Supporting Information

Zhu and Hummer 10.1073/pnas.1009313107

SI Methods

In the string method (1), the system is represented by a set of "coarse coordinates," which can be either Cartesian coordinates or collective variables. The number of such coarse coordinates can be large, but is nonetheless much smaller than the dimensionality of the phase space of the membrane-embedded and solvated protein. A free energy can be defined as a function of these coarse coordinates by integrating out all other degrees of freedom. The objective then is to identify a "minimum free energy pathway" that connects two minima in the configuration space, such that the mean force, or the gradient of the free energy, is everywhere tangent to the pathway in the Cartesian space of the associated atomic coordinates (1). In the string method, the pathway, or the "string," is represented by some discrete points, or "images," in the configuration space. An initial string is refined toward the minimum free energy pathway through iterations of simulations on individual images. In the following sections, we will introduce the details of the string method and the application of the technique in this study.

The Cartesian coordinates of N selected C_α atoms are used to describe the conformation of the protein in this study, denoted by a vector $\vec{x}$ in $3N$-dimensional space. The difference between two conformations $\vec{x}_1$ and $\vec{x}_2$ (assuming proper superimposition) can be simply quantified by their vector difference $\Delta\vec{x} = \vec{x}_2 - \vec{x}_1$. A conventional measure of the conformational difference is the rmsd, which is proportional to the Euclidean distance $|\Delta\vec{x}|$ in the $3N$-dimensional space:

$$\mathrm{rmsd}(\vec{x}_1,\vec{x}_2) = |\Delta\vec{x}|/\sqrt{N}. \qquad \text{[S1]}$$

Multidimensional Curve Fitting. In the string method, the pathway (string) is represented by $M+1$ images $\vec{x}_0,\ldots\vec{x}_M$ in the $3N$-dimensional space. Because of statistical errors during the evolution of the string, the images usually do not exactly lie on a smooth line and therefore typically require curve smoothing in practice. As a simple method, the position of each image can be adjusted toward the midpoint of the two neighboring images (1). In this study we introduce a new smoothing method in which the images are fitted into a single multidimensional curve that then represents the string.

The multidimensional curve is expressed in the same form as in the harmonic Fourier beads (HFB) method (2):

$$\vec{x}^{\mathrm{cur}}(t) = \vec{x}_0 + (\vec{x}_M - \vec{x}_0)t + \sum_{i=1}^{3N}\sum_{j=1}^{P} w_{ij}\sin(j\pi t)\cdot\hat{e}_i, \qquad \text{[S2]}$$

where $\hat{e}_i$ is the axis of the ith coordinate in the $3N$-dimensional space. As the parameter t is varied from 0 to 1, $\vec{x}^{\mathrm{cur}}(t)$ connects $\vec{x}_0$ and $\vec{x}_M$ through a continuous curve. In addition to $\vec{x}_0$ and $\vec{x}_M$, the analytic form of $\vec{x}^{\mathrm{cur}}(t)$ is determined by the linear coefficients $\{w_{ij}\}$ of P sinusoidal basis functions in each dimension. Given the images $\{\vec{x}_k\}$, one then aims to minimize the χ^2 defined as

$$\chi^2 = \sum_{k=1}^{M-1} |\vec{x}^{\mathrm{cur}}(t_k) - \vec{x}_k|^2 \qquad \text{[S3]}$$

by optimizing both $\{w_{ij}\}$ and $\{t_k\}$. Specifically, when $\{w_{ij}\}$ are given, one can perform a one-dimensional nonlinear fit on each t_k ($k = 1,\ldots,M-1$) independently to find the point $\vec{x}^{\mathrm{cur}}(t_k)$ on the curve with the closest distance to $\vec{x}_k$; when $\{t_k\}$ are given, one can find the set of $\{w_{ij}\}$ that minimizes the χ^2 by a linear least-squares fit of P coefficients on $M-1$ data in each of the $3N$ dimensions independently. By alternately optimizing $\{t_k\}$ and $\{w_{ij}\}$, the χ^2 can be iteratively minimized. Upon convergence, $\{w_{ij}\}$ then determine a smooth curve that best approximates $\{\vec{x}_k\}$, and the images after the smoothing are given by $\{\vec{x}^{\mathrm{cur}}(t_k)\}$. We find that the fit results are not sensitive to the initial values assigned before the iterations. Initialization with $\{w_{ij} = 0\}$, e.g., works well for the path fits in this study.

