Supporting Information

Bin Zhang *, **Thomas F. Miller, III** *

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

Atomistic simulations

All atomistic simulations were implemented within the TCL scripting protocol of the NAMD package (1).

All simulations were performed on the Sec channel from the archaeal *Methanococcus jannaschii* species, for which a high resolution crystal structure of has been reported (2). The molecular dynamics (MD) protocol that we employ follows that of Gumbart and Schulten (3). The channel was simulated in an explicit membrane composed of 254 palmitoyloleoylphosphatidylcholine (POPC) lipid molecules and an explicit solvent of 24296 rigid water molecules. Interactions were described using the CHARMM27 force field (4), including the TIP3P model for the water molecules. Counter-ions (Na^+ and Cl^-) were included to achieve electroneutrality and a salt concentration of approximately 50 mM (see Fig. S1). The protonation state of the histidine residues was chosen to be neutral. The initial system contains 115402 atoms in a simulation cell of size $110\,\text{Å} \times 110\,\text{Å} \times 100\,\text{Å}$. The system was described using orthorhombic periodic boundary conditions. Long-range electrostatics were calculated using the particle mesh Ewald (PME) technique (5). From the initial configuration of the system with the translocon in its crystal structure geometry, the system was equilibrated for production runs. This equilibration involved a 2000-step minimization, followed by a 0.5 ns NVT simulation with harmonic restraints (k = 2.0 kcal $\text{mol}^{-1}\text{Å}^{-2}$) applied to all atoms except the lipid tails, followed by a 1 ns NPT simulation with harmonic restraints applied only to the protein backbone, followed by a 3.5 ns NPT simulation with no restraints.

Production runs were performed in the NPT ensemble, with Langevin dynamics (damping coeffient 5 ps^{-1}) to keep the system at 300 K and with Nose-Hoover Langevin barostat (damping period 200 fs, damping time 100 fs) (6, 7) to maintain the pressure at 1 atm. The dynamics were integrated using a multiple-time-stepping approach (8), in which a 1 fs timestep was used for bonded interactions, a 2 fs timestep was used for short-range non-bonded interactions, and a 4 fs timestep was used for long-range electrostatic interactions. Short-range interactions were truncated at a distance of 12 Å, using a smoothing function in the range of distances from 10-12 Å.

Coarse-grained simulations

All CG simulations were implemented within the TCL scripting protocol of the NAMD package (1).

We employ a coarse-grained (CG) representation that combines the MARTINI coarse-graining algorithm for the lipid molecules, water molecules, and ions (9) and the residue-based coarse-graining scheme of Shih *et al.* for the amino-acid residues (10). The CG system was initialized by positioning the CG particles for the amino acid residues and the lipid molecules at the center-of-mass of the corresponding moieties in the atomistic model. The system was then solvated with CG particles representing water molecules and the counter-ions. The ion number and type were set to be the same as in the corresponding atomistic system. The final CG system contained a total of 9882 particles. All CG simulations were run in the NPT ensemble, with Langevin dynamics (damping coefficient 5 ps^{-1}) to keep the system at 323 K and Langevin piston (damping period 5 ps, damping time 2.5 ps) to maintain the pressure to 1 atm. A temperature of 323 K, rather than 300 K, was used because CG simulations at the higher temperature were found to better reproduce all-atom simulations of the lipid structure at 300 K (9, 10).

From the initialized configuration, the system was relaxed in preparation for production runs. A 5000-step minimization was first performed with the protein backbone and lipid heads harmonically restrained (k = 1.0 kcal $\text{mol}^{-1}\text{Å}^{-2}$). Then, a 5 ns CG MD simulation run was performed with a 5 fs timestep and with the protein backbone and the lipid heads harmonically restrained (k = 1.0 kcal $\text{mol}^{-1}\text{Å}^{-2}$ and 0.2 kcal $\text{mol}^{-1}\text{Å}^{-2}$, respectively). Finally, the whole system was released to relax along a 10 ns long trajectory with a 10 fs timestep. All other CG simulations were also performed using a 10 fs timestep.

Collective variables

Detailed definitions and illustrations for all collective variables employed in the paper are presented here. In general, if an α-carbon is used to define a collective variable in the atomistic representation, then the corresponding backbone CG particle is used in the CG representation.

Lateral gate distance. The lateral gate (LG) distance, d_{LG}, is defined as the distance of minimum approach between the line of least-squares fitting for the α-carbons of residues in the TM2b helix (residues Ile^{75}, Gly^{76}, Val^{79}, Thr^{80}, Ile^{84}, Leu^{87}, Ser^{91}, Gly^{92} in the α-subunit) and the corresponding fitting line for residues in the TM7 helix (residues Ile^{257}, Pro^{258}, Ile^{260}, Leu^{261}, Ala^{264}, Leu^{265}, Asn^{268}, Leu^{271}, Trp^{272}, Ala^{275}, Leu^{276}, Arg^{278} in the α-subunit).

An illustrated explanation of d_{LG} is provided in Fig. S2. In Fig. S2A, the residues shown in blue correspond TM2b helix, and residues shown in red correspond to the TM7 helix. In Fig. S2B, the blue line corresponds to h_1, the least-squares-fit line through the α-carbons of the residues used to define the TM2b helix; the red line corresponds to h_2, the least-squares-fit line through the α-carbons of the residues used to define the TM7 helix. In Fig. S2C, the green segment corresponds to d_{LG}, the distance of minimum approach between lines h_1 and h_2; this distance is calculated using

$$d_{\text{LG}} = |r_{12} + \mu e_2 - \lambda e_1|\,, \qquad (r_{12} = r_2 - r_1), \qquad [1]$$

where

$$\begin{aligned} \lambda &= [r_{12}\cdot e_1 - (r_{12}\cdot e_2)(e_1\cdot e_2)]\,/\,[1-(e_1\cdot e_2)^2], \text{ and} \\ \mu &= -[r_{12}\cdot e_2 - (r_{12}\cdot e_1)(e_1\cdot e_2)]\,/\,[1-(e_1\cdot e_2)^2]. \end{aligned} \qquad [2]$$

Here, r_1 is an arbitrary point on h_1 and e_1 is the unit vector that is parallel to h_1; r_2 and e_2 are similarly defined.

Pore-plug distance. The pore-plug (PP) distance, d_{PP}, is defined as the distance between the center-of-mass of the α-carbons for the residues that comprise the isoleucine ring of the channel (Ile^{75}, Val^{79}, Ile^{170}, Ile^{174}, Ile^{260}, and Leu^{406} in the α-subunit) and the center-of-mass of the α-carbons for the residues of the plug domain (Ile^{55}-Ser^{65} in the α-subunit).

An illustrated explanation of $d_{\rm PP}$ is provided in Fig. S3. In Fig. S3A, the residues shown in blue correspond to the pore of the channel, and the residues shown in red correspond to the plug moiety. In Fig. S3B, the blue bead, p_1, corresponds to the center-of-mass of the α-carbons that define the channel pore; the red bead, p_2, corresponds to the center-of-mass of the α-carbons that define the channel plug. In Fig. S3C, the green segment corresponds to $d_{\rm PP}$, the distance between points p_1 and p_2.

Plug-peptide orientation parameter. The plug-peptide orientation parameter, θ, is defined to be the angle between a vector v_1 that points from the peptide substrate to the plug moiety and a vector v_2 that points outward from the the opening of the LG. If $\cos(\theta) > 0$, then the plug is between peptide and the LG, as is shown for the snapshot of the hydrophilic peptide in Fig. 6A of the main text. For $\cos(\theta) < 0$ corresponds to the reversed orientation in which the peptide is between the plug and the LG.

An illustrated explanation of θ is provided in Fig. S4. Fig. S4A shows the two LG helices (TM2b and TM7) in green and the rest of the channel in gray. The TM2b helix is defined in terms of the residues Ile^{75}-Gly^{92} in the α-subunit of the translocon, and the TM7 helix is defined in terms of the residues Ile^{257}-Arg^{278} in the α-subunit of the translocon. Fig. S4B shows the lines h_1 and h_2, which are the least-squares fits through the α-carbons of the residues that compose the helix TM2b and TM7, respectively. Fig. S4C introduces the vector $n_1 = h_1 \times h_2$ (red), and Fig. S4D shows n_2 (blue) which is aligned with the z-axis of the simulation cell (and is perpendicular to the plane of the lipid bilayer). Together, the vectors n_1 and n_2 define the plane of the LG that separates the channel interior and the membrane exterior. Finally, Fig. S4E shows $v_2 = n_1 \times n_2$ (green), which is the vector that points outward from the opening of the LG.

In Fig. S4F, the residues shown in orange, referred to as the "lower residues" of the peptide substrate, are determined as follows. For any configuration of the system, we consider the Cartesian coordinates for the α-carbons of the peptide substrate; the lower residues are defined to be those 15 substrate residues with the lowest values of the Cartesian coordinate along the z-axis. The residues shown in red correspond to the translocon plug moiety (Ile^{55}-Ser^{65} in the α-subunit). In Fig. S4G, the orange bead, p_2, is the center-of-mass for the 15 α-carbons from the lower residues; the red bead, p_3, is the center-of-mass of the α-carbons for the plug moiety. In Fig. S4H, the arrow v_1 (yellow) connects p_2 and p_3, pointing from the peptide substrate to the plug moiety. We then obtain

$$\cos(\theta) = \frac{v_1 \cdot v_2}{|v_1||v_2|}. \qquad [3]$$

Lateral gate surface area. The LG surface area is illustrated in Fig. S5 and is calculated as follows. The z-axis is first uniformly discretized at a resolution of Δz between the bottom, z_0, and top, z_N, of the lipid bilayer; z_0 and z_N are defined in terms of the centers-of-mass for the lipid head groups of each leaf of the bilayer. For each discretized value along the z-axis, z_j, there is a corresponding slab that is parallel to the xy-plane, that is of thickness Δz, and that is centered around z_j.

