

ACCEPTED MANUSCRIPT



A structural model of the active ribosome-bound membrane protein insertase YidC

Stephan Wickles, Abhishek Singharoy, Jessica Andreani, Stefan Seemayer, Lukas Bischoff, Otto Berninghausen, Johannes Soeding, Klaus Schulten, Eli O van der Sluis, Roland Beckmann

DOI: <http://dx.doi.org/10.7554/eLife.03035>

Cite as: eLife 2014;10.7554/eLife.03035

Received: 8 April 2014  
Accepted: 8 July 2014  
Published: 10 July 2014

This PDF is the version of the article that was accepted for publication after peer review. Fully formatted HTML, PDF, and XML versions will be made available after technical processing, editing, and proofing.

This article is distributed under the terms of the [Creative Commons Attribution License](#) permitting unrestricted use and redistribution provided that the original author and source are credited.

Stay current on the latest in life science and biomedical research from eLife.  
[Sign up for alerts](#) at [elife.elifesciences.org](http://elife.elifesciences.org)

1  
2 **A structural model of the active ribosome-bound membrane protein**  
3 **insertase YidC**

4  
5 Stephan Wickles <sup>1</sup>, Abhishek Singharoy <sup>3</sup>, Jessica Andreani <sup>1,2</sup>, Stefan Seemayer <sup>1</sup>,  
6 Lukas Bischoff <sup>1</sup>, Otto Berninghausen <sup>1</sup>, Johannes Söding <sup>1,2</sup>, Klaus Schulten <sup>3</sup>, Eli  
7 O. van der Sluis <sup>1,\*</sup> and Roland Beckmann <sup>1,\*</sup>

8  
9 <sup>1</sup>Gene Center Munich and Center for integrated Protein Science Munich, Department  
10 of Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Straße  
11 25, 81377 Munich, Germany

12  
13 <sup>2</sup>Max-Planck-Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen,  
14 Germany

15  
16 <sup>3</sup>Beckman Institute for Advanced Science and Technology, University of Illinois  
17 at Urbana-Champaign, 405 N. Mathews Ave, Urbana Illinois 61801, USA

18  
19  
20 \* to whom correspondence should be addressed:  
21       beckmann@lmb.uni-muenchen.de, vandersluis@lmb.uni-muenchen.de  
22

23

24 **SUMMARY**

25

26 The integration of most membrane proteins into the cytoplasmic membrane of  
27 bacteria occurs co-translationally. The universally conserved YidC protein  
28 mediates this process either individually as a membrane protein insertase, or in  
29 concert with the SecY complex (Dalbey et al. 2011; Kol et al. 2008). Here, we  
30 present a structural model of YidC based on evolutionary co-variation analysis  
31 (Hopf et al. 2012), lipid-versus-protein-exposure (Lai et al. 2013) and molecular  
32 dynamics simulations. The model suggests a distinctive arrangement of the  
33 conserved five transmembrane domains and a helical hairpin between  
34 transmembrane segment 2 (TM2) and TM3 on the cytoplasmic membrane  
35 surface. The model was used for docking into a cryo-electron microscopy  
36 reconstruction of a translating YidC-ribosome complex carrying the YidC  
37 substrate F<sub>0</sub>C. This structure reveals how a single copy of YidC interacts with the  
38 ribosome at the ribosomal tunnel exit and identifies a site for membrane protein  
39 insertion at the YidC protein-lipid interface. Together, these data suggest a  
40 mechanism for the co-translational mode of YidC-mediated membrane protein  
41 insertion.

42

43

44 **TEXT**

45 At present, a mechanistic understanding of the function of YidC, as well as its  
46 mitochondrial and chloroplast counterparts Oxa1 and Alb3, respectively, is  
47 limited by a lack of structural information. High resolution structures are  
48 available only for the first periplasmic domain (P1) of *Escherichia coli* YidC (Fig.  
49 1a) (Oliver and Paetzel 2008; Ravaud et al. 2008), however, this domain is poorly  
50 conserved, only present in Gram-negative bacteria and not essential for  
51 functionality (Jiang et al. 2003). Furthermore, the region(s) of YidC mediating the  
52 interaction with the ribosome have not been identified, and the oligomeric state  
53 of YidC during co-translational translocation remains controversial (Herrmann  
54 2013; Kedrov et al. 2013; Kohler et al. 2009). Hence, we set out to determine a  
55 molecular model of ribosome-bound YidC during co-translational translocation  
56 of the substrate  $F_0C$  (van der Laan et al. 2004), an integral membrane subunit of  
57 the ATP synthase complex.

58 In order to build an initial structural model of YidC, we predicted contacts  
59 between pairs of residues based on covariation analysis (Hopf et al. 2012; Marks  
60 et al. 2011). For that purpose, we constructed a multiple sequence alignment of  
61 *E. coli* YidC excluding the nonconserved first transmembrane helix (TM1) and  
62 the P1 domain (Fig. 1a) and computed direct evolutionary couplings between  
63 pairs of YidC residues (Kamisetty et al. 2013). The resulting matrix of coupling  
64 strengths (Fig. 1b) contains several diagonal and anti-diagonal patterns of  
65 stronger coupling coefficients, which are indicative of parallel or anti-parallel  
66 helix-helix pairs, respectively. We computed probabilities for each possible helix-  
67 helix contact by aggregating the evidence of stronger coupling coefficients over  
68 the expected interaction patterns and calibrating the resulting raw scores on an  
69 independent dataset of helix-helix interactions to obtain accurate interaction  
70 probabilities. Seven helix-helix contacts attained probabilities above 57% (Fig.  
71 1b-d) while all other possible contacts scored below 15%, demonstrating the  
72 specificity of the method (Fig. 1-figure supplement 1b).

73 We roughly positioned the five TM helices of *E. coli* YidC relative to each other  
74 using the predicted helix-helix contacts as constraints, and rotated them  
75 according to their predicted lipid or protein exposure (Lai et al. 2013) (Fig. 1c).

76 Next, we used MODELLER (Eswar et al. 2008) to create full length models based  
77 on the TM core, secondary structure prediction and the 50 residue-residue  
78 contacts with the highest coupling coefficients (39 excluding intrahelical  
79 contacts, indels and topology violations). In the resulting model (Fig. 1e & f), the  
80 conserved membrane integrated core of YidC forms a helical bundle arranged  
81 like the vertices of a pentagon, in the order 4-5-3-2-6 (clockwise) when viewed  
82 from the cytoplasm (Fig. 1f). Notably, all the predicted interactions between TM  
83 domains can be explained by monomeric YidC suggesting that dimer or oligomer  
84 formation may not be strictly required for YidC activity (see also below).

85 Outside the membrane region, strong helix-helix contacts were predicted within  
86 the cytoplasmic loop between TM2 and TM3, which can be explained the by  
87 formation of a helical hairpin (Fig. 1f). The base of this "helical paddle domain"  
88 (HPD) is structurally constrained by predicted contacts with TM3, its tip on the  
89 other hand is more mobile and appears to interact with lipid headgroups (see  
90 below).

