Supporting information for: Reconciling the roles of kinetic and thermodynamic factors in membrane-protein insertion

James C. Gumbart,[†] Ivan Teo,[‡] Benoît Roux,[*,†,¶] and Klaus Schulten[*,‡]

Biosciences Division, Argonne National Laboratory, Argonne, IL 60439, and Beckman Institute and Department of Physics, University of Illinois at Urbana-Champaign, Urbana IL 61801

E-mail: roux@uchicago.edu; kschulte@ks.uiuc.edu

[*]To whom correspondence should be addressed

[†]Argonne National Laboratory

[‡]University of Illinois at Urbana-Champaign

[¶]Department of Biochemistry and Molecular Biology and Gordon Center for Integrative Science, The University of Chicago, Chicago, IL 60637

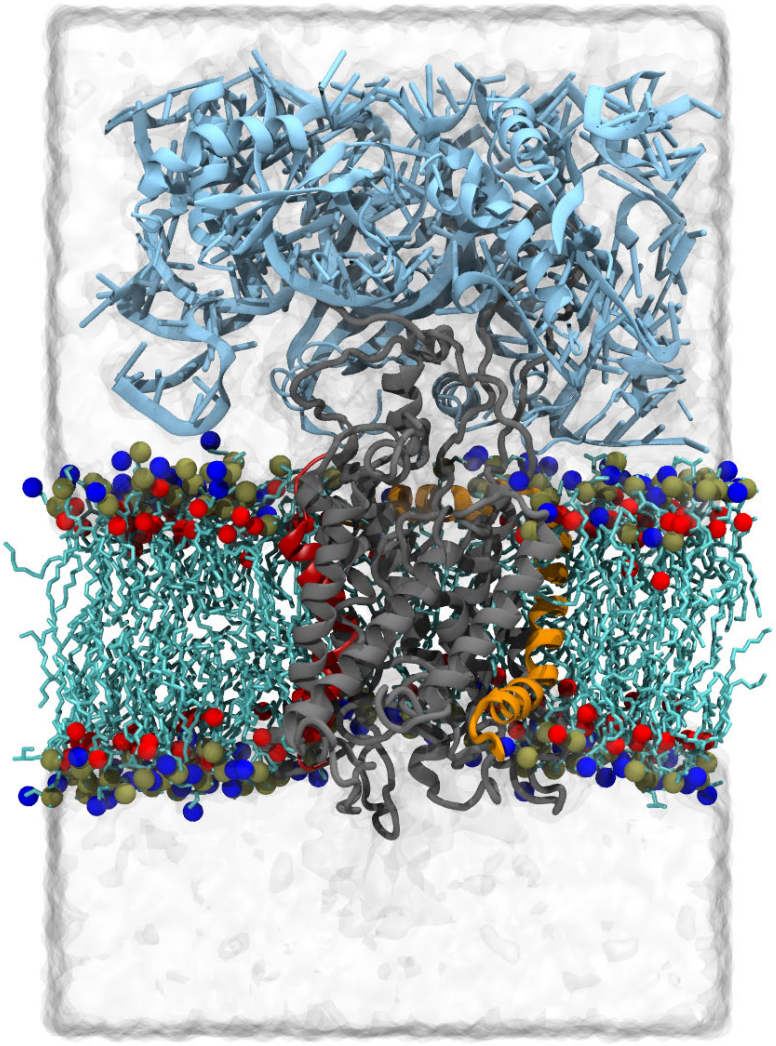

Figure S1: Simulated system used for the Anton simulations. SecY is shown in grey, SecE in orange, and the portion of the ribosome retained in light blue. The lipid tails are in cyan with selected phosphorus, nitrogen, and oxygen atoms of the head groups displayed as brown, blue, and red spheres, respectively. The water box for the periodic system is in light grey.

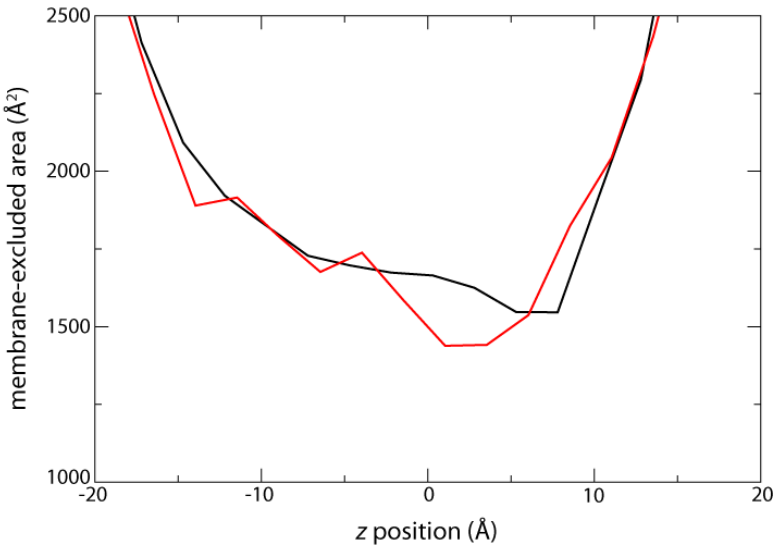

Figure S2: Area inaccessible to lipids as a function of z, the position along the membrane normal. The black curve is derived from simulation of the native SA embedded in SecY with an initially closed lateral gate, while the red is from simulation with an initially open gate (see Fig. 1 in the main text). They are nearly overlapping, with the exception of a small region near the center of the membrane where lipid tails can breach the open gate into the channel. The curves are averaged over the course of each simulation.