Similar to other smoothing techniques (1), in our method the two end images $\vec{x}_0$ and $\vec{x}_M$ remain unchanged, whereas other images $\vec{x}_1,\ldots\vec{x}_{M-1}$ are adjusted by the smoothing. Furthermore, going beyond the HFB method (2), the curve parameters $\{t_k\}$ are also subject to optimization in our method, which is therefore expected to work well even if $\{\vec{x}_k\}$ are not evenly spaced. With the analytical form of the string, we can calculate by numerical integration the arc length between $\vec{x}_0$ and any given $\vec{x}^{\mathrm{cur}}(t)$:

$$s(t) = \int_0^t \sqrt{\left|\frac{d\vec{x}^{\mathrm{cur}}(t)}{dt}\right|^2}\,dt. \qquad \text{[S4]}$$

A parameter $L(t) = s(t)/s(1)$ is defined as $s(t)$ normalized by the total arc length of the string $s(1)$. This reduced path length $L \in [0,1]$ then describes the relative position of $\vec{x}^{\mathrm{cur}}(t)$ along the pathway and serves as a progression parameter for the transition. The smoothness of the fitted curve is determined by P, as incorporating more basis functions will fit the data more closely, but will likely result in a more oscillatory curve. We find that using two sinusoidal basis functions ($P = 2$) in each dimension can already fit the coarse-grained pathway obtained in our previous study (3) reasonably well, with rmsds between the original and fitted data all below 0.15 Å. This setting ($P = 2$) is therefore used in all path fits in this study.

Description of Symmetry. The five subunits of GLIC and ELIC are identical in sequence and highly similar in the crystal structures (4–6). Nevertheless, the structures along individual pathways do not have to satisfy fivefold symmetry (but each asymmetric path will enter the path ensemble in five symmetry-related copies).

The $3N$ coordinates in $\vec{x}$ can also be formally expressed as a 3-by-N matrix X, in which each column represents the three Cartesian coordinates of a single atom. The protein is composed of five subunits, each containing $n = N/5$ atoms. Correspondingly, we can write $X = (X_1 \quad X_2 \quad X_3 \quad X_4 \quad X_5)$, in which X_i is a 3-by-n matrix representing the coordinates of a single subunit. We assume that the symmetry axis lies at the z axis and that the five subunits are arranged in clockwise order when viewed from the $+z$ direction. We now define a linear permutation-rotation operator G:

$$G(\vec{x}) = G(X) \equiv R_z(2\pi/5)\cdot(X_2 \quad X_3 \quad X_4 \quad X_5 \quad X_1), \qquad \text{[S5]}$$

where $R_z(2\pi/5)$ is a 3-by-3 rotation matrix for a $2\pi/5$ (72°) angle with respect to the z axis. The operation corresponds to a permutation of the five subunits followed by a 72° rotation and, in fact, generates a symmetry-related equivalent structure, in which each subunit assumes the conformation of a neighboring subunit in the original structure. If coordinates $\vec{x}$ already exhibit a perfect fivefold symmetry, they will remain unchanged under the G operation. Moreover, any coordinates will be restored after undergoing five such operations, i.e.,

$$G^5(\vec{x}) = \vec{x} \quad \text{[S6]}$$

for any $\vec{x}$. Because permutation and rotation preserve the inner product of vectors, the G operation satisfies

$$G(\vec{x}_1) \cdot G(\vec{x}_2) = \vec{x}_1 \cdot \vec{x}_2, \quad \text{[S7]}$$

for any $\vec{x}_1$ and $\vec{x}_2$.

Based on operator G, a symmetrization operator S is defined as

$$S(\vec{x}) \equiv \frac{1}{5}\sum_{i=1}^{5} G^i(\vec{x}). \quad \text{[S8]}$$

Any $\vec{x}$ can then be divided into a symmetric component $\vec{x}_S$ and an antisymmetric component $\vec{x}_A$:

$$\vec{x}_S = S(\vec{x}), \quad \text{[S9]}$$

$$\vec{x}_A = \vec{x} - \vec{x}_S. \quad \text{[S10]}$$

$\vec{x}_S$ assumes a perfect fivefold symmetry, i.e.,

$$G(\vec{x}_S) = \vec{x}_S. \quad \text{[S11]}$$

By repeatedly applying Eq. **S7**, we have

$$\vec{x}_S \cdot \vec{x}_A = G(\vec{x}_S) \cdot G(\vec{x}_A) = \vec{x}_S \cdot G(\vec{x}_A) = \vec{x}_S \cdot G^i(\vec{x}_A). \quad \text{[S12]}$$

Taking $i = 1,\ldots,5$, and summing up the left- and right-hand sides of Eq. **S12**, we obtain the orthogonality relationship:

$$\vec{x}_S \cdot \vec{x}_A = 0. \quad \text{[S13]}$$

Eqs. **S9** and **S10** therefore represent an orthogonal decomposition of $\vec{x}$ into two subspaces. The magnitude of $\vec{x}_A$ then quantifies the deviation of the conformation from perfect symmetry. Furthermore, using only the symmetric coordinates to describe the system can effectively reduce the dimensionality, as discussed below.

Harmonic Restraints. In our implementation of the string method, the N selected atoms are subject to harmonic restraints in the simulations, with reference positions denoted by a $3N$-dimensional vector $\vec{r}$. In this study, $\vec{r}$ is perfectly symmetric by design, $S(\vec{r}) = \vec{r}$, such that the restraints would not favor a particular conformation over its symmetry-related copies. During the simulation, the potential for the harmonic restraints is given by

$$U^C(\vec{x}) = \frac{K}{2}|\vec{x} - \vec{r}|^2, \quad \text{[S14]}$$

where K is the spring constant. If we decompose $\vec{x}$ into the symmetric part $\vec{x}_S$ and the antisymmetric part $\vec{x}_A$ (Eqs. **S9** and **S10**), it is obvious that such a potential not only restrains $\vec{x}_S$ toward $\vec{r}$, but also acts to attenuate $\vec{x}_A$, which is not always necessary and desirable.

