Free-energy cost for translocon-assisted insertion of membrane proteins

Supporting Information

James Gumbart[1,2], Christophe Chipot[2], and Klaus Schulten[1,2]

[1]Department of Physics, and [2]Beckman Institute, University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

Justification for the use of perturbation theory.

Assuming that a reasonable model of a reaction coordinate can be defined, it is possible to determine the potential of mean force delineating membrane protein insertion using a variety of approaches (1). Whereas this route evidently offers the advantage of providing a detailed picture of the complete reaction pathway, it also suffers from shortcomings inherently rooted in the choice of the reaction coordinate. The most natural reaction coordinate, used in other studies (2–6), is the direction normal to the membrane. On the one hand, enhanced diffusion along this direction by means of importance sampling methods will necessarily be hampered by the slowest degrees of freedom of the molecular system, namely the collective relaxation of the lipid chains as the protein partitions into the membrane, which evidently cannot be embraced by a rudimentary, one-dimensional order parameter. On the other hand, charged amino acids are prone to exert a severe disruption of the lipid organization as they are translocated from the aqueous environment to the hydrophobic core of the bilayer (2). This disruption stems from the fact that ionic species do not migrate into the membrane as naked charged moieties, which would be enthalpically unfavorable, but are accompanied instead by a number of water molecules preserving ionic hydration, while necessarily altering the local structure of the lipid bilayer (7). In stratification strategies, wherein the reaction pathway is broken down into a number of contiguous, possibly overlapping windows, this deleterious effect can be artificially alleviated by reconstructing for each stratum the lipid bilayer around the moving charged amino acid (2). Altogether, this route, however attractive should the complete potential of mean force for membrane protein insertion be sought, can be prohibitive in terms of computational effort.

An appealing alternative, should one be merely interested in the net free-energy difference between the reference and the target states of the translocation, i.e., the end points of the reaction pathway, consists in resorting to perturbation theory (8). Under these premises, the reaction coordinate is a general-extent parameter that couples the unperturbed and the perturbed states of a so-called alchemical transformation, whereby one chemical species is altered into a distinct one. Using perturbation theory to measure the free-energy cost for membrane-protein insertion is anticipated to be appreciably more cost-effective than potential-of-mean-force computations. Free-energy perturbation (FEP) calculations suffer, however, from shortcomings of their own. In particular, they are burdened by acute convergence issues, which can be ascribed primarily to the difficulty of sampling from the reference state microstates representative of the target state. This difficulty is even more salient when the perturbation is large, but can be generally alleviated by means of stratification strategies, wherein the number of nonphysical intermediates is augmented to ensure appropriate overlap of the underlying ensembles of states (1).

Precision of the free-energy calculations.

To increase the statistical precision of the calculation (9), each simulation was run bidirectionally. The configurational ensembles obtained from the forward and backward transformations were subsequently combined by means of the Bennett acceptance ratio (BAR) method (10). The statistical error associated to the measured free-energy difference corresponds to the variance of the BAR estimator (11, 12). In a molecular dynamics trajectory, configurations separated by short-time intervals are necessarily correlated. To account for the correlation between contiguous configurations, the variance of the BAR estimator ought to be weighted by the correlation length of the relevant time series (12). A number of numerical schemes have been devised to estimate this quantity. (13–15) In the present work, use was made of the group renormalization scheme of Flyvbjerg and Petersen (14). As depicted in S1, the difficulty with methods based on a decomposition of the data into blocks is rooted in the range of blocks over which a meaningful information can be inferred. From a certain block size, the variance of the ensemble average levels off and reaches a plateau that spans a range of block sizes, over which the correlation length can be safely determined.