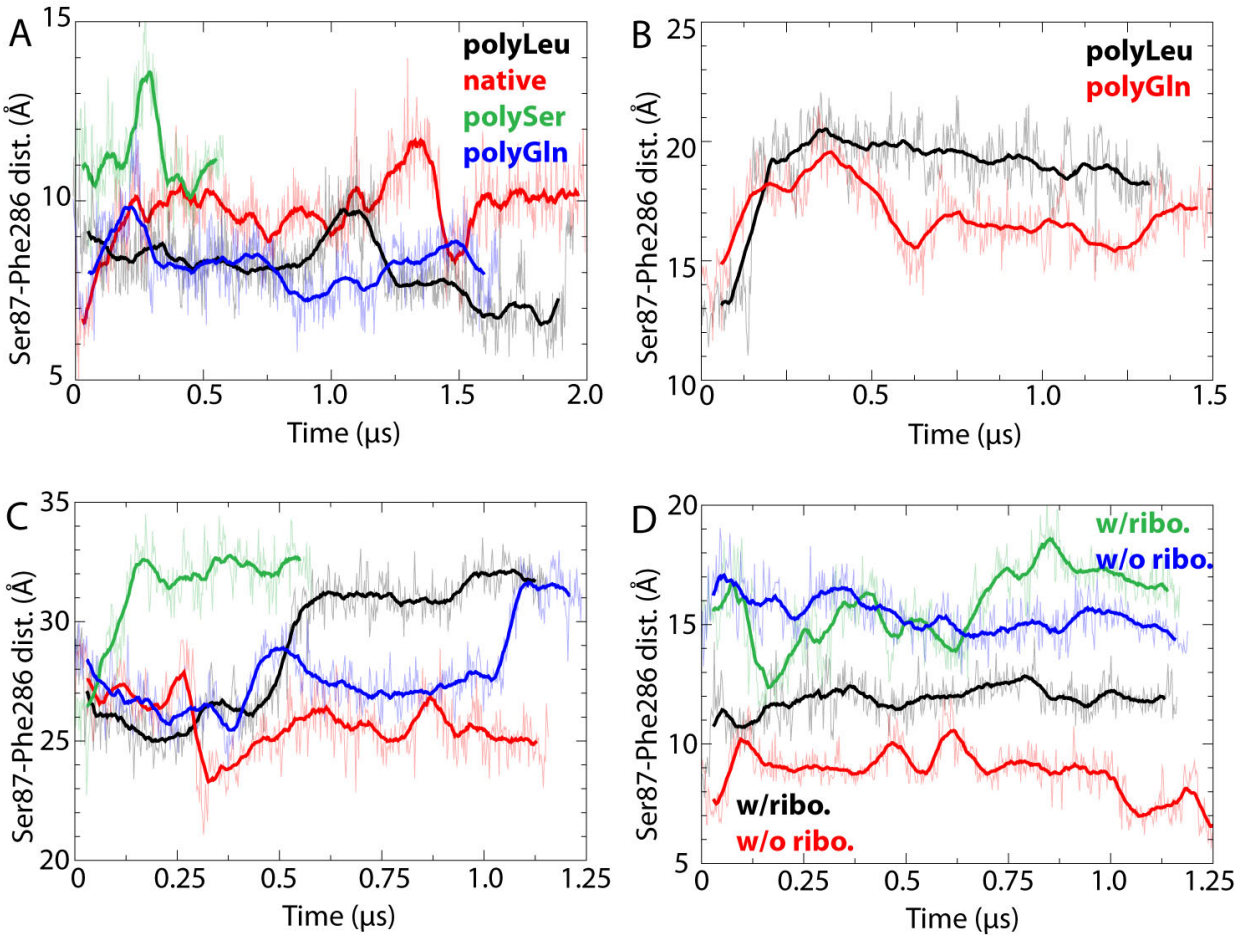

Figure S3: Plots of Ser87-Phe286 distance over time with different nascent helical TM segments embedded in SecY's central pore. (A) Initially closed SecY. Gate opening for hydrophobic helices (polyLeu and SA) are shown in black and red, respectively, with hydrophilic ones (polySer and polyGln) in green and blue. (B) Initial intermediate opening with polyLeu (black) and polyGln (red) inside. (C) Initially open, colored the same as in (A). (D) Empty SecY started in a closed state (red/black) and in a state of intermediate gate opening (blue/green), with and without the ribosome bound.

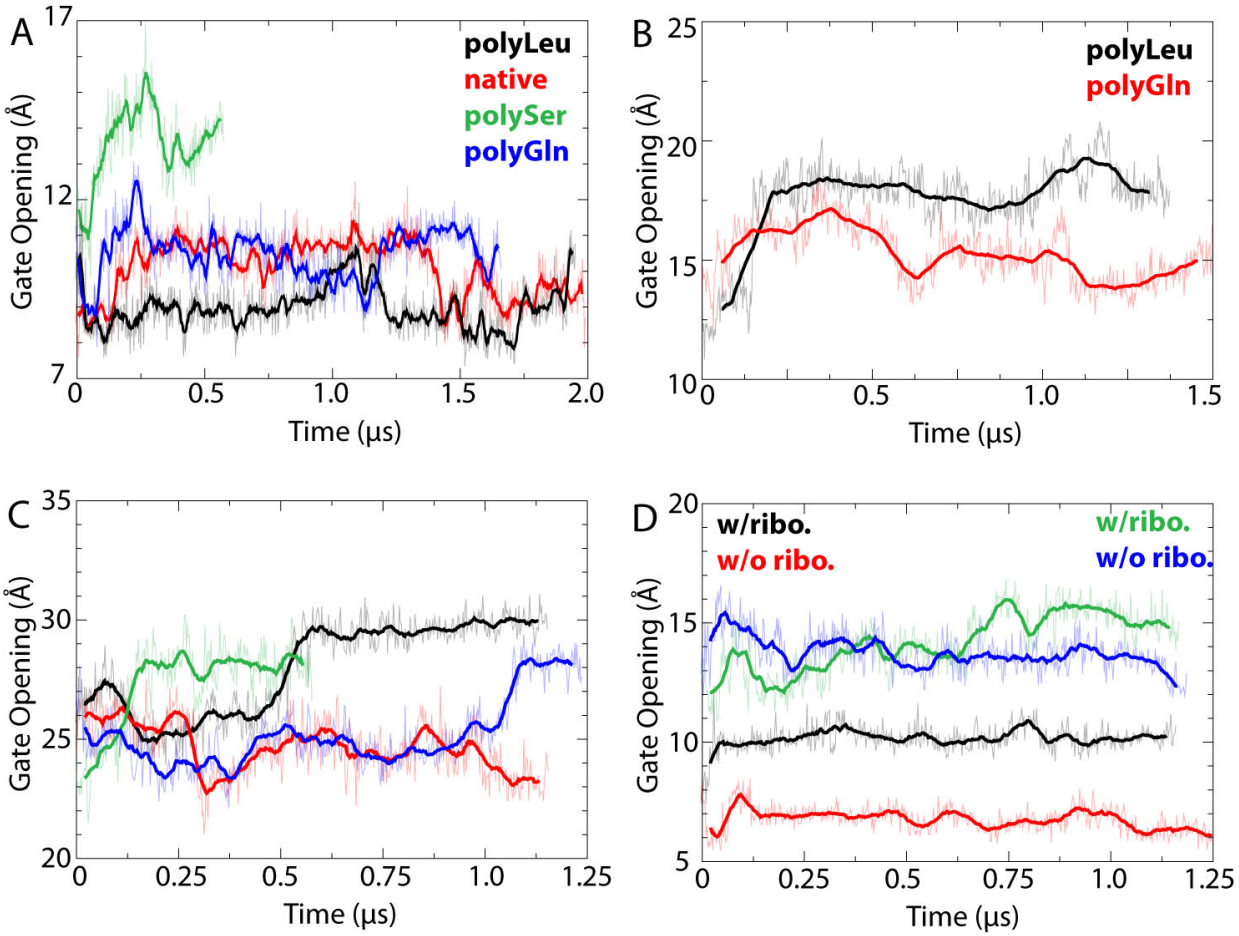

Figure S4: Lateral gate opening over time, identical to Fig. S3, except a different definition of the gate opening is used. Here, gate opening is measured between the centers of mass for the C_α atoms of residues 83, 86, 87, 91, and 94 on TM2b of SecY and 279, 282, 283, 285, and 286 on TM7 (*E.coli* numbering). These residues are a subset of those used to define gate opening in Zhang and Miller (2010) (1).