An alternative restraining potential can be defined as

$$U^C(\vec{x}) = \frac{K}{2}|S(\vec{x}) - \vec{r}|^2. \quad \text{[S15]}$$

The potential is a function of only the symmetric coordinates $\vec{x}_S$ and is independent of the antisymmetric component $\vec{x}_A$. To derive the restraining forces $\vec{f}^C$ arising from this potential, we note that S is a linear operator and can be expressed as a matrix, which is actually the transpose of the Jacobian matrix for the transformation $\vec{y} = S(\vec{x})$. We also note that $S^2(\vec{x}) = S(\vec{x})$ and $S(\vec{r}) = \vec{r}$. Therefore the restraining forces are given by

$$\vec{f}^C = -\nabla_{\vec{x}} U^C(\vec{x}) = K[\vec{r} - S(\vec{x})]. \quad \text{[S16]}$$

Unlike in the former scheme (Eq. **S14**), $\vec{x}_A$ in this case is not affected by the restraints and is free to evolve during the simulation. Furthermore, we can represent the protein conformation by the symmetric coordinates $\vec{x}_S$ only, and exclude $\vec{x}_A$ from the description. Because $\vec{x}_S$ has only $3n$ independent dimensions, in the new scheme the dimensionality of the configuration space is effectively reduced by a factor of 5, which may alleviate some of the challenges arising from very high dimensions.

Implementation of the String Method. Because the initial string is usually only a rough guess of the minimum free energy pathway and may need further refinement, iterations of simulations are required to evolve the string and update the images so that they converge toward the minimum free energy pathway. This can be done by performing a simulation for each image k ($k = 0,\ldots,M$) with harmonic restraints at $\vec{r}_k$, and recording the mean coordinates $\vec{x}_k^{\text{mean}}$ of the restrained atoms during the simulation. The images can then be updated according to the mean forces (1). Some new update methods have also been proposed recently, employing swarms of short unbiased simulations (7) or confined simulations in Voronoi cells (8). In this study, we adopt a conservative update scheme, in which the new string is simply placed at the mean coordinates $\{\vec{x}_k^{\text{mean}}\}$ from the current iteration of simulations. Compared to the reference coordinates $\{\vec{r}_k\}$, $\{\vec{x}_k^{\text{mean}}\}$ should lie closer to the minimum free energy pathway, and the new string should therefore represent a better approximation of the pathway. Over iterations the string will then evolve toward the true pathway.

As mentioned earlier, $G^i(\vec{x})$ is an equivalent structure of $\vec{x}$ due to symmetry and, therefore, should appear with the same probability as $\vec{x}$. Consequently, in sufficiently long simulations, $\vec{x}_k^{\text{mean}}$ is expected to be perfectly symmetric, without an antisymmetric component. In practice, however, due to limited simulation time, $\vec{x}_k^{\text{mean}}$ is generally not symmetric. We therefore remove the antisymmetric component in $\vec{x}_k^{\text{mean}}$ by applying the symmetrization operator:

$$\vec{x}_k^{\text{sym}} = S(\vec{x}_k^{\text{mean}}), \qquad k = 0,\ldots,M. \quad \text{[S17]}$$

The symmetric component $\vec{x}_k^{\text{sym}}$ is then used to calculate the mean restraining force:

$$\vec{f}_k^{\text{mean}} = K(\vec{r}_k - \vec{x}_k^{\text{sym}}). \quad \text{[S18]}$$

$\vec{x}_k^{\text{sym}}$ is effectively an average over the conformations of the five subunits in $\vec{x}_k^{\text{mean}}$, which can help reduce fluctuations in the mean coordinates. More importantly, when the restraints are applied only in the symmetric subspace (Eqs. **S15** and **S16**), one must use $\vec{x}_k^{\text{sym}}$ instead of $\vec{x}_k^{\text{mean}}$ to calculate the mean force, because the antisymmetric component is not subject to the restraining force. A multidimensional curve is then fitted over $\{\vec{x}_k^{\text{sym}}\}$ using the method described earlier, representing the new string. The curve is evenly divided into M segments, and the resulting $M + 1$ points on the curve are used as the restraints $\{\vec{r}_k\}$ in the new iteration of simulations. By construction, these new $\{\vec{r}_k\}$ are also perfectly symmetric.

According to thermodynamic theory, when a system is driven by an external force from one state to another in a quasiequilibrium process, the reversible work done by the external force is

equal to the free energy difference between the two states. Therefore, if we assume that the mean force varies smoothly along the string, we may use the average of the mean restraining forces on two adjacent images to calculate the free energy difference:

$$\Delta G_{k\to k+1} = \frac{1}{2}(\bar{f}_k^{\text{mean}} + \bar{f}_{k+1}^{\text{mean}}) \cdot (\bar{x}_{k+1}^{\text{sym}} - \bar{x}_k^{\text{sym}}). \qquad \text{[S19]}$$

The free energy of each image can then be assigned after cumulating the ΔG values. We note that the free energy defined in the string method is actually a projection onto the configuration space formed by the chosen coarse coordinates. Although all other degrees of freedom have been integrated out in this free energy, the configuration space itself may still have large dimensionality, and a "state" could correspond to a volume rather than a single point in the configuration space. The free energy above is essentially measured between individual points (the images) in the configuration space and thus may not fully capture the differences between the corresponding states. However, because both open and closed states are relatively compact and similar in their overall structure, we assume here that the corrections for the local curvature of the free energy surface in the $3N$-dimensional space are independent of the position along the pathway.

