

Supporting information for: Direct Simulation of Early Stage Sec-Facilitated Protein Translocation

Bin Zhang and Thomas F. Miller, III*

Division of Chemistry and Chemical Engineering, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125

E-mail: tfm@caltech.edu

Abbreviations: TM, transmembrane; LG, lateral gate; SP, signal peptide; MD, molecular dynamics

Materials and Methods

Simulation Protocols

Initialization. The crystal structure from Ref. 1 is used as the starting point for simulations of the SecA-SecYEG complex. To meet the size-constraints of the Anton hardware,^{2,3} only SecA residues within 15 Å of the translocon in the crystal structure are included in the simulations. Specifically, if any atom within a residue of SecA is less than 15 Å from any atom within a residue of the translocon (SecYEG), then all atoms of that SecA residue are included in the simulations; otherwise, all atoms of that SecA residue are deleted from the simulations. The partial atomic charges of the SecA residues are left unchanged following truncation; however, since the force-fields employed in this study exhibit integer values for the net charge of each amino acid residue, the simulation protocol does not introduce any net fractional charges into the simulation cell. In

*To whom correspondence should be addressed

all simulations, the non-hydrogen atoms of SecA are harmonically restrained to their positions in the reported crystal structure with a force constant of $k = 2.0 \text{ kcal/mol}/\text{\AA}^2$.

Residues 42-61 of the SecY protein, which are unresolved in the crystal structure, are constructed in a random loop configuration and then refined against a pseudo-energy function using the MODELLER protocol⁴ that consists of conjugate gradient minimization and molecular dynamics (MD) with simulated annealing;^{5,6} the employed pseudo-energy function includes interactions from the CHARMM22 force field⁷ and restraints based on the statistical distributions of known protein structures.⁸ The small number of crystallographically unresolved residues at the C-terminal end of the SecY protein (Res. 424-431), which are not expected to play a significant role in the nascent protein insertion process or in establishing the structural integrity of the translocon channel, are neglected.¹

The SecA-SecYEG complex is embedded in a membrane composed of 222 palmitoyloleoylphosphatidylcholine lipid molecules and surrounded by 25767 explicit water molecules; Na^+ and Cl^- counterions are included to achieve electroneutrality in the simulation cell at a salt concentration of approximately 50 mM. In orienting the SecA-SecYEG complex relative to the membrane, we obtain coordinates from the Orientations of Proteins in Membranes (OPM) database,⁹ in which the position of the protein relative to the lipid bilayer minimizes the transfer free energy from water to the membrane hydrophobic core.¹⁰ The system is described using orthorhombic periodic boundary conditions. The initial system contains 121107 atoms in a simulation cell of size $105 \text{ \AA} \times 105 \text{ \AA} \times 120 \text{ \AA}$; the membrane bilayer lies parallel to the x - y plane of the simulation cell.

All equilibration and nascent-protein growth simulations (described immediately below) are performed using the GROMACS molecular package, version 4.5.3.¹¹ Interactions are described using the CHARMM36 force field with the TIP3P water model.^{7,12} Long-range electrostatics are calculated using the particle-mesh Ewald (PME) technique,¹³ with the real space contribution to this potential cut off at 12 Å. Short-range van der Waals interactions are smoothly switched off over the distances from 10 to 12 Å. The GROMACS neighbor list for all short-ranged interactions is cut off at 12 Å and updated every 20 fs. All bond distances are constrained using the P-LINCS

algorithm,¹⁴ and a time-step of 2 fs is employed. Simulations are performed either at constant temperature and constant volume (i.e., the NVT ensemble) or at constant temperature and constant pressure (i.e., the NPT ensemble). Constant temperature simulations are fixed at 300 K using the Nose-Hoover thermostat;^{15,16} three separate thermostats, each with a coupling constant of 0.4 ps, are applied to the protein, lipids and water molecules, respectively. Constant pressure simulations are fixed at 1 bar using the Parrinello-Rahman barostat¹⁷ with a coupling constant of 4 ps. Pressure coupling is applied semi-isotropically, such that the *x* and *y* dimensions of the simulation cell remain equal to each other and deform independently of the *z* dimension of the simulation cell.

Equilibration. The initial system is equilibrated using the following four-step process. First, the energy of the system is minimized using the steepest-descent method to eliminate steric clashes that lead to forces in excess of 11.9 kcal/mol/Å; during this minimization, harmonic restraints ($k = 2.0$ kcal/mol/Å²) are applied to the lipid head groups and the heavy atoms of SecYEG. Second, the system is relaxed using a 1-ns simulation in the NVT ensemble with harmonic restraints ($k = 2.0$ kcal/mol/Å²) applied to the lipid head groups and the heavy atoms of SecYEG. Third, the system is relaxed using a 1-ns simulation in the NPT ensemble with harmonic restraints ($k = 2.0$ kcal/mol/Å²) applied to the heavy atoms of SecYEG. Finally, the system is relaxed using a 50-ns simulation in the NPT ensemble without restraints on the lipid or SecYEG.

Additional discussion and testing of the equilibration process is provided in *Supporting Information: Robustness of the insertion trajectory initialization.*

Nascent-Protein Growth. Following equilibration of the SecA-SecYEG complex, we introduce a nascent protein composed of $n + 4$ amino acid residues; the four N-terminal residues are initially positioned in a β-strand configuration and aligned with the axis of the translocon channel, and the center-of-mass positions of the remaining n residues are initially placed at an “insertion point” at the cytosolic mouth of the translocon channel. As is described in the main text, we consider two insertion points. The insertion points are defined relative to the positions of the atoms of SecA, which are restrained in absolute space to the geometry of SecA-SecYEG crystal structure, as described under *Initialization*. Coordinates in absolute space associated with the geometry of the

SecA-SecYEG crystal structure are given in Table S1. In terms of this unique coordinate system, the position of insertion point IP2 is $\{x - a, y - b, z - c\}$, where $\{x, y, z\}$ is the Cartesian center-of-mass for residues 780-785 of the two-helix-finger domain of SecA, and where $a = -6 \text{ \AA}$, $b = 3 \text{ \AA}$, and $c = -1 \text{ \AA}$ are chosen to avoid steric clashes; the coordinates for IP1 are simply shifted $+10 \text{ \AA}$ from IP2 along the z -axis.

Each of the n residues on the C-terminal end of the nascent protein exists in either an off-state, in which non-bonding interactions between each residue and the rest of the system are excluded, or an on-state, in which all interactions are included. Residues in the off-state are tethered to the insertion point via harmonic restraints with $k = 23.8 \text{ kcal/mol}/\text{\AA}^2$. Upon sequentially switching each residue from the off-state to the on-state, the simulation cell is subjected to a partial minimization to avoid large steric clashes associated with the newly introduced nascent protein residue; this minimization is only performed with respect to forces that exceed a magnitude of $23.8 \text{ kcal/mol}/\text{\AA}$, such that only atoms in the immediate vicinity of the newly introduce amino-acid residue are primarily affected. After switching each residue from the off-state to the on-state, it is pulled from the insertion point toward the center-of-mass of the translocon pore residues for a period of 5 ns to create space for the next amino-acid residue in the nascent-protein sequence; this is achieved by harmonically tethering ($k = 2.38 \text{ kcal/mol}/\text{\AA}^2$) the center-of-mass of the nascent-protein residue to a virtual bead that moves with constant velocity $v = 1 \text{ \AA/ns}$. Since these simulations are performed in the NPT ensemble, the simulation cell volume appropriately relaxes upon inclusion of the additional nascent-protein residues. Since the nascent-protein growth involves only the introduction of uncharged amino-acid residues, the total simulation cell remains neutral.

Nascent-Protein Evolution. At nascent-protein lengths corresponding to 15, 30 and 45 amino acid residues, the simulation cell is ported to the Anton computing system for microsecond-timescale relaxation. In these simulations, short-ranged Van der Waals and electrostatic interactions are cutoff at 9.48 \AA , and long-ranged electrostatic contributions are included using the k-space Gaussian Split Ewald method¹⁸ with a cubic $(64 \times 64 \times 64)$ k-point grid, an electrostatic screening parameter of $\sigma = 2.44 \text{ \AA}$, and a Gaussian charge-spreading width of $\sigma_s = 1.72 \text{ \AA}$. Bond-lengths in-

volving hydrogen atoms are constrained using the M-SHAKE algorithm.¹⁹ We employ the RESPA numerical integration scheme²⁰ with a timestep of 2 fs; short-ranged interactions are updated every timestep, and long-range electrostatic interactions are updated every 6 fs. The Berendsen coupling scheme²¹ is applied to keep the simulation at a temperature of 300 K and a pressure of 1 bar; the thermostat employs a coupling timescale of $\tau = 1.0$ ps, and the barostat employs a semi-isotropic coupling timescale of $\tau = 2.0$ ps.

