 °C to an A_{600} of 0.6 and induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside for 3 h. Cells were harvested by centrifugation and lysed using an Avestin Emulsiflex-3C cell homogenizer. The lysate was clarified by centrifugation (30,000 × g) for 30 min at 4 °C. The supernatant was applied to a 5-ml nickel-nitrilotriacetic acid column (Qiagen) that had been equilibrated with 20 mM Tris-HCl, pH 8.0, 100 mM NaCl (buffer A). The column was washed with 30 ml of buffer A containing 20 mM imidazole and eluted with a step gradient (100–500 mM imidazole in buffer A at 100-mM increments) in 5-ml volumes. The majority of the protein eluted from the column in fractions containing 100, 200, and 300 mM imidazole, which were pooled and concentrated using an Ami-

TABLE 1

Data collection, phasing, and refinement statistics

The data collection statistics in brackets are the values for the highest resolution shell.

Crystal Parameters	
Space group	I4 ₁ 22
<i>a</i> , <i>b</i> , <i>c</i> (Å)	126.1, 126.1, 288.5
Data Collection Statistics	
Wavelength (Å)	0.979474
Resolution (Å)	50.0–2.5 (2.6–2.5)
Total Reflections	557,038
Unique reflections	40,327 (3687)
R_{merge}^a	0.079 (0.379)
Mean (<i>I</i>)/σ (<i>I</i>)	44.7 (4.0)
Completeness (%)	99.1 (92.0)
Redundancy	13.8 (9.7)
Phasing Statistics	
Number of Sites	8 (out of a possible 10)
Overall FOM ^b (50–2.5 Å)	0.49
Overall FOM ^b (after density modification)	0.72
Refinement Statistics	
Protein molecules (chains) in A.U.	2
Residues	558
Water molecules	80
Total number of atoms	4,371
R_{cryst}^c / R_{free}^d (%)	21.2 / 24.9
Average B-factor (Å ²) (all atoms)	33.9
Rms deviation on angles (°)	1.515
Rms deviation on bonds (Å)	0.015

$$^a R_{\text{merge}} = \frac{\sum_{hk} \sum_l |I_l(hkl) - \overline{I(hkl)}|}{\sum_{hk} \sum_l I_l(hkl)}, \text{ where } I_l(hkl) \text{ is the observed intensity, and } \overline{I(hkl)}$$

is the average intensity obtained from multiple observations of symmetry-related reflections after rejections.

$$^b \text{FOM} = \text{figure of merit} = \left\langle \frac{\sum P(\alpha) e^{i\alpha}}{\sum P(\alpha)} \right\rangle, \text{ where } \alpha \text{ is the phase angle and } P(\alpha) \text{ is the phase}$$

probability distribution.

con ultra centrifugal filter device (Millipore). Concentrated protein was then applied to a Sephacryl S-100 HiPrep 26/60 size-exclusion chromatography column on an ÄKTA Prime system (GE Health Care) running at 1 ml/min in buffer A. Fractions containing pure YidC_{EC}P1 were pooled and concentrated to 32 mg/ml and stored at –80 °C. Analytical size-exclusion chromatography in line with Multi-Angle Light Scattering analysis is consistent with YidC_{EC}P1 being a monodispersed monomer in solution (data not shown).

Se-Met-incorporated YidC_{EC}P1 was prepared by growing an overnight culture of BL21(DE3) transformed with pYidC_{EC}P1 in M9 minimal medium supplemented with 100 µg/ml ampicillin. 30 ml of overnight culture was used to inoculate 3 × 1 liter of M9 minimal medium (100 µg/ml ampicillin) that was grown at 37 °C to an A_{600} of 0.6. Each 1-liter culture was then directly supplemented with a mixture of the following amino acids: 100 mg of lysine, phenylalanine, threonine; 50 mg of isoleucine, leucine, valine; 60 mg of selenomethionine. After 15 min, protein expression was induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside (final concentration) for 3 h at 37 °C. The purification procedure of Se-Met-incorporated YidC_{EC}P1 was the same as that used for the native protein.

Crystallization—The crystals used for single wavelength anomalous diffraction data collection were grown by the hanging drop vapor diffusion method. The crystallization drops were prepared by mixing 1 µl of protein (32 mg/ml) with 1 µl of reservoir solution and then equilibrating the drop against 1 ml of reservoir solution. The YidC_{EC}P1 construct yielded crystals in the space group I4₁22 with unit cell dimensions 126.1 × 126.1 × 288.4 Å. The crystals have two molecules in the