Computational Details. ***System preparation.*** Protein coordinates were taken from the crystal structure (PDB ID code 3EHZ) of GLIC in the open state (5). We note that in this article we use the same residue numbers as in the PDB file 3EHZ (5), which are shifted by one residue in the PDB file 3EAM (6) of a second GLIC structure. Only the transmembrane domain (TMD) of the protein (from residue 192 to the C terminal) is included in the model, with the extracellular domain (ECD) removed. The protein was embedded in a Palmitoyl-oleoyl-phosphatidylcholine lipid bilayer and solvated by water molecules. Although it was proposed that nAChR contains internal sites with buried cholesterols, such a feature is not expected in prokaryotic pentameric ligand-gated ion channels (pLGICs) (9).

The opening probability of GLIC is regulated by the extracellular pH, and therefore most of the residues responsible for the pH gating are presumably located in the ECD (5, 6, 10) and thus not present in our model. Indeed, in the extracellular half of the TMD, where extracellular pH changes could possibly be sensed, there are only two residues that could become protonated at reduced pH: His234, which is fully buried between four helices of the TMD, and is thus assumed to remain neutral; and Glu242, which is partly exposed to the pore lumen, but close to Lys247 of the TMD and Lys32 of the ECD, and is thus expected to remain charged even at a slightly acidic pH on the extracellular side. The remaining ionizable residues are facing the intracellular side, which should remain close to neutral pH even in activating conditions. In our simulations, we have all Glu residues charged. One possible concern is the protonation state of Glu271 near the intracellular surface, as its side chain is found to rotate about the C_β-C_γ bond toward the surface (but without disruption of the backbone structure). All His residues are neutral, with the proton assigned to the ϵ and δ nitrogens in His234 and His276, respectively.

Ionic strength of the solution can, in principle, be mimicked in the simulations by adding ions in the bulk water. In this study, however, we chose to incorporate the minimum amount of 5 chloride ions necessary to neutralize the net charge of the protein. If cations were present, they might enter the pore in some of the images but not others, without proper equilibration on the simulation time scale. In our simulation setup without cations, this problem is avoided and the sampling in a limited simulation time may be more consistent. Nevertheless, although the impact of ion concentration on the GLIC conformation is not known experimentally, we caution that possible effects of ions in the simulations remain to be tested.

The simulation system (Fig. S4) thus contains the TMD of GLIC, 280 lipid molecules, 13,493 water molecules, and 5 chloride ions, with a total of 88,324 atoms. The unit cell (Fig. S4) of the periodic system has dimensions of ∼100 Å × 100 Å × 80 Å.

Representative atoms. We chose the C_α atoms from residues 193–276 to represent the protein conformation, corresponding to $n = 84$ atoms in each subunit, or $N = 420$ atoms in the pentamer. These residues include the first three helices (M1–M3) in the TMD and the two connecting loops, and represent the "common core" (6) of the TMD. The same set of atoms was used to construct a coarse-grained pathway in our previous study (3). It was recently discovered that the last helix in the TMD, M4, is critical for the glycine receptor, a eukaryotic pLGIC, to form pentamers (11). In contrast, in the crystal structures of ELIC and GLIC, M4 is peripheral and only loosely coupled to the other helices (4–6). In this study, although M4 is not selected into the representative C_α atoms to avoid overrestraining, it is fully incorporated in our model and is free to adapt to the compact conformation of M1–M3 during the simulations. Each image on the string is then described by the $3N$ Cartesian coordinates, and the harmonic restraints in this study are applied only on these N atoms.

Initial string. The procedure to generate the initial string is summarized in Fig. S5. The first coarse-grained pathway was taken from our previous study (3), obtained by constructing a mixed elastic network model (MENM) with identical spring constants between every pair of the C_α atoms within a cutoff distance of 9 Å. We refer the reader to ref. 3 for details of such calculations. Although in the development of the MENM it was assumed that the endpoint structures correspond to local energy minima, which is actually not the case for the isolated TMD in this study, we note that the MENM pathway is independent of the model parameters (3), even for strong driving forces under which the barrier has disappeared. Furthermore, such coarse-grained pathways do not need to be highly accurate as they merely provide an initial guess of the string and will be extensively refined.