Channel axis definition

To analyze configuration changes during nascent-protein insertion, we define a one-dimensional coordinate associated with the translocon channel axis. For the initial configuration of each insertion trajectory, the channel axis coordinate is defined as the z -component of the Cartesian coordinate system for the simulation cell with origin positioned at the center-of-mass of the SecA-SecYEG complex. To avoid artifacts due to fluctuations or drift in the translocon position during the long insertion trajectories, atomic positions in each subsequent configuration of the trajectories are aligned to those of the initial configuration, using a protocol that minimizes mean-square displacement in the α -carbons of SecYEG.²²

Translocon lateral gate width profile

The lateral gate (LG) width profile illustrated in Figure 3A in the main text is calculated as follows. The channel-axis coordinate is uniformly discretized at a resolution of $\Delta\eta = 2\text{\AA}$, and the simulation cell is thus divided into corresponding parallel slabs, $\{\eta_j\}$. For each slab, the width of the LG opening is determined by considering the translocon α -carbon atoms that lie within the slab and that correspond to one of two particular subsets of the translocon atoms. The first subset, which is shown in green in Figure 3A in the main text and which corresponds to one half of the translocon LG, includes residues in transmembrane (TM) helices TM7-8 (V270-I335 in the SecY protein). The second subset, which is shown in yellow in Figure 3A in the main text and which corresponds to the other half of the LG, includes residues M80-S142 in the SecY protein. The width of the LG

opening for each slab, w_j , is defined as the minimum distance from any α -carbon in the first subset to any in the second subset. We note that a similar definition is used to calculate the LG surface area in Ref. 23.

For each insertion trajectory, we obtain the LG width profile as a function of both simulation time and the channel axis coordinate. In making Figures 3B and 3C in the main text and Figure S2, we smooth the LG width profile in both dimensions, with raw input data corresponding to the channel width profile evaluated at time intervals of 240 ps. The smoothed output, obtained using a modified ridge estimator with dimensionless smoothing parameter $s = 1$, is provided on a 50×50 grid, with spacing $\Delta t = 0.076\mu\text{s}$ ($\Delta t = 0.090\mu\text{s}$ for Figure S2) and $\Delta\eta = 0.7\text{\AA}$. As shown in Figure S7, this smoothing does not affect any of the trends discussed in connection with Figure 3.

Hydrophobic contact area

To quantify the extent of the hydrophobic contact between the nascent-protein signal peptide (SP) and the lipid molecules in Figures 5C and 5D in the main text and Figures S4C and S4D, we determine the interfacial contact area using the INTERVOR tool²⁴ in the Visual Molecular Dynamics (VMD) program.²⁵ The area is calculated via Delaunay triangulation of the Voronoi surface that separates two groups of atoms in the system. Atoms in the nascent-protein SP are treated as one group, and atoms in the lipid molecules that are within a cutoff distance $d_0 = 10\text{\AA}$ of the transloccon LG helices (SecY residues 80-100 and 270-290) are treated as the other group. Increasing the cutoff distance to $d_0 = 20\text{\AA}$ results in no significant differences from the data presented in either Figure 5 or Figure S4.

Homology modeling

To enable comparison between the residues of the *Thermotoga maritima* SecY protein studied in this work and the residues of the *S. cerevisiae* Sec61p upon which previous experimental mutagenesis studies have been performed,²⁶ we determine the structure for Sec61p using homology

modeling. All homology modeling is performed using the MODELLER program.⁴

We first construct 50 initial structural models for Sec61p from the alignment of its sequence with the *Methanococcus jannaschii* SecY protein (PDB: 1RHZ);²⁷ the Cartesian coordinates for the atoms in these initial models are locally randomized with respect to a common reference structure. The structure for each model is then optimized with respect to (*i*) homology-derived restraints from the alignment, (*ii*) molecular mechanics force field from CHARMM22,⁷ and (*iii*) statistically derived potentials from a representative set of known protein structures; the final homology-determined structure then corresponds to the optimized model with the lowest discrete optimized protein energy (DOPE) score.²⁸ A comparison between the homology-determined structure for Sec61p and the *Thermotoga maritima* SecY crystal structure is shown in Figure S6A. The *Thermotoga maritima* SecY residues that are homologous to residues E382, E106, and E460 in the *S. cerevisiae* Sec61p are then identified as the closest negatively charged residues in the aligned structures. We note that by alternatively performing this analysis with the *Thermotoga maritima* SecY protein (PDB: 3DIN) as the structural template for Sec61p, the determined homologous residues are unchanged, although the spatial distance between residue E330 in SecY and residue E382 in Sec61p increases.

To further illustrate the spatial distribution of the Sec61p charged residues relative to the bound ribosome, we build another homology-determined structure from the cryo-electron-microscopy-derived structure of SecY in association with the ribosome (PDB: 3KC4, 3KCR).²⁹ The same method explained above is employed, and the resulting structure is shown in Figure S6B.

Robustness of the insertion trajectory initialization

Here, we examine the robustness of the equilibration and insertion protocol employed for the insertion trajectories in the main text. We consider potential biases due to the relatively low resolution (4.5 Å) of the experimental crystal structure that is used for a starting structure in these simulations, as well as potential biases due to the use of harmonic restraints to stabilize the truncated

SecA protein in the reported insertion trajectories. Several important points suggest that this issue does not impact the reliability of the results in the main text.

First, we note that although the experimental resolution of the electron density for the SecA-SecYEG complex is relatively low (4.5 Å),¹ the crystal structure was solved with the assistance of higher resolution crystal structures for both the translocon (Ref. 27, 3.2 Å resolution) and SecA (Ref. 30, 2.2 Å resolution). The electron density map for the SecA-SecYEG complex provides sufficient quality for the unique assignment of all of the SecYEG TM helices and each of the SecA domains.¹ Incorporation of conserved interactions from the higher-resolution structures, which exhibit strong sequence homology with the components of the SecA-SecYEG complex,^{27,30} improves the credibility of the resulting structure for the complex.

Second, equilibration of the system in the MD simulations reduces bias associated with the choice of the initial structure. As described in *Supporting Information: Simulation Protocols*, the initial SecA-SecYEG crystal structure was equilibrated for 50 ns prior to performing the μ s-timescale nascent protein insertion trajectories. During this equilibration, the LG relaxed from a relatively open conformation associated with the SecA-SecYEG crystal structure to a more closed conformation like that of the archaeal crystal structure of the translocon.²⁷ To illustrate this conformational change, the surface area associated with the opening of the translocon LG for the initialization trajectory employed in the main text (T_0) is plotted as a function of simulation time in Figure S8. The LG surface area at each configuration is obtained via quadrature of the LG width profile along the translocon channel axis; the endpoints of this integration are defined by the center-of-mass of the phosphorus atoms in each lipid layer. The LG width profile and channel axis coordinate are defined in *Supporting Information: Materials and Methods*.

To confirm the robustness of the equilibrated starting configurations for the nascent protein insertion trajectories, additional equilibration trajectories are reported in Figure S8. Each of the additional three trajectories is initialized from the SecA-SecYEG crystal structure configuration but with different initial velocities drawn from Maxwell-Boltzmann distribution at a temperature of $T = 300$ K. Unlike the equilibration trajectory employed in the main text (trajectory T_0), the

additional trajectories were obtained using the OPLS force field and without truncation of SecA; however, the results are the same. In each case, the equilibration trajectories eventually relax towards the more closed conformation of the translocon LG.