Structure of YidC

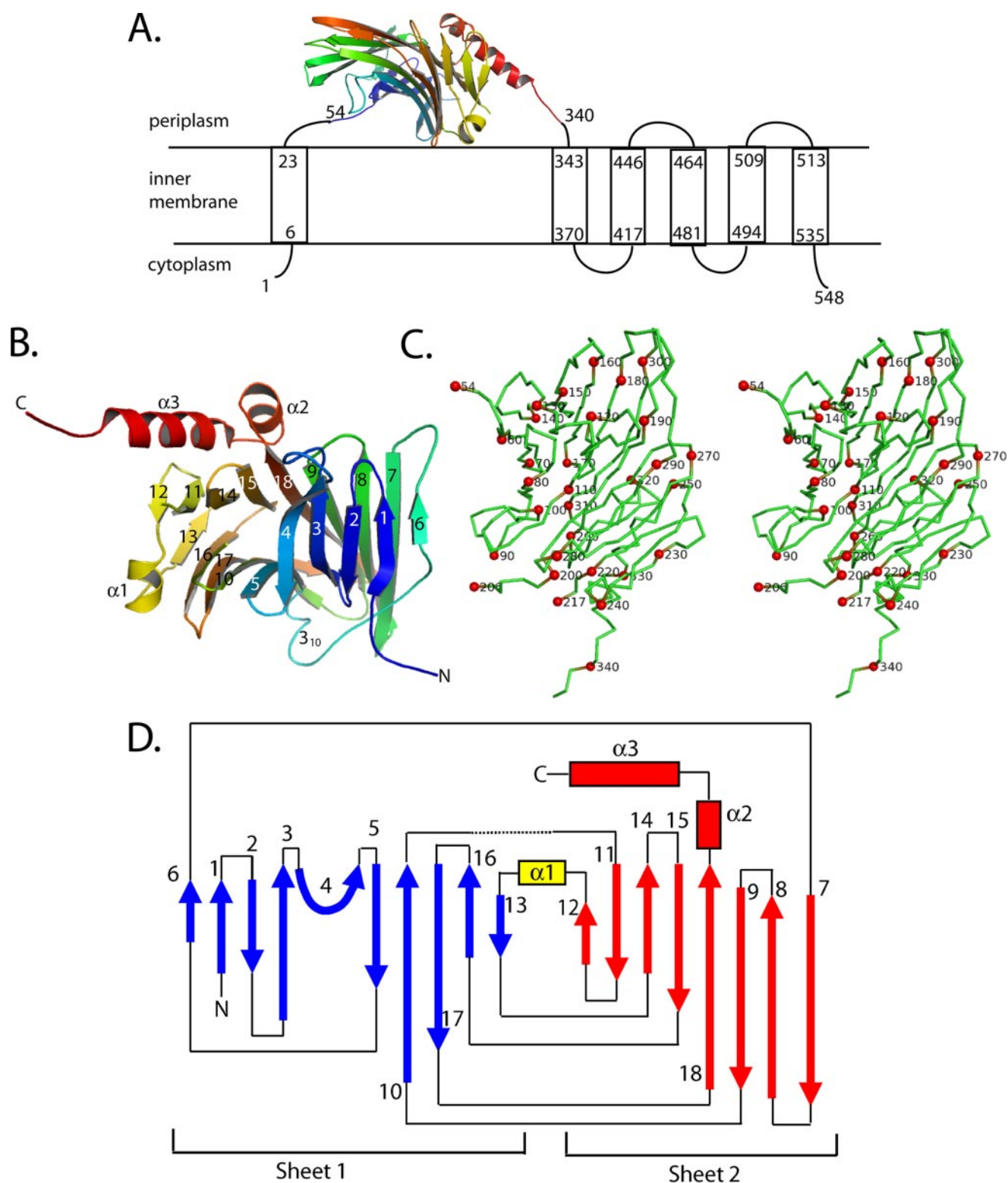


FIGURE 1. The protein fold of YidC_{EC}P1. *A*, the position of YidC_{EC}P1 within the full-length YidC membrane protein, shown schematically within the *E. coli* inner membrane. The YidC_{EC}P1 domain is known to be localized in the periplasm (20); however, its exact orientation with respect to the membrane has not been determined. *B*, a ribbon diagram of YidC_{EC}P1. The structure is colored gradually from N terminus (*blue*) to the C terminus (*red*). The strands are numbered 1–18 and the helices labeled α 1– α 3. *C*, a divergent stereo image of a C α trace of YidC_{EC}P1. Every tenth residue is marked with a *sphere* and labeled. *D*, a protein topology diagram for YidC_{EC}P1. Strands are shown as *arrows*, helices as *boxes*, and loops as *lines*. β -sheet 1 is shown in *blue*, β -sheet 2 in *red*, and helix 1 in *yellow*.

asymmetric unit with a Matthews coefficient of $4.01 \text{ \AA}^3 \text{ Da}^{-1}$ (69.36% solvent). The optimal crystallization reservoir condition was 0.1 M glycine, pH 3.1, 0.2 M $(\text{NH}_4)_2\text{SO}_4$, and 13%

polyethylene glycol 3350. Crystallization was performed at room temperature ($\sim 22^\circ\text{C}$). The cryo-solution condition contained 0.1 M glycine, pH 3.1, 0.2 M $(\text{NH}_4)_2\text{SO}_4$, 15% pol-

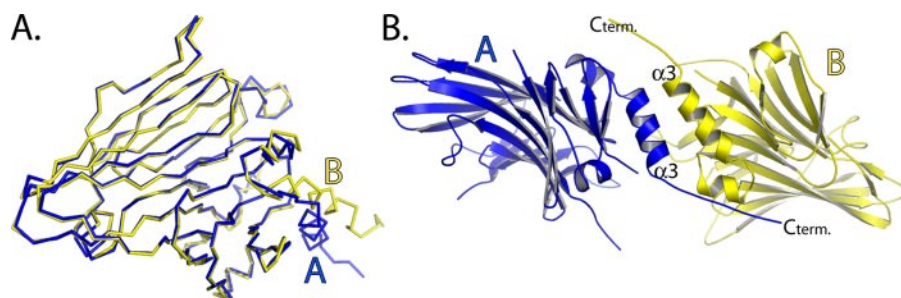


FIGURE 2. The YidC_{EC}P1 asymmetric unit. *A*, superposition of molecules A (blue) and B (yellow) shows a difference in the position of the C-terminal α -helix 3. *B*, molecules A and B within the chosen asymmetric unit are shown as a ribbon diagram. Many of the interactions between molecules A and B occur via the C-terminal helix (α 3).

ethylene glycol 3350, and 20% glycerol. Crystals were incubated in cryo-solution for \sim 5 min before being flash-cooled in liquid nitrogen.

Data Collection—Diffraction data were collected on selenomethionine-incorporated crystals at beamline 8.2.2 of the Advance Light Source, Lawrence Berkeley Laboratory, University of California at Berkeley using a Quantum 315 ADSC area detector. The crystal-to-detector distance was 320 mm. Data were collected with 1° oscillations, and each image was exposed for 3 s. The diffraction data were processed with the program HKL2000 (24). See Table 1 for data collection statistics.