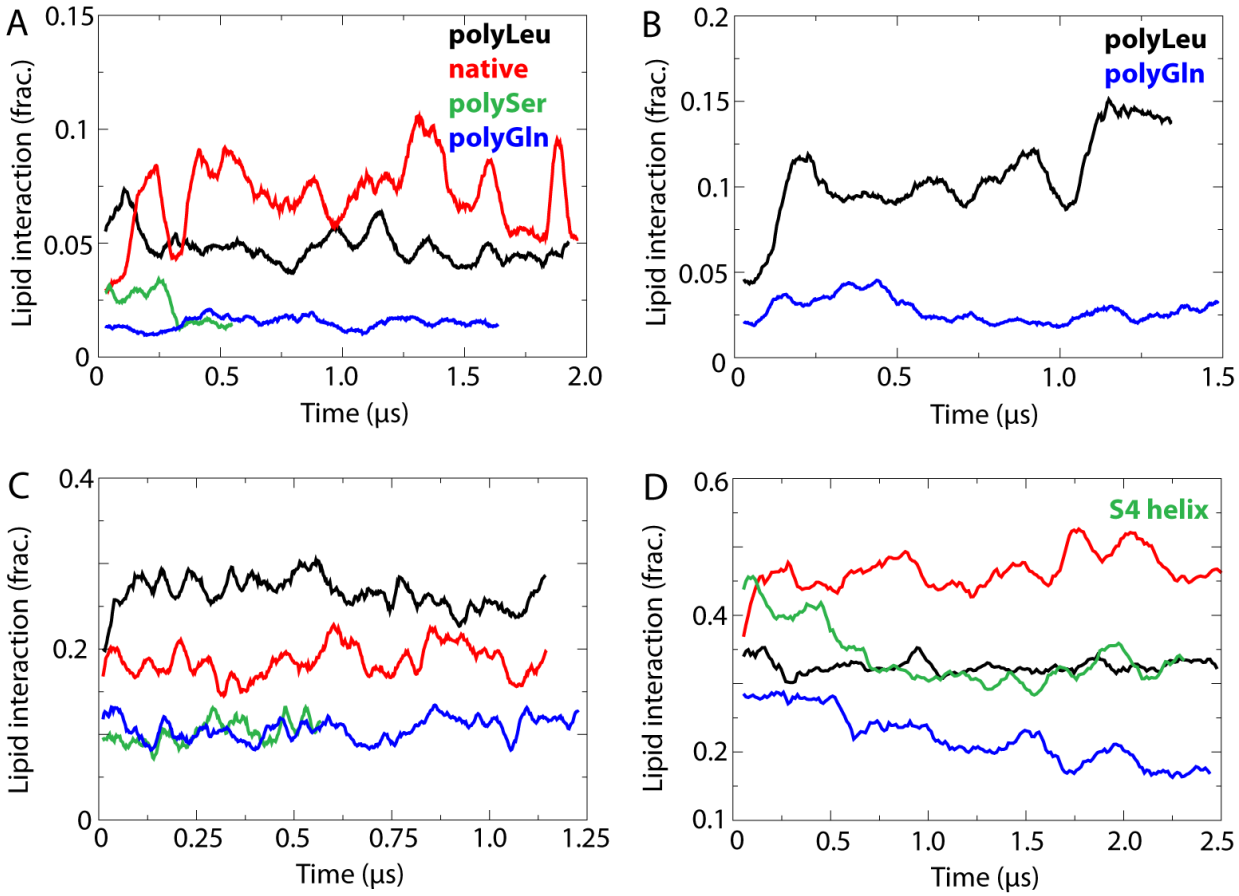

Figure S5: Interaction between lipids and a tested helix as a fraction of its total surface area. (A-C) Lipid-interaction area for helices embedded in the center of the channel, corresponding to those in Fig. S3A-C. (A) Initially closed SecY. (B) SecY with an intermediate opening initially. (C) Initially open SecY. (D) Change in interaction area for helices initially positioned at the lateral gate (see Fig. 2 in the main text). Helices are colored as in (A), except for polySer, which is replaced by the S4 helix here.

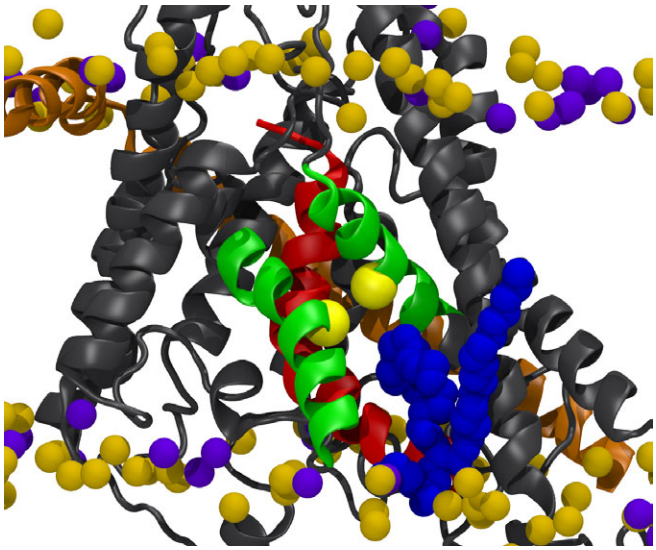

Figure S6: Incursion of a lipid into the closed channel after $\sim$0.5 μs. SecY is shown in grey, SecE in orange, and the lateral gate helices in green. The nascent SA helix (red) is still within the predominantly closed channel, indicated by the proximity of residues Ser87 and Phe286 (yellow spheres). A lipid contacting the SA across the otherwise closed gate is shown in a blue space-filling representation.

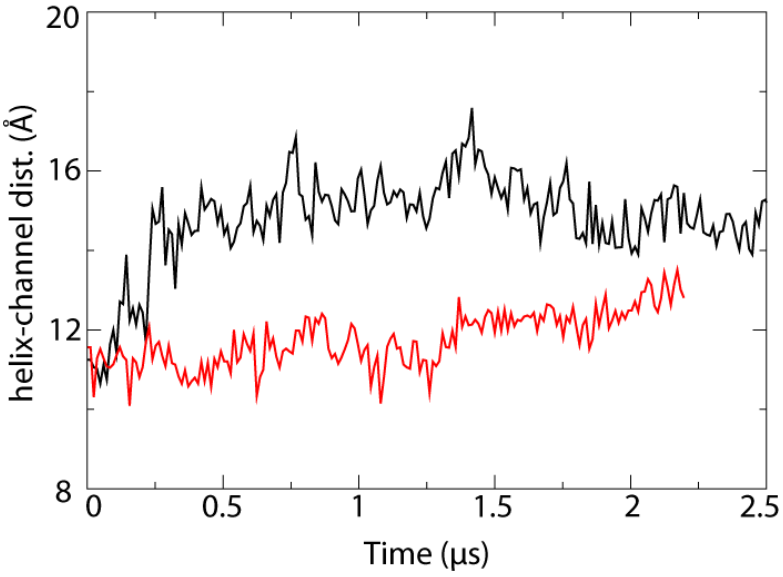

Figure S7: Spontaneous motion of the native SA in SecY. The separation between the center of SecY and the SA is shown over time as in Fig. 2G in the main text. The black curve represents simulation at $T = 353$ K and the red curve at $T = 323$ K. The same trend of the helix moving into the membrane is observed for both, although the one at higher temperature is more rapid.