91 While this manuscript was under review, two crystal structures were published  
92 of *Bacillus halodurans* YidC2 (BhYidC2, 34% sequence identity with *E. coli* YidC)  
93 (Kumazaki et al. 2014), providing us with a unique opportunity to directly assess  
94 the accuracy of our model. Overall, the root mean square deviation (RMSD)  
95 between the TM helices of our model and those of BhYidC2 is 7.5 Å (3W06) and  
96 7.3 Å (3W07), which is within the resolution limits of our method. The global  
97 arrangement of TM helices is the same as in BhYidC2, yet, their tilt angle relative  
98 to the plane of the membrane is slightly different (Fig. 2). The tilt angle of the  
99 HPD also differs, as well as its side that faces the membrane (Video 1), which  
100 may be indicative of a high degree of flexibility of this domain, consistent with its  
101 high crystallographic B-factors (Kumazaki et al. 2014). Notably, the HPD is not  
102 essential for YidC function in *E. coli* since the deletion of the entire domain is  
103 possible without compromising cell viability (Jiang et al. 2003).

104 A qualitative difference between our model and BhYidC2 that may have more  
105 mechanistic importance is the relative position of TM3. In the structure of  
106 BhYidC2 a hydrophilic groove is formed on the cytoplasmic side of the TM  
107 bundle that has been proposed to form a binding site for YidC substrates  
108 (Kumazaki et al. 2014). Interestingly, the opening state of this groove differs

109 between the two crystal forms, *i.e.* it is more open in 3W06 than in 3W07 (Video  
110 1), largely due to movement of the N-terminal half of TM3 (Fig. 2-figure  
111 supplement 1). In our model on the other hand, this hydrophilic groove is even  
112 more closed than in 3W07 because we imposed covariation-based constraints  
113 between TM3 and TM5 (Pro<sup>425</sup>-Pro<sup>499</sup>) and between TM3 and TM6 (Cys<sup>423</sup>-Gln<sup>528</sup>  
114 & Phe<sup>433</sup>-Thr<sup>524</sup>) (Fig. 2 and Video 1). Strikingly, in BhYidC2 the distances  
115 between the C $\beta$  atoms of these three pairs are outliers compared to other  
116 residue-residue pairs (20.5 Å/20.9 Å/14.9 Å vs an average of 8.2 Å, Fig. 2-figure  
117 supplement 2). Thus, given that (i) the position of TM3 differs in the two crystal  
118 forms, and (ii) that covariation analysis predicts with high accuracy a closer  
119 interaction of TM3 with TM6 and one contact with TM5, we conclude that  
120 movement of TM3 is a genuine feature of YidC. This movement and the  
121 accompanying dynamics of the hydrophilic groove may represent a crucial step  
122 in the functional cycle of the YidC insertase.

123 In summary, the overall structure of our YidC model agrees well with the  
124 BhYidC2 crystal structure, and a comparison of both structures reveals dynamic  
125 regions in YidC that may be of mechanistic importance. This further illustrates  
126 the power of covariation analysis not merely for structure prediction but also for  
127 obtaining dynamic insights (Hopf et al. 2012).

128 Next, in order to further characterize and validate our obtained YidC model, we  
129 assessed its stability and biochemical properties in the bacterial membrane by  
130 employing traditional molecular dynamics (MD) simulations. Overall, the model  
131 was found to be very stable during the simulation. While the five TM helices  
132 enable a rigid protein core, the polar loop regions tend to swim on the  
133 membrane surface (Fig. 3a). An analysis of inter-residue interactions within the  
134 TM region (Fig. 3b) provides a firm basis to the observed stability of YidC:  
135 hydrophobic residues on the exterior of the TM bundle stabilize interactions  
136 with the apolar lipid tails. The YidC core, in turn, is stabilized both via short and  
137 long-range interactions between the five helices. Residues towards the  
138 cytoplasmic side of the core are primarily polar or charged and, therefore,  
139 engaged in strong electrostatic or charge-dipole interactions. In contrast,  
140 residues on the periplasmic side are primarily aromatic and involved in stacking  
141 and other nonpolar dispersion interactions.

142 In order to verify the functional relevance of residues suggested by the MD  
143 simulations, we created alanine mutants and subjected them to an *in vivo*  
144 complementation assay. Some of the most stabilizing residues, T362 in TM2 and  
145 Y517 in TM6, both of which are located at the same height in the membrane,  
146 completely inactivated YidC when mutated to alanine (Fig. 3d & Fig. 3-figure  
147 supplement 1). Both mutants were stably expressed, indicating that the lack of  
148 complementation was not caused by instability of YidC (Fig. 3-figure supplement  
149 2). Several residues close to this pair show intermediate activity levels (F433,  
150 M471 and F505), whereas residues further away do not show an effect (Fig. 3-  
151 figure supplement 1). Taken together, we provide a model for the overall  
152 arrangement of the conserved domains of YidC that is in good agreement with  
153 our covariation analysis, lipid exposure prediction, MD simulation, *in vivo*  
154 complementation analysis as well as the recent crystal structures.

155 Interestingly, we observed that YidC induces thinning of the lipid bilayer during  
156 the MD simulation. A significant thinning of 7-10 Å results from the hydrophobic  
157 mismatch between the TM helices and the membrane (Fig. 3e). The thinning is  
158 similar in the upper and lower leaflet, and the thinnest region is in proximity of  
159 TM3 and TM5. Since membrane inserting YidC substrates have been chemically  
160 cross-linked to both these helices (Klenner and Kuhn 2012; Klenner et al. 2008;  
161 Yu et al. 2008), we argue that thinning of this region in particular may be  
162 relevant for the molecular mechanism of YidC-dependent membrane insertion.

163 In addition, the distribution of hydrophilic and hydrophobic residues within YidC  
164 revealed the presence of a hydrophilic environment on the cytoplasmic side of  
165 the YidC TM bundle (Fig. 3f), which continues into the mentioned hydrophobic  
166 cluster of aromatic residues towards the periplasmic side. It is tempting to  
167 speculate that this hydrophilic environment may receive the polar termini and  
168 loops of YidC substrates during the initiation of translocation, thus facilitating  
169 their transfer across the hydrophobic core of the (thinned) lipid bilayer (see  
170 below). Notably, essentially the same conclusions have been drawn on the basis  
171 of the BhYidC2 crystal structures and accompanying cross-linking studies  
172 (Kumazaki et al. 2014).

173 In order to provide a molecular model of YidC in its active state, we reconstituted  
174 purified full length YidC (extended with the C-terminus of *R. baltica* YidC (Seitl et