Structure Determination and Refinement—The YidC_{EC}P1 structure was solved by single wavelength anomalous dispersion using a data set collected at the peak wavelength (0.9794 Å), the program SHELX (25) within ccp4i (26), and Autosol within PHENIX version 1.3 (27). SHELXC found eight of the possible ten selenium sites. The program Autobuild within PHENIX version 1.3 (27) automatically constructed \sim 90% of the polypeptide chain and performed density modification. The rest of the model was built using the program Coot (28). The structure was refined using the program Refmac5 (29) and the program CNS (30). The final models were obtained by restrained refinement in Refmac5 with Translation Liberation Screw Rotation (TLS) restraints obtained from the TLS motion determination server (31). The data collection, phasing, and refinement statistics are summarized in Table 1.

Structural Analysis—Secondary structural analysis was performed with the programs DSSP (32), HERA (33), and Promotif (34). The programs SUPERIMPOSE (35) and SUPERPOSE (36) were used to overlap coordinates for structural comparison. The program CONTACT within the program suite CCP4 (26) was used to measure the hydrogen bonding and van der Waals contacts. The program CASTp (37) was used to analyze the molecular surface and search for potential substrate binding sites. The program SURFACE RACER 1.2 (38) was used to measure the solvent-accessible surface of the protein and individual atoms within the protein. A probe radius of 1.4 Å was used in the calculations. The Protein-Protein Interaction Server (39, 40) was used to analyze the interactions between the molecules in the asymmetric unit. The stereochemistry of the structure was analyzed with the program PROCHECK (41). The DALI server was used to find proteins with similar protein folds (42).

Figure Preparation—Figures were prepared using PyMOL (43). The alignment figure was prepared using the programs ClustalW (44) and ESPript (45).

RESULTS

Structure of YidC_{EC}P1—We have produced a C-terminal His₆-tagged soluble construct of the major periplasmic domain of *E. coli* YidC (YidC_{EC}P1) that spans residues Asp²⁶-Leu³⁴⁰ (Fig. 1A). High resolution size-exclusion chromatography analysis and multiangle light scattering analysis of the YidC_{EC}P1 reveal that the protein is very soluble in the absence of detergents and is monomeric in nature (data not shown). YidC_{EC}P1 was crystallized, and the structure was solved by single wavelength anomalous diffraction and refined to 2.5 Å resolution. There are two molecules in the asymmetric unit, and the refined structure includes residues 57–340. In addition, there is electron density for 3 residues at the C terminus that corresponds to the affinity tag used to purify the protein. There is no visible electron density observed for a presumably mobile loop that spans residues 207–216. Additionally, no electron density is observed for the N-terminal residues 26–56. To facilitate crystallization and improve the diffraction quality of the crystals the following mutations were introduced into YidC_{EC}P1: E228A, K229A, E231A, K232A, and K234A. These lysine and glutamate residues were targeted for mutation to alanine in an attempt to reduce the degree of conformational entropy associated with longer side chains that may impede crystallization (46, 47). The mutant YidC_{EC}P1 protein behaves identically to the wild-type YidC_{EC}P1 with regard to its solubility, chromatographic behavior, and light-scattering properties. In addition, no difference was seen between the mutant and wild-type YidC_{EC}P1 proteins when analyzed by CD spectroscopy (data not shown).

The YidC_{EC}P1 Protein Fold—YidC_{EC}P1 is a large distorted β -sandwich motif (super sandwich) constructed from 18 β -strands in two β -sheets and three α -helices (Fig. 1). Sheet 1 of YidC_{EC}P1 is a mixed β -sheet that contains β -strands 1–6, 10, 13, 16, and 17, and sheet 2 is completely antiparallel and contains strands 7–9, 11, 12, 14, 15, and 18. Strand 4 within sheet 1 is highly twisted in order to maintain anti-parallel interactions between strands 2 and 5. Helix 1 (residues 235–239, between β -strands 12 and 13) sits at the mid-point of the structure between β -sheet 1 and β -sheet 2 and packs against the edge of the β -sandwich. Helix 2 (314–317) and helix 3 (325–336) are positioned near the C terminus and pack against β -sheet 2. There are also two 3_{10} -helices spanning residues 123–125 and 156–158, respectively.

YidC_{EC}P1 has the approximate dimensions of $38 \times 60 \times 45$ Å with a significant groove formed along the face of the twisted β -sheet 1. The groove formed along the opposing β -sheet (sheet 2) of the sandwich is partially occupied by α -helix 3. Examination of surface electrostatics indicates that YidC_{EC}P1 does not appear to have any major hydrophobic surface that