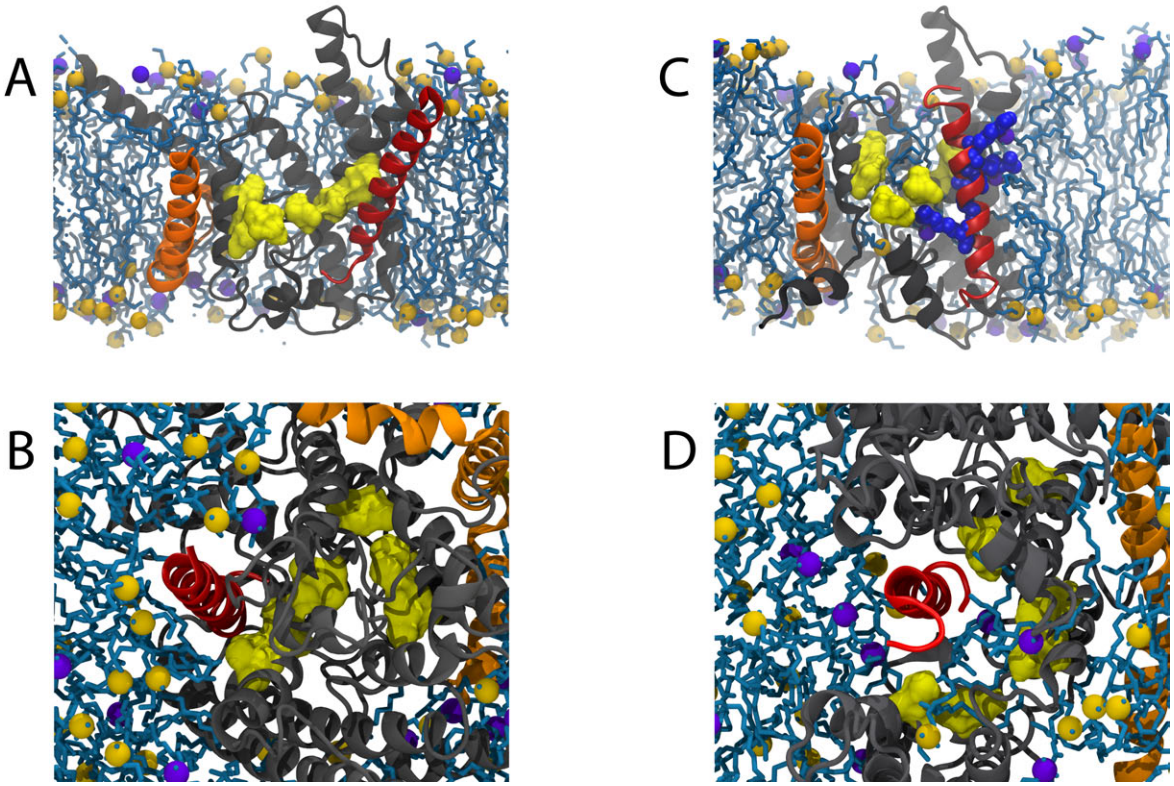

Figure S8: Spontaneous motion of a helix in SecY. SecYE, the tested helix, and the membrane are displayed as in Fig. 2 in the main text. (A,B) Final state (t=2.5 μs) for the SA shown in the membrane plane (A) and from the cytoplasmic side (B). (C,D) Final state of the S4 helix. The four arginine residues are shown in a blue space-filling representation, highlighting their interactions with lipid head groups.

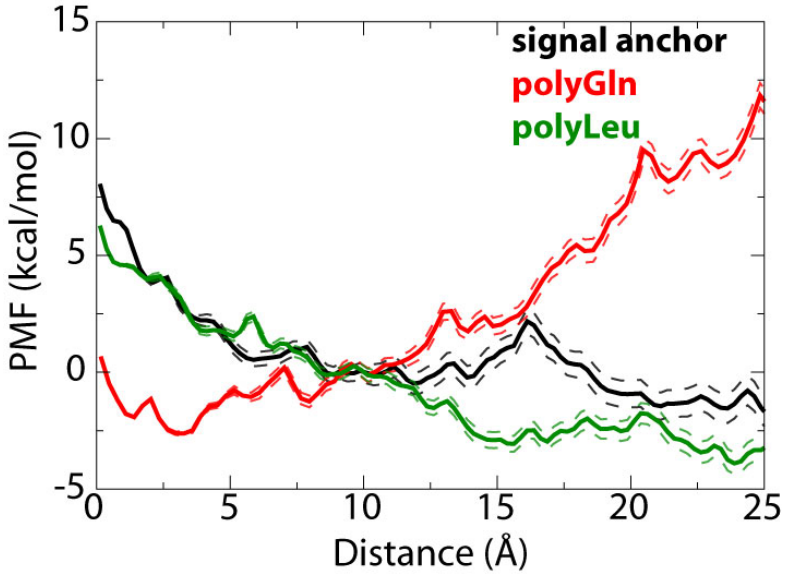

Figure S9: Potentials of mean force from Fig. 3 in the main text. The dashed lines above and below each heavy line represent one standard deviation away from the mean. The statistical error was calculated using the method of Zhu and Hummer (2).

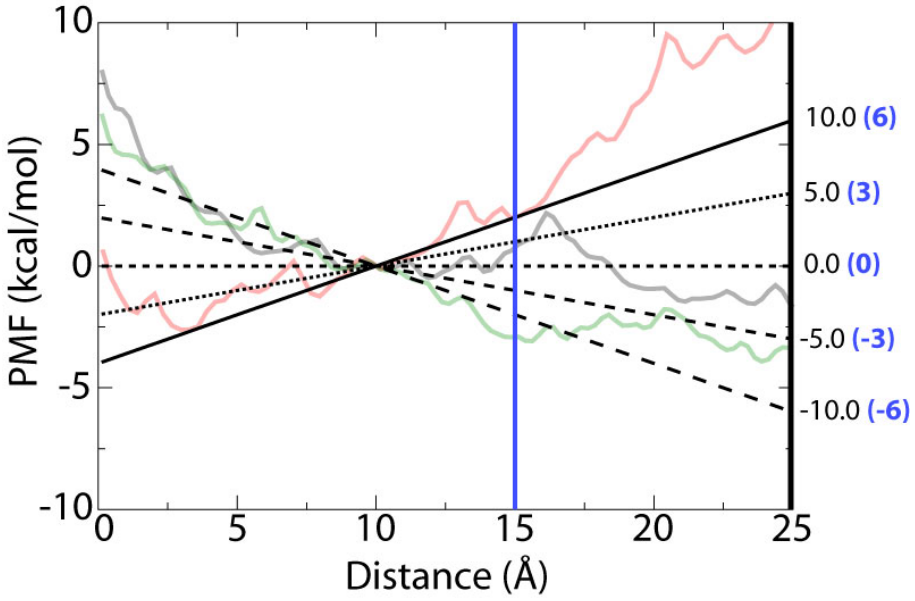

Figure S10: Simplified potentials of mean force. Linear PMFs (black lines, solid and dashed) generated as inputs for the kinetic model are shown as a function of distance from the center of the channel, $r_{\mathrm{SecY}} = 0$. The solid blue and black vertical lines correspond to the reference points chosen for the membrane, $r_{\mathrm{mem.}} = 15$ and $r_{\mathrm{mem.}} = 25$ Å, respectively (see Fig. 5 in the main text and Fig. S11). The numbers to the right give $\Delta G(\mathrm{SecY} \rightarrow \mathrm{mem.})$ at the corresponding reference point for each linear PMF. The PMFs for the SA, polyGln, and polyLeu helices determined from umbrella-sampling simulations are shown in the background, colored as in Fig. 3 in the main text.