175 al. 2014)) with ribosome nascent chains (RNCs) exposing the first TM helix of  
176  $F_{0c}$ , and subjected the complex to cryo-EM and single particle analysis to a  
177 resolution of  $\sim 8 \text{ \AA}$  (Fig. 4a & b). In agreement with previous structural studies  
178 (Kohler et al. 2009; Seitz et al. 2014), YidC binds to the ribosomal exit site,  
179 however, the improved resolution now allows for a more detailed interpretation.  
180 Firstly, we were able to separate the weaker electron density of the detergent  
181 micelle from that of YidC (Fig. 4a). Secondly, the presence of elongated structural  
182 features (Fig. 4d-f) allowed us to dock our molecular model in a distinct  
183 orientation (cross correlation coefficient 0.865). Following placement of the  
184 YidC-core model, two prominent densities in the membrane region, one next to  
185 TM3 and one next to TM5, remained unaccounted for. These could be attributed  
186 to either TM1 of YidC or to the TM helix of the nascent chain (NC)  $F_{0c}$ . Given that  
187 (i) YidC substrates are known to crosslink to TM3 (Klenner and Kuhn 2012;  
188 Klenner et al. 2008; Yu et al. 2008), and (ii) that the density neighboring TM3 is  
189 aligned with the ribosomal exit tunnel and (iii) that at the same relative position  
190 nascent chains have been observed inside the SecY channel (Frauenfeld et al.  
191 2011) (Fig. 4-figure supplement 1), the most plausible assignment to the density  
192 near TM3 appeared to be the TM helix of  $F_{0c}$ . To verify this, and to exclude that  
193 the density neighboring TM5 corresponds to the nascent chain, we reconstituted  
194 single cysteine mutants of YidC either in TM3 (M430C and P431C) or in TM5  
195 (V500C and T503C) with RNCs of a single cysteine mutant of  $F_{0c}$ (G23C), and  
196 exposed them to disulphide crosslinking. Upon exposure to the oxidator DTNB,  
197 only in the TM3 mutants a DTT-sensitive  $\sim 90 \text{ kDa}$  product appeared that reacted  
198 with antibodies against the nascent chain (NC-tRNA $\sim 30 \text{ kDa}$ , Fig. 4c) as well as  
199 YidC ( $\sim 60 \text{ kDa}$ , Fig. 4c). Thus, the adduct represented indeed the inserting  $F_{0c}$   
200 TM domain crosslinked to TM3 of YidC. RNCs lacking a cysteine in the nascent  
201 chain (Fig. 4-figure supplement 2) or YidC mutants with cysteines in TM5 did not  
202 yield any crosslinks (Fig. 4c). Hence, we conclude that the unaccounted electron  
203 density next to TM3 represents the TM of the nascent chain, and that the density  
204 neighbouring TM5 represents TM1 of YidC (Fig. 4d-f).  
205 We attribute the remaining unaccounted electron density in the periplasmic  
206 region to the P1 domain; however, because it is substantially smaller than the  
207 crystal structure of P1, we did not include it in our molecular model. Flexibility

208 relative to the conserved membrane region of YidC is the most likely explanation  
209 for this finding. We did not observe density for the HPD, in agreement with its  
210 flexibility observed in both, the crystal structures of BhYidC2 and the MD  
211 simulations (Fig. 3c).

212 In order to validate our molecular model of co-translationally active YidC, we  
213 mutated residues that would be in direct contact with the ribosome (Fig. 5a & b)  
214 and analyzed their effect on functionality in the *in vivo* complementation test.  
215 Indeed, mutation of residues Y370A and Y377A (contacting ribosomal RNA helix  
216 59) and D488K (contacting ribosomal protein uL23) severely interfere with YidC  
217 activity (Fig. 5c & Fig. 5-figure supplement 1) thereby emphasizing their  
218 functional significance. All these mutants were stably expressed, indicating that  
219 the lack of complementation was not caused by instability of YidC (Fig. 5-figure  
220 supplement 2). Given that YidC in general is known to be very tolerant to point  
221 mutations (Jiang et al. 2003), this provides further support for the overall  
222 correctness of our model of ribosome-bound YidC during membrane protein  
223 insertion.

224 Finally, it is notable that we observe only a single monomer of YidC bound to the  
225 active ribosome. This is in agreement with recent literature showing clearly that  
226 both YidC (Herrmann 2013; Kedrov et al. 2013; Seitzl et al. 2014) and the SecY  
227 complex (Frauenfeld et al. 2011; Park et al. 2014; Park and Rapoport 2012;  
228 Taufik et al. 2013) can be fully active as monomers. However, the comparison of  
229 models for active YidC and active SecY (Fig. 5e & Fig. 4-figure supplement 1)  
230 reveals an important difference between the two proteins that has mechanistic  
231 implications. While SecY is known to translocate hydrophilic nascent chains  
232 through its central aqueous channel (Cannon et al. 2005; Driessen and Nouwen  
233 2008; Rapoport 2007) and insert TM domains through a lateral gate (Gogala et  
234 al. 2014; Van den Berg et al. 2004), our model suggests that the YidC substrates  
235 are inserted at the protein-lipid interface. Two principal findings of our work  
236 suggest how YidC may facilitate this process: (i) it provides a hydrophilic  
237 environment within the membrane core for receiving the hydrophilic moieties  
238 (termini or loops) of a substrate, and (ii) it reduces the thickness of the lipid  
239 bilayer: initial interaction of the hydrophilic moieties of YidC substrates with the  
240 hydrophilic environment of YidC would allow for a partial insertion into the

241 membrane, while facilitating exposure of the hydrophobic TM domains to the  
242 hydrophobic core of the bilayer. The latter in turn may compensate for the  
243 energetic penalty of driving the hydrophilic moieties across the (already  
244 thinned) bilayer. Further biochemical and structural studies that capture the  
245 earlier stages of this translocation process are eagerly awaited to fully elucidate  
246 this mechanism.  
247

248

## 249 **METHODS**

250

### 251 **Covariation analysis**

252 We constructed a multiple sequence alignment of YidC excluding the unconserved  
253 first transmembrane helix (TM1) and the periplasmic P1 domain. We searched for  
254 homologous sequences of *E. coli* YidC starting from the PFAM seed alignment of  
255 family PF02096 (Punta et al. 2012) and using the sensitive homology detection  
256 software HHblits (Remmert et al. 2012). First, 5 iterations of HHblits were run  
257 against the clustered Uniprot database with no filtering, to retrieve as many  
258 homologous sequences as possible. Then, we post-processed the alignment using  
259 HHfilter to generate a non-redundant alignment at 90% sequence identity. This  
260 resulted in an alignment containing 2366 sequences aligned across YidC helices  
261 TM2-TM6. Using this multiple sequence alignment, we computed direct evolutionary  
262 couplings between pairs of YidC residues using the method of Kamisetty et al  
263 (Kamisetty et al. 2013).

264 To compute probabilities for each possible helix-helix contact, we aggregated the  
265 evidence of stronger coupling coefficients over the expected interaction patterns for  
266 helix-helix contacts, taking into account the expected periodicity of  $\sim 3.5$  residues per  
267 alpha helix turn. We built three non-redundant datasets of mainly-alpha proteins from  
268 the CATH database (Sillitoe et al. 2013). For each protein, we slid a square pattern (of  
269 size  $17 \times 17$  residues = 289 cells) over the matrix of coupling strengths. For each  
270 pattern position, we used Bayes theorem to calculate the raw probability for a helix-  
271 helix interaction, given the 289 coupling strengths. The distributions of coupling  
272 strengths for interacting and non-interacting helix residues were fitted on dataset #1  
273 (1118 proteins). We assigned different weights to the pattern cells, depending on their  
274 position within the pattern and the direction of the helix-helix interaction (parallel or  
275 antiparallel); these weights were optimized on dataset #2 (204 proteins). Finally, we  
276 calibrated the resulting raw scores on dataset #3 (85 proteins) to obtain accurate  
277 interaction probabilities. For cross-validation purposes, we also performed  
278 optimization on dataset #3 and calibration on dataset #2. Optimization on either  
279 dataset #2 or dataset #3 results in the same choice of weights for the pattern cells. The  
280 final posterior probabilities were obtained as the average of the values calibrated on  
281 datasets #2 and #3, weighted by dataset size. The calibration plots for datasets #2 and

282 #3 are shown in Fig. 1-figure supplement 1a. The histogram of final posterior  
283 probabilities obtained for YidC is shown in Fig. 1-figure supplement 1b, which  
284 illustrates the specificity of the helix-helix predictions.