Structure of YidC

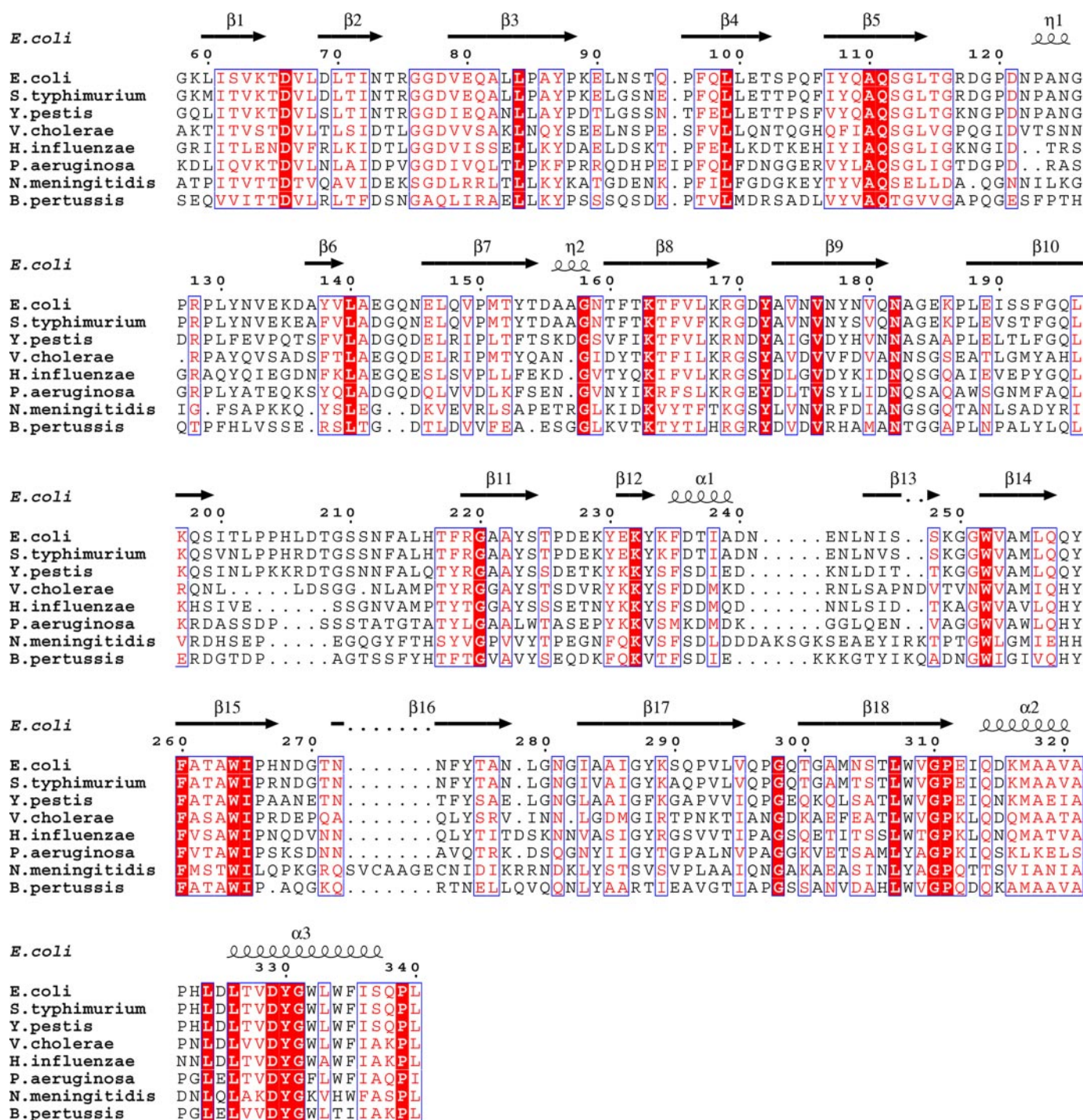


FIGURE 3. Sequence alignment of the major periplasmic region of YidC (YidC_{P1}). The secondary structure as calculated by DSSP (32) is shown above the alignment. The sequences were acquired from the Swiss-Prot data base with the accession numbers for each sequence in parentheses: *E. coli* (P25714); *Salmonella typhimurium* (Q8ZKY4); *Yersinia pestis* (Q8Z9U3); *Vibrio cholerae* (Q9KYY4); *Haemophilus influenzae* (A4N1U4); *Pseudomonas aeruginosa* (Q9HT06); *Neisseria meningitidis* (Q9JW48); *Bordetella pertussis* (P65622).

could accommodate interactions with the acyl chains of the membrane lipids but could possibly interact with the lipid head groups. This is consistent with the solubility of YidC_{EC}P1 in the absence of detergents.

Protein-Protein Interactions Observed in the YidC_{EC}P1 Crystals—Superposition of the two molecules in the asymmetric unit shows that the only significant structural difference between molecule A and molecule B is a shift in the orientation

of the C-terminal α -helix 3 (Fig. 2). The difference in the orientation of helix 3 is likely due to crystal-packing interactions. As mentioned earlier, to facilitate crystallization, five mutations were introduced into a region of YidC_{EC}P1 to replace a cluster of lysine and glutamate residues. The structure shows that these residues are located on or near β -strand 12. Molecule A makes a significant number of crystal contacts between its C terminus and the residues that were mutated in a symmetry-related mol-

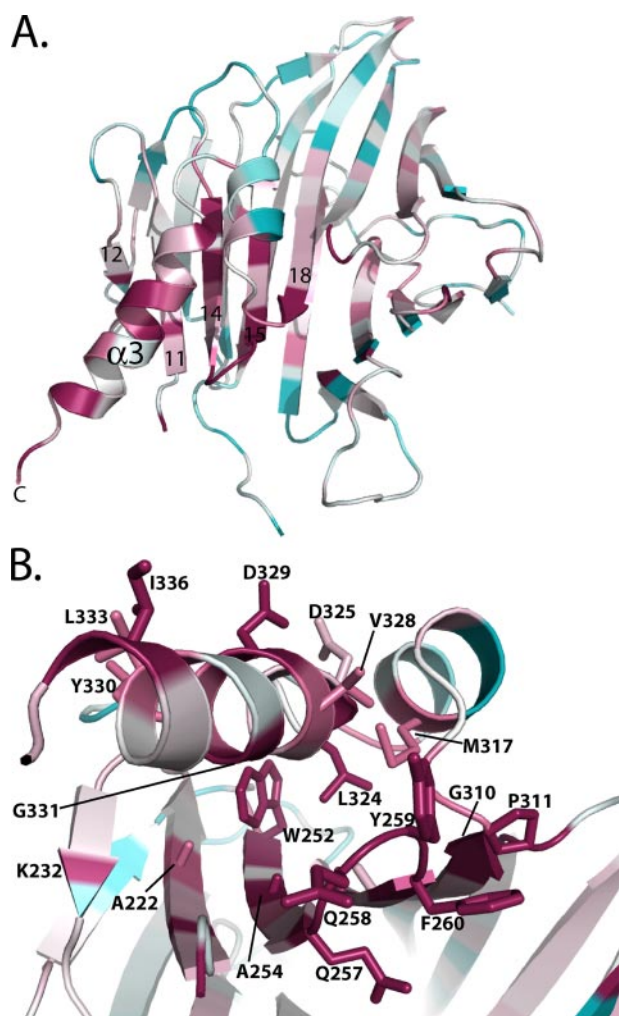


FIGURE 4. Conserved amino acids within the large periplasmic domain of YidC in Gram-negative bacteria mapped onto the structure of YidC_{EC}-P1. A, the regions that are most conserved are rendered in purple, the least conserved in blue. Significant conservation is seen in α -helix 3 and the end of β -sheet 2, which α -helix 3 packs against. B, a close-up view of the most conserved region, as seen from the end of α -helix 3. The side chains are shown as sticks, and the most conserved residues in this region are labeled.

ecule A. This type of interaction is not observed in molecule B, giving a possible explanation for the differences seen in the orientation for the C-terminal α -helix 3.