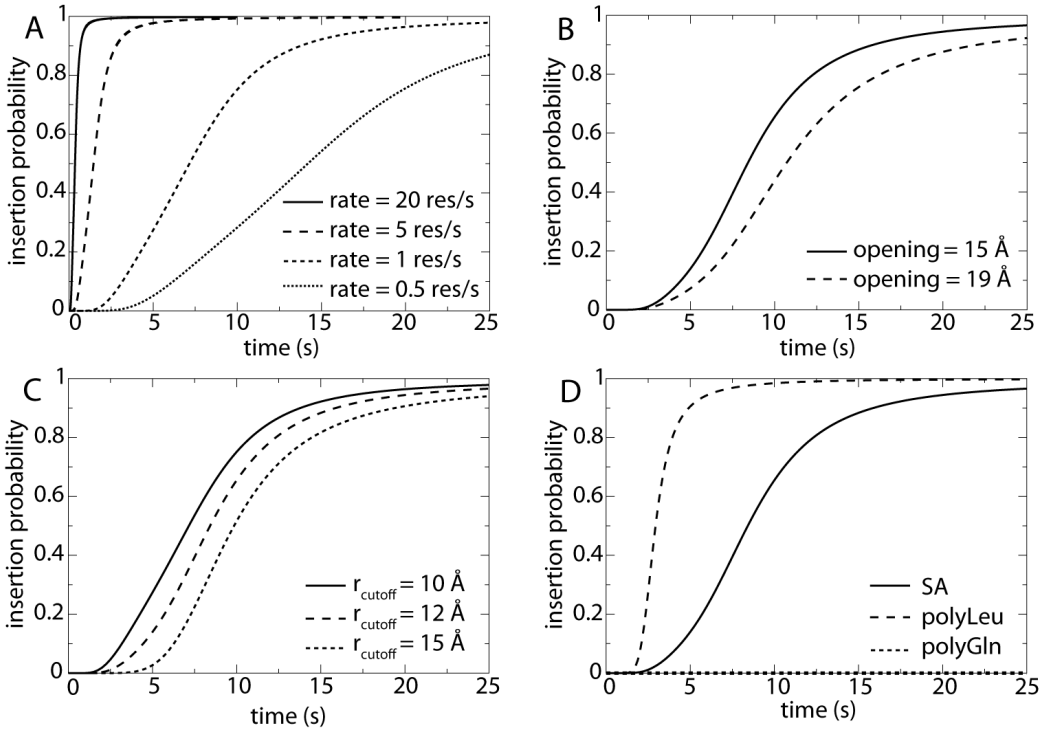

Figure S11: Probability of membrane insertion. The baseline parameters used are a translation rate of 1 res/s, a gate opening of 15 Å, and $r_{\text{cutoff}} = 12$ Å. (A-C) Insertion of SA as a function of (A) translation rate, (B) gate opening, and (C) $r_{\text{cutoff}} = 12$ Å. (D) Insertion of SA compared to the polyLeu and polyGln helices.

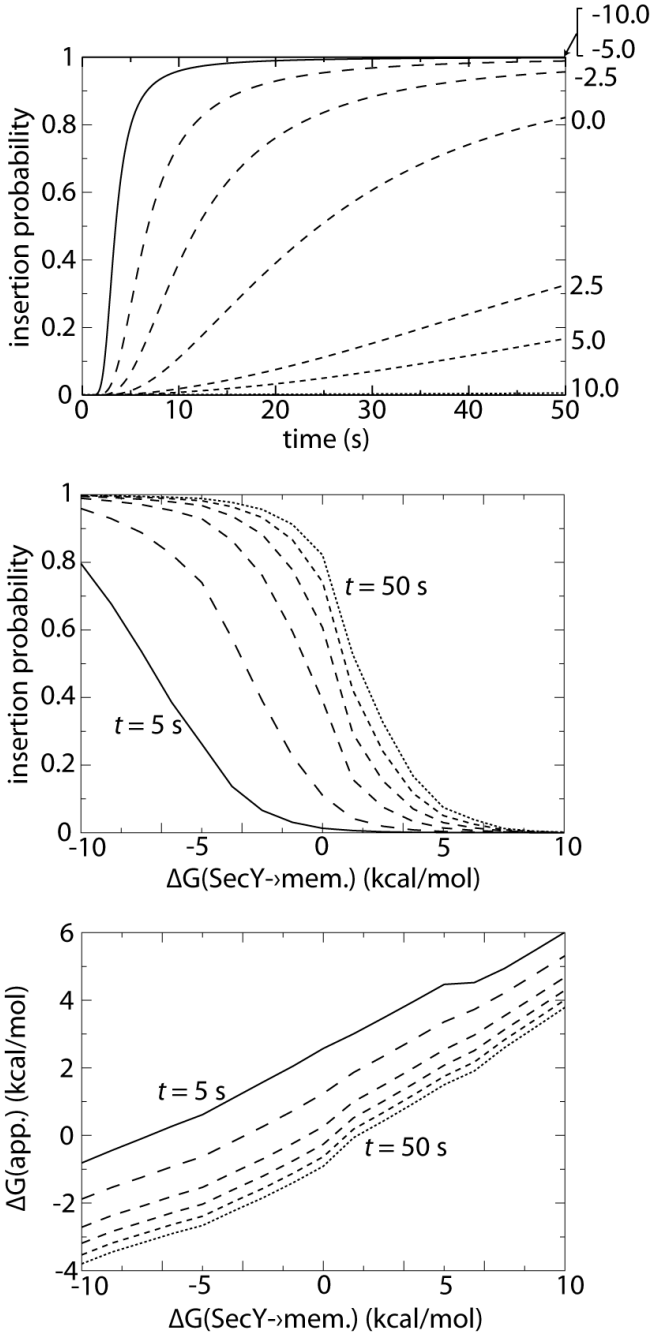

Figure S12: Membrane-insertion probability based on simplified PMFs for the membrane reference point $r_{\text{mem.}} = 25$ Å. All parts are otherwise identical to Fig. 5 in the main text.

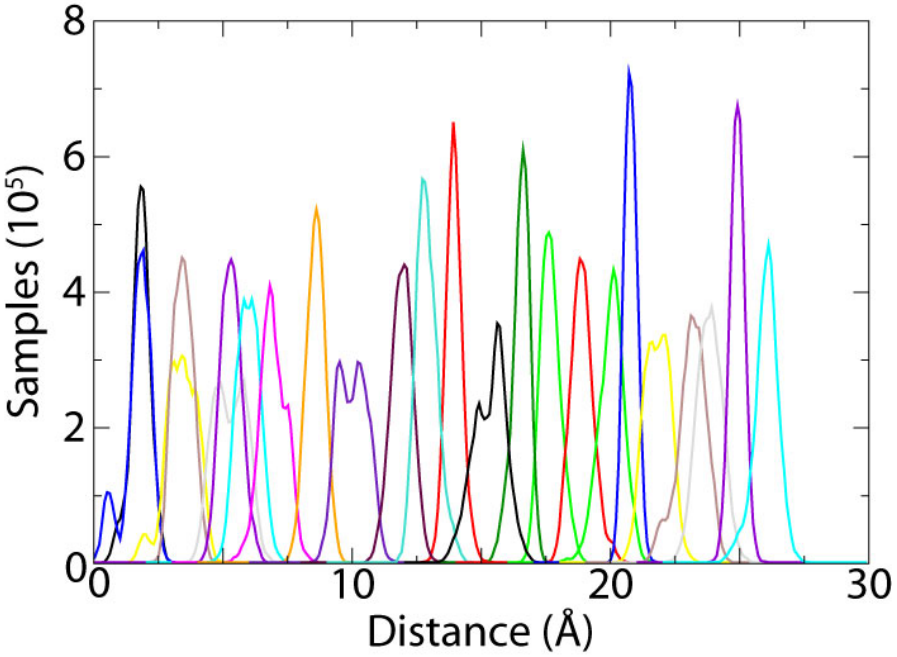

Figure S13: Histograms for 24 umbrella-sampling simulations used to reconstruct the PMF for the native SA.

Potential of mean force

The radial potential of mean force (PMF) used for the finite-element models is composed of two parts. The first component U_1 is due to interactions between the helix and the environment, determined from all-atom umbrella sampling MD simulations or from simplified linear fits. The second component U_2 is due to the restraint imposed by the remainder of the ribosome-bound nascent chain. Because the helix inside SecY can temporarily arrest translocation, the downstream nascent chain can accumulate in the cytoplasm (3). This nascent chain is modeled as a freely-jointed chain such that, in the absence of an underlying potential, the mean end-to-end length is given by

$$\langle r^2 \rangle = N_r L^2 \tag{1}$$

where N_r is the number of residues in the chain and $L = 3.8$ Å, the Kuhn length, is the average residue length in an extended polypeptide chain. The length r is taken to represent the range in which the helix can diffuse away from the channel. We assume that the chain is growing at a constant rate so that $N_r(t) = t/\tau$, thus making r proportional to $\sqrt{t}$. The characteristic time τ is that required to synthesize one residue by the ribosome and is set to 1 s in most calculations (3, 4), although other possibilities are considered in Fig. 5A in the main text. The length r has a probability distribution (5)

$$\rho_{\text{free}}(r, N_r) = \exp(-3r^2/2N_r L^2) \tag{2}$$

with an associated potential of

$$U_{\text{free}}(r, N_r) = 3r^2/(2\beta N_r L^2). \tag{3}$$

Hence, with $T = 300$ K and $\tau = 1$ s, the resulting potential $U(r,t)$ is

$$U(r,t) = U_1(r) + U_{\text{free}}(r, N_r) = U_1(r) + U_2(r,t) = U_1(r) + \frac{\alpha r^2}{t} \tag{4}$$

where $\alpha = 0.062$ (kcal/mol)·(s/Å^2).

