2.8 Å-crystal structure of *Escherichia coli* YidC revealing all core regions, including flexible C2 loop

Yoshiki Tanaka*, Akiya Izumioka*, Aisyah Abdul Hamid, Akira Fujii, Takamitsu Haruyama, Arata Furukawa, and Tomoya Tsukazaki*

Graduate School of Science and Technology, Nara Institute of Science and Technology, Nara 630-0192, Japan

*These authors contributed equally to this work.

*Correspondence: ttsukazaki@mac.com (T.T.)
SUMMARY

YidC/Alb3/Oxa1 family proteins are involved in the insertion and assembly of membrane proteins. The core five transmembrane regions of YidC, which are conserved in the protein family, form a positively charged cavity open to the cytoplasmic side. The cavity plays an important role in membrane protein insertion. In all reported structural studies of YidC, the second cytoplasmic loop (C2 loop) was disordered, limiting the understanding of its role. Here, we determined the crystal structure of YidC including the C2 loop at 2.8 Å resolution with $R/R_{\text{free}} = 21.8/27.5$. This structure and subsequent molecular dynamics simulation indicated that the intrinsic flexible C2 loop covered the positively charged cavity. This crystal structure provides the coordinates of the complete core region including the C2 loop, which is valuable for further analyses of YidC.
INTRODUCTION

Membrane proteins are translated by ribosomes and properly integrated into the membrane, which is a fundamental mechanism conserved in all organisms [1]. The bacterial membrane protein YidC, an essential factor for cell growth, is involved in the integration of membrane proteins [2]. YidC proteins are conserved in the mitochondria and thylakoid as Oxa1 and Alb3, respectively [3, 4]. Recently, YidC-like proteins were identified in the endoplasmic reticulum membrane [5] and archaea [6]. YidC has been proposed to function as an insertase for membrane protein biogenesis and chaperon promoting the proper folding of membrane proteins in the membrane. During membrane protein integration via the SecYEG complex, the protein-conducting channel YidC cooperates with SecYEG to assist with folding of newly synthesized proteins in the membrane. The translating ribosome directly interacts with SecYEG [7-9] or YidC [10], enabling co-current protein translation and integration into the membrane.

Conserved regions of YidC/Oxa1/Alb3 family proteins contain five transmembrane (TM) regions. Based on currently available high-resolution structures, Bacillus halodurans YidC (BhYidC) at 2.4 Å [11], Escherichia coli YidC (EcYidC) at 3.2 Å [12], and Thermotoga maritima YidC (TmYidC) at 3.8 Å resolutions [13], the architectures of the core five TM helixes are essentially identical (Fig. 1B). The core TMs form a positively charged cavity open to the membrane and cytoplasmic side. The conserved Arg is positioned in the cavity. In contrast, loops of the cytoplasmic region 1 (C1) in the crystal structures of EcYidC and BhYidC show different orientations, but the C1 loop in TmYidC is disordered. The C2 loop is completely disordered in all avertable structures. Hence, the C1 and C2 loops appear to have intrinsically high mobility. Only YidCs from gram-negative bacteria possess the additional TM1 and periplasm region (P1), when compared to other YidC orthologs. TM1 and P1 are not crucial for protein activity [14], which is consistent with the non-identical architecture of the P1 domains in EcYidC and TmYidC. Additionally, TM1 is disordered in each crystal structure. TM1 may be extremely flexible in the membrane and may function as an anchor to the membrane or signal peptide for targeting the membrane. In Bacillus subtilis, Arg in the cavity is essential for membrane integration of MifM, a substrate of YidC, and cell growth [11]. In contrast, this residue is not essential for E. coli viability [15], but shows a
conditional defect in cell growth [12]. Together with molecular dynamics (MD) simulation results, this cavity provides a hydrophilic environment and appears to always be filled by water molecules [11, 16]. Mutations that reduce the hydrophilicity of the cavity caused defects in membrane insertion by YidC [17]. Additionally, the cavity cross-linked to MifM provides insight into the insertion model of single spanning membrane proteins by YidC [11]. Other studies of cross-linking showed that YidC interacts with the SecYEG and SecDF complex [18, 19]. YidC complexed with SecYEG and SecDF was shown to form a Sec holo translocon complex [20, 21].

Currently, some information exists regarding the function of YidC, which can be used to perform functional analyses. However, the available structures lack information for the C2 loop, which is important for interacting with ribosomes [22], making evaluation of its molecular mechanism difficult. In this study, we determined the 2.8 Å crystal structure of EcYidC including the C2 loop and conducted MD simulation, which revealed the flexibility of the C2 loop. The complete structure of the core YidC region provides a structural basis for further analysis.