Conserved Regions of YidC_{EC}-P1—Amino acid sequence alignment of eight YidC variants from various Gram-negative bacterial species reveals a number of conserved residues located throughout YidC_{EC}-P1 (Fig. 3). Most notably, the region at the extreme C terminus of the construct corresponding to α -helix 3 is well conserved. PFAM (48) analysis reveals that the residues 61–350 of YidC define a conserved domain PFAM-B_1222 that is remarkably consistent with the region of YidC_{EC}-P1 observed in the electron density (residues 57–340). Sixty-one YidC variants were extracted from the domain PFAM-B_1222, aligned using ClustalW (44), and analyzed using the program CONSURF (49) that maps conserved residues onto a three-dimensional structure. As shown in Fig. 4, a significant number of conserved residues map to α -helix 3 and to β -strands 11, 12, 14, 15, 18 that cluster on the face of β -sheet 2 and pack against α -helix 3 (Fig. 4A). Closer inspection reveals that many of the

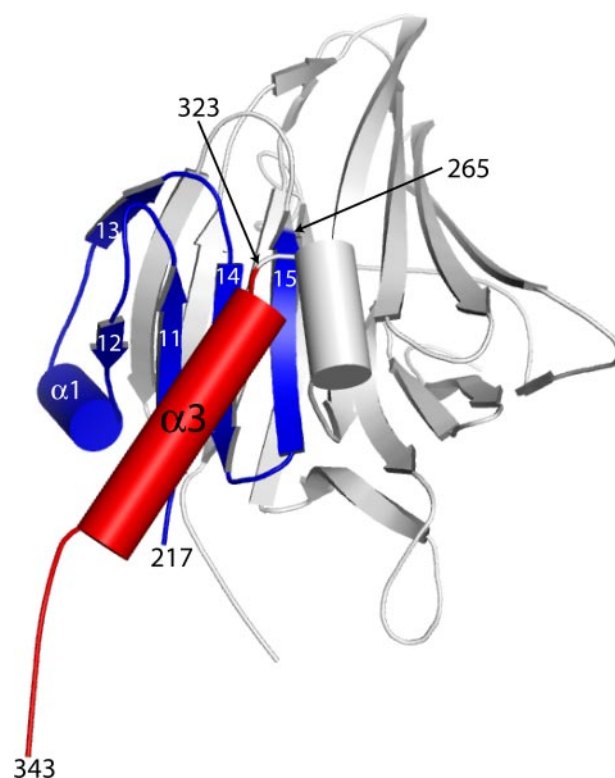


FIGURE 5. The functional regions of YidC_{EC}-P1. The blue region (215–265) has been shown to be responsible for binding to SecF of the SecDFYajC heterotrimer (18). The red region (323–346) has been shown to have insertase function (18, 21).

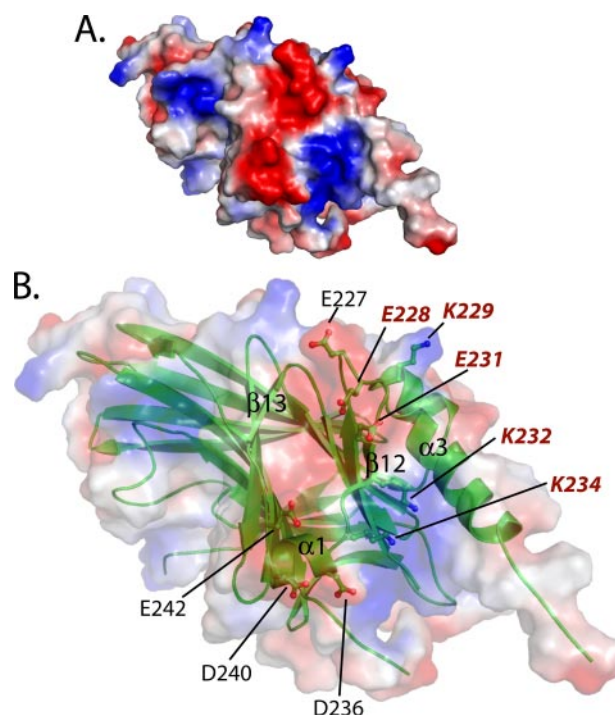


FIGURE 6. Charged surface patches on the edge of the YidC_{EC}-P1 β -sandwich correspond to a region that prior experiments suggest interacts with SecF (18, 21). A, the electrostatic molecular surface of YidC_{EC}-P1. B, the same view as A but rendered with a semitransparent surface revealing a ribbon diagram of YidC_{EC}-P1 and the side chains for charged residues at the edge of the β -sandwich shown as sticks and labeled. The residues mutated to alanine for purposes of improved crystal quality are labeled in red. These residues occur in a cluster on or near β -strand 12.