The system was first minimized and equilibrated for 1 ns with the protein fixed. Another nanosecond of equilibration was followed in which the protein was allowed to move with harmonic restraints applied on the selected C_α atoms. Then a "pulling" simulation was initiated in which the system was driven to the closed state following the coarse-grained pathway (3). Specifically, 100 frames were retrieved from the coarse-grained pathway, and the selected C_α atoms were restrained sequentially at each of the frames for 0.2 ns. This 20-ns simulation then provided a trajectory from the open to the closed conformation. The trajectory was evenly divided into $M = 10$ segments, generating 11 images (including the starting and end frames). The first and last images represent the open and closed states, respectively. Each image was further equilibrated for 2 ns with corresponding harmonic restraints from the coarse-grained pathway.

In our initial coarse-grained pathway (3), the closed-state conformation was taken from the ELIC crystal structure (4). More equilibration was therefore needed to relax GLIC in this state. We thus removed the harmonic restraints on the last image and performed a 10-ns unrestrained simulation to fully relax the protein. Then the selected C_α coordinates were averaged over the last 5 ns of this simulation after proper alignment. Taking these average coordinates (after symmetrization) as an improved model for the closed conformation, along with the open conformation from the crystal structure (5), we constructed another pathway between the two conformations using the mixed elastic network models (3) as described earlier. The corresponding frames from this new pathway then served as new restraints

for the images. This initial string was subsequently refined according to an iterative procedure, as described below.

String refinement. In the first stage, the string was refined with restraints in the $3N$-dimensional space, using the update scheme described earlier. In each iteration, we obtained the mean coordinates $\vec{x}_k^{\text{mean}}$ ($k = 0,\ldots,10$) from individual simulations and calculated the symmetrized coordinates $\vec{x}_k^{\text{sym}}$ (Eq. **S17**). Each $\vec{x}_k^{\text{sym}}$ was then aligned to the symmetrized crystal structure $\vec{r}_{\text{cry}}$ to remove any overall rotation that might occur during the simulations. A $3N$-dimensional curve was then fitted over the aligned $\{\vec{x}_k^{\text{sym}}\}$, representing the new string. The curve was evenly divided into 10 segments, yielding 11 coordinates $\{\vec{r}_k\}$ to be used as the references of the restraints in the next iteration of simulations. As discussed in the main text, the open conformation of the TMD alone did not appear to be metastable and would drift away from the crystal structure $\vec{r}_{\text{cry}}$ over the iterations. We therefore made a projection of $\vec{r}_{\text{cry}}$ to the line connecting $\vec{r}_0$ and $\vec{r}_1$ and used the projected point $\vec{r}'_{\text{cry}}$ to replace $\vec{r}_0$ as the reference for the first image, thus eliminating the drift of the image along the string while allowing it to evolve in the directions perpendicular to the string. We performed 26 iterations of 1-ns simulations, followed by 13 iterations of longer simulations of 2 ns each (Table S1). A spring constant of $K = 1$ kcal/mol/Å^2 was used for the harmonic restraints (Eq. **S14**) in these simulations.

In the second stage, we applied the harmonic restraints only in the $3n$ symmetric dimensions, with no restraint in the antisymmetric subspace. This was done in practice by using Eq. **S15** instead of Eq. **S14** as the restraining energy and Eq. **S16** as the restraining force during the simulations. Unlike in the first stage, the symmetric restraints here cannot prevent the overall translation or tilting of the entire protein. We thus applied additional restraints in the x and y directions on the center of the N selected C_α atoms to eliminate the overall translation. We also applied five restraints in the z direction on the centers of the n atoms in each individual subunit to eliminate tilting. The same method as in the first stage was used to update the string after the simulations in each iteration. We performed 15 iterations of 1-ns simulations followed by 16 iterations of 2-ns runs (Table S1). Then a last round of simulations, each 40 ns, was performed to obtain the final string and to calculate the free energy.

During the iterative string refinement, we monitored the free energy profile for channel closure. Fig. S3 compares the free energies obtained from short (1–2 ns) runs at the end of each refinement stage to the final profile from long (40 ns) simulations. Within the statistical uncertainties expected for the short runs, the free energy profile appears to be converged. The spatial evolution of the string is quantified in Fig. S6. Over a small number of iterations the string evolves in a diffusion-like manner in the high-dimensional conformation space, whereas over many iterations the distance to an initial string appears stabilized, and the string appears to be converged overall.

Unrestrained simulations. To validate the minimum free energy pathway obtained from the string method, we also performed three free simulations without restraints (Table S2). Starting from the first image (open state) in the initial string, two simulations, each 100 ns, were initiated with different random seeds. We also extended the 10-ns equilibration of the closed state, as mentioned earlier, to 100 ns.

Simulation protocol. All simulations were performed using the CHARMM (c35b1) force field (12–14), the TIP3P water model (15), and the NAMD2 program (16), under the periodic boundary conditions with constant temperature (300 K) and pressure (1 atm). Full electrostatics was calculated using the particle mesh Ewald method (17). For simulations involved in the string method, the mean coordinates $\vec{x}_k^{\text{mean}}$ were calculated by averaging over the corresponding coordinates in every time step after discarding the first 0.5 ns of the simulation. All molecular images in this article were rendered in VMD (18).