The width of the LG opening for each slab is determined by considering the backbone CG particles within the z_j slab of the simulation cell and within two particular subsets of the translocon residues. The first subset, shown in red in Fig. S5A, includes residues in TM7-9 (Lys^{250}-Gly^{400} in the α-subunit). If $\cos(\theta) < 0$, the second subset, shown in green in Fig. S5A, includes residues Met^{70} to Ile^{160} in the α-subunit. If $\cos(\theta) > 0$, the second subset, shown in both green and yellow in Fig. S5A, includes residues Leu^{40} to Ile^{160} in the α-subunit. The width of the LG opening width for a given slab, w_j, is defined as the minimum distance from any CG particle in the first subset to any particle in the second subset; this definition accounts for the effect of the plug-substrate orientation on the LG surface area.

The LG surface area is finally obtained from the sum $\sum_{j=1}^{N} w_j \Delta z$, where $N = 20$. This is illustrated in Figs. S5B and C.

Initializing the peptide substrate

The hydrophobic (Leu$_{30}$) and hydrophilic (Gln$_{30}$) peptides were initialized as idealized α-helices with all-atom resolution using PyMOL (11). The idealized α-helices for both peptides were built with Ramachandran angles of ($\phi = -60^\circ$, $\psi = -45^\circ$). Simulations including the peptide substrate were initialized by inserting the idealized α-helix into a configuration for the channel with the plug displaced (i.e., a configuration of the translocon with $d_{\rm PP} = 20$ Å and $d_{\rm LG} = 6$ Å that was drawn from the restrained MD simulations used to calculate Fig. 3A in the text). The helix was positioned in the channel by aligning it with the z-axis of the simulations cell (perpendicular to the lipid bilayer) and placing the center-of-mass of the helix at the same position as the center-of-mass of the channel pore residues (Ile^{75}, Val^{79}, Ile^{170}, Ile^{174}, Ile^{260}, and Leu^{406} in the α-subunit). Having initialized the channel-substrate system with full atomistic resolution, the system was mapped onto the coarse-grained representation as described in the SI Text: "Coarse-grained simulations" and then equilibrated for 700-800 ns. As is discussed in the text, the centers-of-mass for the backbone CG particles of the substrate and the channel pore residues were tethered to each other with a weak harmonic restraint of 0.5 kcal mol^{-1}Å^{-2} to allow for arbitrary long simulations without the possibility of peptide diffusion out of the channel.

The long (over 700 ns) equilibration timescale for these simulations allows for extensive sampling of the peptide and translocon configuration space, such that the calculated free energy profiles are not dependent on the details of the initialization protocol described above. The slowest relaxation timescale that was found during equilibration corresponds to the relative orientation of the plug residue and the peptide substrate. This point is illustrated in Fig. S6, in which the plug-peptide orientation parameter is plotted as a function of the simulation time for the sampling trajectories. In part A, it is seen that the trajectories with the hydrophobic substrate relax relatively quickly (within about 400 ns) with respect to the initial orientation of the substrate and the plug. However, for many of the trajectories with the hydrophilic substrate (part B), the initial configuration appears to be a metastable conformation that eventually relaxes on a longer timescale. For the trajectories for which the LG distance is restrained to $d^{\circ}_{\rm LG}/\text{Å} = 9$, 9.5, 10, 11, 13, and 15, it was found that that the plug moiety and peptide substrate undergo an abrupt, kinetically frustrated reorientation on the timescale of hundreds of nanoseconds. Recognizing this clear tendency for reorientation, the $d^{\circ}_{\rm LG}/\text{Å} = 12$ trajectory was reinitialized from the $d^{\circ}_{\rm LG}/\text{Å} = 11$ trajectory at 600 ns, and the $d^{\circ}_{\rm LG}/\text{Å} = 14$ trajectory was reinitialized from the $d^{\circ}_{\rm LG}/\text{Å} = 13$ trajectory at 800 ns. This initially frustrated reorientation seems to be thermodynamically favorable for the hydrophilic substrate, since reorientation in the reverse direction was never observed.

Finally, we emphasize that the definition of the free energy profiles reported in Figs. 4 and 7 in the main text make no assumptions about the configuration of the peptide within the channel. The error bars in these plots, obtained from 160 ns block-averages of the production data, and the trajectories in Fig. S6 suggest that the thermal distribution of configurations has been thoroughly sampled. However, as in any molecular simulation of a complex system, it is possible that 1.5+ μs trajectories are not adequate to discover all thermodynamically dominant configuration of the system; at the very least, the calculations reported in Figs. 4 and 7 offer a meaningful characterization of long-lived basins of stability.

Side-chain transfer free energies for the CG residues

A measure of the accuracy of the residue-based CG models is obtained by considering the hydrocarbon/water transfer free energies for the amino acid residues. Recent simulation studies have demonstrated that the MARTINI (12) and the Sansom (13) residue-based coarse-graining methods, which are closely related to the CG potential employed here (10), give rise to side-chain transfer free energies that exhibit strong correlation and reasonable absolute agreement with experimental results. For the CG potential employed in the current study (10), we have also calculated the transfer free energies for all of the amino acid side-chains.

The transfer free energies for CG amino acid side-chains is obtained from the difference of the side-chain solvation free energies (ΔG) in water and in non-polar solvent. Each solvation free energy is calculated using the free energy perturbation formula (14).

$$\Delta G^{\mathrm{solv}} = -k_{\mathrm{B}}T \sum_{i=1}^{N} \ln \left\langle e^{-\beta[\mathcal{H}(\mathbf{x},\mathbf{p};\lambda_{i+1})-\mathcal{H}(\mathbf{x},\mathbf{p};\lambda_i)]} \right\rangle_i , \quad [4]$$

where

$$\mathcal{H}(\mathbf{x},\mathbf{p};\lambda) = \lambda \mathcal{H}_{\mathrm{solv}}(\mathbf{x},\mathbf{p}) + (1-\lambda)\mathcal{H}_{\mathrm{vac}}(\mathbf{x},\mathbf{p}), \quad [5]$$

and where the angle brackets correspond to the ensemble average for the system with Hamiltonian $\mathcal{H}(\mathbf{x},\mathbf{p};\lambda_i)$. The classical Hamiltonians for the system with and without interactions between the solvent and the side-chain are given by $\mathcal{H}_{\mathrm{solv}}$ and $\mathcal{H}_{\mathrm{vac}}$, respectively. We employed the NAMD implementation of this method.

Each solvation free energy was obtained from $N = 20$ ensemble averages of a system containing 1878 CG solvent molecules and one CG side-chain particle, with $\lambda_i = \{0, 0.00001, 0.0001, 0.001, 0.05, 0.10, 0.15, \ldots, 0.85, 0.90, 0.95, 0.99, 0.999, 0.9999, 0.99999\}$; the additional values of λ_i in the limits approaching $\lambda \to 0$ or 1 were included to avoid numerical instabilities. Each ensemble average was calculated from a 2 ns MD trajectory at constant temperature (300 K) and constant pressure (1 atm) with a timestep 40 fs. The first 40 ps of the trajectory was discarded as equilibration. The potential energy parameters for the non-polar solvent particles are the same as those for the CG particles in saturated lipid tails (i.e., the hydrophobic-apolar CG particle type). It was confirmed that the calculated free energies are converged with respect to equilibration time, the MD timestep, and the number of discretizations in λ.

The results for the transfer free energies of the amino-acid side-chains in the CG model are presented in Fig. S7. The statistical error for the simulations was typically of the size of the plotted symbols. The experimental results correspond to cyclohexane/water side-chain transfer free energies (15). Comparison of these results reveals reasonable correlation between the CG model and the experimental results. With regard to the amino acid residues that form the peptide substrates in our simulations, the Leu transfer free energies are within 1 kcal mol^{-1} of the experimental results, whereas the CG model underestimates the hydrophilicity of the Gln residues by approximately 2 kcal mol^{-1}. Although there is much room to improve the CG model, the accuracy suggested in Fig. S7 is not inconsistent with the level of accuracy observed in fully atomistic models. In particular, precise atomistic simulations have demonstrated that the choice of all-atom water potential leads to deviations of up to 1.5 kcal mol^{-1} in amino acid side-chain solvation free energies (16), and the choice of molecular mechanics force field for the amino acid changes the calculated solvation free energy by over 1 kcal mol^{-1} in many cases (17). The calculations presented in Fig. S7, along with the comparison between the atomistic and CG free energy profiles in Figs. 3 and S8, suggest that the CG potentials employed in this study form a reasonable basis for the qualitative interpretation of the simulation results.

For other tests reported in the SI, it is useful to have a CG side-chain particle that is of intermediate hydrophobicity with respect to the Leu and Gln residues. We developed such a side-chain particle with the same potential energy functional form as the other CG side-chain particles (10). The interaction parameters for this "Int" CG side-chain particle are presented in Table S1, and the corresponding transfer free energy is reported as the red diamond in Fig. S7.

Scaffolding contribution to the free energy profile

Fig. 3 in the main text presents the free energy profile for the translocon as a function of d_{LG} and d_{PP}. Here, in Fig. S8A, these results are re-plotted with error estimates. The red/yellow-shaded surface corresponds to the atomistic free energy profile, and the blue/orange-shaded surface below it corresponds to the CG free energy profile. The error bars correspond to the standard deviation of the mean free energy profile obtained from the five block averages of the simulation data.

To investigate the impact of the scaffolding interactions on the calculated CG free energy profile, it was re-calculated for the CG model without scaffolding. Following the same protocol as was used to obtain Fig. 3B, an additional set of 80 CG MD sampling trajectories was performed without scaffolding interactions, each of which was of length 20 ns and was harmonically restrained for the collective variables $d_{\mathrm{LG}}/\mathring{A} \in [6, 13]$ and $d_{\mathrm{PP}}/\mathring{A} \in [11, 21]$. Using the WHAM algorithm, the free energy profile without scaffolding was constructed and is plotted as the grey surface in Fig. S8A. All three profiles are vertically shifted to have a minimum at 0 kcal mol^{-1}. Comparison of the CG free energy profiles with and without scaffolding in Fig. S8A indicates that the features of the CG profile are not dramatically altered by the inclusion of the scaffolding interactions.