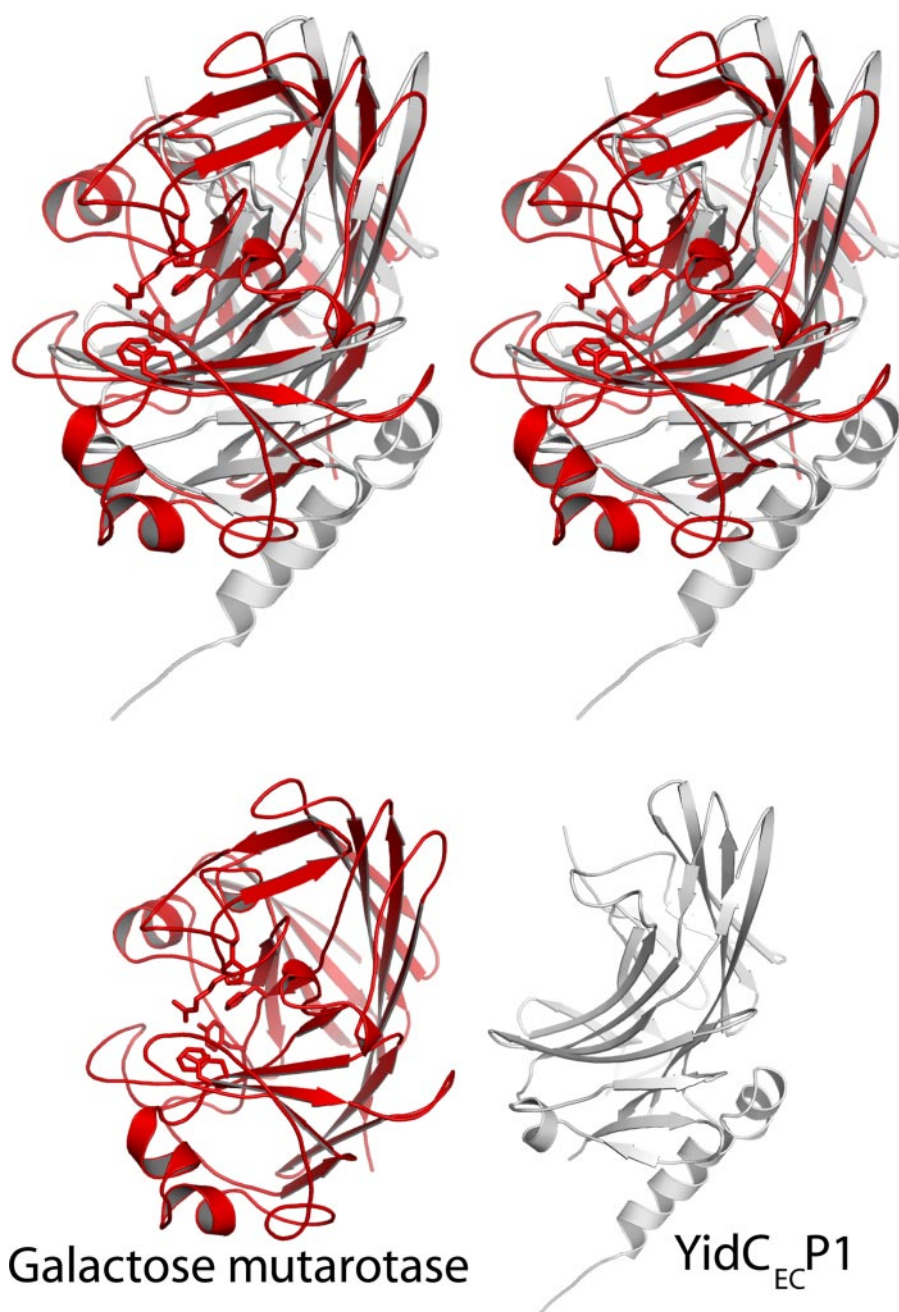


FIGURE 7. YidC_{EC}P1 shows structural similarity to a group of sugar-binding proteins. YidC_{EC}P1 (white) is shown superimposed on the structure of galactose mutarotase (red) (Protein Data Bank code 1L7J) (50). This image is shown in divergent stereo. The side chains for the residues involved in sugar binding in galactose mutarotase are rendered as sticks.

conserved residues are involved in interactions between α -helix 3 and β -sheet 2 (Fig. 4B), suggesting that the interaction may be biologically significant.

Mapping Functional Regions—Previous studies of YidC have mapped two functional regions to the YidC periplasmic domain. First, deletion analysis of YidC has revealed that residues 323–346 of the first periplasmic domain are essential for cell viability and insertase activity (18, 21). This region corresponds to the conserved α -helix 3 at the C terminus of YidC_{EC}P1 (Fig. 5). Second, Xie *et al.* (18) have shown that residues 215–265 of *E. coli* YidC are sufficient for binding to SecF. As depicted in Fig. 5, this region maps to β -strands 11–15 and

α -helix 1 at the edge of the β -sandwich. The approximate molecular surface area for the proposed SecF binding region that includes β -strands 12 and 13 and α -helix 1 is 450 Å². The surface of this region of the structure includes the residues that were mutated for purposes of improving the crystal quality. If these surface residues are modeled back to their wild-type residues, it can be seen that there is a significant negatively charged patch of molecular surface adjacent to a positively charged patch corresponding to the region that was found to be important for SecF binding (Fig. 6). This suggests that the interaction between YidC and SecF may be electrostatic in nature. It is worth noting that the conserved regions of the structure (Fig. 5) correspond well with the regions known to be functionally significant (Fig. 6).

Search for Structural Homologues—Interestingly, despite very low sequence identity, YidC_{EC}P1 shows a significant degree of structural similarity with galactose mutarotase from *Lactococcus lactis* (50). The root mean square deviation for superposition of YidC_{EC}P1 on galactose mutarotase is 3.3 Å for 202 equivalent C α atoms, with 8% sequence identity for those residues compared (Fig. 7). As noted by Thoden and Holden (50), other proteins with related structures include copper amine oxidase (51), hyaluronate lyase (52), chondroitinase (53), β -galactosidase (54), and maltose phosphorylase (55). With the exception of copper amine oxidase, most of the related proteins contain sugar binding sites. The sugar binding pockets for these proteins do not seem to be

conserved in YidC_{EC}P1. Additionally, OpgG, which is located in the *E. coli* periplasm and is required for the biosynthesis of osmotically regulated glucans (56), shares structural similarity with YidC_{EC}P1.

DISCUSSION

In this study we present the first structure of the major periplasmic domain of the protein YidC of *E. coli*. The domain consists of a large stable β -sandwich with a short α -helix at the midpoint of the fold and edge of the sandwich and two α -helices at the C terminus that lay on the curved face of β -sheet 2.