Umbrella Sampling of Na^+ Ion Conduction. Umbrella sampling has been previously applied to characterize ion conduction in narrow channels (19, 20), and a similar protocol was adopted here. The GLIC simulation system above was slightly modified by adding a Na^+ ion and a Cl^- ion. The Na^+ was originally placed in the intracellular side of the bulk water and was pulled in the $+z$ direction across the pore to the extracellular side in a ~7.5-ns simulation, thus generating the initial coordinates for the subsequent umbrella-sampling simulations. We employed a total of 153 umbrella windows with a uniform spacing of 0.5 Å, thus covering a total distance of 76 Å in the z direction. In each window a simulation of 1 ns was performed, with data from the last 0.9 ns used for analysis. Similarly, we also pulled a Na^+ ion in the $-z$ direction, from the extra- to the intracellular side across the pore, followed by umbrella sampling as described above. We thus have two simulations with different initial coordinates for each umbrella window. When applying the weighted histogram analysis method (WHAM) (21), we combined the data in the two simulations to obtain a unified histogram for each window. In the two simulation setups corresponding to open and closed states, the overall protein conformation was maintained by applying the same restraints as in the string method calculations described earlier. For the ion, the harmonic restraint with a spring constant of 10 kcal/mol/Å^2 was applied in the z direction. In the xy dimensions, in contrast, the ion is not subject to the restraint if it lies within 6 Å of the symmetry axis, and the restraint only acts to pull the ion toward the symmetry axis when the deviation is larger than 6 Å.

The one-dimensional free energy was calculated using WHAM (21). The local diffusion coefficient of the ion in each umbrella window was obtained by integrating the autocorrelation function of the z coordinate, as described in ref. 22. The diffusion coefficients from different windows were then subject to smoothing and interpolation, resulting in a smooth and continuous curve. The free energy and the position-dependent diffusion coefficients for the open state are plotted in Fig. S1.

Based on the Nernst–Planck theory, the maximum ion conductance can be calculated from the umbrella-sampling results for narrow pores (19, 20). In contrast, here we aim to estimate ion conductance under a low ion concentration, such that ion–ion interaction can be ignored and the conductance is linearly dependent on the concentration. Moreover, we only focus on the scenario in which a small voltage V is applied on the pore with identical 1D bulk density p_0 on the two sides. Within the linear-response range, the voltage-driven flux is equal to the diffusional flux under a concentration difference of

$$\Delta p = p_0 \frac{qV}{k_B T}, \qquad \textbf{[S20]}$$

where q is the charge of the ion. According to the Smoluchowski equation, the stationary-state flux is given by

$$j = \frac{\Delta p \exp(U_0/k_B T)}{\int_{z_1}^{z_2} dz \exp[U(z)/k_B T]/D(z)}, \qquad \textbf{[S21]}$$

where $U(z)$ is the free energy, $U_0 \equiv U(z_1) = U(z_2)$ is the free energy in the bulk region, and $D(z)$ is the diffusion coefficient. With the electric current given by $I = qj$, we obtain the conductance:

$$\gamma = I/V = \frac{q^2 p_0 \exp(U_0/k_B T)}{k_B T \int_{z_1}^{z_2} dz \exp[U(z)/k_B T]/D(z)}. \quad \text{[S22]}$$

In the bulk region, the lateral restraint $u(x,y)$ described earlier confines the ion in a cylindrical space with effective cross-section area of $S = \iint \exp[-u(x,y)/k_B T]dxdy$, which determines the offset constant between the 3D and 1D potentials of mean force (20). For a given volume density ρ of ions in the bulk, the equivalent 1D density is then given by $p_0 = \rho S$. We have $S \approx 125$ Å^2 for the lateral restraint used in our simulations. At a Na^+ concentration of 140 mM, or $\rho = 8.43 \times 10^{-5}$/Å^3, the corresponding 1D bulk density p_0 is then ~0.01/Å. Using this value along with the obtained $U(z)$ and $D(z)$, the conductance of the open GLIC pore is estimated to be ~0.8–1.8 pS. We note that the histograms in some of our umbrella windows did not appear to be fully converged, and therefore our calculated free energy and ion conductance may bear some uncertainty and remain to be further improved. Moreover, this simplified model only accounts for a single-ion mechanism and ignores ion–ion interactions and thus cannot describe the possibly nonlinear dependence of the conductance on the ion concentration when the latter is not sufficiently low.

For the GLIC pore in the closed conformation, similar simulations were performed as described above, except that the pulling simulation along the $-z$ direction and the subsequent umbrella sampling were not carried out. Therefore only one simulation of 1 ns was done in each umbrella window. Although suffering a poorer convergence compared to the open-state calculations due to insufficient sampling, the obtained free energy has a barrier of clearly more than 10 kcal/mol in the pore region. This translates into a conductance at least 2,000 times lower than the open-state conformation.

1. Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G (2006) String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J Chem Phys* 125:24106.
2. Khavrutskii IV, Arora K, Brooks CL, 3rd (2006) Harmonic Fourier beads method for studying rare events on rugged energy surfaces. *J Chem Phys* 125:174108.
3. Zhu F, Hummer G (2009) Gating transition of pentameric ligand-gated ion channels. *Biophys J* 97:2456–2463.
4. Hilf RJ, Dutzler R (2008) X-ray structure of a prokaryotic pentameric ligand-gated ion channel. *Nature* 452:375–379.
5. Hilf RJ, Dutzler R (2009) Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel. *Nature* 457:115–118.
6. Bocquet N, et al. (2009) X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation. *Nature* 457:111–114.
7. Pan AC, Sezer D, Roux B (2008) Finding transition pathways using the string method with swarms of trajectories. *J Phys Chem B* 112:3432–3440.
8. Vanden-Eijnden E, Venturoli M (2009) Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J Chem Phys* 130:194103.
9. Brannigan G, Henin J, Law R, Eckenhoff R, Klein ML (2008) Embedded cholesterol in the nicotinic acetylcholine receptor. *Proc Natl Acad Sci USA* 105:14418–14423.
10. Nury H, et al. (2010) One-microsecond molecular dynamics simulation of channel gating in a nicotinic receptor homologue. *Proc Natl Acad Sci USA* 107:6275–6280.
11. Haeger S, et al. (2010) An intramembrane aromatic network determines pentameric assembly of Cys-loop receptors. *Nat Struct Mol Biol* 17:90–98.
12. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616.
13. Feller SE, MacKerell AD (2000) An improved empirical potential energy function for molecular simulations of phospholipids. *J Phys Chem B* 104:7510–7515.
14. MacKerell AD, Jr., Feig M, Brooks CL, III (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 Comput Chem* 25:1400–1415.
15. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
16. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802.
17. Essmann U, et al. (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–8593.
18. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. *J Mol Graph* 14:27–38.
19. Allen TW, Andersen OS, Roux B (2006) Ion permeation through a narrow channel: Using gramicidin to ascertain all-atom molecular dynamics potential of mean force methodology and biomolecular force fields. *Biophys J* 90:3447–3468.
20. Allen TW, Andersen OS, Roux B (2006) Molecular dynamics—potential of mean force calculations as a tool for understanding ion permeation and selectivity in narrow channels. *Biophys Chem* 124:251–267.
21. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM (1992) The weighted histogram analysis method for free-energy calculations on biomolecules .1. The method. *J Comput Chem* 13:1011–1021.
22. Hummer G (2005) Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New J Phys* 7:34.

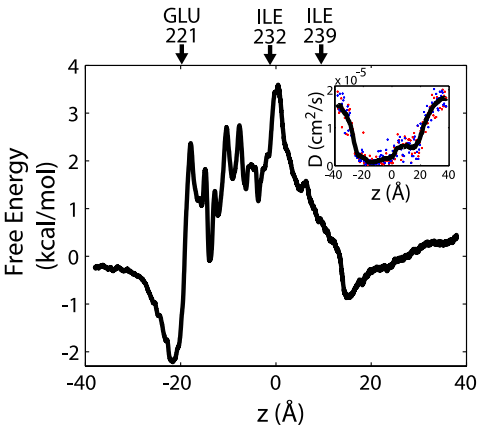

Fig. S1. Free energy of the Na^+ ion as a function of the z coordinate. The protein is in the open conformation. The side-chain positions of some pore-lining residues are indicated (as shown in Fig. 2). (*Inset*) The position-dependent diffusion coefficients (dots: the diffusion coefficients calculated in each umbrella window; black line: the smoothed curve).

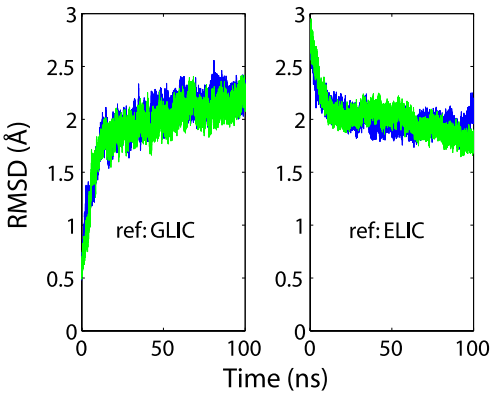

Fig. S2. rmsds of the two unrestrained simulations starting from the open conformation with respect to the GLIC (5) and ELIC (4) crystal structures, respectively. In each frame, the coordinates of the 420 selected C_α atoms were symmetrized and then used to calculate the rmsd. The figure shows that the protein conformation moves away from the GLIC crystal structure and evolves toward ELIC during both simulations. At the end of the simulations the protein adopts conformations with rmsds slightly closer to ELIC than GLIC crystal structures, fully consistent with the evolution of the reduced path length toward $L \approx 0.6$ in Fig. 5*C*, and the free energy plateau about halfway through the transition in Fig. 5*A*.

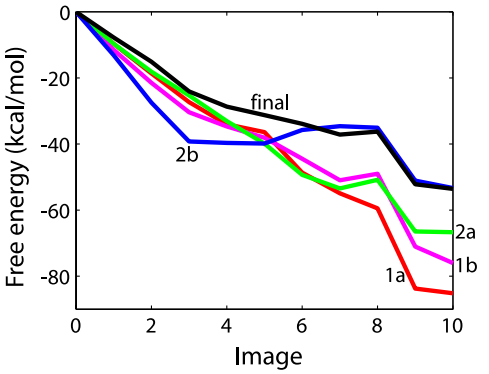

Fig. S3. Free energy calculated from some short (1-ns or 2-ns) restrained simulations during the iterative string refinement. The lines (*1a*, *1b*, *2a*, and *2b*) represent the last simulations in the corresponding stages as summarized in Table S1. The free energy obtained from the final 40-ns simulation is also plotted (black).