285

### 286 **YidC initial model building**

287 The conserved TM helices of *E.coli* YidC were positioned according to the  
288 covariation based helix-helix contact prediction, and rotated based on their predicted  
289 lipid or protein exposure (Lai et al. 2013), resulting in a starting model of the  
290 conserved TM core of YidC. Additional information based on direct residue-residue  
291 interactions (covariance analysis) and secondary structure predictions by Jpred 3  
292 (Cole et al. 2008) were used as structural restraints in MODELLER (Eswar et al.  
293 2008). From a total of 10 output models that differed mainly in the relative orientation  
294 of the loop regions, the model that satisfied the imposed constraints best was used for  
295 further studies.

296

### 297 **Molecular dynamics simulation**

#### 298 *System preparation*

299 All simulations were performed with the MD software NAMD 2.9 using the  
300 CHARMM36 force field for the proteins and lipids (Klauda et al. 2010). The TIP3P  
301 model is used to simulate water (Jorgensen et al. 1983). The YidC model was inserted  
302 into the membrane, solvated, and ionized using the Membrane Builder tools on  
303 CHARMM-GUI (Jo et al. 2008). The lipid composition is chosen to be 3 POPE to 1  
304 POPG, as has been successfully used for modelling bacterial membranes in several  
305 past MD simulations (Ash et al. 2004; Mondal et al. 2013). An initial membrane  
306 surface of area 110 Å x 110 Å was constructed along the XY plane. The protein lipid-  
307 construct was solvated with 25 Å thick layers of water along the Cartesian Z  
308 directions, and ionized to charge neutralization using Monte Carlo sampling of Na<sup>+</sup>  
309 and Cl<sup>-</sup> ions at 0.15 M concentration. The overall system size is 0.15 M. Prior to  
310 simulation the system was subjected to 10000 steps of conjugate gradient energy  
311 minimization, followed by 100 ps of thermalization and 25 ns of equilibration. During  
312 the first 10 ns of the equilibration stage, the protein was kept fixed, allowing the  
313 lipids, ions and water molecules to equilibrate. Subsequent 15 ns of equilibration  
314 included the protein as well. We then performed 500 ns of MD simulation at 300K.

315 The final 100 ns was repeated thrice to examine the statistical significance of the  
316 result.

### 317 *Simulation parameters*

318 The systems were kept at constant temperature using Langevin dynamics for all non-  
319 hydrogen atoms with a Langevin damping coefficient of  $5\text{ps}^{-1}$ . A constant pressure of  
320 1atm was maintained using the Nose-Hoover Langevin piston with a period of 100fs  
321 and damping timescale of 50fs. Simulations were performed with an integration time  
322 step of 1fs where bonded interactions were computed every time step, short-range  
323 non-bonded interactions every two time steps, and long range electrostatic  
324 interactions every four time steps. A cutoff of 12 Å was used for van der Waals and  
325 short-range electrostatic interactions: a switching function was started at 10 Å for van  
326 der Waals interactions to ensure a smooth cutoff. The simulations were performed  
327 under periodic boundary conditions, with full-system, long-range electrostatics  
328 calculated by using the PME method with a grid point density of  $1/\text{Å}$ . The unit cell  
329 was large enough so that adjacent copies of the system did not interact via short-range  
330 interactions.

### 331 *Flexibility analysis*

332 The overall flexibility of the transmembrane helices relative to their average  
333 configuration was compared. Positional variance of the helix residues was quantified  
334 as a measure of their flexibility. Positional variance was computed by summing the  
335 deviation of individual backbone atom position and dividing by the number of  
336 backbone atoms in the loop. This measure is slightly different from the usual root  
337 mean square fluctuation (RMSF) as contributions from overall displacements of the  
338 helices and their motions relative to the rotation/translation and internal motions of  
339 the protein are included to probe flexibility.

### 340 *Interaction energy, hydrogen bonds, and membrane thickness analysis*

341 To further understand the details of the structure and dynamics of the YidC model we  
342 performed interaction energy, hydrogen bond, and membrane thinning analysis. These  
343 analyses were carried out on the MD trajectory using standard tools available on  
344 VMD. In particular, interaction energies were computed for each trajectory frame of  
345 the final 100 ns simulation using the NAMD Energy plugin on VMD. The numbers  
346 were then time averaged over the entire 100ns, locally averaged for every residue  
347 over a cut-off distance of 10 Å, and plotted on the structure in Fig. 3b. Hydrogen  
348 bonds are defined solely on the basis of geometric parameters (bond angle:  $20^\circ$ ; bond-

349 length: 3.8 Å) between donors and acceptors. Thickness at a given point on the  
350 membrane surface was probed by finding the nearest lipid head group and measuring  
351 the minimum distance between the phosphate on that lipid head and one on the  
352 opposite leaflet.

353

#### 354 **Purification of ribosome nascent chain complexes (RNCs)**

355 RNC constructs encoding residues 1-46 of F<sub>OC</sub> (preceded by an N-terminal His-tag  
356 and 3C rhinoprotease cleavage site, and followed by an HA-tag and TnaC stalling  
357 sequence) were cloned into a pBAD vector (Invitrogen) by standard molecular  
358 biology techniques, and expressed and purified as described in Bischoff *et al.*,  
359 (submitted). Briefly, *E.coli* KC6  $\Delta$ *smpB* $\Delta$ *ssrA* (Seidelt et al. 2009) carrying the  
360 plasmid for F<sub>OC</sub> was grown in LB with 100 µg/ml ampicilin at 37°C to an OD<sub>600</sub> = 0.5  
361 and expression was induced for 1h by adding 0.2 % arabinose. Cells were lysed and  
362 debris was removed by centrifugation for 20 min at 16.000 rpm in a SS34-rotor  
363 (Sorvall). The cleared lysate was spun overnight through a sucrose cushion at 45.000  
364 rpm in a Ti45 rotor (Beckmann), the ribosomal pellet was resuspended for 1 h at 4°C  
365 and RNCs were purified in batch by affinity purification using Talon (Clontech).  
366 After washing the Talon beads with high salt buffer the RNCs were eluted and loaded  
367 onto a linear 10 % - 40 % sucrose gradient. The 70S peak was collected, RNCs were  
368 concentrated by pelleting, resuspended in an appropriate volume of RNC Buffer (20  
369 mM HEPES pH 7.2, 100 mM KOAc, 6 mM MgOAc<sub>2</sub>, 0.05% (w/v) dodecyl  
370 maltoside), flash frozen in liquid N<sub>2</sub> and stored at -80°C. The complete sequence of  
371 the nascent chain is:

372 MGHHHHHHHHDYDIPTTLEVLFGQPGTMENLNMDLLYMAAAVMMGLAAI  
373 GAAIGIGILGGKFLEGAARQPDLIYPYDVPDYAGPNILHISVTSKWFNIDNKIV  
374 DHRP