The structure reveals that residues 323–346, which have previously been shown to be essential for cell viability and insertase

activity (18), correspond to α -helix 3. Remarkably, Jiang *et al.* (21) have performed alanine scanning mutagenesis experiments on residues 324–342 of the YidC periplasmic domain and shown that there is no single residue side chain (beyond C β) within this region that is essential for cell viability. Consistent with this result, modeling using alanine side chains within residues 324–342 of the structure of YidC_{EC}P1 suggests that the structure of α -helix 3 would not be significantly altered. Thus, it is reasonable to assume that α -helix 3 of the YidC periplasmic domain is dependent on secondary structure, rather than on individual side chain interactions, for insertase activity. Interestingly, secondary structural analysis of the evolutionarily related proteins Oxa1 of mitochondria and ALB3 of chloroplasts predicts an α -helix located N-terminal to the first transmembrane domain. It is tempting to speculate that this structural element represents a conserved feature related to insertase function.

Xie *et al.* (18) have shown that residues 215–265 of the YidC periplasmic domain fused to maltose-binding protein are sufficient to interact with SecF of the SecDFYajC heterotrimer. The structure presented here shows that this region (215–265) consists of β -strands 11, 12, 14, and 15, which contribute to β -sheet 2, and β -strand 13 that contributes to β -sheet 1. This region also includes the short α -helix 1 (residues 235–239) (Fig. 5). This represents the edge of the β -sandwich structure and corresponds to a negatively charged surface just adjacent to a positively charged surface, suggesting that the interactions between SecF and YidC may be predominately electrostatic (Fig. 6).

An interesting question is how the YidC periplasmic domain is oriented with respect to the membrane. Because the structure does not reveal a hydrophobic surface and the construct does not require detergent for solubility, it is reasonable to assume the domain is probably loosely tethered to the membrane. This notion is consistent with the observation that residues 26–55 do not appear in the electron density and are thus assumed to be flexible. Furthermore, PsiPred analysis (57) does not predict secondary structure for residues 28–59. It is possible that this region forms a flexible “tether” or “linker” (Fig. 1A).

Previously it has been shown that full-length YidC purifies as a mixture of monomers and dimers in the presence of detergent (58). High resolution size-exclusion chromatography analysis in tandem with multiangle light dynamic light scattering analysis indicates that YidC_{EC}P1 behaves as a monomer in solution in the absence of detergent with and without the mutations introduced for purposes of crystallization. Furthermore, although two molecules are present within the asymmetric unit, these molecules do not appear to interact in a manner that would suggest the presence of a strong dimer except for the fact that monomers in the chosen asymmetric unit (the one with the most buried surface between the monomers) interact via α -helix 3, the helix proposed to have insertase function (Fig. 2B). It is possible that YidC oligomerization is mediated by interactions between the transmembrane segments, as has been proposed for the interaction between SecF and SecD (59). It is also possible that the periplasmic domain studied here forms dimers or even higher order oligomers when in close proximity to a membrane.

The function of the major periplasmic domain of YidC remains unclear. *In vivo* assays of YidC-mediated protein insertase activity have shown that up to 90% of the periplasmic

domain (residues 25–323) can be deleted without inhibiting function (21). Further, although it has been shown *in vitro* that the YidC periplasmic domain interacts with SecF, this interaction is not required for insertion of Sec-dependent or Sec-independent substrates (18). Taken together, these observations suggest that the YidC periplasmic domain, which is conserved in all Gram-negative bacteria, may have a function unrelated to protein insertion into the inner membrane. The structure presented here reveals that the YidC periplasmic domain shares significant structural similarity with proteins that are involved in binding sugars such as galactose mutarotase (Fig. 7). Although the YidC periplasmic domain does not appear to bear a sugar binding motif similar to galactose mutarotase, it is tempting to speculate that it could still interact with sugars present in the periplasm. Alternatively, it is conceivable that the YidC periplasmic domain could interact with other periplasmic proteins such as chaperones that facilitate protein folding or secretion following translocation across, or into, the inner membrane. The availability of a three-dimensional structure of the YidC periplasmic domain, as well as stable, soluble protein will facilitate experiments designed to address these ideas.

Acknowledgments—We thank Dr. Corie Ralston at beamline 8.2.2 of the Advance Light Source, Lawrence Berkeley Laboratory, University of California at Berkeley. We also thank Dr. Jaeyong Lee and Dr. Igor D'Angelo for help with data collection.

REFERENCES

- Driessen, A. J., Manting, E. H., and van der Does, C. (2001) *Nat. Struct. Biol.* **8**, 492–498
- Van den Berg, B., Clemons, W. M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C., and Rapoport, T. A. (2004) *Nature* **427**, 36–44
- Nouwen, N., Berrelkamp, G., and Driessen, A. J. (2007) *J. Mol. Biol.* **372**, 422–433
- Vrontou, E., and Economou, A. (2004) *Biochim. Biophys. Acta* **1694**, 67–80
- Duong, F., and Wickner, W. (1997) *EMBO J.* **16**, 2756–2768
- Matsuyama, S., Fujita, Y., and Mizushima, S. (1993) *EMBO J.* **12**, 265–270
- Duong, F., and Wickner, W. (1997) *EMBO J.* **16**, 4871–4879
- Economou, A., Pogliano, J. A., Beckwith, J., Oliver, D. B., and Wickner, W. (1995) *Cell* **83**, 1171–1181
- Urbanus, M. L., Scotti, P. A., Froderberg, L., Saaf, A., de Gier, J. W., Brunner, J., Samuelson, J. C., Dalbey, R. E., Oudega, B., and Luirink, J. (2001) *EMBO Rep.* **2**, 524–529
- Samuelson, J. C., Chen, M., Jiang, F., Moller, I., Wiedmann, M., Kuhn, A., Phillips, G. J., and Dalbey, R. E. (2000) *Nature* **406**, 637–641
- Nagamori, S., Smirnova, I. N., and Kaback, H. R. (2004) *J. Cell Biol.* **165**, 53–62
- van der Laan, M., Urbanus, M. L., Ten Hagen-Jongman, C. M., Nouwen, N., Oudega, B., Harms, N., Driessen, A. J., and Luirink, J. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5801–5806
- van der Laan, M., Bechtluft, P., Kol, S., Nouwen, N., and Driessen, A. J. (2004) *J. Cell Biol.* **165**, 213–222
- Samuelson, J. C., Jiang, F., Yi, L., Chen, M., de Gier, J. W., Kuhn, A., and Dalbey, R. E. (2001) *J. Biol. Chem.* **276**, 34847–34852
- Scotti, P. A., Urbanus, M. L., Brunner, J., de Gier, J. W., von Heijne, G., van der Does, C., Driessen, A. J., Oudega, B., and Luirink, J. (2000) *EMBO J.* **19**, 542–549
- Serek, J., Bauer-Manz, G., Struhalla, G., van den Berg, L., Kiefer, D., Dalbey, R., and Kuhn, A. (2004) *EMBO J.* **23**, 294–301
- Nouwen, N., and Driessen, A. J. (2002) *Mol. Microbiol.* **44**, 1397–1405
- Xie, K., Kiefer, D., Nagler, G., Dalbey, R. E., and Kuhn, A. (2006) *Biochem-*