It is clear that for trajectory sT₂ in Figure S8, the timescale for the closing of the LG is slower than that observed in trajectories T₀ and sT₁. The origin of this slow timescale is the intercalation of lipid molecules between the LG helices, which hinders LG closing. To overcome this slow timescale, we removed the three lipid molecules that were in closest proximity to the LG helices in trajectories sT₁, sT₂, and sT₃ at the simulation time of 80 ns; as is seen in Figure S8, elimination of these lipid molecules enables the “hung” trajectory sT₂ to then relax to more closed LG configurations. In no case was it found that lipid molecules intercalate between the LG helices after the closed LG conformation was reached. The results in Figure S8 thus suggest that equilibration of the system relaxes the initial bias of the crystal structure and consistently leads to partial closing of the translocon LG.

Third, we find that the harmonic restraints that are applied to SecA in the main text, which are necessary for the μ s-timescale stability of the truncated SecA structure, do not appear to bias the conformational distribution of the translocon. In Figure S9A, trajectories sT₄-sT₆ are performed exactly as trajectories sT₁-sT₃, except that trajectories sT₄-sT₆ do not include harmonic restraints on the heavy atoms of SecA, whereas sT₁-sT₃ do include these restraints (as in the main text). As before, to avoid slow relaxation timescales associated with the intercalation of molecules in the LG, we delete the three nearest lipid molecules to the LG helices after 80 ns of simulation time. The trajectories performed without SecA restraints are qualitatively unchanged from those that employ restraints; in all cases, the trajectories relax to the partially closed configurations that are consistent with the equilibration trajectory employed in the main text. In Figure S9B, we show the root mean squared displacement (RMSD) of the SecY protein in these six trajectories; again, no major effect associated with the SecA harmonic restraints is found.

Fourth, the translocon LG conformational changes that are emphasized in the main text are large and qualitatively distinct from biases associated with the SecA harmonic restraints. Fig-

ure S9C shows the LG surface area as a function of time in the nascent protein insertion trajectory T₁. This trajectory shows a pronounced opening of the LG that accompanies insertion of the nascent protein into the translocon channel. This trend is in marked contrast to the LG closing and conformational relaxation that is observed in the trajectories that are initialized from the SecA-SecYEG crystal structure (Figure S9A), and it is distinct from any effects associated with the transient intercalation of lipid molecules between the LG helices in the initialized system. Furthermore, the conformational opening of the translocon LG that is observed in Figure S9C is far more pronounced than any observed bias in Figure S9A or Figure S9B that is associated with the use of harmonic restraints on the SecA residues.

In summary, concerns over the reliability of the starting configuration of the simulations are mitigated by (*i*) the fact that the solution of the 4.5 Å resolution SecA-SecYEG crystal structure was obtained with input from much higher-resolution structures of the component proteins, and (*ii*) the observation in Figure S8 that equilibration of the SecA-SecYEG complex relaxes the initial conformation of the crystal structure and leads to more closed conformations of the translocon LG. Furthermore, concerns about the use of harmonic restraints to stabilize the SecA structure are mitigated by (*iii*) the observation in Figure S9A and Figure S9B that harmonic restraints have little effect on the conformation of the translocon LG, and (*iv*) the fact that results and conclusions that are emphasized in the main text involve trends and conformational changes (such as those in Figure S9C) that are far more pronounced than any observed biases due to the initial crystal structure or the use of harmonic restraints on the SecA residues.

Insertion trajectories with different periods of growth/evolution

For comparison with trajectories T₁-T₄ that are discussed in the main text, we performed two additional insertion simulations in which each nascent-protein growth period involves the addition of only two additional nascent-protein residues at a pace of one residue per five nanoseconds and in which the intervening nascent-protein evolution periods span 100 ns. The details for these two

additional insertion trajectories (T_5 and T_6) are presented in Table S2. We note that the protocol employed for these additional trajectories allows equilibration after smaller periods of growth than the trajectories discussed in the main text; however, since the evolution periods in trajectories T_5 and T_6 span only 100 ns, the net pace of nascent-protein insertion in trajectories T_5 and T_6 is faster than the pace of insertion for trajectories T_1 - T_4 . As is seen in Figure S10, this faster pace of insertion in the additional trajectories leads to simulation results that do not exhibit the anticipated features of protein translocation or membrane integration. The figure compares trajectories T_5 and T_6 with trajectories T_1 and T_2 ; all four trajectories share the same insertion point, trajectories T_1 and T_5 model nascent proteins with the same sequence, and trajectories T_2 and T_6 model nascent proteins with the same sequence. As is extensively discussed in the main text, the slower insertion protocol employed in trajectories T_1 and T_2 leads to docking of the nascent-protein SP at the translocon LG and associated conformational changes in the LG. However, the more rapidly inserted trajectories T_5 and T_6 do not exhibit these mechanistic features; instead, the nascent protein becomes jammed at the cytosolic mouth of the translocon and exhibits little conformational sampling on the timescale of the simulations.

Water occupancy and dynamics at the translocon cytosolic mouth

Here, we examine the occupancy and exchange dynamics of water molecules at the cytosolic mouth of the translocon. A water molecule is defined to occupy the cytosolic mouth if its oxygen atoms lies within 18 Å of the channel axis and with -20 and 20 Å along the channel axis. Note that a similar definition was used to identify \mathcal{N} , the number of nascent-protein residues that partition from the translocon channel into the membrane interior. A snapshot of the occupying water molecules is show in Figure S11A.

Figure S11B presents the number of water molecules in the cytosolic mouth as a function of simulation time for trajectories T_1 and T_2 . In both trajectories, a large number of water molecules occupies the cytosolic mouth throughout the simulations. For trajectory T_1 , the number of water

molecules remains relatively unchanged; for T₂, this number gradually decreases during nascent-protein insertion. The presence of water molecules in the cytosolic mouth is consistent with the fact that the translocon interior is hydrophilic. The occupancy number trends for the two trajectories can be correlated with the relative degrees of membrane integration exhibited by the nascent protein in the two trajectories. In T₁, the nascent protein is hydrophobic and exhibits a significant amount of integration (Figure 4 in the main text); it creates little competition for the space that is occupied by the water molecules. On the other hand, the hydrophilic nascent protein remains confined within the translocon channel; water molecules are thus driven out of the cytosolic mouth to make room for the nascent protein.

To further characterize the exchange dynamics of water molecules between the cytosolic mouth and bulk regions of the simulation cell, Figure S11C presents the time autocorrelation function (TACF) for the occupancy of individual water molecules. This TACF reports the propensity for a given water molecule to occupy the cytosolic mouth at time t , given that it is in occupancy at the earlier time 0; the TACF was obtained by time-averaging along the insertion trajectories. As is seen in the figure, the TACF decays on the timescale of 20 ns, which is significantly shorter than the length of the simulated insertion trajectories. This result suggests that the water dynamics is slaved to the motion of the nascent protein, adapting relatively quickly to its conformational changes.

Finally, we explore the mechanism by which water molecules exchange between the cytosolic mouth and bulk regions of the simulation cell. In particular, we investigate whether water exchange is expedited by the truncation of the SecA protein in the simulations presented in the main text. Figure S12 presents sterically accessible regions for water molecules in the interfacial region of SecA and the translocon, which is obtained using the Molaxis program.³¹ The regions are colored according to the radius, r , of the largest spherical particle that can be locally accommodated, with red indicating $r < 1.4\text{\AA}$, green indicating $1.4 \leq r < 2.7\text{\AA}$, and blue indicating $r \geq 2.7\text{\AA}$. Note that only a single continuous channel provides a pathway for exchange of water molecules between the cytosolic mouth region and the surrounding cytosolic bulk region; all other channels encounter prohibitively narrow openings that are colored by red. The primary exchange channel

corresponds to a physical opening that is observed in the untruncated structure SecA-SecYEG crystal structure,¹ suggesting that the truncation protocol employed in these simulations does not bias the water exchange dynamics.

References

- (1) Zimmer, J.; Nam, Y.; Rapoport, T. A. *Nature* **2008**, *455*, 936–943.
- (2) Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Ierardi, D. J.; Kolossvary, I.; Klepeis, J. L.; Layman, T.; McLeavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C. *SIGARCH Comput. Archit. News* **2007**, *35*, 1–12.
- (3) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341–346.
- (4) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A. *Curr Protoc Bioinformatics* **2006**, Chapter 5, Unit 5.6.
- (5) Sali, A.; Blundell, T. L. *J Mol Biol* **1993**, *234*, 779–815.
- (6) Fiser, A.; Sali, A. *Bioinformatics* **2003**, *19*, 2500–2501.
- (7) MacKerell, A. D.; Bashford, D.; Bellott,; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J Phys Chem B* **1998**, *102*, 3586–3616.
- (8) Fiser, A.; Do, R. K.; Sali, A. *Protein Sci* **2000**, *9*, 1753–1773.