375

#### 376 **Purification of YidC**

377 For purification and reconstitution studies, *E.coli* YidC extended with the C-terminus  
378 from *R. baltica* (Seitl et al. 2014) was re-cloned into pET-16 (Novagen) with an N-  
379 terminal His-tag followed by a 3C rhinovirus protease site. Expression and  
380 purification was performed essentially as described (Lotz et al. 2008). Briefly, *E.coli*  
381 C43(DE3) cells (Miroux and Walker 1996) harboring the YidC construct were grown  
382 at 37°C to an OD<sub>600</sub> = 0.6 and expression was induced by adding 0.5 mM IPTG. YidC

383 was solubilized with Cymal-6 (Anatrace) and purified by affinity chromatography  
384 using TALON (Clontech). The N-terminal His-tag of the eluted protein was cleaved  
385 off with 3C protease during overnight dialysis at 4°C, followed by gel filtration  
386 chromatography (Superdex 200, GE Healthcare). Fractions of the monodisperse peak  
387 were pooled, concentrated to ~ 1 mg/ml in YidC Buffer (20 mM NaPO<sub>4</sub> pH 6.8, 100  
388 mM KOAc, 10 % glycerol, 0.05 % Cymal-6) and directly used for further structural  
389 or biochemical assays.

390

### 391 **Disulphide crosslinking**

392 For disulphide crosslink analysis, F<sub>OC</sub><sup>(G23C)</sup>-RNCs and single cysteine mutants of  
393 YidC were purified separately and reconstituted by incubating 100 pmol of RNCs  
394 with 500 pmol of freshly purified YidC for 30 min at 37°C. The endogenous cysteine  
395 in YidC at position 423 was replaced by serine. Disulphide crosslinking was induced  
396 by adding 1 mM 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) for 10 min at 4°C and  
397 quenched by adding 20 mM N-Ethylmaleimide (NEM) for 20 min at 4°C.  
398 Crosslinked RNC-YidC complexes were separated from non-crosslinked YidC using  
399 a 10 % - 40 % linear sucrose gradient, and the 70S peak was harvested and analysed  
400 by SDS-PAGE followed by western blotting.

401

### 402 **Complementation assay**

403 For *in vivo* complementation studies, wildtype *E. coli* YidC was recloned into  
404 pTrc99a (Pharmacia), and mutants were created by standard molecular cloning  
405 techniques. *E. coli* FTL10 cells (Hatzixanthis et al. 2003) harboring pTrc99a plasmids  
406 encoding the YidC variants were grown overnight at 37°C in LB medium  
407 supplemented with 100 µg/ml ampiciline, 50 µg/ml kanamycin and 0.2% arabinose.  
408 YidC depletion was carried out by transferring the cells to LB medium supplemented  
409 with 100 µg/ml ampiciline, 50 µg/ml kanamycin and 0.2% glucose, followed by and  
410 additional incubation for 3h at 37°C. Cell suspensions of all constructs were adjusted  
411 to OD<sub>600</sub> = 0.1 and either loaded onto SDS-PAGE gels for subsequent Western  
412 blotting, or further diluted to OD<sub>600</sub> = 10<sup>-5</sup>. Each dilution was spotted on LB agar  
413 plates supplemented 100 µg/ml ampiciline, 50 µg/ml kanamycin and either 0.2%  
414 arabinose or 0.2% glucose, and incubated overnight at 37°C.

415

### 416 **Electron microscopy and image processing**

417 For cryo-EM analysis, F<sub>0</sub>c-RNC:YidC complexes were reconstituted by incubating  
418 10 pmol of RNCs with 100 pmol of freshly purified YidC for 30 min at 37°C in a  
419 final volume of 50 µl of RNC buffer. Samples were applied to carbon-coated holey  
420 grids according to standard methods (Wagenknecht et al. 1988). Micrographs were  
421 collected under low-dose conditions on a FEI TITAN KRIOS operating at 200 kV  
422 using a 4k x 4k TemCam-F416 CMOS camera and a final pixel size of 1.035Å on the  
423 object scale.

424 Image processing was done using the SPIDER software package (Shaikh et al. 2008).  
425 The defocus was determined using the TF ED command in SPIDER followed by  
426 automated particle picking using Signature (Chen and Grigorieff 2007). The machine-  
427 learning algorithm MAPPOS (Norouzi et al. 2013) was used to subtract “false  
428 positive” particles from the data set and initial alignment was performed using an  
429 empty 70S ribosome as reference. The complete data set (876376 particles) was  
430 sorted using competitive projection matching in SPIDER followed by focused sorting  
431 for ligand density (Leidig et al. 2013), and refined to a final resolution of ~8.0 Å  
432 (Fourier shell correlation (FSC) cut-off 0.5). The final dataset consisted of 58960  
433 particles showing electron density for P-site tRNA and ligand density at the tunnel  
434 exit.

435

436

437 **Acknowledgements**

438 We would like to thank C. Ungewickell for assistance with cryo-electron  
439 microscopy, Susan Vorberg for assistance with covariation analyses, T. Palmer  
440 for providing *E. coli* strain FTL10, A. Driessen and A. Kuhn for providing YidC  
441 antibodies, J. Philippou-Massier and U. Gaul for use of the robotic high-throuput  
442 facility, A. Heuer for assistance with animations and B. Beckert & A. Kedrov for  
443 discussions.

444 SW and LB were supported by the International Max Planck Research School, SS  
445 by grant GRK1721 from the DFG, JA by a Humboldt Research Fellowship of the  
446 Alexander-von-Humboldt Foundation and the Bavarian Network for Molecular  
447 Biosystems (BioSysNet), AS by a Beckman Postdoctoral Fellowship, KS by the  
448 Center for Macromolecular Modeling and Bioinformatics (NIH 9P41GM104601,  
449 NIH R01-GM67887) and the Center for the Physics of Living Cells (NSF PHY-  
450 0822613), JS by the Deutsche Forschungsgemeinschaft (DFG) trough grants  
451 SFB646, GRK1721, and QBM, by the Bundesministerium für Bildung und  
452 Forschung through grant CoreSys and the Bavarian Network for Molecular  
453 Biosystems (BioSysNet), and RB by the Center for Integrated Protein Science and  
454 the European Research Council (Advanced Grant CRYOTRANSLATION).

455

456 **Author contributions**

457 SW performed purifications, reconstitutions, cryo-EM image processing and  
458 model building, LB contributed to purifications and image processing, JA, SS and  
459 JS performed covariation analyses, AS and KS performed MD simulations and  
460 related analyses, OB performed cryo-EM data collection, EvdS designed  
461 experiments and supervised the project together with RB. All authors  
462 contributed to data interpretation and writing of the manuscript.

463

464 **REFERENCES**

465

466

467 1 Ash, W. L., Zlomislic, M. R., Oloo, E. O. & Tieleman, D. P. 2004. Computer  
468 simulations of membrane proteins. *Biochim Biophys Acta* **1666**, 158-189,  
469 doi:10.1016/j.bbamem.2004.04.012.

470 2 Cannon, K. S., Or, E., Clemons, W. M., Jr., Shibata, Y. & Rapoport, T. A. 2005.  
471 Disulfide bridge formation between SecY and a translocating polypeptide  
472 localizes the translocation pore to the center of SecY. *J Cell Biol* **169**, 219-  
473 225, doi:10.1083/jcb.200412019.