- istry **45**, 13401–13408
19. Froderberg, L., Houben, E. N., Baars, L., Luirink, J., and de Gier, J. W. (2004) *J. Biol. Chem.* **279**, 31026–31032
 20. Saaf, A., Monne, M., de Gier, J. W., and von Heijne, G. (1998) *J. Biol. Chem.* **273**, 30415–30418
 21. Jiang, F., Chen, M., Yi, L., de Gier, J. W., Kuhn, A., and Dalbey, R. E. (2003) *J. Biol. Chem.* **278**, 48965–48972
 22. Luirink, J., Samuelsson, T., and de Gier, J. W. (2001) *FEBS Lett.* **501**, 1–5
 23. van Bloois, E., Nagamori, S., Koningstein, G., Ullers, R. S., Preuss, M., Oudega, B., Harms, N., Kaback, H. R., Herrmann, J. M., and Luirink, J. (2005) *J. Biol. Chem.* **280**, 12996–13003
 24. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307–326
 25. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr.* **58**, Pt. 10 Pt. 2, 1772–1779
 26. Collaborative Computational Project, No. 4 *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 760–763
 27. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) *Acta Crystallogr.* **58**, Pt. 11, 1948–1954
 28. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr.* **60**, Pt. 12 Pt. 1, 2126–2132
 29. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) *Acta Crystallogr.* **57**, Pt. 1, 122–133
 30. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr.* **54**, Pt. 5, 905–921
 31. Painter, J., and Merritt, E. A. (2006) *Acta Crystallogr.* **62**, Pt. 4, 439–450
 32. Kabsch, W., and Sander, E. (1983) *Biopolymers* **22**, 2577–2637
 33. Hutchinson, E. G., and Thornton, J. M. (1990) *Proteins* **8**, 203–212
 34. Hutchinson, E. G., and Thornton, J. M. (1996) *Protein Sci.* **5**, 212–220
 35. Diederichs, K. (1995) *Proteins* **23**, 187–195
 36. Maiti, R., Van Domselaar, G. H., Zhang, H., and Wishart, D. S. (2004) *Nucleic Acids Res.* **32**, (Web Server issue), W590–W594
 37. Liang, J., Edelsbrunner, H., and Woodward, C. (1998) *Protein Sci.* **7**, 1884–1897
 38. Tsodikov, O. V., Record, M. T., Jr., and Sergeev, Y. V. (2002) *J. Comput. Chem.* **23**, 600–609
 39. Jones, S., and Thornton, J. M. (1995) *Prog. Biophys. Mol. Biol.* **63**, 31–65
 40. Jones, S., and Thornton, J. M. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 13–20
 41. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291
 42. Holm, L., and Sander, C. (1998) *Nucleic Acids Res.* **26**, 316–319
 43. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA
 44. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680
 45. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) *Bioinformatics* **15**, 305–308
 46. Derewenda, Z. S., and Vekilov, P. G. (2006) *Acta Crystallogr.* **62**, Pt. 1, 116–124
 47. Longenecker, K. L., Garrard, S. M., Sheffield, P. J., and Derewenda, Z. S. (2001) *Acta Crystallogr.* **57**, Pt. 5, 679–688
 48. Finn, R. D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, E. L., and Bateman, A. (2006) *Nucleic Acids Res.* **34**, Data base issue, D247–D251
 49. Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., and Ben-Tal, N. (2003) *Bioinformatics* **19**, 163–164
 50. Thoden, J. B., and Holden, H. M. (2002) *J. Biol. Chem.* **277**, 20854–20861
 51. Parsons, M. R., Convery, M. A., Wilmot, C. M., Yadav, K. D., Blakeley, V., Corner, A. S., Phillips, S. E., McPherson, M. J., and Knowles, P. F. (1995) *Structure* **3**, 1171–1184
 52. Li, S., Kelly, S. J., Lamani, E., Ferraroni, M., and Jedrzejewski, M. J. (2000) *EMBO J.* **19**, 1228–1240
 53. Fethiere, J., Eggimann, B., and Cygler, M. (1999) *J. Mol. Biol.* **288**, 635–647
 54. Jacobson, R. H., Zhang, X. J., DuBose, R. F., and Matthews, B. W. (1994) *Nature* **369**, 761–766
 55. Egloff, M. P., Uppenberg, J., Haalck, L., and van Tilbeurgh, H. (2001) *Structure* **9**, 689–697
 56. Hanouille, X., Rollet, E., Clantin, B., Landrieu, I., Odberg-Ferragut, C., Lip-pens, G., Bohin, J. P., and Villeret, V. (2004) *J. Mol. Biol.* **342**, 195–205
 57. McGuffin, L. J., Bryson, K., and Jones, D. T. (2000) *Bioinformatics* **16**, 404–405
 58. van der Laan, M., Houben, E. N., Nouwen, N., Luirink, J., and Driessen, A. J. (2001) *EMBO Rep.* **2**, 519–523
 59. Nouwen, N., Piwowarek, M., Berrelkamp, G., and Driessen, A. J. (2005) *J. Bacteriol.* **187**, 5857–5860