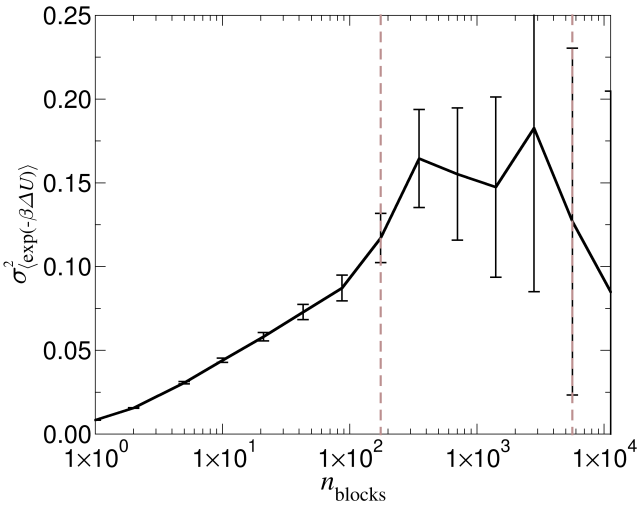

Figure S1: Estimate of the variance of $\langle \exp(-\beta \Delta U) \rangle$ as a function of the number of blocks, employing the approach of Flyvbjerg and Petersen. (14) The data utilized here corresponds to one intermediate state in the creation of the pseudo-infinite poly–L α–helix in the lipid bilayer on the basis of a 154–intermediate stratification strategy with windows of uneven widths. The total number of steps accrued for data collection in this stratum is equal to 225,000. Shown here are estimates from between 1 and 11,250 blocks of 225,000 to 20 samples, respectively. It is apparent that the variance levels off and reaches a plateau over a range of blocks, comprised here between approximately 175 and 5,625. The correlation length ought to be measured within that range.

Table S1: Stratification and sampling strategies explored to optimize the number of strata and the simulation time per stratum.

		simulation time per stratum (ns)		ΔG (kcal/mol)		
simulation	strata	equilibration	data collection	forward	backward	BAR
1	50	0.20	1.00	+35.38	—	—
2	50	0.24	0.24	+24.61	—	—
3	100	0.12	0.12	+27.60	—	—
4	50	0.36	0.36	+24.78	−16.50	+19.79
5	100	0.18	0.18	+25.13	−15.87	+19.89
6	50	0.48	0.48	+18.63	−13.75	+15.06
7	100	0.24	0.24	+19.95	−12.04	+15.28
8	50	0.60	0.60	+17.51	−17.28	+16.88

Multiple strategies were tested to optimize the number of strata and the simulation time per stratum. The time chosen was that necessary to achieve thermodynamic micro-reversibility in the case of the mirror mutation into leucine of an arginine residue borne by a transmembrane, pseudo-infinite poly-leucine α-helix spanning a fully hydrated DPPC lipid bilayer, in the midst of its hydrophobic core and in the aqueous environment. Fifty independent strata involving 0.60 ns of thermalization followed by 1.20 ns of data collection, i.e., a total of 90 ns of sampling, appears to constitute a reasonable compromise between computational effort and accuracy. A similar strategy has been adopted for the homologous point mutation in SecY. The choice of an equal number of molecular-dynamics steps per stratum for equilibration and evaluation of the ensemble average is dictated by the imperious necessity to let both the lipid environment and the protein scaffold of the translocon relax appropriately as the the leucine residue located at the center of the bilayer is progressively transmuted into a charged arginine. The Bennett acceptance ratio (BAR) estimates were obtained by combining the forward and backward transformations (10).

Probability distribution functions.