Finite-element kinetic models

Diffusion within the lipid bilayer of the nascent chain around the SecY channel is described by a time-evolving probability density. This probability density is defined over a 2D circular plane segment $\mathscr{S}$ centered on the membrane-normal axis of the channel. The circumferential boundary of $\mathscr{S}$ is reflective so as to close the system, making possible the use of the Boltzmann approach described later. The radius of $\mathscr{S}$ is set to 2000 Å, a large value compared to the extent of the channel, so as to minimize boundary effects. This radial extent is further validated by the fact that the potentials of mean force typical in this study present an effectively insurmountable barrier at a radius much smaller than 2000 Å.

For the purpose of simulating diffusion in this system, we employ a discretization of the plane into a collection of N non-overlapping regions $A = \{a_k \subset \mathscr{S} : k = 1,2,\ldots,N,\ \bigcup_k a_k = \mathscr{S},\ a_i \cap a_j = \emptyset\ \forall\ i \neq j\}$ that form a grid over the system. This discretization involves defining on the plane a set of random positions $\mathscr{W} = \{\mathbf{w}_k : k = 1,2,\ldots,N\}$, called cell centers, and subsequently calculating the Voronoi tessellation ($\mathbf{r} \in a_k \iff |\mathbf{r} - \mathbf{w}_k| < |\mathbf{r} - \mathbf{w}_i|\ \forall\ i \neq k$) induced by $\mathscr{W}$. For the latter purpose, the DelaunayTri() and VoronoiDiagram() classes of commands were applied to $\mathscr{W}$ in Matlab (6).

Prior to Voronoi tessellation, $\mathscr{W}$ is initialized as follows. First, the positions of the cell centers are randomly assigned on $\mathscr{S}$. A topology-preserving algorithm (7) is then used to adapt the positions of the cell centers such that they follow a spatial distribution function $\rho(\mathbf{r})$, which is pre-set to allocate cell center densities according to the local level of detail required to describe different regions in the system. For example, a higher level of detail is required for regions near the complicated surface of the channel than for the membrane far away from the channel. The algorithm used efficiently reproduces the distribution obtained by the k-means clustering algorithm (7). In particular, the adaptive distribution approaches one that induces a centroidal Voronoi tessellation (CVT) (8). Heuristically, since a spatial variable $X(\mathbf{r})$ at position $\mathbf{r} \in a_k$ will be approximated as its evaluation $X(\mathbf{w}_k)$ at the corresponding cell center, a CVT is advantageous because the cell center $\mathbf{w}_k$ would then coincide with the centroid of a_k.

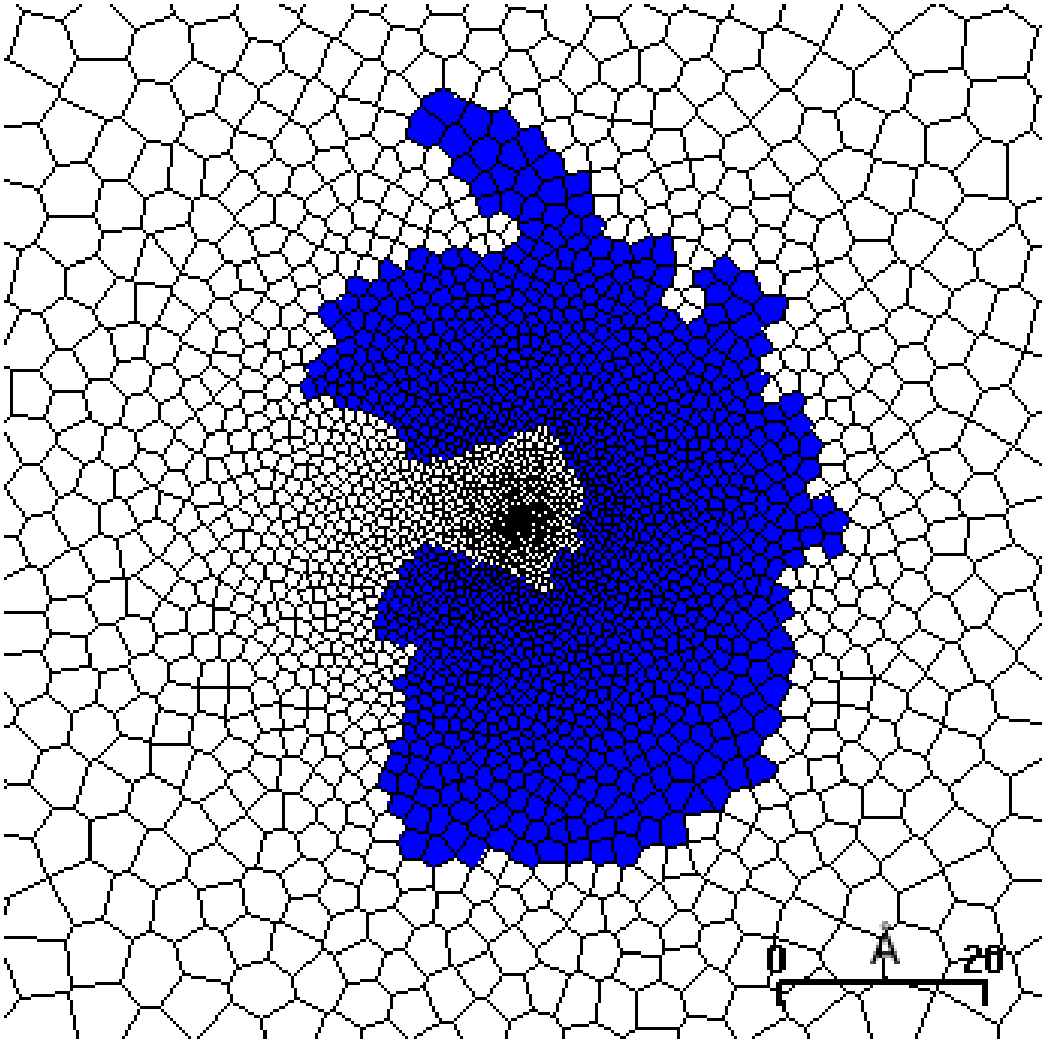

Figure S14: Voronoi tessellation of membrane plane segment, enlarged to show SecY cross-section (in blue). The density of cells is higher within the channel so as to describe diffusive behavior inside the channel in greater detail.

The geometry of the SecY channel is modeled as a reflective boundary by the following means: From the PDB file of the laterally open SecY channel, a 2D projection of the protein atoms on a membrane-parallel plane was taken. This projection was then overlaid on the Voronoi tessellation, and those cells whose centers fell within a certain radius of any atom were identified as reflective. This radius is given by the CHARMM van der Waal's radius of the atom plus 3 Å, which approximates the closest distance of approach of the nascent helix center to the channel walls.