- (9) Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. *Bioinformatics* **2006**, *22*, 623–625.
- (10) Lomize, A. L.; Pogozheva, I. D.; Lomize, M. A.; Mosberg, H. I. *Protein Sci* **2006**, *15*, 1318–1333.
- (11) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J Chem Theory Comput* **2008**, *4*, 435–447.
- (12) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, J., A. D.; Pastor, R. W. *J Phys Chem B* **2010**, *114*, 7830–7843.
- (13) Darden, T.; York, D.; Pedersen, L. *J Chem Phys* **1993**, *98*, 10089–10092.
- (14) Hessa, T.; Meindl-Beinker, N. M.; Bernsel, A.; Kim, H.; Sato, Y.; Lerch-Bader, M.; Nilsson, I.; White, S. H.; von Heijne, G. *Nature* **2007**, *450*, 1026–1030.
- (15) Nose, S. *J Chem Phys* **1984**, *81*, 511–519.
- (16) Hoover, W. G. *Phys Rev A* **1985**, *31*, 1695.
- (17) Parrinello, M.; Rahman, A. *J Appl Phys* **1981**, *52*, 7182–7190.
- (18) Shan, Y.; Klepeis, J. L.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *J Chem Phys* **2005**, *122*, 54101–54113.
- (19) Krautler, V.; Van Gunsteren, W. F.; Hunenberger, P. H. *J Comput Chem* **2001**, *22*, 501–508.
- (20) Tuckerman, M.; Berne, B. J.; Martyna, G. J. *J Chem Phys* **1992**, *97*, 1990–2001.
- (21) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J Chem Phys* **1984**, *81*, 3684–3690.
- (22) Kabsch, W. *Acta Crystallogr Sect A: Found Crystallogr* **1978**, *34*, 827–828.

- (23) Zhang, B.; Miller, T. F. *Proc Natl Acad Sci USA* **2010**, *107*, 5399–5404.
- (24) Loriot, S.; Cazals, F. *Bioinformatics* **2010**, *26*, 964–965.
- (25) Humphrey, W.; Dalke, A.; Schulten, K. *J Mol Graph* **1996**, *14*, 33–38, 27–38.
- (26) Goder, V.; Junne, T.; Spiess, M. *Mol Biol Cell* **2004**, *15*, 1470–1478.
- (27) Berg, B. v. d.; Clemons, W. M.; Collinson, I.; Modis, Y.; Hartmann, E.; Harrison, S. C.; Rapoport, T. A. *Nature* **2004**, *427*, 36–44.
- (28) Shen, M. Y.; Sali, A. *Protein Sci* **2006**, *15*, 2507–2524.
- (29) Gumbart, J.; Trabuco, L. G.; Schreiner, E.; Villa, E.; Schulten, K. *Structure* **2009**, *17*, 1453–1464.
- (30) Osborne, A. R.; Clemons, J., W. M.; Rapoport, T. A. *Proc Natl Acad Sci USA* **2004**, *101*, 10937–10942.
- (31) Yaffe, E.; Fishelovitch, D.; Wolfson, H. J.; Halperin, D.; Nussinov, R. *Proteins* **2008**, *73*, 72–86.

Table S1: Cartesian positions for the α -carbon atoms of SecA residues 780-785.

Residue ID	Position (\AA)		
	x	y	z
780	-2.073	-9.178	22.870
781	-3.665	-11.383	20.201
782	-6.463	-10.542	17.788
783	-7.417	-6.878	17.970
784	-4.841	-4.080	18.208
785	-2.543	-3.605	21.212

Table S2: Summary of the additional insertion trajectories.

Trajectory	IP	Mature Domain	Growth Cycles	Total Length
T ₅	1	L ₃₀	15	2.5 μs
T ₆	1	Q ₃₀	15	2.5 μs

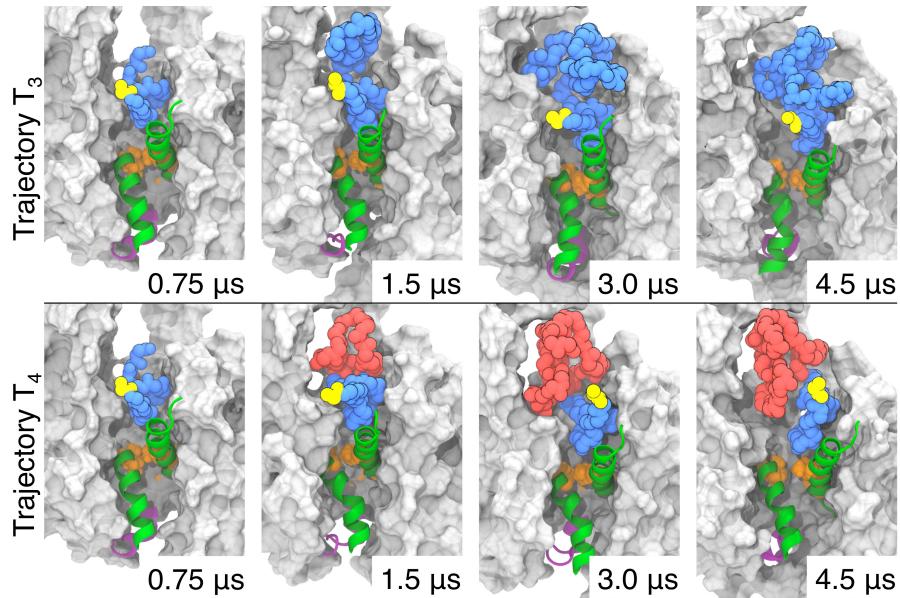


Figure S1: Structural features of the nascent protein and translocon at various times along the insertion trajectories T_3 and T_4 . This figure is comparable to Figure 2 in the main text and employs the same representation. As in Figure 2, the trajectories exhibit localization of the SP residues in the LG region. Both trajectories shown here exhibit loop configurations of the nascent protein SP, with the N-terminus exposed to lipid head-groups and with some hydrophobic segments of the SP buried within the translocon channel. The trajectories in Figure 2 exhibit a different nascent-protein conformation, with the SP N-terminus buried in the translocon interior and with the hydrophobic segments of the SP exposed to the membrane.