Results and Discussion

2.8 Å-crystal structure of YidC

We used the Helical Data Collection Method at beamline BL32XU at SPring-8 with a microbeam (Hirata 2013) by using more than 100 crystals to determine the crystal structure of YidC at higher resolution than the previous 3.2 Å resolution structure, which was determined using a single crystal [12]. The YidC model was refined to 2.8 Å resolution with $R/R_{\text{free}} = 21.8/27.5$ (Fig. 1A, Table 1).

Gram-negative bacterial YidC consists of a periplasm domain (P1), cytoplasmic loops (C1 and C2), and six TM helices (TM1–M6). One of these helices, TM1, could not be assigned even in the 2.8 Å structure. Comparison of the 2.8 Å structure with the 3.2 Å structure showed that the overall structures had nearly the same RMSD value of 0.49 Å for Ca atoms, whereas the 2.8 Å structure revealed the previously disordered regions of 49–55 and 204–215 in the P1 region and C2 loop (480–492) (Table 1 and Figure S1). The C2 loop is thought to be important for interacting with the ribosome [22], while the P1 region is
2.8 Å-crystal structure of *E. coli* YidC

Tanaka *et al.*

not essential [14]. Collectively, the 2.8 Å resolution structure completely revealed the core region of YidC. Figure 1B and Sup. Fig. 1 show summaries of the core regions of the crystal structures of YidC, clearly demonstrating that the C2 loop was modeled only this study. Previous structural studies, including MD simulations and surface representations, did not adequately take into account the C2 loop structure. Although the positively charged cavity of YidC possessing the conserved Arg366 has been proposed to be open to the cytoplasm and membrane interior based on previous structures [11, 12, 15], our results showed that the C2 loop is at the cytoplasmic entrance of the cavity and appears to restrict exceeding exposure of the cavity. We modeled the C2 loop but the average B-factor of the C2 loop (133) in the structure is higher than that of the overall structure (57.5) and even that of the C1 loop (85.6), suggesting that the C2 loop is considerably more flexible than other core regions (Fig. 2a,b). Because crystal structures represent snapshots of stable conformations, the C2 loop may be at the entrance in the resting state of YidC and become dislocated when YidC is in an active form. The interaction between the ribosome and C2 loop may trigger dislocation of the C2 loop.

**Flexible C2 loop covers the positively charged cavity**

Next, we performed a 100-ns MD simulation of YidC embedded in a POPC bilayer using the complete model of the core region, residues 325–532. The root mean square deviation (RMSD) plot for Cα atoms in the MD simulation (Fig. 2a red line) showed that TM2, 3, 4, and 6 retained the starting model with approximately <1.5 Å and that the C1 loop fluctuated, corresponding to the results of previous MD simulations [11, 16]; residues in the C2 loop modeled in this structural study showed high RMSDs. Comparison of the B-factors of Cα atoms in the crystal structure with the RMSD in the MD simulation showed a good correlation between the fluctuation patterns of these plots (Fig. 2a). The regions with high values for the B-factor and RMSD contained near the position of the 400th residue in the C1 loop and near the position of the 490th residues in the C2 loop and cytoplasmic side of TM5. The TM5 helix lacks a hydrogen bond because of Pro499 positioned in the middle, which is considered to cause the high mobility of the cytoplasmic side (residues 486–498). In accordance with the C2 loop
fluctuation, the cytoplasmic region of TM5 underwent significant structural changes during simulation. In contrast, although conserved Pro371 and Pro 431 in the middle of TM2 and TM3, respectively, cannot form hydrogen bonds to generate an α-helix, fluctuations such as in TM5 were not observed in MD simulation. Because TM5 is located outermost and has fewer intermolecular interactions than other α-helices, TM5 showed the highest mobility. The C2 loop covers the hydrophilic cavity in the crystal structure and initial model in MD simulation, but after 40 ns of MD simulation, the cytoplasmic side of TM5 was dislocated by 3 Å to open the cavity and the C1 loop was shifted 4 Å towards the outside (Fig. 2a,d). To quantify the degree of cavity opening, we measured and plotted the distance between the Ca atoms of Pro371 in TM2 and Pro419 in TM3 as well as Ca atom of Asp488 (Fig. 2c,d). The results showed that the values for 371–488 and 419–488 at 43.84 ns, one of largest points, were 35.3 and 21.3 Å, which are significantly larger than the values of 28.7 and 15.3 Å found in the crystal structure, respectively. In contrast, the minimum values were 27.9 and 14.9 Å at 6.56 ns, which are even lower than those in the crystal structure. During 100-ns MD simulation, we observed entrance fluctuation of the cavity (Sup Movie. 1), which protects and exposes the hydrophobic cavity containing conserved the important residue Arg366 depending on the situation. For example, when a substrate of YidC interacts with the C1 and/or C2 loops during the insertion process, the loops may shift outwards to open the cavity induced by a trigger, including ribosome binding to YidC; this causes the cavity to enables substrate interactions with the inside of YidC.