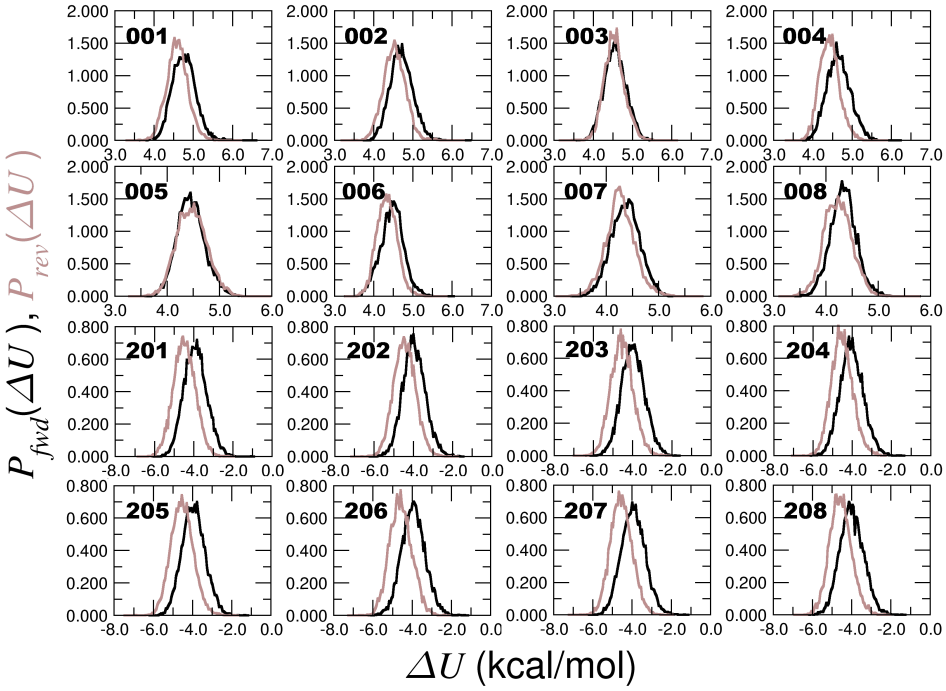

Figure S2: Probability distribution functions underlying the immersion in bulk water of a pseudo-infinite poly-leucine α-helix. Shown in this figure are the distributions characterizing the forward (dark lines) and the backward (light lines) transformations in the first and in the last eight intermediate states of a 208–window stratification strategy.

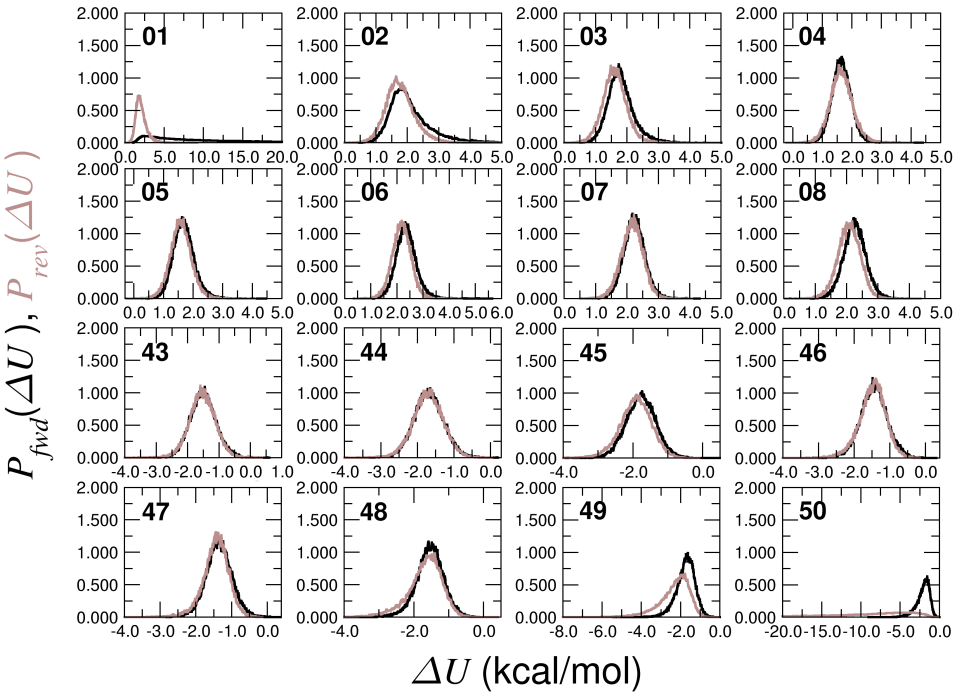

Figure S3: Probability distribution functions underlying the mirror mutation into leucine of an arginine residue borne by a transmembrane, pseudo-infinite poly-leucine α-helix spanning a fully hydrated DPPC lipid bilayer, in the midst of its hydrophobic core and in the aqueous environment. Shown in this figure are the distributions characterizing the forward (dark lines) and the backward (light lines) transformations in the first and in the last eight intermediate states of a 50–window stratification strategy.