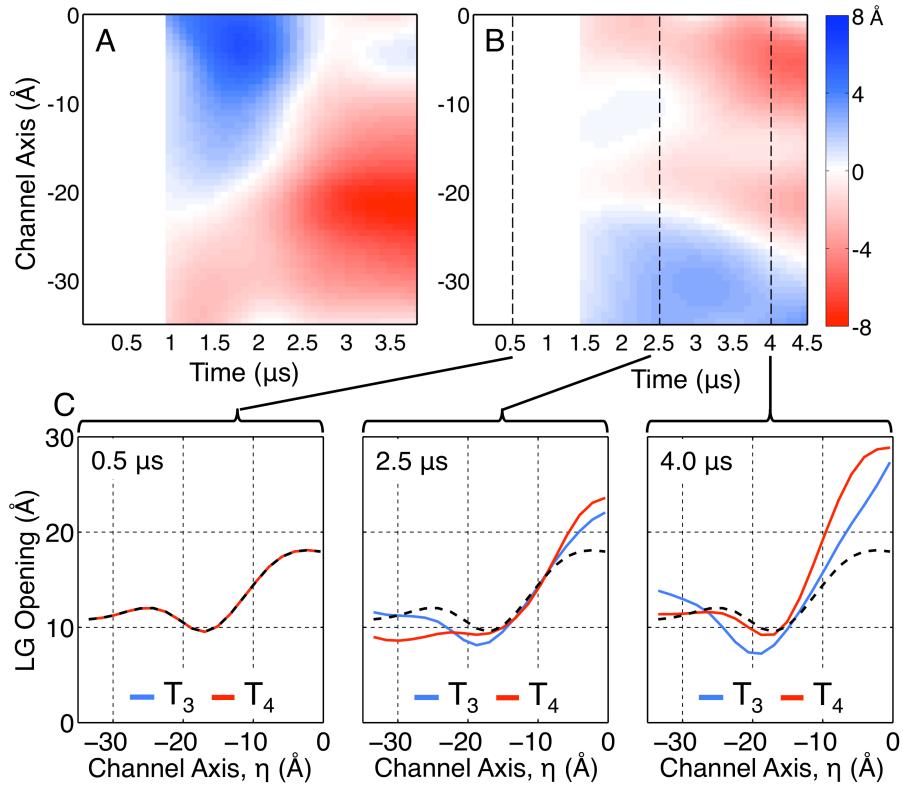


Figure S2: The LG width profiles for trajectories T_3 and T_4 . (A) For reference, the difference of the LG width profiles between trajectories T_1 and T_2 , which is reproduced from Figure 3C in the main text. (B) The difference of the LG width profiles between trajectories T_3 and T_4 . (C) The LG width profiles for trajectories T_3 (blue) and T_4 (red) at various times. The data at time $0.5 \mu\text{s}$ is repeated in the dashed black curve. The translocon LG undergoes similar conformational changes in both trajectories T_3 and T_4 ; this is consistent with the observation from Figure S1 that the nascent protein adopts a loop configuration in both trajectories, such that both trajectories exhibit similar interactions between the nascent protein and the translocon LG.

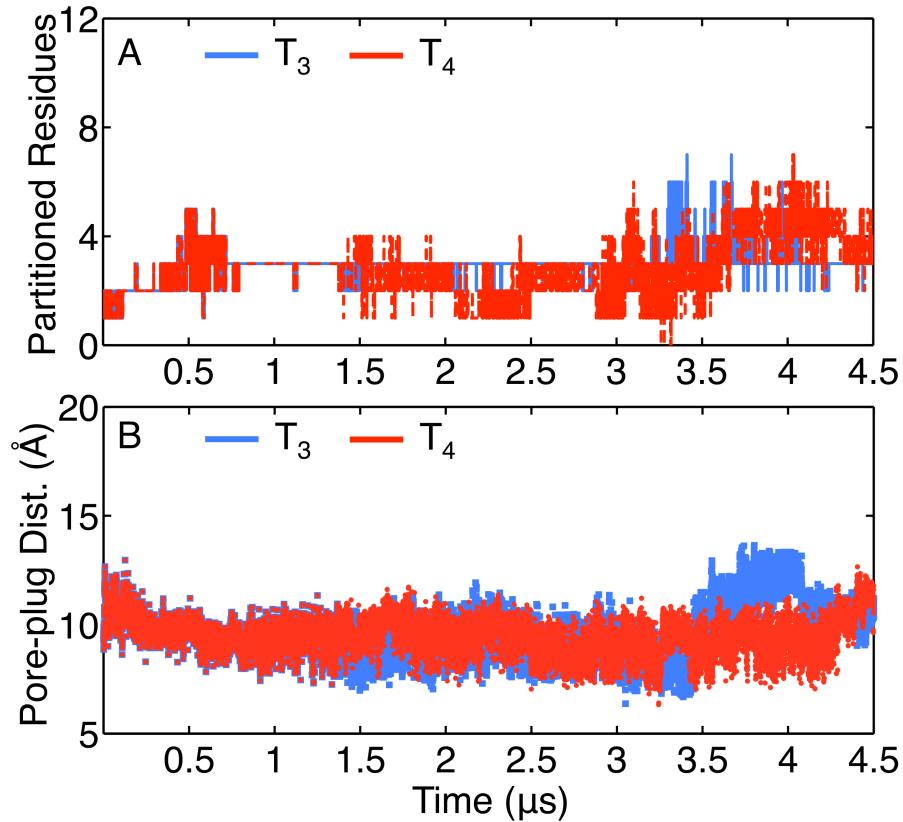


Figure S3: Conformational dynamics of the nascent protein and the translocon plug moiety during insertion trajectories T_3 and T_4 . This figure is comparable to the results for trajectories T_1 and T_2 in Figure 4 of the main text. (A) The time-evolution of the number of membrane-integrated residues, \mathcal{N} , for the insertion trajectories T_3 and T_4 . Neither of the two trajectories presented here exhibit the extensive degree of membrane integration that is observed for trajectory T_1 in Figure 4. The observed values of $\mathcal{N} \approx 3$ in the current figure arise from the close contact of the SP N-terminal residues with the phosphate lipid head groups. This result can again be understood as a consequence of the loop configuration that is assumed by the nascent-protein SP in both trajectories T_3 and T_4 ; since this configuration buries the protein mature domain within the translocon channel interior, little membrane integration is possible. (B) The time-evolution of the pore-plug distance for trajectories T_3 and T_4 . Little displacement of the plug moiety is observed during the course of these simulations.

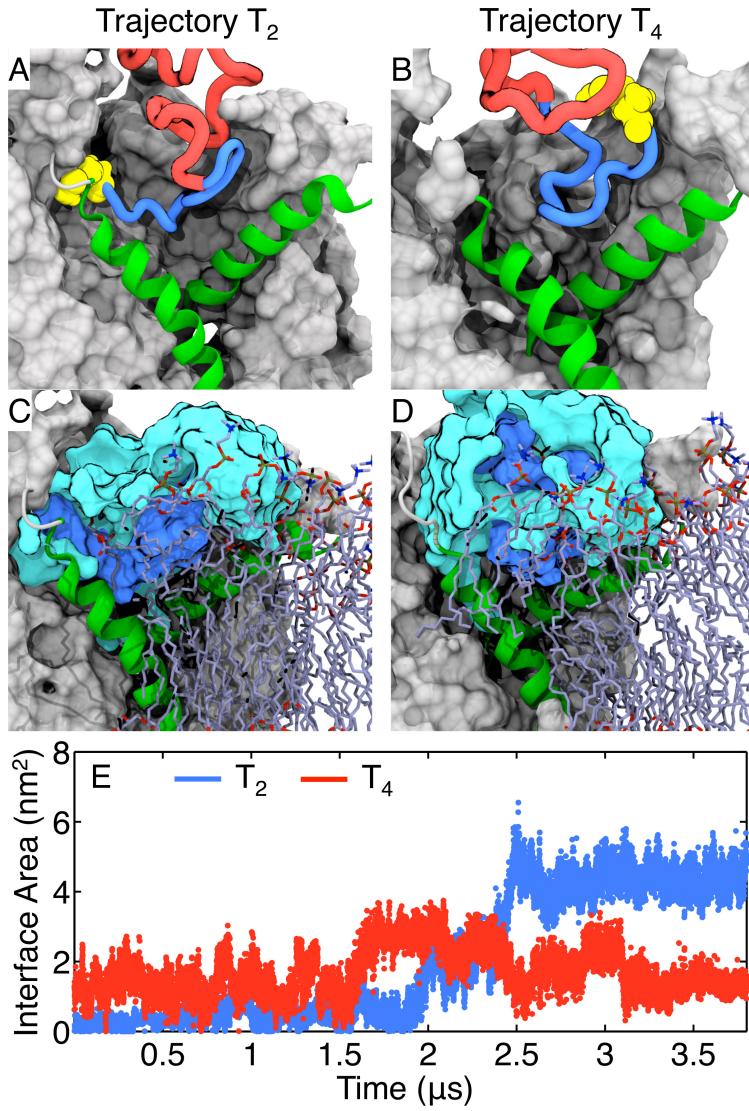


Figure S4: Conformation of the nascent-protein SP and its solvation environment during insertion trajectories T_2 and T_4 . This figure is directly comparable to the results for trajectories T_1 and T_3 in Figure 5 of the main text; the same representations and color schemes are employed. (A, B) The configuration of the nascent-protein SP is shown after 3.5 μ s of simulation time in the two trajectories. (C, D) The corresponding solvation environment for the nascent-protein SP in these two snapshots of the system. In parts A and C, it is seen that the SP adopts a configuration in which the N-terminus is buried inside the translocon, whereas in parts B and D, the SP adopts a loop configuration with its N-terminus exposed to the phosphate lipid head groups. (E) The hydrophobic contact area between the SP and the surrounding lipid molecules is plotted as a function of time for trajectories T_2 (blue) and T_4 (red). Similar to the results presented in Figure 5, the SP configuration with the N-terminus buried in the channel interior (A, C) experiences more extensive hydrophobic contact with the lipid bilayer than the SP configuration with the N-terminus exposed to the lipid head groups (B, D).