For a more detailed comparison, the difference between the free energy profiles for the CG model with and without scaffolding is plotted in Fig. S8B, and the statistical uncertainty of this difference is plotted in 8C. The difference between the CG free energy profiles appears to be more sensitive to changes in d_{PP} than d_{LG}, and for most of the domain, the difference between the CG surfaces does not exceed the statistical uncertainty by more that 5 kcal mol^{-1}. These results indicate that the scaffolding interactions do give rise to some changes in the calculated CG free energy profile, although the differences are relatively small in comparison to the other features on the surface.

Free-Energy Surface Cross-Sections

Figure S9 presents free-energy profiles as a function of the LG coordinate for fixed values of the pore-plug distance. The red curve in part (B) corresponds to the cross-section of the free-energy profile at 12 Å that is indicated by the red band in part (A). The blue curve in part (B) is similarly obtained from the cross-section at 19 Å. The curves in part (B) are vertically shifted to be 0 kcal mol^{-1} at their minimum.

Additional trajectories

Trajectories without scaffolding. To confirm that the closing of the Sec translocon with the hydrophilic peptide inside is not an artifact of the scaffolding interactions, we performed additional CG MD trajectories without scaffolding for the translocon initialized from open configurations ($d_{\mathrm{LG}} = 14$ Å) of the LG. For both the hydrophobic substrate and the hydrophilic substrate, three independent trajectories of length 500 ns are presented in Fig. S10; as in Fig. 5, the initial configurations for the trajectories were drawn from the substrate-containing trajectories with scaffolding that were restrained with respect to d_{LG}.

The results in Fig. S10 are broadly consistent with simulations that include scaffolding in Fig. 5. For the trajectories with the hydrophilic substrate, the initially open LG distance closes during the simulation (Fig. S10A) to the same extent that was seen in Fig. 5. Furthermore, the LG surface area for the hydrophilic trajectories trend downwards over the simulated timescale, although to varying degrees, and the substrate is found to remain within the translocon channel.

The trajectories with the hydrophobic substrate show a greater range of behavior. In one case (royal blue), the initially open LG remains fully open over the course of the simulation, as was seen in Fig. 5. In a second case (dark blue), the hydrophobic peptide exits the channel, and the LG distance closes behind it in such a way that the LG surface area remains relatively large. In a third case (light blue), the peptide partially exits the channel, and the LG surface area closes in such a way that the LG distance remains relatively large. In interpreting these results for the hydrophobic substrate, it must be remembered that the CG model without scaffolding does not fully preserve structural features of the translocon on these long timescales (Fig. 2C and Fig. S11); it is possible that unphysical distortions in the channel are facilitating the exit of the substrate.

Trajectories with substrate of intermediate hydrophobicity. The trajectories presented in Fig. 5 in the main text illustrate the metastability of the translocon in the presence of strongly hydrophobic and strongly hydrophilic substrates. Here, we explore the metastability of the translocon with a peptide substrate of intermediate hydrophobicity. Six CG MD trajectories of length 500 ns were performed with the substrate Int_{30}, a linear peptide composed of 30 CG amino acids with side-chains that exhibit a transfer free energy between that of the Leu and Gln residues (see SI Text: "Side-chain transfer free energies for the CG residues," Fig. S7, and Table S1). As in Fig. 5, scaffolding interactions for the translocon were employed.

Fig. S12 presents the CG MD trajectories performed with the Int_{30} substrate. These additional trajectories were initialized from the same configurations as the six trajectories reported in Fig. 5 (which include closed, partially open, and fully open configurations for the LG). The color scheme in Fig. S12 identifies which of the additional trajectories shares the same initial configuration as a given trajectory in Fig. 5.

Fig. S12 indicates that the Int_{30} substrate supports both the metastable open and closed configurations of the translocon LG. For the two trajectories that are initialized with $d_{\mathrm{LG}} = 6$ Å, this distance remains relatively unchanged over the course of the simulations and the LG surface remains closed in the range of 400-450 Å^2. Similarly, for the two trajectories that are initialized with $d_{\mathrm{LG}} = 15$ Å, this distance remains relatively unchanged over the course of the simulations, but the LG surface areas relaxes into the range of 600-650 Å^2 that is sufficiently open to allow for large substrate exposure to the membrane (Fig. 5C).

Of the two trajectories that were initialized from the partially open LG ($d_{\mathrm{LG}} = 11$ Å), one (plotted in blue) exhibits LG surface area values in the open range (600-650 Å^2), whereas the other (plotted in red) exhibits LG surface area values in the closed range (400-450 Å^2). After 500 ns of simulation time, all six of the trajectories in Fig. S12 exhibit LG surface areas either in the range of 400-450 Å^2 or in the range of 600-650 Å^2, although the LG distance seems to relax on a slower timescale. Consistent with Fig. 5, the trajectories in Fig. S12 suggest that even in the presence of a substrate of intermediate hydrophobicity, the LG surface area exhibits two-state behavior with respect to opening and closing of the LG.

Mutations in the translocon pore residues

Here, we explore how mutations in the translocon pore residues alter the free energy cost for opening the LG. These pore residues have been demonstrated to have significant impact on the functioning of the Sec translocon (2, 18). In a first set of calculations, we replace the six amino acid residues that comprise the hydrophobic translcon pore moiety (Ile^{75}, Val^{79}, Ile^{170}, Ile^{174}, Ile^{260}, and Leu^{406} in the α-subunit) with either six hydrophilic (Gln) residues or six intermediate (Int) residues, and we calculate the corresponding changes in the free energy profiles in Fig. 7. (For a discussion of the Int residues, see SI Text: "Side-chain transfer free energies for the CG residues," Fig. S7, and Table S1.) The mutated free energy profiles are calculated from the simulation data used to construct Fig. 7, using

$$F_{\mathrm{mut}}(A_{\mathrm{LG}}) = -k_{\mathrm{B}}T \ln P_{\mathrm{mut}}(A_{\mathrm{LG}}) \qquad [6]$$

where

$$P_{\mathrm{mut}}(A_{\mathrm{LG}}) \propto \langle \delta(A_{\mathrm{LG}} - A_{\mathrm{LG}}(\mathbf{x})) e^{-(U_{\mathrm{mut}}(\mathbf{x}) - U_{\mathrm{wt}}(\mathbf{x}))/(k_{\mathrm{B}}T)} \rangle_{U_{\mathrm{wt}}}, \qquad [7]$$

$A_{\mathrm{LG}}(\mathbf{x})$ is the LG surface area as a function of the positions x of the CG particles, $U_{\mathrm{wt}}(\mathbf{x})$ is the potential energy surface for the wild-type system, and $U_{\mathrm{mut}}(\mathbf{x})$ is the potential energy surface for the mutated system (21). The angle brackets indicate thermal averaging on the wild-type potential energy surface. The free energy surfaces for the wild-type and mutant translocons are plotted in Fig. S13. Although the variance of the exponential term in Eq. 7 leads to diminished statistical certainty in the free energy profiles for the mutants, clear trends seem to emerge.

Figs. S13A and B illustrate the effect of hydrophobicity of the pore residue side-chains. Part A shows the results for the translocon with the hydrophobic substrate, and part B shows the results for the hydrophilic substrate. For the case of the hydrophobic substrate, making the pore more hydrophilic favors the open LG. This trend can be easily understood in terms of the preference of the hydrophobic substrate for the membrane environment over the increasingly hydrophilic environment of the mutated translocon channel. However, part B shows that for the case of the hydrophilic substrate, the changes are less dramatic. Indeed, over the range of 400-600 Å^2 for the surface area, which we find to be the range of LG opening in our MD trajectories, there is very little difference in the free energy profiles for the hydrophilic substrate.

We can use a similar analysis to study the effect of the pore-residue bulkiness. We consider two translocon mutants in which all six hydrophobic pore residues are replaced with CG particles for which the Lennard-Jones radius for the CG side-chain particle is either increased by 5% or decreased by 10%, and the free energy profiles are re-calculated using Eqs. 6 and 7. Fig. S13C shows the resulting profiles for the translocon with the hydrophobic substrate, and Fig. S13D shows the resulting profiles for the hydrophilic substrate. Fig. S13C suggests that increasing the bulkiness of the pore residues leads to the relative stabilization of the open LG. Physically, this is reasonable. With the bulkier pore residues, less room is available inside the channel, which favors the open configurations of the LG in which the hydrophobic substrate partially extends into the membrane (see Fig. 6B in main text). Fig. S13C suggests that the bulkiness of the pore residue side-chains does not significantly change the free energy profile in the range of 400-600 Å^2 for the surface area. This can be rationalized by observing that since the hydrophilic substrate is tucked behind the plug residue (Fig. 6A), the opening of the LG does not relieve the confined environment of the substrate.

The simulations presented in Figs. S13 indicate that the employed CG model is sensitive to the molecular details of the translocon system.

Alternative collective variable definitions

To investigate the robustness of our results with respect to the definitions of the LG distance and LG surface area collective variables, we have modified those definitions and recalculated the free energy profiles in Figs. 4 and 7 in the main text. The new free energy profiles, which are presented in Fig. S14, are extracted from the simulation data for the translocon with the the peptide substrate using the WHAM algorithm.