In this study, we solved the 2.8 Å structure of all core regions of YidC, including the C2 loop. The C2 loop may cover the cavity, but the cavity does not always open to the cytoplasm as previously reported. The structure provides highly accurate information of *E. coli* YidC, which is the most studied member of the YidC/Alb3/Oxa1 family. These findings can facilitate the functional analyses and MD simulations of YidC at the amino acid residue level.

**Figure Legends**

Figure 1 | Crystal structures of YidC
2.8 Å-crystal structure of *E. coli* YidC

(a) 2.8 Å crystal structure of EcYidC. P1, TM2, C1, TM3, TM4, and TM5 are colored in blue, light blue, green, yellow, orange, pink, and red, respectively. Magnified views of the C2 area (right panels) including the 2Fo-Fc electron density map at 1.0 σ (blue) and with omitting the map at 1.7 σ (green). (b) Comparison of the 2.8 Å crystal structure with previously reported structures, PDB ID 3WVF, 3WO6, and 5Y83. The structures are represented as surface models and colored as in A. Each conserved Arg residue is colored in purple.

**Figure 2 | MD simulation of EcYidC**

(a) Plots of B-factor (blue) in the crystal structure and Cα RMSD in 100-ns MD simulation (red) with the residue number of EcYidC core region. The lines are the moving average (p, 3) respectively. (b) The cartoon representation of the crystal structure of EcYidC colored by B-factor value; high (red) to low (blue). (c) Plots of changing Cα distances between residues 371 and 488 (blue); and residues 419–488 (red) versus time. (d) The 2.8 Å crystal structure and MD structure snapshots of the core region at 6.56 and 43.84 ns.

**Figure S1 | 2.8 Å crystal structure of YidC elucidates its C2 loop.**

(a) Superimposition of 2.8 Å structure of EcYidC on BhYidC (PDB ID 3WO6) and TmYidC (PDB ID 5Y83). (b) 2.8 Å structure of EcYidC. (c) 3.2 Å structure of EcYidC (PDB ID 3WVF).

**Material and Methods**

**Data collection and structure determination**

The X-ray diffraction datasets of YidC crystals were collected at 100 K at beamlines BL32XU at SPring-8 using X-ray wavelength of 1.00 Å. The complete datasets were obtained by merging multiple small-wedge (10° each, Δφ = 0.500°) datasets collected from single crystal, and was all data were
2.8 Å-crystal structure of *E. coli* YidC

Tanaka *et al.*

collected automatically by the Zoo system. The collected diffraction images were processed using KAMO [23] with XDS [24]. The cluster of all datasets resulted in the best merging result consisting of 97 datasets. The initial phase was calculated by molecular replacement using PHASER [25] with the previously determined YidC from *Escherichia coli* structure (PDB ID 3WVF) [12]. The structural model of YidC was stepwise-refined using COOT [26] and PHENIX [27, 28] to $R_{\text{work}}/R_{\text{free}} = 0.218/0.275$ with space group $P1$ at 2.8 Å resolution. A Ramachandran plot was constructed using Molprobity [28] and molecular graphics were generated using CueMol2 (http://www.cuemol.org/).

**Molecular dynamics (MD) simulations**

Simulation was carried out using GROMACS version 2016.4 simulation suite [29]. The simulation was started from the 2.8 Å crystal structure of the core region of EcYidC described in this article. The Charmm36 force field [30] was applied. A monomer of the protein was embedded into a POPC bilayer generated by The CHARMM-GUI Membrane Builder [31, 32] and solvated in a $80 \times 80 \times 120 \text{ Å}^3$ box of the simple point charge model [33] water molecules. The water molecules were replaced with Na$^+$ and Cl$^-$ ions to neutralize the simulation system. The simulations were performed with minimization of 50,000 steps for a target Fmax of no greater than 1000, 500, 100, 50, and 20 kJ mol$^{-1}$ nm$^{-1}$. Next, the simulations were performed with an equilibrium of 100 ps in the NVT ensemble and 1 ns in the NPT ensemble. MD simulations were carried out for 100 ns (0.002 ps $\times$ 50,000,000 steps) for the core region of YidC and lipid bilayer in water. The results were analyzed by GROMACS analysis and a movie was generated using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC, Portland, OR, 8
2.8 Å-crystal structure of *E. coli* YidC