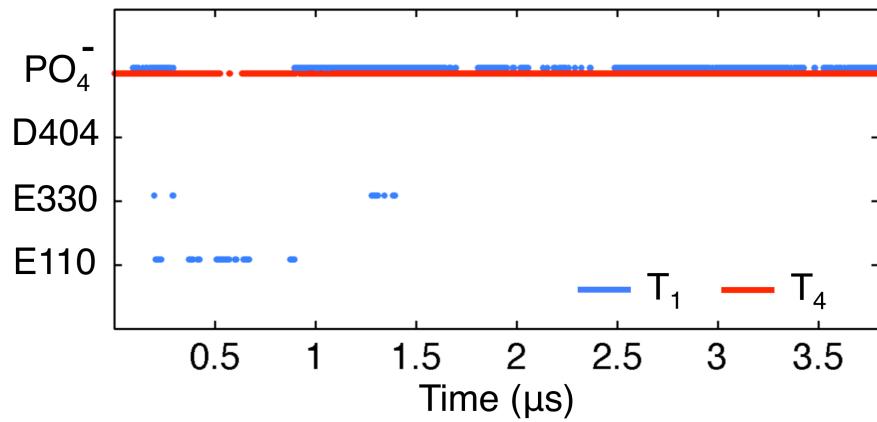


Figure S5: The time-evolution of the salt bridges along trajectories T_1 and T_4 . This figure is directly comparable to the results for trajectories T_2 and T_3 in Figure 6 of the main text. As seen in Figure 6, the current figure indicates that salt-bridge contacts involving the N-terminus of the nascent protein form almost immediately and persist throughout the duration of the insertion simulations.

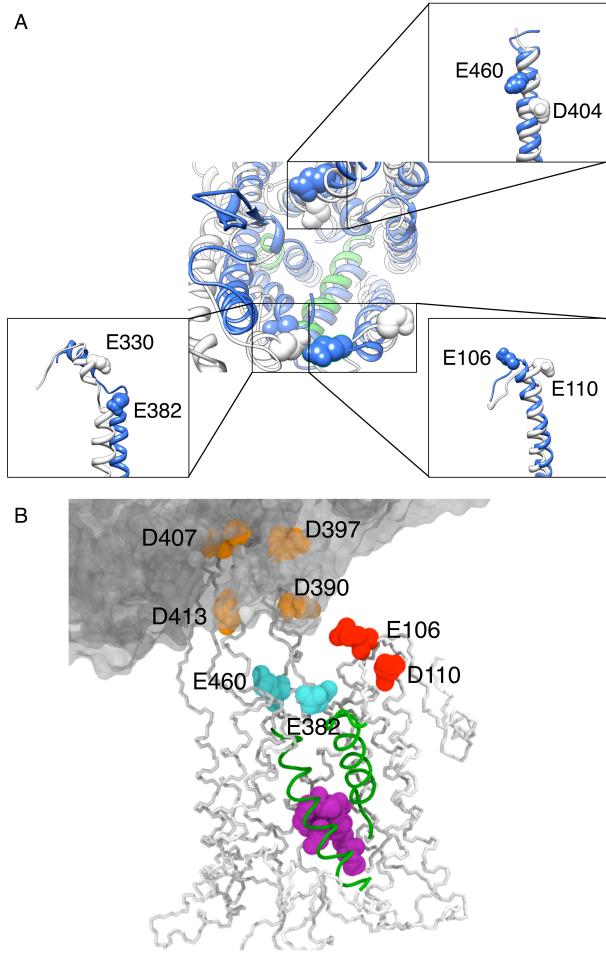


Figure S6: Homology-determined structures for the *S. cerevisiae* Sec61p protein, which has been characterized in experimental mutagenesis studies.²⁶ The procedure for constructing these structures is described in the *Homology modeling* section. (A) Comparison of the *Thermotoga maritima* SecY crystal structure (white and green) and the homology-determined *S. cerevisiae* Sec61p structure (blue). The two structures are aligned via minimization of the root mean squared displacement of the α -carbon atoms between residues 80-140 of the two molecules. The negatively charged residues of *Thermotoga maritima* SecY that we consider in this paper (shown in white) are in close proximity to the negatively charged residues that are studied via mutagenesis in Ref. 26 (blue). (B) The homology-determined structure for Sec61p in complex with the ribosome (see text for details). The ribosome is rendered in a gray transparent surface, and the translocon is drawn in white with the LG helices in green and the plug moiety in violet. Charged residues on the translocon are labeled and shown in space-filling representation. From the results of Ref. 26, it follows that mutation of residues in orange decreases the integration of protein with either Type II or Type III orientation; mutation of residues in red enhances the integration of protein with Type II orientation and suppresses the integration with Type III orientation; and mutation of residues in cyan suppresses the integration of protein with Type II orientation and enhances the integration with Type III orientation. This figure, along with the presented insertion simulations, emphasizes that the position of charged residues impacts the regulation of integral membrane protein topogenesis.

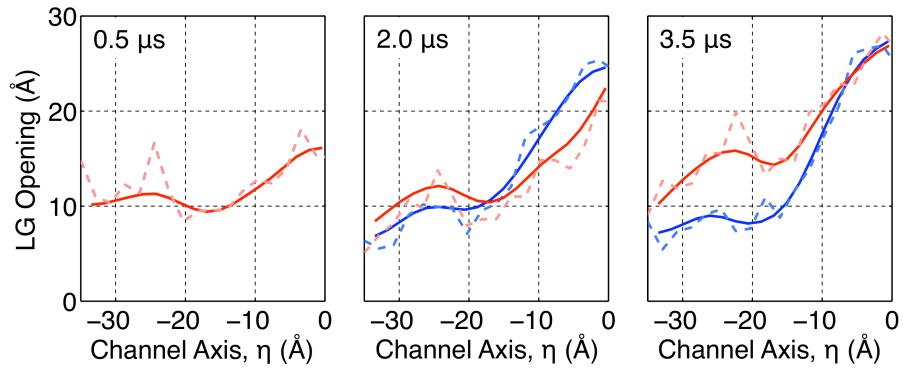


Figure S7: Comparison of the raw (dashed) and smoothed (solid) LG width profiles. See the *LG width profile* section for details. The data presented here correspond to the results for trajectory T₁ that appear in Figure 3B of the main text.

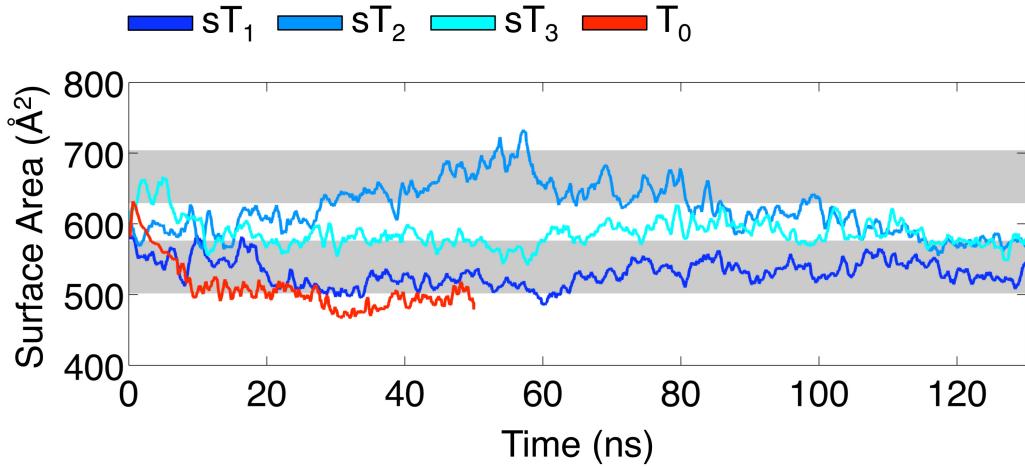


Figure S8: Comparison of additional equilibration trajectories from the SecA-SecYEG crystal structure (sT₁-sT₃) with equilibration trajectory that is employed in the main text (T₀). The surface area of the translocon LG is plotted as a function of simulation time. In all cases, the trajectories relax to partially closed conformations for the translocon LG. The shaded gray areas indicate values for the surface area that are consistent with the open and partially closed conformations for the translocon LG. See text for details.