Lateral gate distance. As is described in the SI Text: "Collective variables", Fig. 4 is calculated using a definition for the LG distance that employs least-squares-fit lines to the backbone CG particles of residues in the TM2b helix and in the TM7 helix. In this original definition, the line h_1 is fit to residues Ile^{75}, Gly^{76}, Val^{79}, Thr^{80}, Ile^{84}, Leu^{87}, Ser^{91}, Gly^{92} in the α-subunit, and the line h_2 is fit to residues Ile^{257}, Pro^{258}, Ile^{260}, Leu^{261}, Ala^{264}, Leu^{265}, Asn^{268}, Leu^{271}, Trp^{272}, Ala^{275}, Leu^{276}, Arg^{278} in the α-subunit. In Fig. S14A, we demonstrate how the calculated free energy profile changes upon increasing the number of residues used in these least-squares fits. Specifically, we change the definition of h_1 to be the least-squares-fit line to the backbone CG particles in all residues of TM2b (Ile^{75} - Gly^{92} in the α-subunit), and we change the definition of h_2 to be the least-squares-fit line to the backbone CG particles in all residues of TM7 (le^{257} - Arg^{278} in the α-subunit). To facilitate a direct comparison between the two definitions, the x-axis of the free energy profiles using the new LG distance definition (light blue and light red) are shifted to the left (smaller values) by a constant value of 1.9 Å. It is clear from the figure that aside from the constant shift, the results are essentially unchanged.

Lateral gate surface area. As is described in the SI Text: "Collective variables", Fig. 7 is calculated using a definition for the LG surface area that includes only the backbone CG particles. Here, we use the same method for calculating the LG surface area, except that we also include the side-chain CG particles. The resulting calculation the free energy profile is presented in Fig. S14B. The x-axis of the free energy profiles using the new LG surface area definition (light blue and light red) are shifted to the right (larger values) by a constant value of 140 Å^2. Again, the results are changed very little with respect to the statistical certainty.

It is interesting to note that for the system with the hydrophobic substrate, there is a more pronounced barrier in the free energy profile calculated with the new definition of the LG surface area (Fig. S14B). This could be related to the observed metastability of the closed state of the translocon in the presence of the hydrophobic substrate (Figs. 5 and S12). Metastability requires that in some kinetically relevant coordinate (which might or might not be the LG surface area), there exists a barrier that is large in comparison to $k_{\text{B}}T$. The appearance of such a barrier in the free energy profile for the new LG surface area might indicate that the new definition better describes the kinetic bottleneck for LG opening. Naturally, a complete investigation (22) of the dynamics of LG opening and membrane insertion is needed to solidify this conclusion.

1. Phillips JC et al. (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781–1802.
2. Berg Bvd et al. (2004) X-ray structure of a protein-conducting channel. Nature 427:36–44.
3. Gumbart J, Schulten K (2006) Molecular dynamics studies of the archaeal translocon. Biophys J 90:2356–2367.
4. Mackerell AD et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616.
5. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N.log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092.
6. Martyna GJ, Tobias DJ, Klein ML (1994) Constant pressure molecular dynamics algorithms. J Chem Phys 101:4177–4189.
7. Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: The Langevin piston method. J Chem Phys 103:4613–4621.
8. Tuckerman ME, Berne BJ, Martyna GJ (1991) Molecular dynamics algorithm for multiple time scales: Systems with long range forces. J Chem Phys 94:6811–6815.
9. Marrink SJ, de Vries AH, Mark AE (2004) Coarse grained model for semiquantitative lipid simulations. J Phys Chem B 108:750–760.
10. Shih AY, Arkhipov A, Freddolino PL, Schulten K (2006) Coarse grained protein-lipid model with application to lipoprotein particles. J Phys Chem B 110:3674–3684.
11. Delano WL (2002) The PyMOL Molecular Graphics System (Delano Scientific, Palo Alto CA).
12. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH (2007) The MARTINI force field: Coarse grained model for biomolecular simulations. J Phys Chem B 111:7812–7824.
13. Bond PJ, Wee CL, Sansom MS (2008) Coarse-grained molecular dynamics simulations of the energetics of helix insertion into a lipid bilayer. Biochem 47:11321–11331.
14. Chipot C, Shell MS, Pohorille A (2007) Free Energy Calculations: Theory and Applications in Chemistry and Biology, eds Pohorille A, Chipot C (Springer, Heidelberg).
15. Radzicka A, Wolfenden R (1988) Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. Biochem 27:1664–1670.
16. Shirts MR, Pande VS (2005) Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. J Chem Phys 122:134508–134521.
17. Shirts MR, Pitera JW, Swope WC, Pande VS (2003) Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. J Chem Phys 119:5740–5761.
18. Smith MA, Clemons W. M., J, DeMars CJ, Flower AM (2005) Modeling the effects of prl mutations on the Escherichia coli SecY complex. J Bacteriol 187:6454–6465.
19. Plath K, Mothes W, Wilkinson BM, Stirling CJ, Rapoport TA (1998) Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. Cell 94:795–807.
20. Rapoport TA (2007) Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature 450:663–669.
21. Frenkel D, Smit B (1996) Understanding Molecular Simulation: From Algorithms to Applications (Academic, San Diego).
22. Miller TF, Vanden-Eijnden E, Chandler D (2007) Solvent coarse-graining and the string method applied to the hydrophobic collapse of a hydrated chain. Proc Natl Acad Sci USA 104:14559–14564.

Table S 1. Potential energy parameters for the interaction of the particle of intermediate hydrophobicity with the other CG particle types.

CG Particle Type	CG Particle Type*	Lennard-Jones Energy-scale, ϵ (kcal mol^{-1})	Lennard-Jones Length-scale,† $R_{\min}$ (Å)
Int	C	-0.621	5.300
Int	Nx	-0.812	5.300
Int	No	-0.813	5.300
Int	Nd	-0.812	5.300
Int	Na	-0.812	5.300
Int	Qx	-0.812	5.300
Int	Qo	-0.621	5.300
Int	Qd	-0.717	5.300
Int	Qa	-0.717	5.300
Int	P	-0.717	5.300
Int	Nxx	-0.812	5.300
Int	Nxg	-0.812	5.300
Int	Ca	-0.621	5.300
Int	Qdr	-0.717	5.300
Int	Nxn	-0.812	5.300
Int	Qad	-0.717	5.300
Int	Pc	-0.717	5.300
Int	Nxq	-1.195	5.300
Int	Qae	-0.717	5.300
Int	Ph	-0.717	5.300
Int	Qdh	-0.717	5.300
Int	Ci	-0.621	5.300
Int	Int	-1.195	5.300
Int	Cl	-0.621	5.300
Int	Qdk	-0.717	5.300
Int	Cm	-0.621	5.300
Int	Cf	-0.621	5.300
Int	Cp	-0.621	5.300
Int	Ps	-0.717	5.300
Int	Pt	-0.717	5.300
Int	Cw	-0.621	5.300
Int	Nxy	-0.812	5.300
Int	Cv	-0.621	5.300
Int	Nap	-0.812	5.300
Int	CDB	-0.621	5.300

*CG particle-type names are consistent with Ref. (10).

† $R_{\min}$ is related to the usual Lennard-Jones lengthscale via $R_{\min} = \sigma 2^{1/6}$.

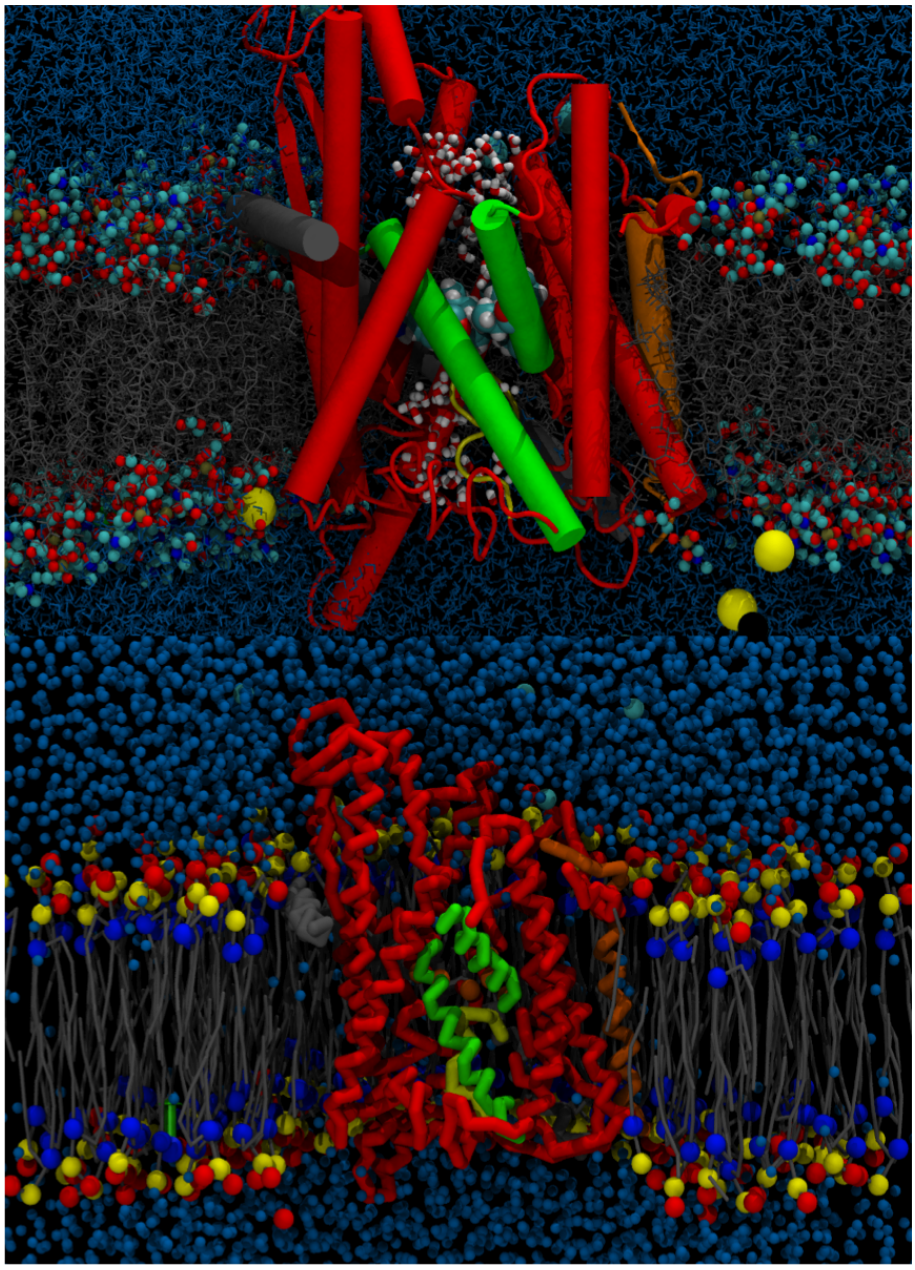

Fig. S 1. Snapshots of the all-atom (top) and coarse-grained (bottom) simulation systems.