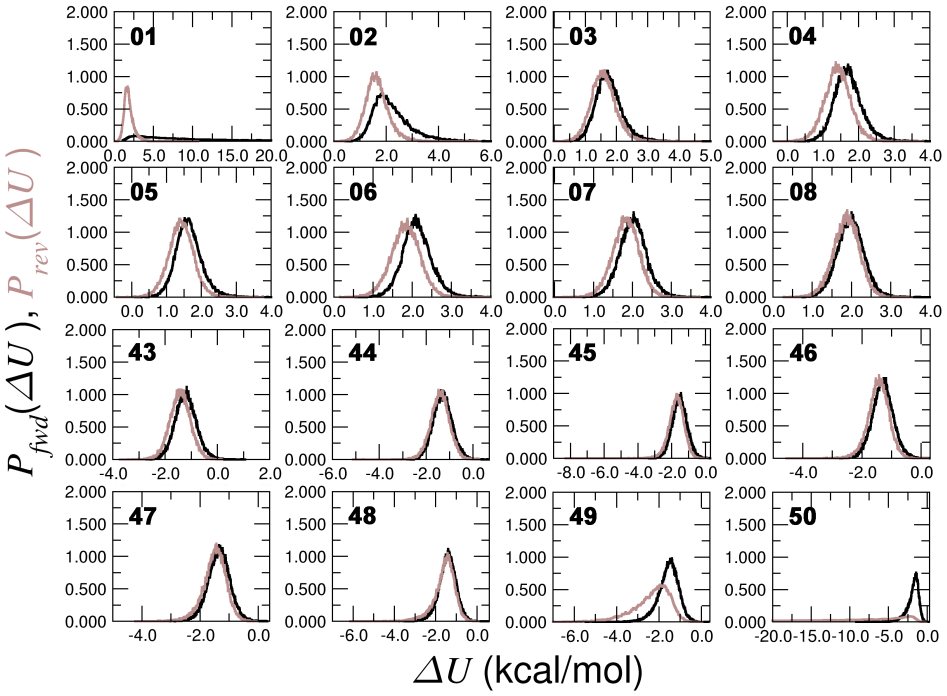

Figure S4: Probability distribution functions underlying the mirror mutation into leucine of an arginine residue borne by an integral, pseudo-infinite poly-leucine α-helix spanning the translocon, in the midst of the protein and in the aqueous environment. Shown in this figure are the distributions characterizing the forward (dark lines) and the backward (light lines) transformations in the first and in the last eight intermediate states of a 50–window stratification strategy.

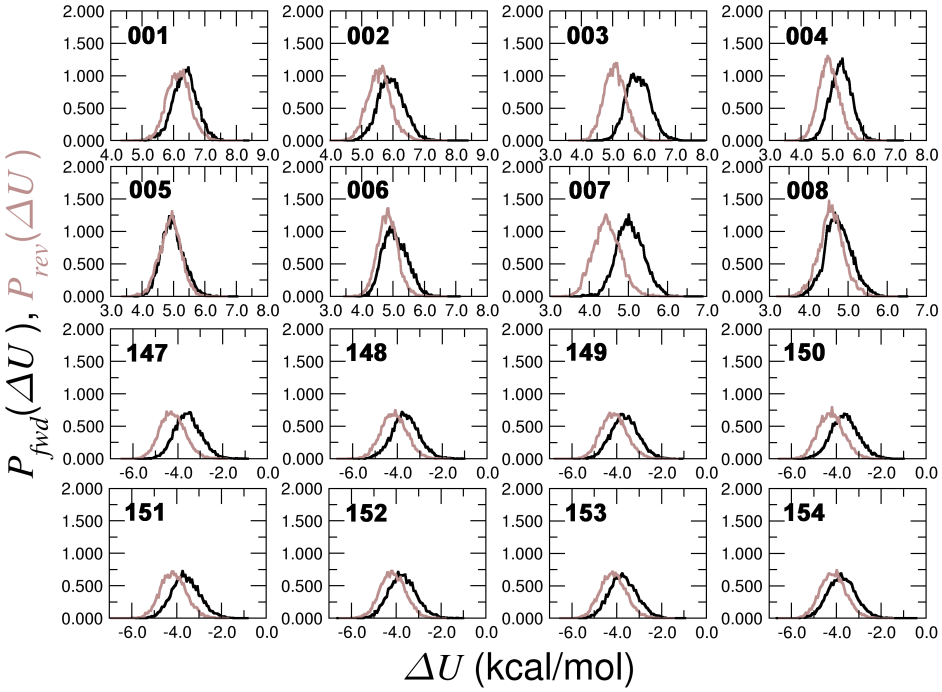

Figure S5: Probability distribution functions underlying the insertion of a transmembrane, pseudo-infinite poly-leucine α-helix into a fully hydrated DPPC lipid bilayer. Shown in this figure are the distributions characterizing the forward (dark lines) and the backward (light lines) transformations in the first and in the last eight intermediate states of a 154–window stratification strategy.