474 3 Chen, J. Z. & Grigorieff, N. 2007. SIGNATURE: a single-particle selection  
475 system for molecular electron microscopy. *J Struct Biol* **157**, 168-173,  
476 doi:10.1016/j.jsb.2006.06.001.

477 4 Cole, C., Barber, J. D. & Barton, G. J. 2008. The Jpred 3 secondary structure  
478 prediction server. *Nucleic Acids Res* **36**, W197-201,  
479 doi:10.1093/nar/gkn238.

480 5 Dalbey, R. E., Wang, P. & Kuhn, A. 2011. Assembly of bacterial inner  
481 membrane proteins. *Annu Rev Biochem* **80**, 161-187,  
482 doi:10.1146/annurev-biochem-060409-092524.

483 6 Driessen, A. J. & Nouwen, N. 2008. Protein translocation across the  
484 bacterial cytoplasmic membrane. *Annu Rev Biochem* **77**, 643-667,  
485 doi:10.1146/annurev.biochem.77.061606.160747.

486 7 Eswar, N., Eramian, D., Webb, B., Shen, M. Y. & Sali, A. 2008. Protein  
487 structure modeling with MODELLER. *Methods Mol Biol* **426**, 145-159,  
488 doi:10.1007/978-1-60327-058-8\_8.

489 8 Frauenfeld, J. *et al.* 2011. Cryo-EM structure of the ribosome-SecYE  
490 complex in the membrane environment. *Nat Struct Mol Biol* **18**, 614-621,  
491 doi:10.1038/nsmb.2026.

492 9 Gogala, M. *et al.* 2014. Structures of the Sec61 complex engaged in nascent  
493 peptide translocation or membrane insertion. *Nature* **506**, 107-110,  
494 doi:10.1038/nature12950.

495 10 Hatzixanthis, K., Palmer, T. & Sargent, F. 2003. A subset of bacterial inner  
496 membrane proteins integrated by the twin-arginine translocase. *Mol*  
497 *Microbiol* **49**, 1377-1390.

498 11 Herrmann, J. M. 2013. The bacterial membrane insertase YidC is a  
499 functional monomer and binds ribosomes in a nascent chain-dependent  
500 manner. *J Mol Biol* **425**, 4071-4073, doi:10.1016/j.jmb.2013.08.003.

501 12 Hopf, T. A. *et al.* 2012. Three-dimensional structures of membrane  
502 proteins from genomic sequencing. *Cell* **149**, 1607-1621,  
503 doi:10.1016/j.cell.2012.04.012.

504 13 Jiang, F. *et al.* 2003. Defining the regions of Escherichia coli YidC that  
505 contribute to activity. *J Biol Chem* **278**, 48965-48972,  
506 doi:10.1074/jbc.M307362200.

507 14 Jo, S., Kim, T., Iyer, V. G. & Im, W. 2008. CHARMM-GUI: a web-based  
508 graphical user interface for CHARMM. *J Comput Chem* **29**, 1859-1865,  
509 doi:10.1002/jcc.20945.

510 15 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M.  
511 L. 1983. Comparison of Simple Potential Functions for Simulating Liquid

- 512 Water. *Journal of Chemical Physics* **79**, 926-935, doi:Doi  
513 10.1063/1.445869.
- 514 16 Kamisetty, H., Ovchinnikov, S. & Baker, D. 2013. Assessing the utility of  
515 coevolution-based residue-residue contact predictions in a sequence- and  
516 structure-rich era. *Proc Natl Acad Sci U S A* **110**, 15674-15679,  
517 doi:10.1073/pnas.1314045110.
- 518 17 Kedrov, A. *et al.* 2013. Elucidating the native architecture of the YidC:  
519 ribosome complex. *J Mol Biol* **425**, 4112-4124,  
520 doi:10.1016/j.jmb.2013.07.042.
- 521 18 Klauda, J. B. *et al.* 2010. Update of the CHARMM all-atom additive force  
522 field for lipids: validation on six lipid types. *J Phys Chem B* **114**, 7830-  
523 7843, doi:10.1021/jp101759q.
- 524 19 Klenner, C. & Kuhn, A. 2012. Dynamic disulfide scanning of the  
525 membrane-inserting Pf3 coat protein reveals multiple YidC substrate  
526 contacts. *J Biol Chem* **287**, 3769-3776, doi:10.1074/jbc.M111.307223.
- 527 20 Klenner, C., Yuan, J., Dalbey, R. E. & Kuhn, A. 2008. The Pf3 coat protein  
528 contacts TM1 and TM3 of YidC during membrane biogenesis. *FEBS Lett*  
529 **582**, 3967-3972, doi:10.1016/j.febslet.2008.10.044.
- 530 21 Kohler, R. *et al.* 2009. YidC and Oxa1 form dimeric insertion pores on the  
531 translating ribosome. *Mol Cell* **34**, 344-353,  
532 doi:10.1016/j.molcel.2009.04.019.
- 533 22 Kol, S., Nouwen, N. & Driessen, A. J. 2008. Mechanisms of YidC-mediated  
534 insertion and assembly of multimeric membrane protein complexes. *J Biol*  
535 *Chem* **283**, 31269-31273, doi:10.1074/jbc.R800029200.
- 536 23 Kumazaki, K. *et al.* 2014. Structural basis of Sec-independent membrane  
537 protein insertion by YidC. *Nature*, doi:10.1038/nature13167.
- 538 24 Lai, J. S., Cheng, C. W., Lo, A., Sung, T. Y. & Hsu, W. L. 2013. Lipid exposure  
539 prediction enhances the inference of rotational angles of transmembrane  
540 helices. *BMC Bioinformatics* **14**, 304, doi:10.1186/1471-2105-14-304.
- 541 25 Leidig, C. *et al.* 2013. Structural characterization of a eukaryotic  
542 chaperone--the ribosome-associated complex. *Nat Struct Mol Biol* **20**, 23-  
543 28, doi:10.1038/nsmb.2447.
- 544 26 Lotz, M., Haase, W., Kuhlbrandt, W. & Collinson, I. 2008. Projection  
545 structure of yidC: a conserved mediator of membrane protein assembly. *J*  
546 *Mol Biol* **375**, 901-907, doi:10.1016/j.jmb.2007.10.089.
- 547 27 Marks, D. S. *et al.* 2011. Protein 3D structure computed from evolutionary  
548 sequence variation. *PLoS One* **6**, e28766,  
549 doi:10.1371/journal.pone.0028766.
- 550 28 Miroux, B. & Walker, J. E. 1996. Over-production of proteins in *Escherichia*  
551 *coli*: mutant hosts that allow synthesis of some membrane proteins and  
552 globular proteins at high levels. *J Mol Biol* **260**, 289-298,  
553 doi:10.1006/jmbi.1996.0399.
- 554 29 Mondal, S., Khelashvili, G., Shi, L. & Weinstein, H. 2013. The cost of living  
555 in the membrane: a case study of hydrophobic mismatch for the multi-  
556 segment protein LeuT. *Chem Phys Lipids* **169**, 27-38,  
557 doi:10.1016/j.chemphyslip.2013.01.006.
- 558 30 Norousi, R. *et al.* 2013. Automatic post-picking using MAPPOS improves  
559 particle image detection from cryo-EM micrographs. *J Struct Biol* **182**, 59-  
560 66, doi:10.1016/j.jsb.2013.02.008.