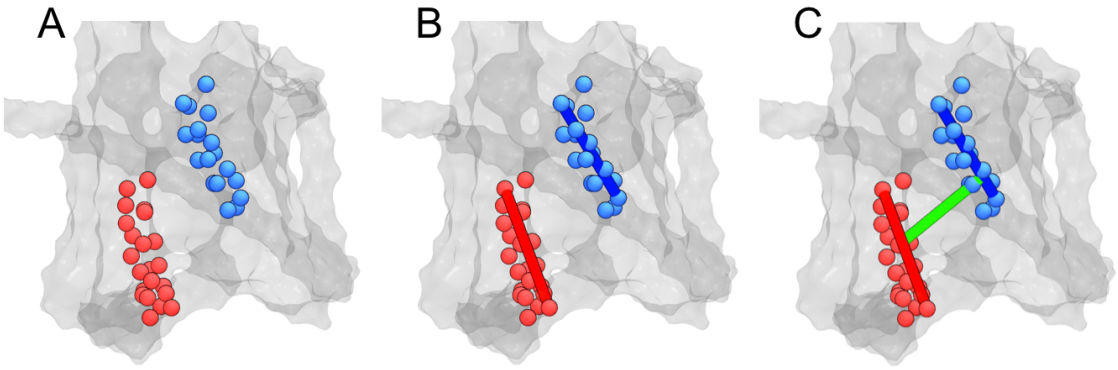

Fig. S 2. The LG distance collective variable. (A) Residues forming the two LG helices: TM2b (blue), and TM7 (red). (B) Least-squares-fit lines for each helix. (C) The minimum distance (green) between the two fit lines.

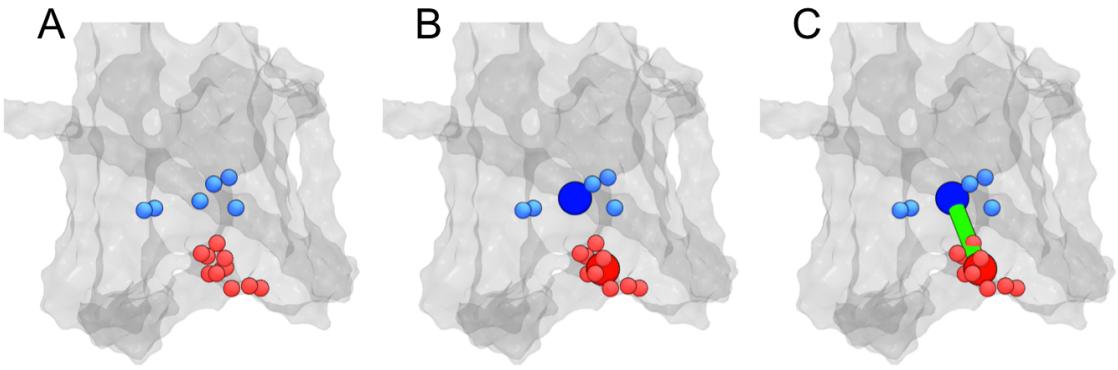

Fig. S 3. The pore-plug distance collective variable. (A) Residues forming the pore (blue) and plug (red). (B) The center-of-mass of the pore residues (dark blue bead), and the center-of-mass of the plug residues (dark red bead). (C) The center-of-mass distance between the pore and the plug (green line)

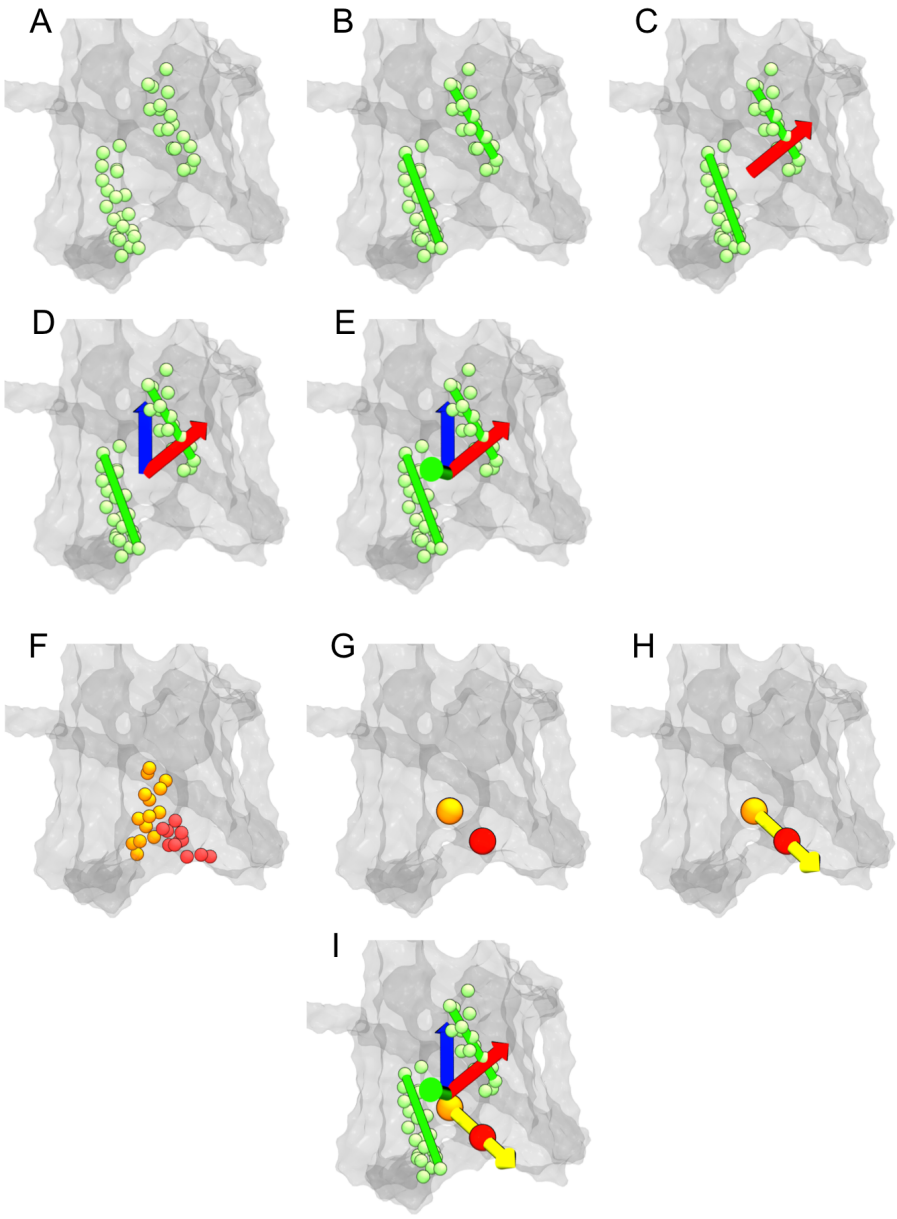

Fig. S 4. The plug orientation order parameter. (A) Residues forming the two LG helices. (B) The least-square fit lines for the TM2b and TM7 helices, h_1 and h_2, respectively. (C) The vector $n_1 = h_1 \times h_2$ (red). (D) The vector n_2 that is aligned with the z-axis of the simulation cell (blue). (E) Vector $v_2 = n_1 \times n_2$ pointing outward from the opening of the LG (green). (F) Residues forming the lower half the peptide substrate (orange), and residues forming the plug moiety (red). (G) Centers-of-mass for the lower peptide residues (orange bead) and the plug residues (red bead). (H) The vector v_1 pointing from the peptide substrate to the plug moiety (yellow). (I) Combined figure showing the relative direction of the v_1 and v_2 vectors, which define the plug-peptide orientation parameter (Eq. 3).

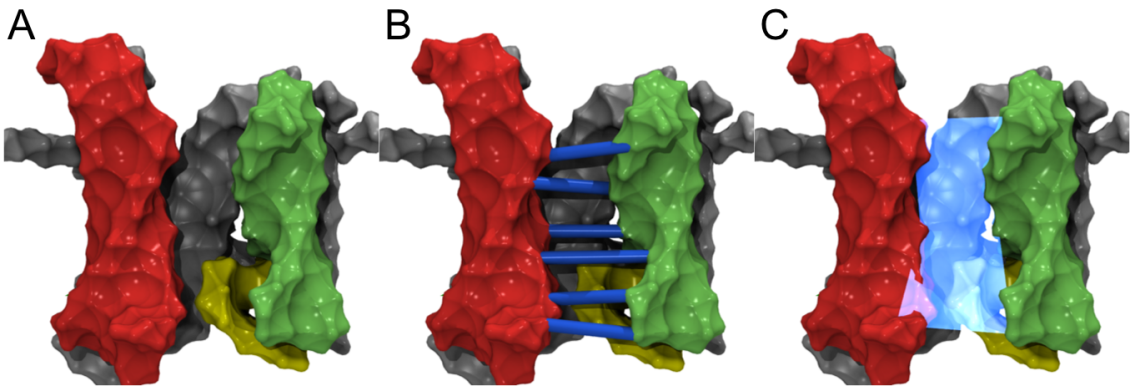

Fig. S 5. Illustration of the LG surface area. (A) TM7-9 (Lys^{250}-Gly^{400} in the in the α-subunit) are shown in red, TM2b-4 (Met^{70} to Ile^{160} in the α-subunit) are shown in green, and the plug moiety (Leu^{40} to Met^{70} in the α-subunit) is shown in yellow. (B) The width of the LG opening at various points along the z-axis. (C) The surface area is obtained by integrating the width of the LG opening over the z-axis, as is described in the text.