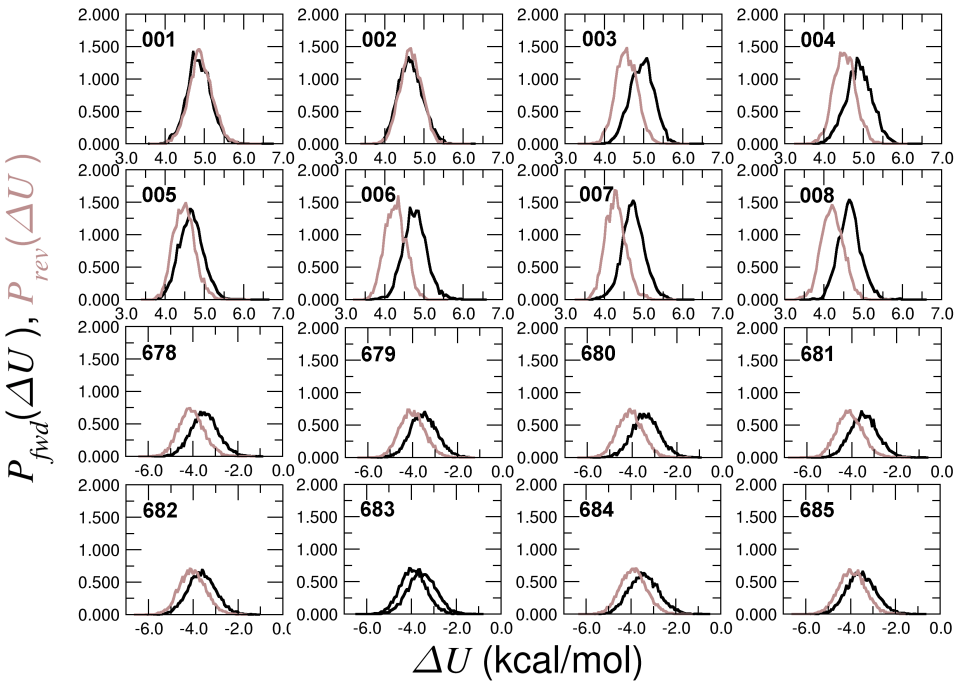

Figure S6: Probability distribution functions underlying the insertion of a transmembrane, pseudo-infinite poly-leucine α-helix into the translocon. Shown in this figure are the distributions characterizing the forward (dark lines) and the backward (light lines) transformations in the first and in the last eight intermediate states of a 685–window stratification strategy.