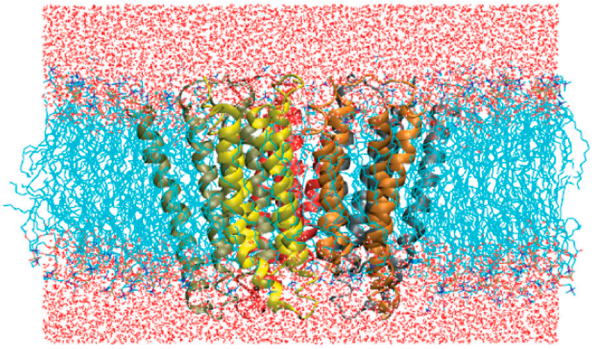

Fig. S4. The unit cell of the simulation system, with the extracellular side up. The five subunits of GLIC are rendered in different colors.

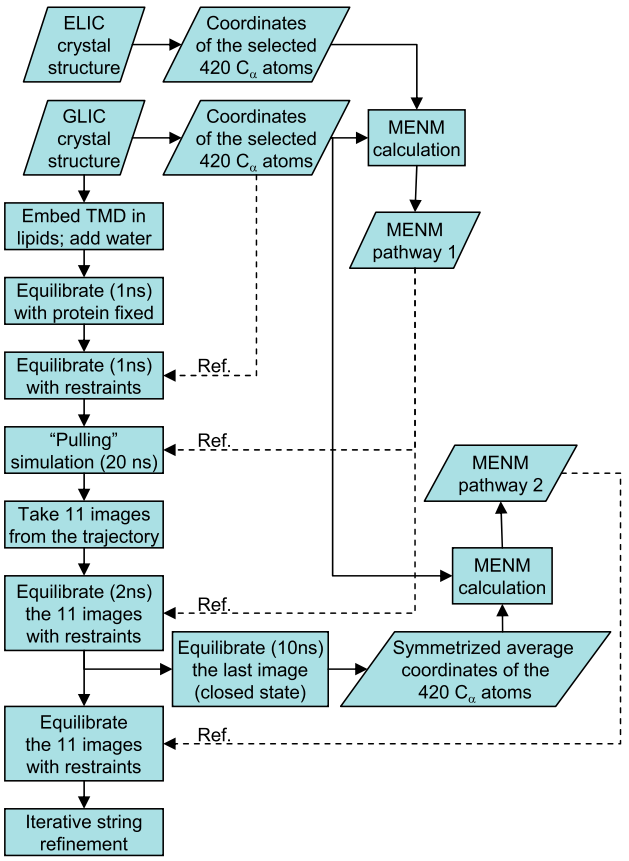

Fig. S5. Flow chart of the procedure used to generate the initial string.

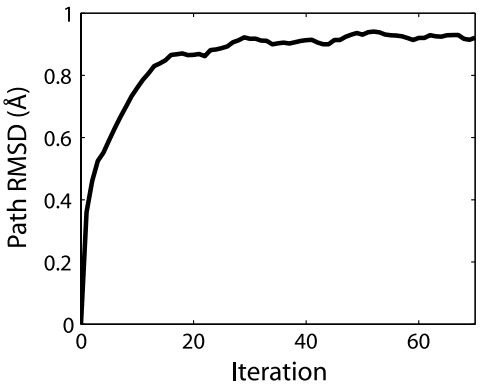

Fig. S6. Deviation of the string from the initial coarse-grained pathway (see *SI Methods*) over the iterations. The average rmsd between two pathways, represented by the fitted curves $\bar{x}^{\mathrm{cur}}(t)$ and $\bar{x}^{\mathrm{cur0}}(t)$ (Eq. **S2**), respectively, is defined as $\mathrm{rmsd} = [\int_0^1 |\bar{x}^{\mathrm{cur}}(t) - \bar{x}^{\mathrm{cur0}}(t)|^2 dt/N]^{1/2}$.

Table S1. Simulation time for each image during the iterative string refinement and the final sampling

	String refinement				Sampling
	Stage 1a	Stage 1b	Stage 2a	Stage 2b	
Simulation time	26 × 1 ns	13 × 2 ns	15 × 1 ns	16 × 2 ns	40 ns
Total/image	52 ns		47 ns		40 ns
Total time	572 ns		517 ns		440 ns

During stage 1 (a and b), harmonic restraints were applied on all the 3*N* coordinates, whereas during stage 2 (a and b) and the final sampling, only the symmetric degrees of freedom were restrained, as explained in *Methods*. The entire simulation time for each image is 139 ns, amounting to a total of 1.529 μs for the 11 images.

Table S2. Summary of the three unrestrained simulations

	1	2	3
Initial coordinates	Open conf.	Open conf.	Closed conf.
Simulation time	100 ns	100 ns	100 ns