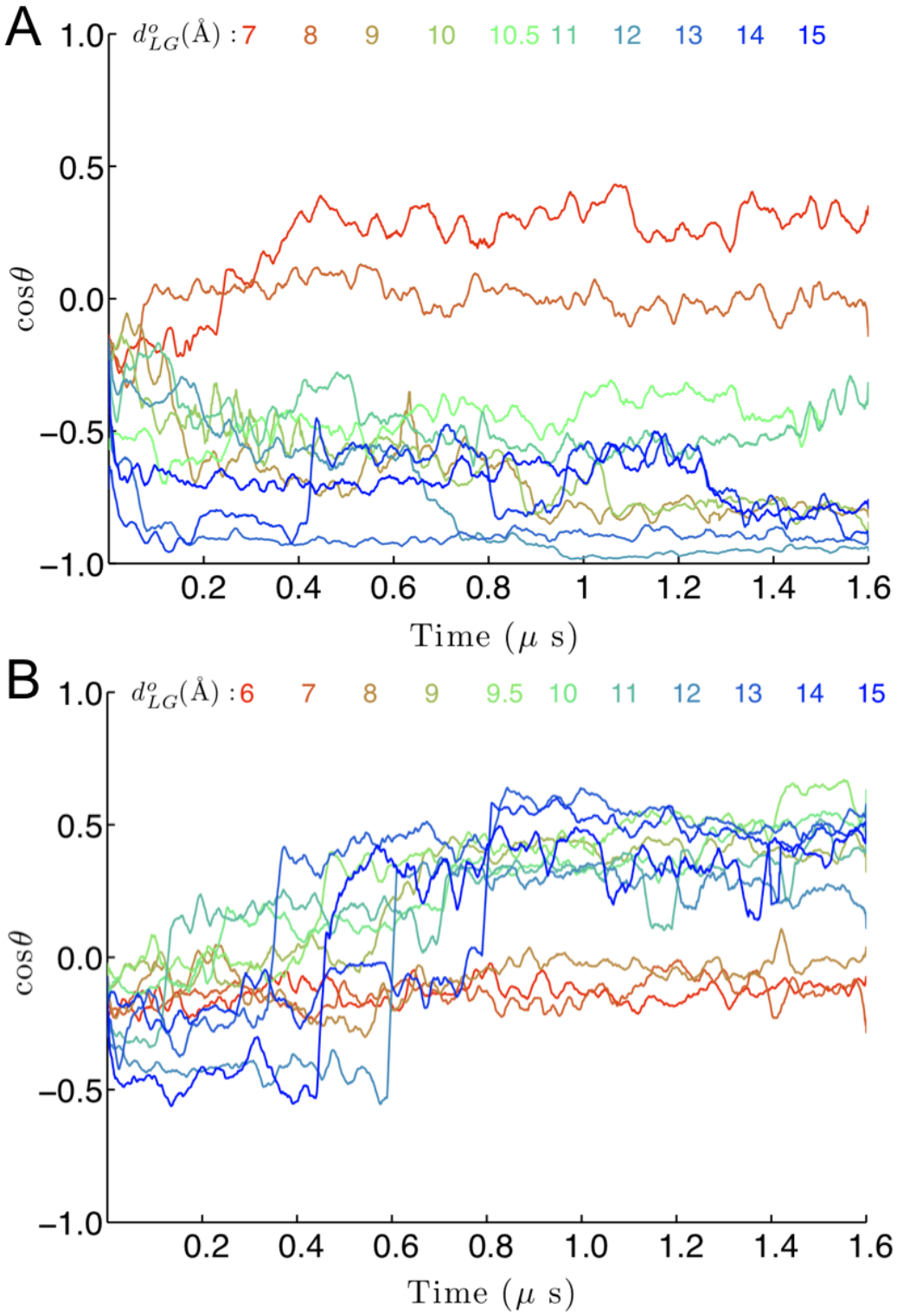

Fig. S 6. The CG MD sampling trajectories for the translocon in the presence of (A) the hydrophobic substrate and (B) the hydrophilic substrate, plotted as a function of the plug-peptide orientation parameter. Each trajectory is restrained to a different value for the LG distance, indicated by color. All curves correspond to 10 ns rolling averages of the simulation data. See text for details.

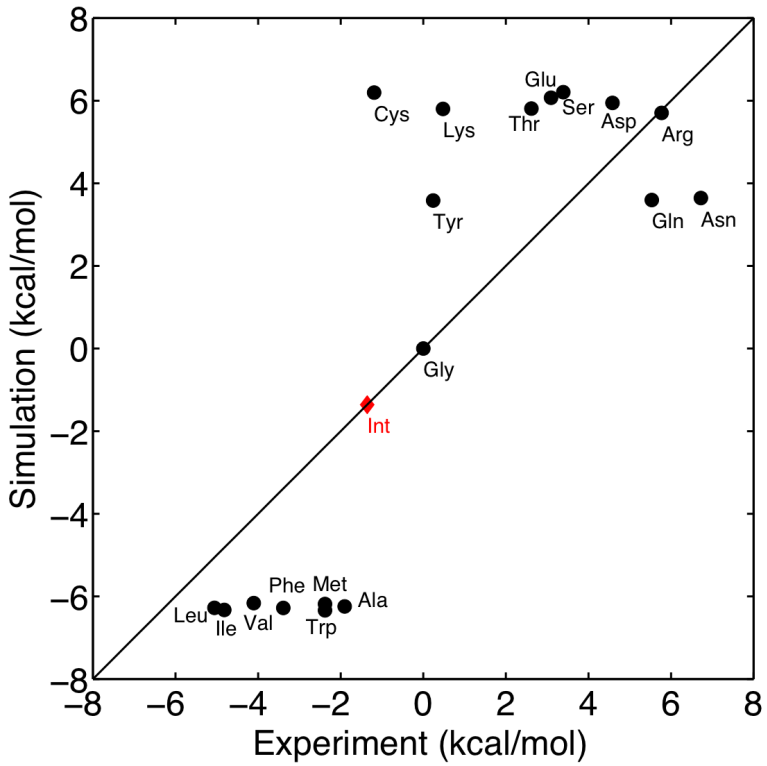

Fig. S 7. Correlation plot between the oil/water transfer free energies for the amino-acid side chains obtained experimentally and from the CG simulations. The red diamond indicates the transfer free energy for an additional CG side-chain particle of intermediate hydrophobicity. Details in text.

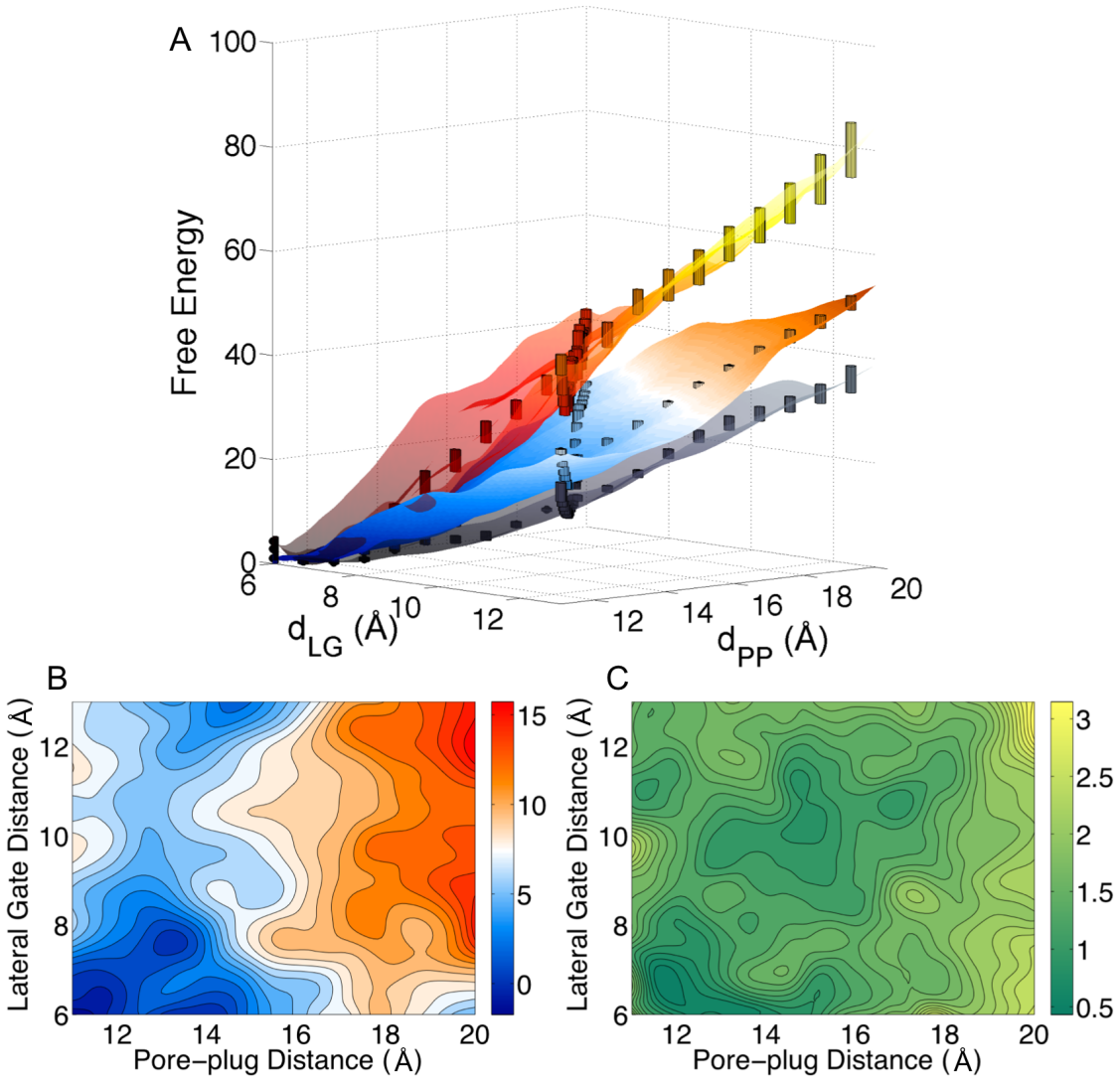

Fig. S 8. Free energy profiles for the translocon as a function of the LG and PP distances, calculated using the atomistic potential (A, red/yellow-shaded) the CG potential with scaffolding (A, blue/orange-shaded), and the CG potential without scaffolding (A, grey). (B) The difference between the free energy profiles obtained using the CG potential with and without scaffolding. (C) The statistical uncertainty in this difference. All energies in kcal mol^{-1}.