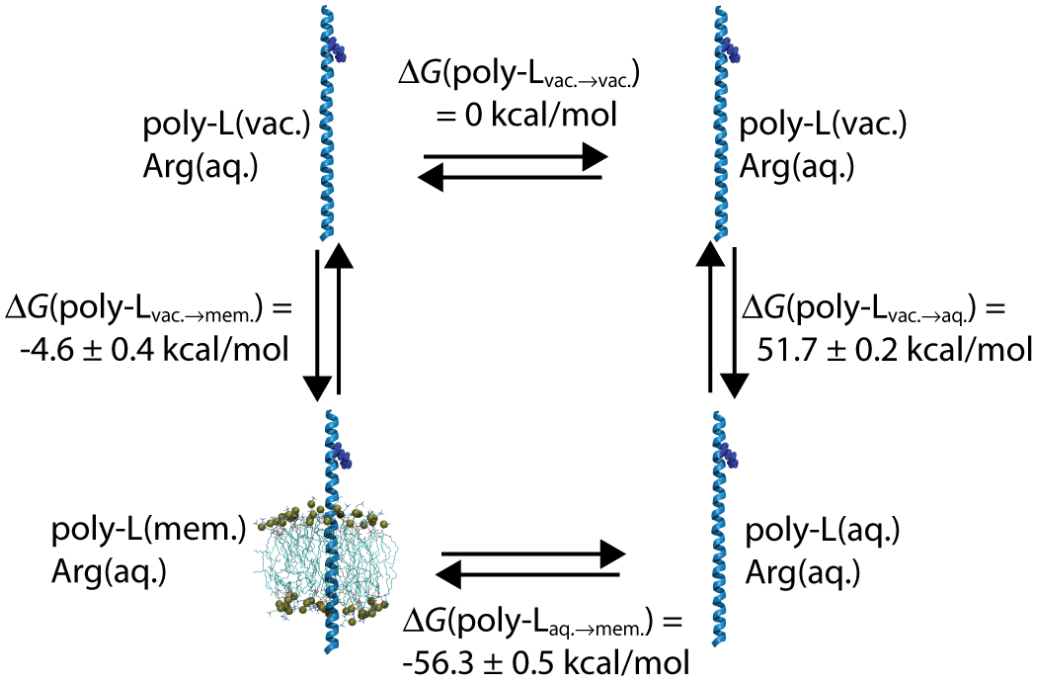

Figure S7: Thermodynamic cycle for TM insertion starting from vacuum. To calculate the transfer free energy for the poly–leucine helix from water to membrane as given in Fig. 2 of the main text, FEP simulations of the transfer from vacuum to membrane and from vacuum to water were carried out. As the arginine residue was actually not present in these transformations, it is denoted as always being in an aqueous state. The free energy of transfer from water to SecY was determined in a similar manner.

Supplemental methods.

System construction.

As noted in the main text, three systems were prepared and used for all FEP simulations. The pseudo-infinite poly–leucine α-helix common to all three is composed of 73 amino acids and is 118 Å long. For the simulation in an aqueous medium, a water box was built around the helix and ionized to 1.0 M strength, using K^+ and Cl^- ions. For the simulation in a membrane, a DPPC bilayer of 99 lipids was constructed and again solvated and ionized to 1.0 M. The DPPC bilayer used in our simulations has an average hydrophobic thickness of roughly 35 Å. Given the α-helical rise of 1.5 Å/residue, this thickness corresponds to an effective helix length of about 23 residues, somewhat longer than the 19 residues typically used in engineered helices, which are based on a thinner (30 Å) DOPC bilayer (16–18). The atom count is 23,000 for the aqueous system, 50,000 atoms for the membrane system, and 150,000 for the SecY system.

Construction of a laterally open translocon.

Simulation of the α–helix in SecY required first the construction of a channel with an open lateral gate. In previous simulations forces were applied perpendicular to the gate to open it (19). A structure of SecY with a partially open gate has, however, since been solved (20), albeit from a species (*Thermotoga maritima*) different than the original crystallographic structure (*Methanococcus jannaschii*) (21). To develop an *M. jannaschii* SecY model with an open lateral gate, the eigenvector routine in the colvars module of NAMD was used (22). A $3N$-long vector was defined, connecting the coordinates of the carbonyl carbon atoms of the TM residues of *M. jannaschii* SecY in the closed state with those of their homologues in *T. maritima* SecY in the partially open state (317 atoms in total). The closed structure of SecY was then driven along this vector in a quasi-equilibrium manner using adaptive biasing forces (22). While in the partially open crystal structure of SecY, the centers of lateral gate helices 2b and 7 are separated by 22 Å, in our fully opened structure the separation is 33 Å. The pseudo-infinite poly–leucine helix was then inserted in this open structure of SecY, which was immersed in a DPPC bilayer, solvated, and ionized as before. The number of atoms required is about 150,000 for the complete SecY/helix system.

Simulation protocols.

All simulations were carried out in the $NP_z\mathcal{A}T$ ensemble, where N denotes the number of particles, P_z, the normal pressure, $\mathcal{A}$, the surface area and T, the temperature. Simulations were run using the molecular dynamics program NAMD 2.7b1 (23) and the CHARMM potential energy function (24) with the CMAP correction (25). The normal pressure and the temperature were fixed at 1 bar

and 330 K, respectively, employing the Langevin piston algorithm (26) and softly damped Langevin dynamics. Periodic boundary conditions were applied in the three directions of Cartesian space. Short-range Lennard-Jones and Coulomb interactions were truncated smoothly by means of a 12-Å spherical cutoff with a switching function applied beyond 10 Å. The particle-mesh Ewald method (27) was employed to compute long-range electrostatic interactions. The Verlet I r–RESPA multiple time-step integrator (28) was used with a time step of 2 fs and 6 fs for updating short- and long-range forces, respectively. Covalent bonds involving hydrogen atoms were constrained to their equilibrium length.