Tanaka *et al.*

172 USA) [34].

173
Table 1. Data collection and refinement statistics.

<table>
<thead>
<tr>
<th><strong>Data collection</strong></th>
<th><strong>E. coli YidC</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength</td>
<td>1.00</td>
</tr>
<tr>
<td>Resolution range</td>
<td>41.84 - 2.8 (2.9 - 2.8)</td>
</tr>
<tr>
<td>Space group</td>
<td>P1</td>
</tr>
<tr>
<td>Unit cell</td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>42.91, 76.19, 91.95</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>78.03, 82.61, 78.05</td>
</tr>
<tr>
<td>Total reflections</td>
<td>228455 (23403)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>27282 (2707)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>8.4 (8.6)</td>
</tr>
<tr>
<td>Redundancy (%)</td>
<td>99.69 (99.74)</td>
</tr>
<tr>
<td>Mean I/σ(I)</td>
<td>5.64 (1.26)</td>
</tr>
<tr>
<td>Wilson B-factor</td>
<td>45.28</td>
</tr>
<tr>
<td>R-merge</td>
<td>0.3327 (1.48)</td>
</tr>
<tr>
<td>R-pim</td>
<td>0.1198 (0.529)</td>
</tr>
<tr>
<td>CC1/2</td>
<td>0.977 (0.729)</td>
</tr>
</tbody>
</table>

**Refinement**

| No. Reflections     | 27242 (2702)    |
| Reflections used for R-free | 1995 (197) |
| Rwork / Rfree       | 0.218/0.275 (0.298/0.356) |
| Number of atoms     | 7650            |
| Protein             | 7626            |
| Monoolein           | 19              |
| solvent             | 5               |
| Ramachandran favored (%) | 96.35        |
| Ramachandran allowed (%) | 3.54          |
| Ramachandran outliers (%) | 0.1           |
| Average B-factor    | 59.46           |
| Protein             | 59.41           |
| Monoolein           | 84.2            |
| solvent             | 37.5            |
| R.m.s. derivations  |                 |
| Bond lengths (Å)    | 0.004           |
| Bond angles (°)     | 1.02            |
| PDB ID              | 6AL2            |
ACKNOWLEDGMENTS

We thank K. Abe and S. Suzuki for secretarial assistance and K. Kobayashi and K. Yoshikaie for technical support. This work was supported by the JSPS/MEXT KAKENHI (grant nos. JP26119007, JP26291023, JP18H02405, JP17H05669, JP17K19528, JP16K14713, JP15H01537, and JP15K06972), Mitsubishi Foundation, Noguchi Institute, Naito Foundation, Mochida Memorial Foundation for Medical and Pharmaceutical Research, Foundation for Nara Institute of Science and technology, and PRESTO, JST (grant no. JPMJPR12L3).

AUTHOR CONTRIBUTIONS


DECLARATION OF INTERESTS

The authors declare no competing interests.
2.8 Å-crystal structure of *E. coli* YidC

Tanaka et al.


[18] I. Sachelaru, N.A. Petriman, R. Kudva, P. Kuhn, T. Welte, B. Knapp, F. Drepper, B. Warscheid, H.G. Koch, YidC occupies the lateral gate of the SecYEG translocon and is sequentially displaced by a nascent membrane protein, J Biol...
2.8 Å-crystal structure of E. coli YidC

Tanaka et al.


This study (EcYidC 2.8 Å) PDB ID 3WVF (EcYidC 3.2 Å) PDB ID 3WO6 (BhYidC 2.4 Å) PDB ID 5Y83 (TmYidC 3.8 Å)

Fig. 1
### Fig. 2

#### (a)

- **Distance (Å)**
  - 35.3 Å
  - 27.9 Å

#### (b)

- **Core region (325-532)**
- **B-factor**
  - 10
  - 120

#### (c)

- **RMSD (Å)**
  - 28.7 Å
  - 15.3 Å

#### (d)

- **Pro499**
  - 28.7 Å
  - 15.3 Å
  - 6.56 ns
  - 43.84 ns

- **Crystal structure**
  - 28.7 Å
  - 15.3 Å
  - 6.56 ns
  - 43.84 ns
Fig. 2