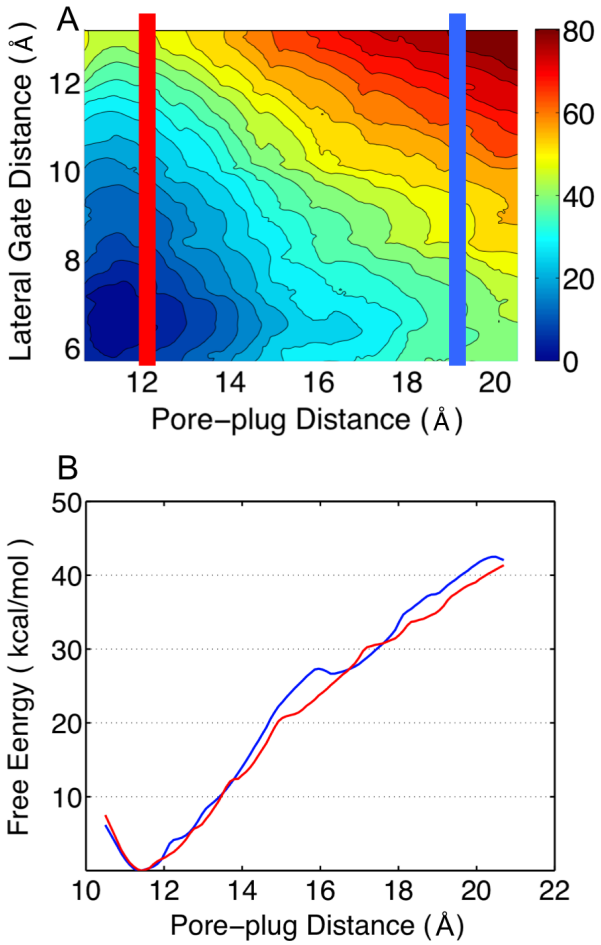

Fig. S 9. Free-energy profiles as a function of the LG coordinate for fixed values of the pore-plug distance. The red curve in part (B) corresponds to the cross-section of the free-energy profile at 12 Å that is indicated by the red band in part (A). The blue curve in part (B) is similarly obtained from the cross-section at 19 Å. The curves in part (B) are vertically shifted to be 0 kcal mol^{-1} at their minimum.

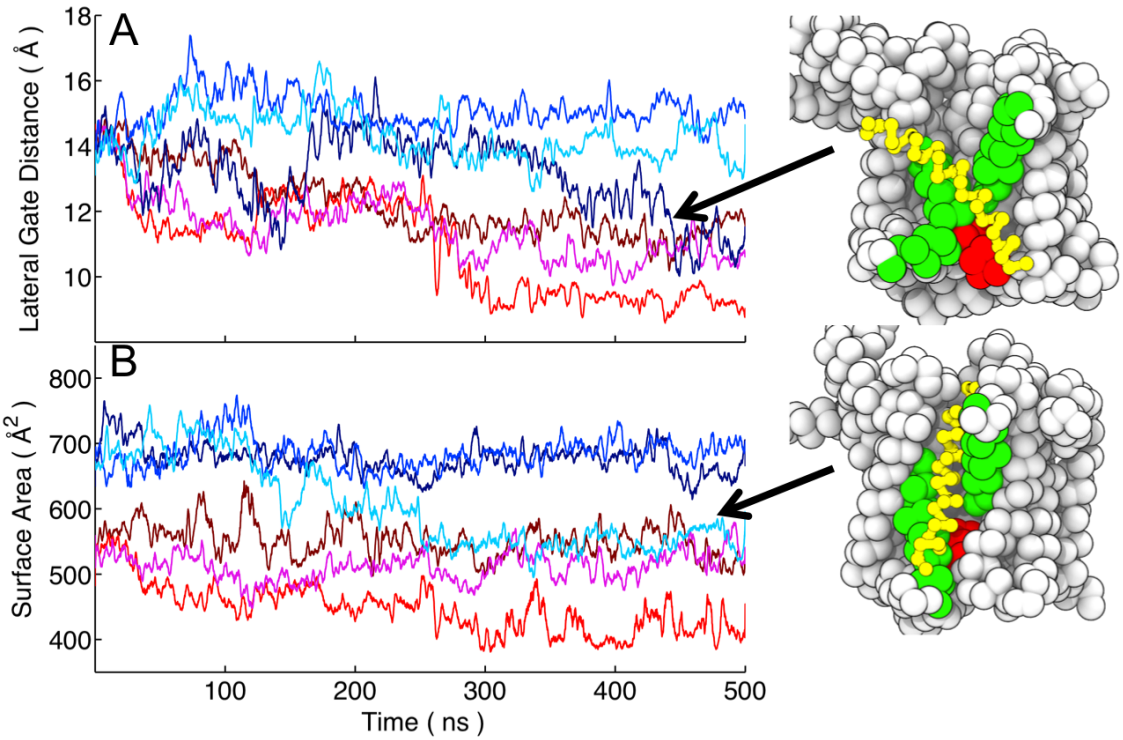

Fig. S 10. CG MD trajectories without scaffolding for the translocon containing either the hydrophobic (blue-shaded) or hydrophilic (red-shaded) substate are initialized from open configurations of the LG. Three independent trajectories for each substrate are performed. (A) The LG distance d_{LG} for the trajectories is plotted as a function of simulation time. (B) The LG surface area for the trajectories is plotted as a function of simulation time. The lines indicate 1 ns rolling averages. Also shown are snapshots from the two trajectories in which the hydrophobic substrate partially exits from the channel; the substrate is indicated in yellow, the LG helices are indicated in green, and the plug moiety is indicated in red.

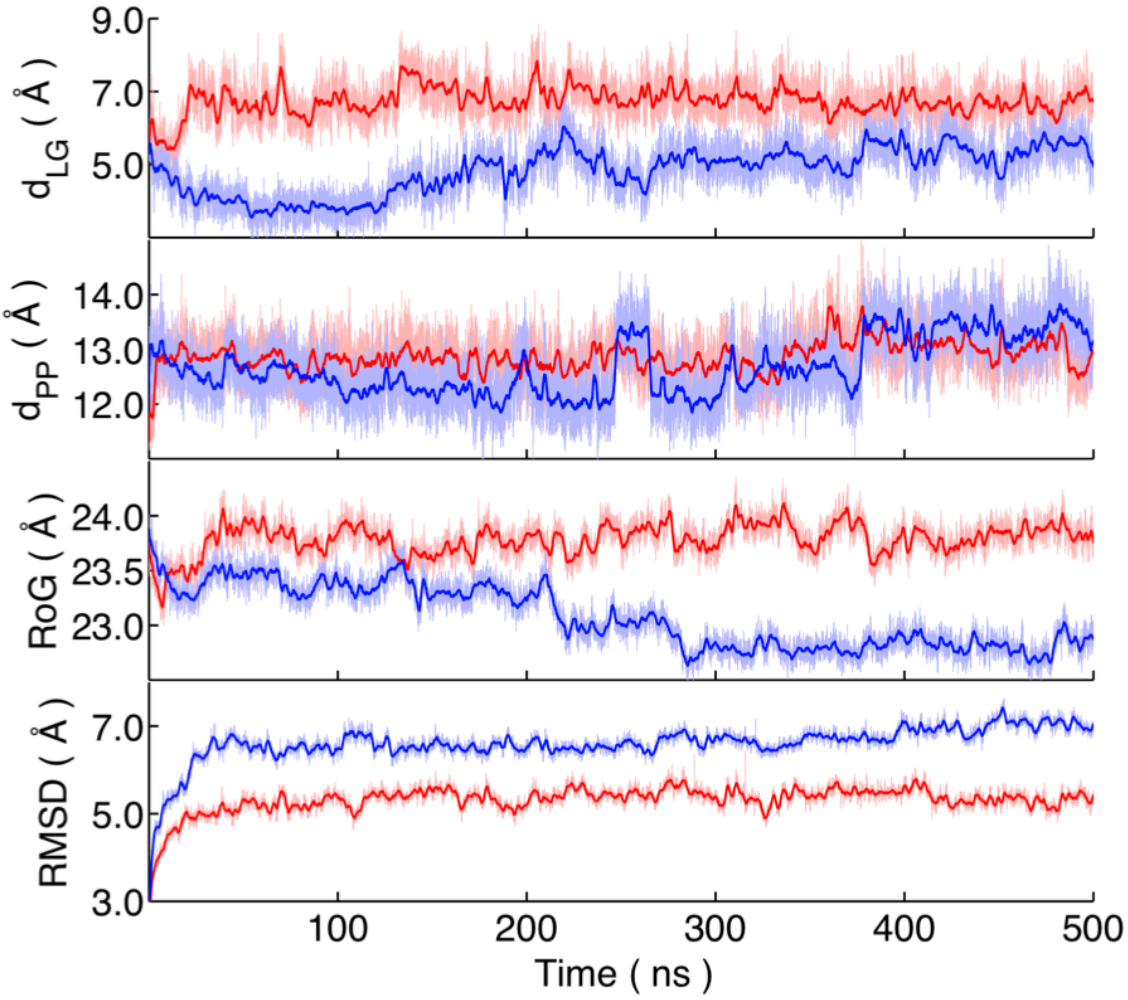

Fig. S 11. CG trajectories with scaffolding (red) and without scaffolding (blue) are plotted as a function of several collective variables. The data for the radius of gyration (RoG) and the backbone RMS displacement are the same as in Fig. 2C. Heavier lines indicate 1 ns rolling averages.