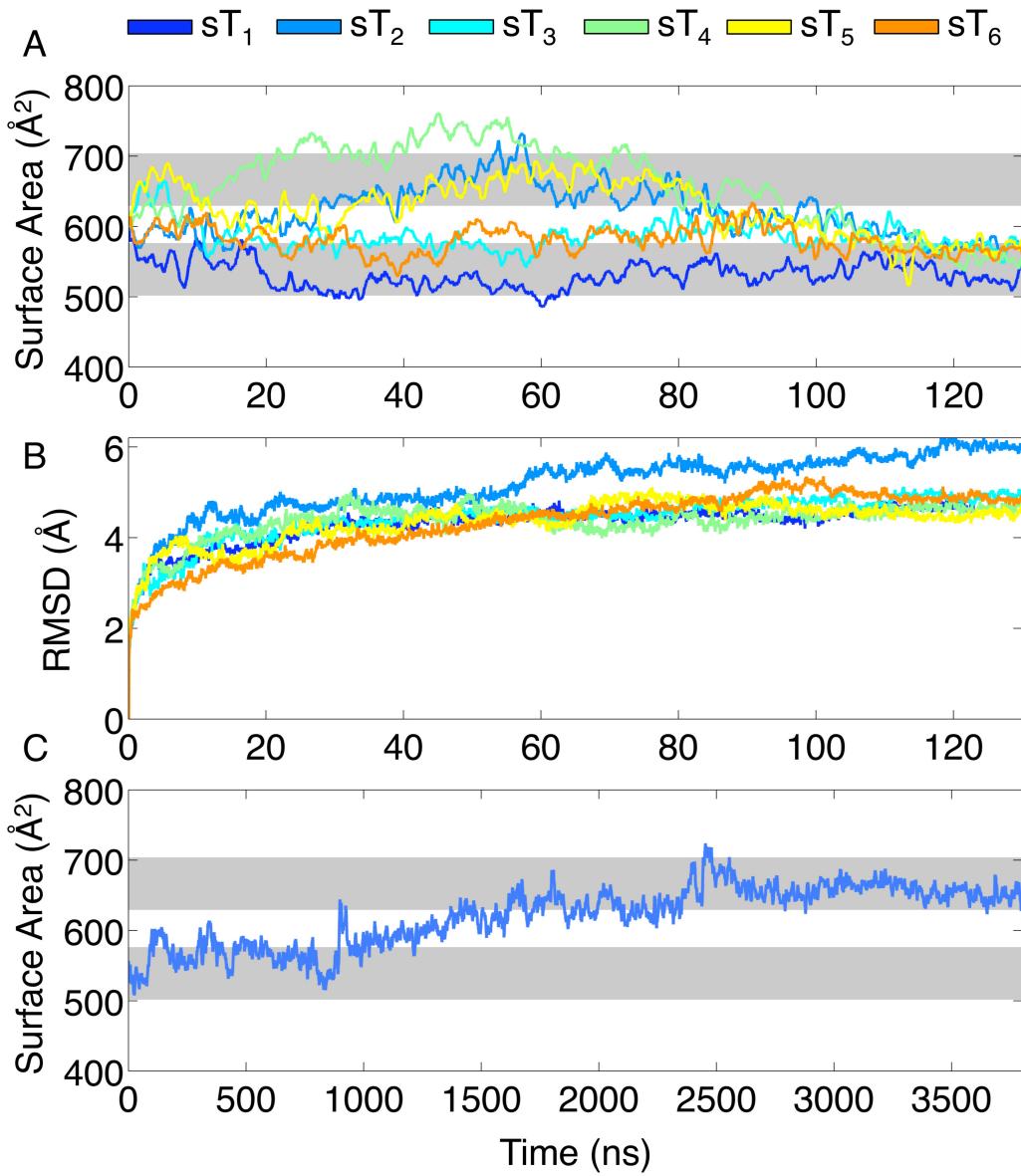


Figure S9: Comparison of additional equilibration trajectories from the SecA-SecYEG crystal structure that are performed with (sT₁-sT₃) and without (sT₄-sT₆) harmonic restraints applied to the SecA residues. (A) The translocon LG surface area is plotted as a function of time. (B) The RMSD of the SecY protein with respect to the crystal structure geometry¹ is plotted as a function of time. (C) The translocon LG surface area plotted during insertion trajectory T₁ from the main text. Comparison of the results in part C with those of part A illustrate that conformational changes associated with relaxation from the initial structure are qualitatively different than the trends associated with nascent protein insertion that are emphasized in the main text. Furthermore, comparison of part C with parts A and B suggests that any biases due to the use of harmonic restraints on SecA are small in comparison to the qualitative conformational changes that accompany nascent protein insertion.

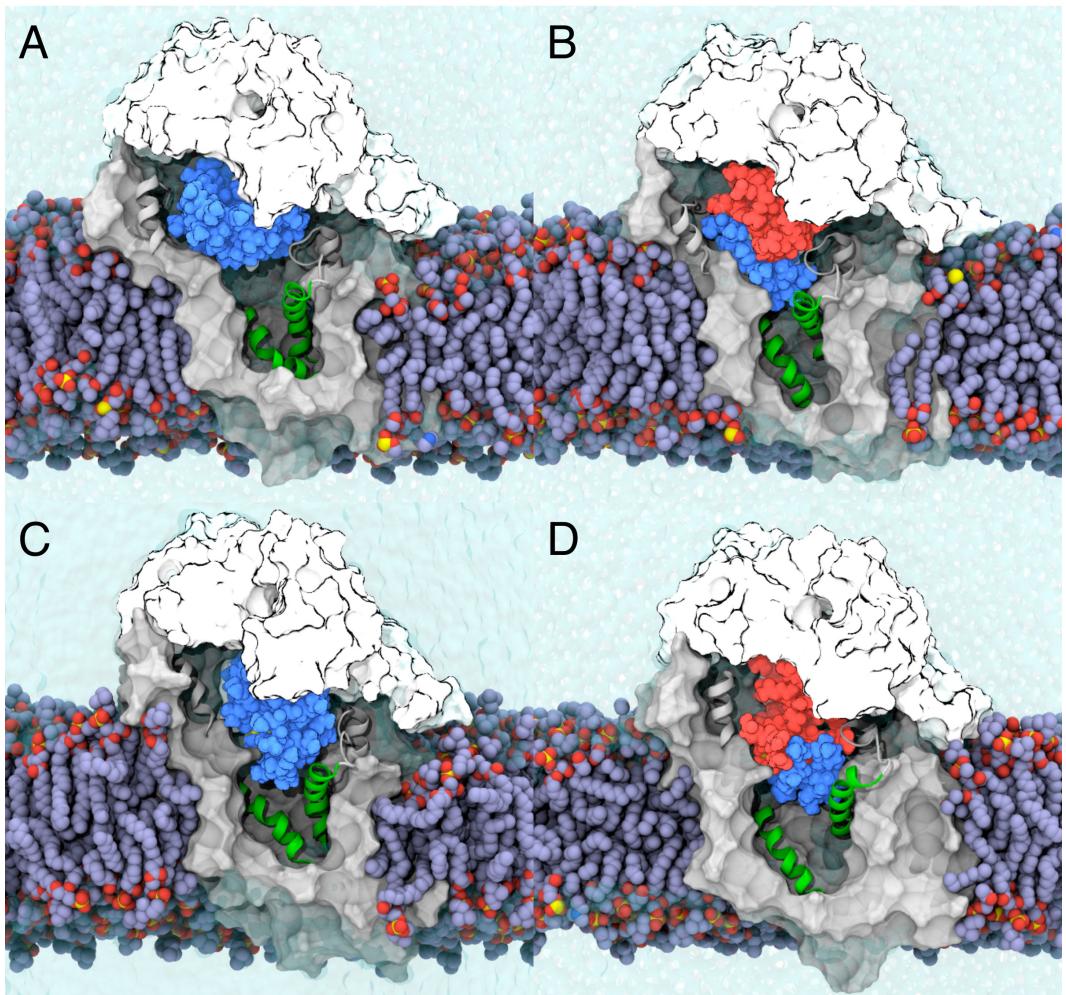


Figure S10: The final configuration from the insertion simulations associated with (A) trajectory T₅, (B) trajectory T₆, (C) trajectory T₁, and (D) trajectory T₂. The slower insertion protocol employed in trajectories T₁ and T₂ leads to docking of the nascent-protein SP at the translocon LG and associated conformational changes in the LG. However, the more rapidly inserted trajectories T₅ and T₆ do not exhibit these mechanistic features; instead, the nascent protein becomes jammed at the cytosolic mouth of the translocon.