In all simulations the C_α atoms of the poly–leucine helix were restrained to their initial positions to avoid deformations or shifts that would unnecessarily complicate the estimation of the free energy. In simulations in which a membrane is present, the center of mass of all lipid phosphorus atoms was restrained in the direction normal to the bilayer, ensuring that the hybrid arginine/leucine residue remained at the geometric center of the membrane. For SecY, restraints were placed on C_α atoms of the lateral gate to maintain its open state.

Free-energy calculations.

Perturbation theory allows the free energy separating two states to be measured by gradually converting one state into the other through creation and/or annihilation of specific parts of the simulation system (12). That the height of the free-energy barriers measured in potential-of-mean-force calculations (2, 6) depends inherently on the choice of a surrogate reaction coordinate constitutes the main motivation for resorting to perturbation theory. The FEP calculations carried out here can be classified into two sets, namely (i) the insertion of the poly–leucine helix into water, into the lipid bilayer, and into SecY, and (ii) the translocation of the arginine side chain borne by poly–leucine towards the hydrophobic core of the bilayer and towards the interior of the translocon. In the first set of simulations, the insertion process was modeled by creating the helix in the relevant surroundings. In the second set, translocation of the charged amino acid was modeled by concomitant replacement of a leucine residue located in the midst of either the lipid bilayer or the translocon by an arginine residue, while the fully hydrated charged amino acid borne by the α–helical segment, approximately 15 Å above the interfacial region, was mutated into a leucine residue.

The first set of free-energy calculations consisted of inserting the pseudo-infinite poly–leucine into water, a lipid bilayer, and into SecY. On account of the significance of the perturbation, which implies, on the one hand, the reorganization of the lipid phase and, on the other hand, the adaptation of the host membrane protein to its guest, a minute stratification strategy was devised to ensure proper overlap of the ensembles of states and guarantee that micro-reversibility be satisfied. A total of 154 intermediate states were introduced for the creation of poly–leucine in the hydrated DPPC bilayer, representing

a total simulation time of 93.0 broken down as follows: 10 windows at 0.6 ns each for $\lambda = 0 - 0.01$, 50 windows at 0.45 ns each for $\lambda = 0.01 - 0.06$, 14 windows at 1.5 ns each for $\lambda = 0.06 - 0.20$, 10 windows at 1.2 ns each for $\lambda = 0.2 - 0.3$, 10 windows at 0.9 ns each for $\lambda = 0.3 - 0.4$, 10 windows at 0.6 ns each for $\lambda = 0.4 - 0.5$, and 50 windows at 0.36 ns each for $\lambda = 0.5 - 1.0$. For the homologous process in SecY, 685 windows and an aggregate simulation time of 128.4 ns were used. This time was divided between the windows as follows: 200 windows at 0.24 ns each for $\lambda = 0 - 0.20$, 10 windows at 1.2 ns each for $\lambda = 0.2 - 0.3$, 10 windows at 0.9 ns each for $\lambda = 0.3 - 0.4$, 450 windows at 0.12 ns each for $\lambda = 0.5 - 0.85$, and 15 windows at 0.36 ns each for $\lambda = 0.85 - 1.0$. Insertion of the helix into water followed a protocol identical to insertion into membrane. In all windows, simulation time was divided equally between equilibration and sampling.

The second set of free-energy calculations was targeted at modeling the translocation of the arginine residue of the pseudo-infinite poly–leucine towards the interior of the lipid bilayer, as well as the center of SecY. The admittedly lesser disruption of the environment, either of lipid or protein nature, upon replacement of the non-polar, leucine amino acid by the charged arginine justifies the choice of a coarser stratification strategy. The alchemical transformations were carried out using 50 intermediate states, representing a total simulation time of 60.0 ns, either in the hydrated DPPC bilayer, or in SecY. Stratification and sampling strategies were optimized on the basis of a series of preliminary simulations detailed in Table S1.

To avoid singularities in the Lennard-Jones potential when the general extent, coupling parameter, λ, tends towards either 0 or 1, a scaled-shifted soft-core potential was introduced (29). Electrostatic interactions were linearly coupled to the simulation over values of λ between 0 and 0.5, after which point they remained fully coupled; Lennard-Jones interactions were linearly coupled over $\lambda = 0$ to $\lambda = 1$. To probe micro-reversibility of the transformations and improve the precision of the free-energy estimates, each calculation was carried out bidirectionally. Forward and backward simulations were combined in the framework of the Bennett acceptance ratio (BAR) (10), which yields minimum-variance free-energy differences. Statistical precision was determined from the BAR estimator (12).