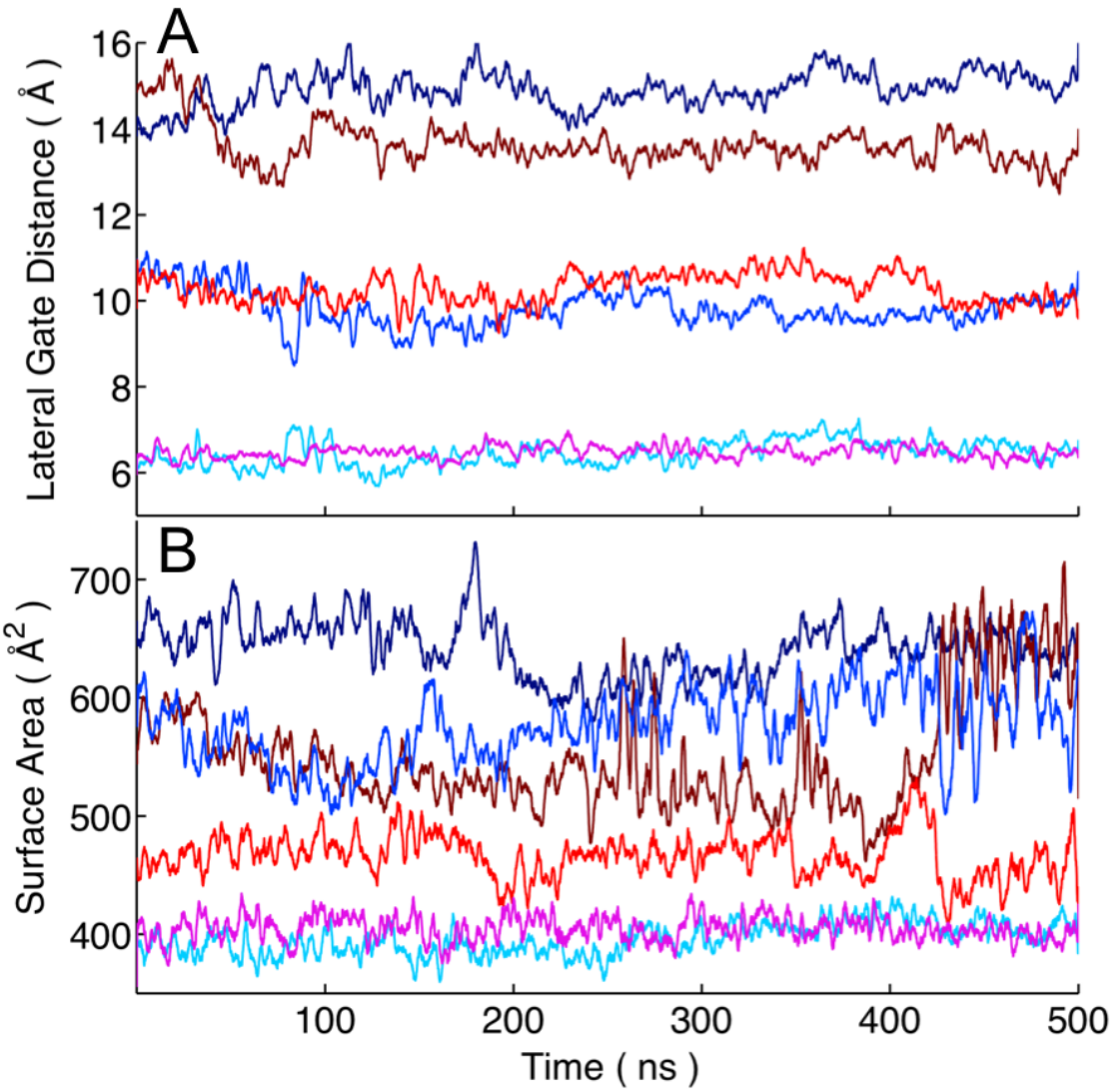

Fig. S 12. CG MD trajectories for the translocon with a substrate of intermediate hydrophobicity, plotted as a function of (A) LG distance and (B) LG surface area. These trajectories are initialized from the same configurations as the six trajectories reported in Fig. 5 (main text), and the color scheme identifies which of the trajectories shares the same initial configuration as a given trajectory in Fig. 5. The lines indicate 1 ns rolling averages.

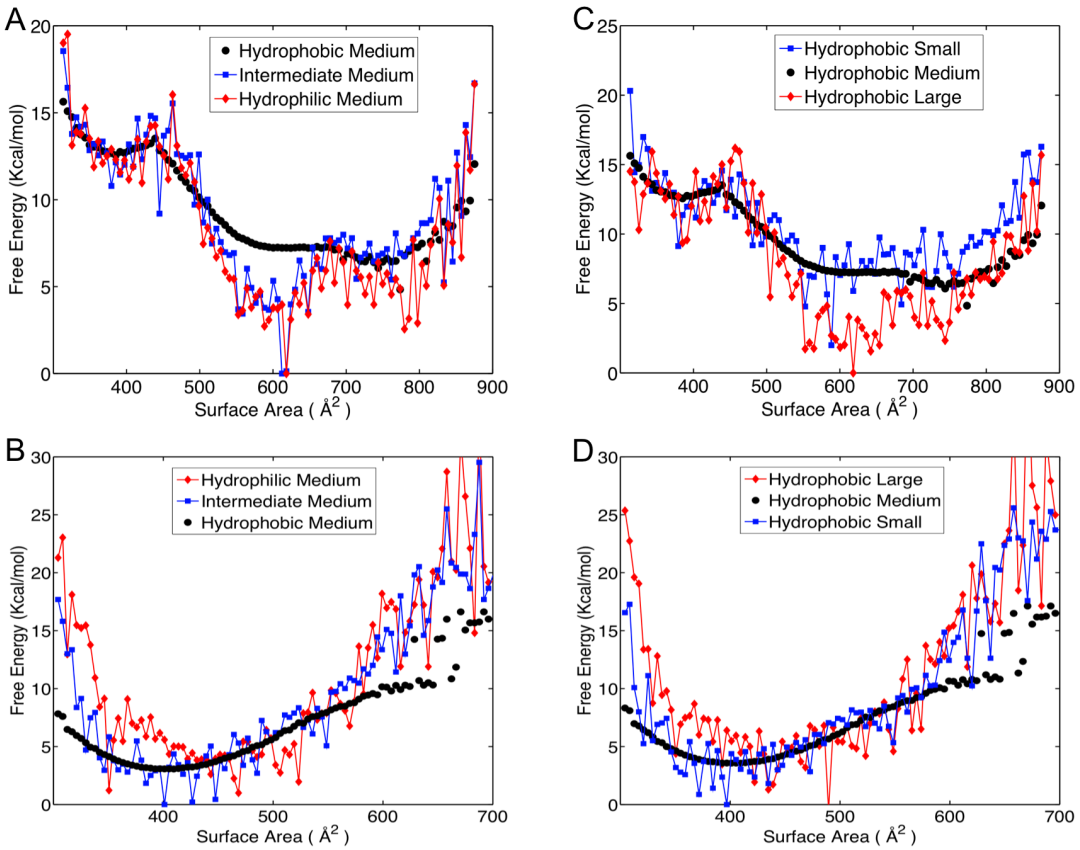

Fig. S 13. Mutations in the translocon pore residues alter the free energy profiles in Fig. 7 of the main text. (A) and (B) The dependence of the free energy profiles on the pore residue hydrophobicity for the translocon with (A) hydrophobic and (B) hydrophilic substrate. (C) and (D) The dependence of the free energy profiles on the pore residue bulkiness for the translocon with (C) hydrophobic and (D) hydrophilic substrate. See text for details.

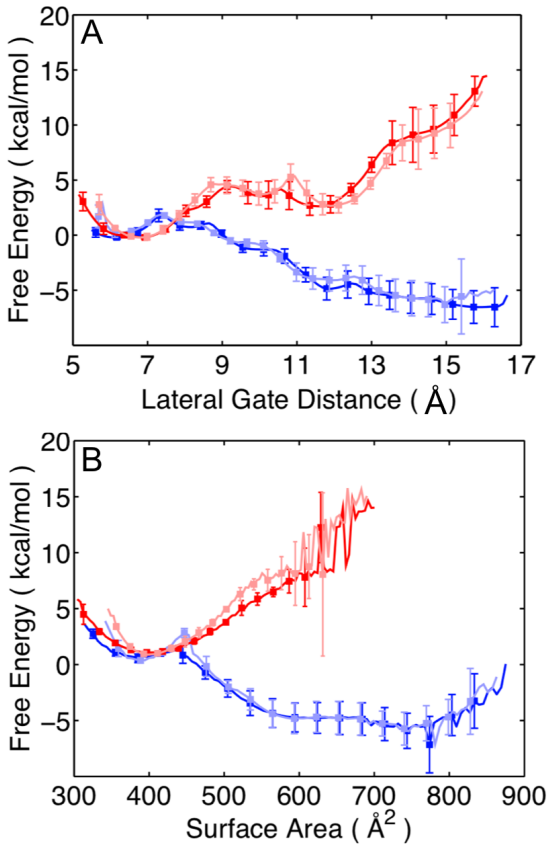

Fig. S 14. Robustness tests for the definition of (A) the LG distance collective variable and (B) the LG surface area collective variable. The dark curves are reproduced from Fig. 7 in the main text. The lighter curves use alternative definitions for the collective variables, described in the text.