- 561 31 Oliver, D. C. & Paetzel, M. 2008. Crystal structure of the major periplasmic  
562 domain of the bacterial membrane protein assembly facilitator YidC. *J Biol*  
563 *Chem* **283**, 5208-5216, doi:10.1074/jbc.M708936200.
- 564 32 Park, E. *et al.* 2014. Structure of the SecY channel during initiation of  
565 protein translocation. *Nature* **506**, 102-106, doi:10.1038/nature12720.
- 566 33 Park, E. & Rapoport, T. A. 2012. Bacterial protein translocation requires  
567 only one copy of the SecY complex in vivo. *J Cell Biol* **198**, 881-893,  
568 doi:10.1083/jcb.201205140.
- 569 34 Punta, M. *et al.* 2012. The Pfam protein families database. *Nucleic Acids*  
570 *Res* **40**, D290-301, doi:10.1093/nar/gkr1065.
- 571 35 Rapoport, T. A. 2007. Protein translocation across the eukaryotic  
572 endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**,  
573 663-669, doi:10.1038/nature06384.
- 574 36 Ravaud, S., Stjepanovic, G., Wild, K. & Sinning, I. 2008. The crystal  
575 structure of the periplasmic domain of the Escherichia coli membrane  
576 protein insertase YidC contains a substrate binding cleft. *J Biol Chem* **283**,  
577 9350-9358, doi:10.1074/jbc.M710493200.
- 578 37 Remmert, M., Biegert, A., Hauser, A. & Soding, J. 2012. HHblits: lightning-  
579 fast iterative protein sequence searching by HMM-HMM alignment. *Nat*  
580 *Methods* **9**, 173-175, doi:10.1038/nmeth.1818.
- 581 38 Seidelt, B. *et al.* 2009. Structural insight into nascent polypeptide chain-  
582 mediated translational stalling. *Science* **326**, 1412-1415,  
583 doi:10.1126/science.1177662.
- 584 39 Seitzl, I., Wickles, S., Beckmann, R., Kuhn, A. & Kiefer, D. 2014. The C-  
585 terminal regions of YidC from Rhodospirellula baltica and Oceanicaulis  
586 alexandrii bind to ribosomes and partially substitute for SRP receptor  
587 function in Escherichia coli. *Mol Microbiol* **91**, 408-421,  
588 doi:10.1111/mmi.12465.
- 589 40 Shaikh, T. R. *et al.* 2008. SPIDER image processing for single-particle  
590 reconstruction of biological macromolecules from electron micrographs.  
591 *Nat Protoc* **3**, 1941-1974, doi:10.1038/nprot.2008.156.
- 592 41 Sillitoe, I. *et al.* 2013. New functional families (FunFams) in CATH to  
593 improve the mapping of conserved functional sites to 3D structures.  
594 *Nucleic Acids Res* **41**, D490-498, doi:10.1093/nar/gks1211.
- 595 42 Taufik, I., Kedrov, A., Exterkate, M. & Driessen, A. J. 2013. Monitoring the  
596 activity of single translocons. *J Mol Biol* **425**, 4145-4153,  
597 doi:10.1016/j.jmb.2013.08.012.
- 598 43 Van den Berg, B. *et al.* 2004. X-ray structure of a protein-conducting  
599 channel. *Nature* **427**, 36-44, doi:10.1038/nature02218.
- 600 44 van der Laan, M., Bechtluft, P., Kol, S., Nouwen, N. & Driessen, A. J. 2004.  
601 F1F0 ATP synthase subunit c is a substrate of the novel YidC pathway for  
602 membrane protein biogenesis. *J Cell Biol* **165**, 213-222,  
603 doi:10.1083/jcb.200402100.
- 604 45 Wagenknecht, T., Grassucci, R. & Frank, J. 1988. Electron microscopy and  
605 computer image averaging of ice-embedded large ribosomal subunits  
606 from Escherichia coli. *J Mol Biol* **199**, 137-147.
- 607 46 Yu, Z., Koningstein, G., Pop, A. & Lührink, J. 2008. The conserved third  
608 transmembrane segment of YidC contacts nascent Escherichia coli inner

609 membrane proteins. *J Biol Chem* **283**, 34635-34642,  
610 doi:10.1074/jbc.M804344200.  
611  
612  
613

614 **FIGURE LEGENDS**

615

616 Figure 1

617 **Evolutionary covariation based structural model of *E. coli* YidC**

618 **a:** Membrane topology of YidC, with helix colouring as in all subsequent Figures.

619 **b:** Matrix of coupling strengths between pairs of YidC residues based on an  
620 alignment of 2366 non-redundant sequences. Helix-helix pairs with posterior  
621 probabilities higher than 57% are outlined in boxes; the 50 residue-residue pairs  
622 with highest coupling coefficients are indicated with red crosses.

623 **c:** Overall arrangement of TM helices viewed from the cytoplasm based on the  
624 prediction of helix-helix pairs (black lines) and exposure to lipid (yellow) or  
625 protein (green). The first residue of each helix is indicated with an asterisk.

626 **d:** Linear representation of YidC with the seven most probable helix-helix pairs  
627 indicated by arches, with thicknesses approximating posterior probabilities.

628 **e & f:** Side view and cytoplasmic view, respectively, of the *E. coli* YidC model  
629 based on covariation analysis, with predicted residue-residue pairs indicated by  
630 yellow pseudobonds.

631

632 Figure 1-figure supplement 1:

633 **a: Calibration plots for the prediction of helix-helix interactions.**

634 Calibration plots for dataset #2 (left), dataset #3 (middle) and combined  
635 datasets #2 and #3 (right). The empirical fraction of true positives is plotted  
636 depending on the uncalibrated probability (raw score) obtained from our  
637 method. Points correspond to empirical averages over bins of 60 predictions  
638 (ordered by increasing uncalibrated probability). Lines correspond to maximum  
639 likelihood fits of the calibration plots using a transformed Bernoulli distribution  
640 with 4 parameters.

641 **b: Histogram of posterior probabilities for helix-helix interactions.**

642 Distribution of predicted calibrated posterior probabilities for YidC (TM2 – TM6)  
643 which contains 7 predicted helices, thus 21 possible helix-helix contacts. The  
644 histogram of predicted probabilities shows the specificity of the predictions:  
645 there is a large gap between 15% and 55% probability and most possible  
646 contacts have probability < 15%.

647 Figure2:

648 **Covariation-based model vs homology model**

649 Comparison of the *E. coli* YidC covariation-based model (a & b) to a homology  
650 model of *E. coli* YidC based on the crystal structure of BhYidC2 (3W06) (c & d).  
651 Predicted residue-residue pairs are indicated by yellow pseudobonds. Note that  
652 extracellular helix 1 (white) was not present in our multiple sequence alignment  
653 and is thus not included in the model.

654

655 Figure 2-figure supplement 1:

656 **Local deviations among YidC structures**

657 a: Smoothed C $\alpha$  distances between the two BhYidC2 crystal forms (3W06 vs  
658 3W07, red), between our model of *E. coli* YidC and 3W06 (green) and between  
659 our model and 3W07 (blue). b: Overall root mean square deviations (RMSD)  
660 between (the TM helices of) our model of *E. coli* YidC and the two BhYidC2  
661 crystal forms.