The time evolution of the probability $p_i(t)$ of finding the nascent helix within cell a_i is calculated using two independent methods. The first method involves solving the discretized Smoluchowski equation on the grid. The second method, in which the Boltzmann distribution is calculated at each

time step, is validated by comparison with the more precise, but also more computationally expensive, Smoluchowski method.

Smoluchowski model

Given the centers $\mathbf{w}_k$ and the induced Voronoi tessellation A, a rate matrix R was constructed with elements R_{ij} giving the rate of transfer from cell j to cell i. R_{ij} is calculated from the Smoluchowski equation

$$\partial_t p(\mathbf{r},t) = \nabla \cdot De^{-\beta U(\mathbf{r},t)} \nabla e^{\beta U(\mathbf{r},t)} p(\mathbf{r},t) \tag{5}$$

where D is the diffusion coefficient, $\beta = 1/kT$, and $U(\mathbf{r},t)$ is the time-dependent potential of mean force as described in section Figure 13.

Integrating the Smoluchowski equation over the region a_i gives

$$\begin{aligned} \int_{a_i} \partial_t p(\mathbf{r},t) d\mathbf{r} &= \int_{a_i} \nabla \cdot De^{-\beta U(\mathbf{r},t)} \nabla e^{\beta U(\mathbf{r},t)} p(\mathbf{r},t) d\mathbf{r} & (6) \\ &= \int_{\partial a_i} De^{-\beta U(\mathbf{r},t)} \mathbf{n}(\mathbf{r}) \cdot \nabla e^{\beta U(\mathbf{r},t)} p(\mathbf{r},t) d\mathbf{r} & (7) \\ &= \int_{\partial a_i} De^{-\beta U(\mathbf{r},t)} \frac{\partial e^{\beta U(\mathbf{r},t)} p(\mathbf{r},t)}{\partial \mathbf{n}(\mathbf{r})} d\mathbf{r} & (8) \end{aligned}$$

where $\mathbf{n}(\mathbf{r})$ is the unit normal vector on the boundary of a_i.

Finally, the above expression is discretized by converting integrals into sums and derivatives into differences as follows:

$$V_i \dot{p}_i(t) = \sum_j De^{-\beta U_{ij}(t)} \frac{e^{\beta U_j(t)} p_j(t) - e^{\beta U_i(t)} p_i(t)}{n_{ij}} \cdot L_{ij} \tag{9}$$

where V_i is the area of a_i, j indexes the immediate neighbors of a_i, U_i is the potential of mean force in a_i, U_{ij} is the potential of mean force at the midpoint of the line between $\mathbf{w}_i$ and $\mathbf{w}_j$, n_{ij} is the distance between $\mathbf{w}_i$ and $\mathbf{w}_j$, and L_{ij} is the length of the interface between a_i and a_j.

Thus, it can be inferred from the above expression that the transition rate from a_j to a_i is given by

$$R_{ij} = D\frac{e^{\beta(U_i-U_j)/2}L_{ij}}{V_i n_{ij}} \tag{10}$$

where we have further approximated $U_{ij} = (U_i + U_j)/2$. This approximation proves advantageous in that the resulting rate matrix obeys detailed balance.

With the rate matrix R defined, we can now obtain the solution of the equation

$$\dot{\mathbf{p}} = R\mathbf{p}. \tag{11}$$

Using the Matlab eigen-solving command eig(), we calculate the eigenvalues λ_i and corresponding eigenvectors v_i, where $i = 1, 2, \ldots, N$. Finally, the probability flowing into cell a_k at time $t + \tau$ is given by $\mathbf{p}(t+\tau) = \sum_l A_l v_l \exp(\lambda_l \tau)$ where A_l are calculated by setting the initial condition $\mathbf{p}_0 = \mathbf{p}(t)$.

The simulation runs on two distinct time steps δt and $\Delta t > \delta t$. The time-dependent potential is updated every Δt steps, and R is re-calculated. On the other hand, $\mathbf{p}$ is evaluated every δt.

Validation of the Smoluchowski model

We validate our model against the analytic solution for free diffusion in two dimensions. We solve the discretized Smoluchowski equation as described above, using a uniform potential and reflective boundary conditions at a large radius of 2000 Å so as to approximate diffusive behavior in an infinite plane. With the particle initially at the origin, the probability, p_{theory}, takes the form of a spreading Gaussian distribution centered on the origin, i.e.,

$$p_{\text{theory}}(r,t) = (4\pi D t)^{-1} \exp(-r^2/4Dt). \tag{12}$$

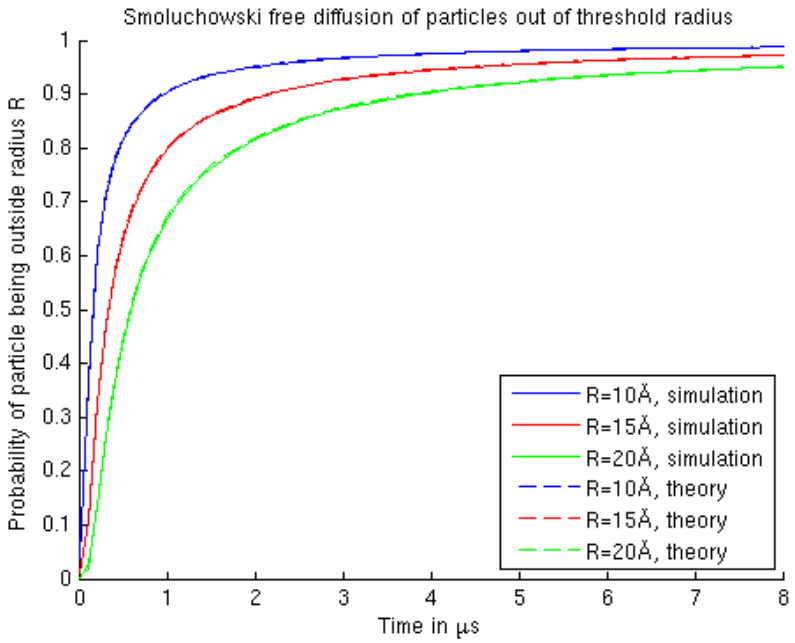

Figure S15: Results for free diffusion in two dimensions with system radius 2000 Å; 5000 Voronoi cells in the distribution with a density of $\rho(r|r < 20\,\text{Å}) = \text{constant}$, $\rho(r|r > 20\,\text{Å}) \propto (r - 15\,\text{Å})^{-1}$; and a diffusion constant $D = 250\,\text{Å}^2/\mu\text{s}$. Proportion of particles outside radius r_{cutoff} were calculated for $r_{\text{cutoff}} = 10$ Å, 15 Å, and 20 Å.

Hence, the probability of finding the particle outside a radius r_{cutoff} is

$$P_{\text{theory}}(t \mid r > r_{\text{cutoff}}) = 1 - \int_0^{r_{\text{cutoff}}} dr(2\pi r) p_{\text{theory}}(r,t) \tag{13}$$

$$= 1 - \int_0^{r_{\text{cutoff}}} dr(2\pi r)(4\pi Dt)^{-1} \exp(-r^2/4Dt) \tag{14}$$

$$= \exp(-r_{\text{cutoff}}^2/4Dt) \tag{15}$$

The results from the Smoluchowski model for typical values of r_{cutoff} used in this study agree well with theory (see Figure 15). However, the simulations used in this study are not based on the Smoluchowski model due to its computationally expensive calculations. Instead, data generated by the Smoluchowski model was used to test the validity of the Boltzmann model, described below, for the spatial and temporal scales typical of the simulations used here.