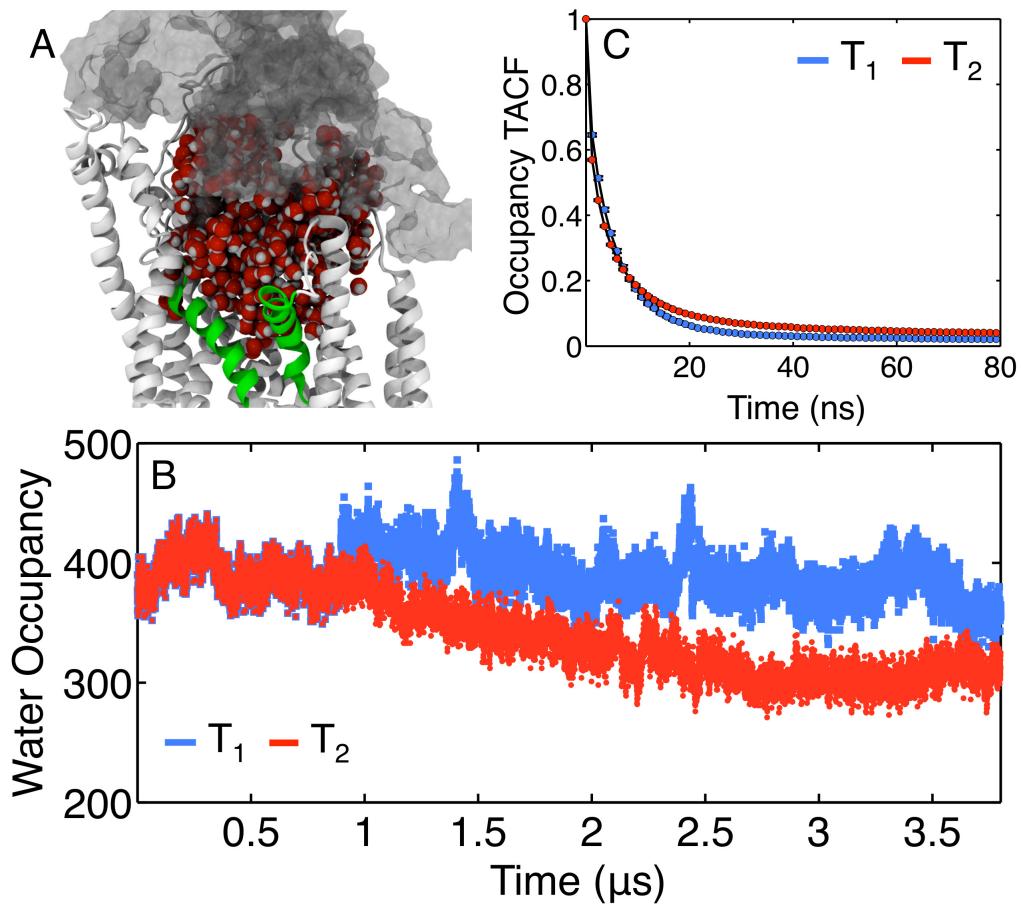


Figure S11: Summary of the occupancy and exchange dynamics of water molecules in the cytosolic mouth of the translocon channel. (A) A simulation snapshot indicating water molecules that occupy the cytosolic mouth. (B) The number of water molecules that occupy the cytosolic mouth as a function of simulation time for trajectories T₁ and T₂. (C) Time autocorrelation functions for water occupancy of the cytosolic mouth, obtained from insertion trajectories T₁ and T₂.

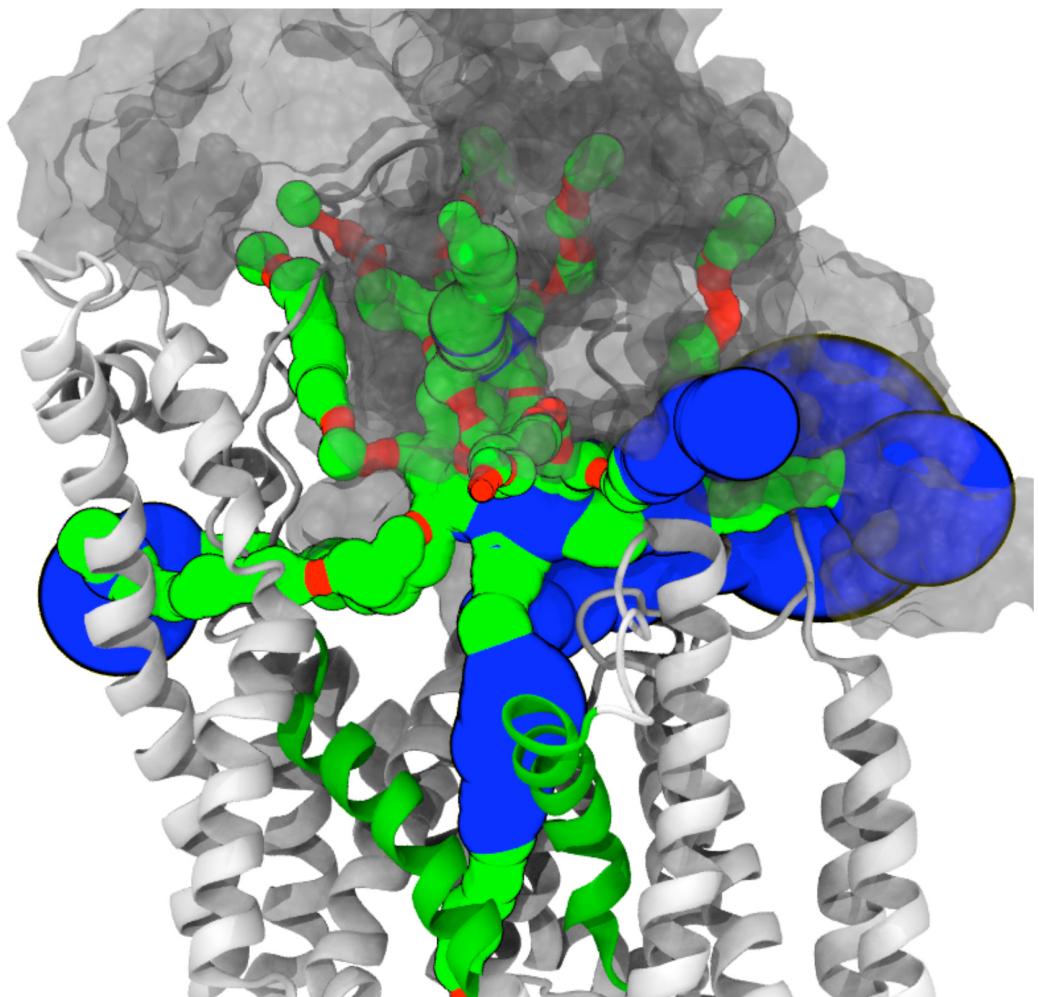


Figure S12: Sterically accessible regions for water molecules in the interfacial region between SecA and the translocon, with the translocon indicated in cartoon representation and with the SecA surface indicated in gray. Spatial regions are colored according to the radius, r , of the largest spherical particle that can be locally accommodated, with red indicating $r < 1.4\text{\AA}$, green indicating $1.4 \leq r < 2.7\text{\AA}$, and blue indicating $r \geq 2.7\text{\AA}$. Only a single continuous channel provides a pathway for exchange of water molecules between the cytosolic mouth region and the surrounding cytosolic bulk region; all other channels encounter prohibitively narrow openings.