References

[1] Chipot C, Pohorille A (2007) *Free Energy Calculations. Theory and applications in chemistry and biology.* (Springer Verlag).

[2] Dorairaj S, Allen T W (2007) On the thermodynamic stability of a charged arginine side chain in a transmembrane helix. *Proc Natl Acad Sci USA* 104:4943–4948.

[3] MacCallum J L, Bennett W F D, Tieleman D P (2007) Partitioning of amino acid side chains into lipid bilayers: results from computer simulations and comparison to experiments. *J Gen Physiol* 129:371–377.

[4] Choe S, Hecht K A, Grabe M (2008) A continuum method for determining membrane protein insertion energies and the problem of charged residues. *J Gen Physiol* 131:563–573.

[5] Johansson A C V, Lindahl E (2008) Position-resolved free energy of solvation for amino acids in lipid membranes from molecular dynamics simulations. *Proteins: Struct, Func, Bioinf* 70:1332–1344.

[6] Johansson A C V, Lindahl E (2009) Protein contents in biological membranes can explain abnormal solvation of charged and polar residues. *Proc Natl Acad Sci USA* 106:15684–15689.

[7] Wilson M A, Pohorille A (1996) Mechanism of unassisted ion transport across membrane bilayers. *J Am Chem Soc* 118:6580–6587.

[8] Zwanzig R W (1954) High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J Chem Phys* 22:1420–1426.

[9] Kofke D, Cummings P (1998) Precision and accuracy of staged free-energy perturbation methods for computing the chemical potential by molecular simulation. *Fluid Phase Equil* 150:41–49.

[10] Bennett C H (1976) Efficient estimation of free energy differences from Monte Carlo data. *J Comp Phys* 22:245–268.

[11] Hahn A M, Then H (2009) Characteristic of bennett's acceptance ratio method. *Phys Rev E* 80:031111.

[12] Pohorille A, Jarzynski C, Chipot C (2010) Good practices in free-energy calculations. *J Phys Chem B* 114:10235–10253.

[13] Straatsma T P, Berendsen H J C, Postma J P M (1986) Free energy of hydrophobic hydration: A molecular dynamics study of noble gases in water. *J Chem Phys* 85:6720.

[14] Flyvbjerg H, Petersen H G (1989) Error estimates on averages of correlated data. *J Chem Phys* 91:461–466.

[15] Frenkel D, Smit B (2002) *Understanding molecular simulations: From algorithms to applications* (Academic Press, San Diego).

[16] S Jayasinghe K H, White S H (2001) Energetics, stability, and prediction of transmembrane helices. *J Mol Biol* 312:927–934.

[17] Hessa T, *et al.* (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* 433:377–381.

[18] Jaud S, *et al.* (2009) Insertion of short transmembrane helices by the Sec61 translocon. *Proc Natl Acad Sci USA* 106:11588–11593.

[19] Gumbart J, Schulten K (2007) Structural determinants of lateral gate opening in the protein translocon. *Biochemistry* 46:11147–11157.

[20] Zimmer J, Nam Y, Rapoport T A (2008) Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* 455:936–943.

[21] van den Berg B, *et al.* (2004) X-ray structure of a protein-conducting channel. *Nature* 427:36–44.

[22] Hénin J, Forin G, Chipot C, Klein M L (2010) Exploring multidimensional free energy landscapes using time-dependent biases on collective variables. *J Chem Theor Comp* 6:35–47.

[23] Phillips J C, *et al.* (2005) Scalable molecular dynamics with NAMD. *J Comp Chem* 26:1781–1802.

[24] MacKerell Jr A D, *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616.

[25] MacKerell Jr A D, Feig M, Brooks III C L (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comp Chem* 25:1400–1415.

[26] Feller S E, Zhang Y H, Pastor R W, Brooks B R (1995) Constant pressure molecular dynamics simulation — the Langevin piston method. *J Chem Phys* 103:4613–4621.

[27] 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.

[28] Tuckerman M E, Berne B J, Martyna G J (1992) Reversible multiple time scale molecular dynamics. *J Phys Chem B* 97:1990–2001.

[29] Zacharias M, Straatsma T P, McCammon J A (1994) Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J Chem Phys* 100:9025–9031.