Boltzmann Model

The Boltzmann model assumes that the system reaches equilibrium within one time step δt. We justify this assumption by comparing the Smoluchowski and Boltzmann results of an example simulation using a linear radial potential of mean force superimposed on a time-dependent harmonic potential, typical of those used in this study. From Figure 16, it is apparent that the Smoluchowski model behavior closely agrees with the results of the Boltzmann model.

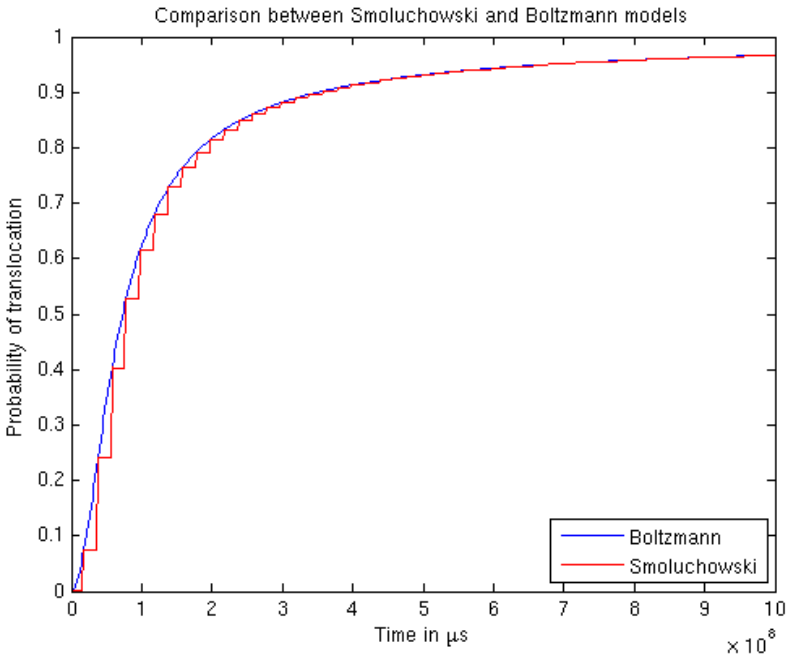

Figure S16: Results of Smoluchowski and Boltzmann simulations in a system of radius 2000 Å centered on the pore axis of the SecY channel. The parameters used include a time step $\delta t = 2$ s, a PMF update time step $\Delta t = 20$ s for the Smoluchowski model, and a diffusion coefficient $D = 250\,\text{Å}^2/\mu\text{s}$. The potential of mean force used is $U(r,t) = (0.1\,\text{kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-1})r + (0.0619\,\text{kcal}\cdot\text{s}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2})r^2/t$. The graph displays the proportion of particles found outside radius $r_{\text{cutoff}} = 15$ Å.

At every timestep δt in the Boltzmann model, the potential of mean force is updated and the corresponding Boltzmann distribution is calculated:

$$p(\mathbf{r},t) = K(t)\exp\left(-\beta U(\mathbf{r},t)\right) \tag{16}$$

where $K^{-1}(t) = \int_{\mathbb{S}} \exp\left(-\beta U(\mathbf{r},t)\right) d\mathbf{r}$ is a normalizing constant.

We tested the sensitivity of the results from the Boltzmann model to the Voronoi grid distribution

used by running a representative case ($r_{\text{cutoff}} = 15$ Å, translation rate=1 res/s, linear PMF slope=-0.2 kcal/mol·Å, channel opening=15 Å) on different grid distributions. As one can see from Figure 17A, the results are robust to changes in grid density gradients. In Figure 17B, the results begin to diverge noticeably when the total cell count falls below 1000, which is much smaller than the 5000 used in production runs. These tests demonstrate that the grid tessellation used for production runs are more than adequate to produce converged results.

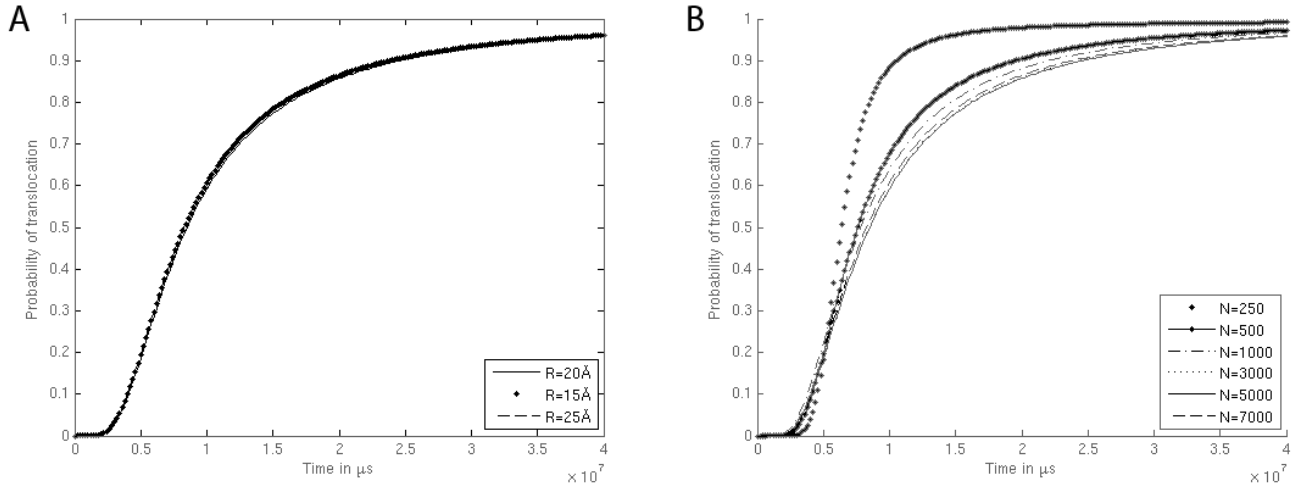

Figure S17: (A) Comparison of results using grid densities of the form $\rho(r|r < R) = \rho_0$, $\rho(r|r > R) \propto (r - (3/4)R)^{-1}$ for R=15 Å, 20 Å, 25 Å while keeping ρ_0 constant. This comparison tests for sensitivity to grid density gradients. (B) Comparison of results using R=20 Å and varying numbers of cells N. This comparison determines the threshold resolution beyond which results begin to diverge.

References

(1) Zhang, B.; Miller, T. F. Hydrophobically stabilized open state for the lateral gate of the Sec translocon. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 5399–5404.

(2) Zhu, F.; Hummer, G. Convergence and error estimation in free energy calculations using the weighted histogram analysis method. *J. Comp. Chem.* **2012**, *33*, 453–465.

(3) Cheng, Z.; Gilmore, R. Slow translocon gating causes cytosolic exposure of transmembrane and lumenal domains during membrane protein integration. *Nat. Struct. Mol. Biol.* **2006**, *13*, 930–936.

(4) Hessa, T.; Monné, M.; von Heijne, G. Stop-transfer efficiency of marginally hydrophobic segments depends on the length of the carboxy-terminal tail. *EMBO Rep.* **2003**, *4*, 178–183.

(5) Kuhn, W.; Grün, F. Statistical behavior of the single chain molecule and its relation to the statistical behavior of assemblies consisting of many chain molecules. *J. Polym. Sci.* **1946**, *1*, 183–199.

(6) MATLAB, *version 7.10.0 (R2010a)*; The MathWorks Inc.: Natick, Massachusetts, 2010.

(7) Martinetz, T. M.; Berkovich, S. G.; Schulten, K. "Neural Gas" for Vector Quantization and its Application to Time-Series Prediction. *IEEE Trans. Neur. Netw.* **1993**, *4*, 558–569.

(8) Q. Du, V. F.; Gunzburger, M. Centroidal Voronoi tessellations: Applications and algorithms. *SIAM Review* **1999**, *41*, 637–676.