662

663 Figure 2-figure supplement 2:

664 **Top 50 scoring residue-residue pairs in covariation analysis**

665 Table showing the 50 residue-residue pairs with the highest covariation scores,  
666 and the distances between the C $\beta$  atoms in the final model of the 39 pairs that  
667 were used as constraints for model building. For comparison, the corresponding  
668 distances in 3W06 are also given. The eleven residue-residue pairs that were  
669 excluded for model building are in italics, with the reason for their exclusion  
670 indicated on the right.

671

672 Video 1:

673 **Conformational states of YidC**

674 Animation showing conformational differences in YidC starting from BhYidC2  
675 crystal form 1 (3W06), towards crystal form 2 (3W07) and ending with our  
676 covariation based YidC model. Views are from within the membrane (left) and  
677 from the cytoplasm (right). Note the movement of the HPD and the closing of the  
678 hydrophilic groove between TM3 (orange) and TM5 (green).

679

680 Figure 3:  
681 **Molecular dynamics simulation of the YidC model**  
682 **a:** Side view (left) and cytoplasmic view (right) of the stable YidC model after a  
683 500 ns MD simulation in a lipid bilayer composed of 3:1 POPE:POPG.  
684 **b:** Ribbon representation of the stable model according to inter-helix energy (in  
685 kcal/mol), blue: -7.5 to -1; white: -1 to -0.002; red:  $\geq -0.002$ . Residues that  
686 inactivate YidC upon mutagenesis are indicated by spheres.  
687 **c:** Ribbon representation of the stable model according to flexibility (in  $\text{\AA}^2$ ), blue:  
688 0.04 to 0.09; white: 0.09-1; red:  $\geq 1.0$ .  
689 **d:** *In vivo* complementation assay of YidC mutants T362A (TM2) and Y517A  
690 (TM6).  
691 **e:** Thickness of the cytoplasmic and periplasmic leaflet of the simulated bilayer  
692 after 500 ns, highlighting the membrane thinning effect in the vicinity of YidC.  
693 The membrane surface is defined by positions of polar head groups in the lipids,  
694 and thickness at a given point on the surface is taken to be the shortest distance  
695 between the head groups from opposite leaflets. The thickness values are  
696 averaged over the MD trajectory and presented as a contour plot on the  
697 membrane surface with a color-scale from red, indicating thicker region  
698 representing bulk bilayer lipids, to blue showing thinned regions close to YidC  
699 suggesting hydrophobic mismatch.  
700 **f:** Distribution of hydrophobic (red) and hydrophilic residues (blue) in YidC at  
701 various heights of the membrane, highlighting the hydrophilic environment in  
702 the center of YidC on the cytoplasmic side.

703

704 Figure 3-figure supplement 1:

#### 705 **Complementation of MD-based mutants**

706 *In vivo* complementation assay of YidC mutants identified as structurally  
707 important by MD simulations. Positions in YidC that were also identified by  
708 covariation analyses are indicated in the right column.

709

710 Figure 3-figure supplement 2:

#### 711 **Expression of MD-based mutants**

712 Western blot of whole FTL10 cells grown on arabinose or glucose, showing the  
713 stable expression of inactive YidC mutants that were identified by MD  
714 simulations.

715

716 Figure 4:

717 **Cryo-EM structure of RNC bound YidC and structural model of the active**  
718 **state**

719 **a:** Side view of the  $\sim 8 \text{ \AA}$  resolution cryo-EM based electron density of the  
720 RNC:YidC complex, with the small subunit depicted in yellow, the large subunit  
721 in grey, P-site tRNA and nascent chain in green, YidC in red and the detergent  
722 micelle in blue.

723 **b:** As in **a**, but sliced through the ribosomal exit tunnel.

724 **c:** Validation of the active state model by disulphide crosslinking. RNCs carrying  
725 the mutant  $F_{0c}^{(G23C)}$  were reconstituted with the indicated single cysteine YidC  
726 mutants, oxidized, applied to a linear sucrose gradient and harvested from the  
727 70S peak before SDS-PAGE and western blotting. Immunodetection was  
728 performed with antibodies raised against the HA-tag (located in the nascent  
729 chain inside the ribosomal exit tunnel) and anti-YidC antibodies. YidC, nascent  
730 chain-tRNA (NC-tRNA) and the expected crosslink product (NC-tRNA x YidC) are  
731 indicated.

732 **d-f:** Structural model of YidC during membrane protein insertion, viewed from  
733 two sides within the membrane (**d** & **e**) and from the cytoplasm (**f**). The  
734 detergent micelle was removed for clarity, the TM helix of  $F_{0c}$  is depicted in  
735 magenta, and the disulphide crosslink between YidC and  $F_{0c}$  with -SS-.

736

737 Figure 4-figure supplement 1:

738 **Comparison of the active states of YidC and SecY**

739 Left: Molecular model of YidC during co-translational translocation of  $F_{0c}$ , and  
740 the contour of active SecY. Middle: Composite model of active YidC with  $F_{0c}$   
741 replaced by the hydrophilic part of nascent FtsQ as found in active SecY. Right:  
742 Molecular model of SecY during co-translational translocation of FtsQ. For  
743 clarity, the N-terminal signal anchor of FtsQ was omitted.

744

745 Figure 4-figure supplement 2:

746 **Negative control for RNC-YidC crosslinking**

747 Crosslinking was performed with a cysteine-less F<sub>0c</sub> RNC as described in the  
748 legend to Figure 3c. A poorly reproducible unknown product is indicated with an  
749 asterisk.

750

751 Figure 5:

752 **Contacts between active YidC and the ribosome**

753 **a & b:** Close-up views from within the membrane region highlighting the contact  
754 between H59 of the ribosome and the 2/3 loop of YidC (a) and ribosomal protein  
755 uL23 and the 4/5 loop of YidC (b). Residues that inactivate YidC upon  
756 mutagenesis or deletion are indicated by magenta spheres.

757 **c:** *In vivo* complementation assay of YidC point mutants D488A, D488K, deletion  
758 mutant  $\Delta$ 487-489 and the double mutants Y370A/Y377A and Y370F/Y377F.

759 **d:** Periplasmic view of the active ribosome-bound YidC model, with the YidC  
760 contour outlined in red. The polypeptide exit tunnel is indicated with an asterisk.

761 **e:** Cartoon based comparison of active SecY (left) and active YidC (right) during  
762 membrane insertion of FtsQ and F<sub>0c</sub>, respectively. The ribosome is depicted in  
763 grey, the aqueous channel in SecY as well as the hydrophilic environment within  
764 YidC are shaded blue, hydrophobic TM domains of the substrates are depicted  
765 magenta, hydrophilic parts in green and the P1 domain by a dashed oval.

766

767 Figure 5-figure supplement 1:

768 **Complementation of ribosome interaction mutants**

769 *In vivo* complementation assay of YidC mutants involved in ribosome binding.

770

771 Figure 5-figure supplement 2:

772 **Expression of ribosome interaction mutants**

773 Western blot of whole FTL10 cells grown on arabinose or glucose, showing the  
774 stable expression of inactive YidC mutants that interact with the ribosome.

775









