

Energetics of glycerol conduction through aquaglyceroporin GlpF

Morten Ø. Jensen^{*†‡§}, Sanghyun Park^{*§}, Emad Tajkhorshid^{*§}, and Klaus Schulten^{*†}

^{*}Beckman Institute, University of Illinois at Urbana–Champaign, Urbana, IL 61801; and [†]Membrane and Statistical Physics Group, Department of Chemistry, Technical University of Denmark, DK-2800 Lyngby, Denmark

Edited by Donald L. D. Caspar, Florida State University, Tallahassee, FL, and approved March 20, 2002 (received for review December 4, 2001)

Aquaglyceroporin GlpF selectively conducts water and linear polyalcohols, such as glycerol, across the inner membrane of *Escherichia coli*. We report steered molecular dynamics simulations of glycerol conduction through GlpF, in which external forces accelerate the transchannel conduction in a manner that preserves the intrinsic conduction mechanism. The simulations reveal channel-glycerol hydrogen bonding interactions and the stereoselectivity of the channel. Employing Jarzynski's identity between free energy and irreversible work, we reconstruct the potential of mean force along the conduction pathway through a time series analysis of molecular dynamics trajectories. This potential locates binding sites and barriers inside the channel; it also reveals a low energy periplasmic vestibule suited for efficient uptake of glycerol from the environment.

Aquaporins (1), a family of water transporting membrane proteins, are present in all life forms, and defects in their function cause physiological disorders (2). Among more than 150 members identified to date (2), the *Escherichia coli* glycerol facilitator (GlpF) belongs to the aquaglyceroporin subclass, which is permeable to both water and glycerol. GlpF also stereoselectively conducts longer linear polyalcohols (3, 4). At physiological conditions, all aquaporins exclude charged solutes, including protons, and thereby preserve the electrochemical potential across the cell membrane (2).

A 2.2-Å resolution x-ray structure of GlpF revealed a homotetrameric architecture with glycerol and water present inside the channel (5). Each monomeric channel has two half-membrane spanning repeats related by a quasi-two-fold symmetry. About half of each repeat is α -helical; the other half adopts a particular nonhelical structure (5, 6). The N termini of the helical repeats meet at the Asn-Pro-Ala (NPA) motifs located at the channel center. The NPA motifs are conserved among all aquaporins (2), and their spatial arrangement is critical for the biological function of the channel (6–8). The nonhelical repeats expose the backbone carbonyl groups of residues 64–66 and 195–201 toward the channel interior, where they serve as hydrogen acceptors for the substrate (6). The channel diameter measures less than 3.5 Å at its narrowest point, the selectivity filter (SF), lined with residues Trp-48, Phe-200, and Arg-206 (5, 9). In the constriction region of the channel, approximately 25 Å long, the substrate is translocated in a single file following a curvilinear pathway (6), a feature critical for excluding proton conduction (10). The hydroxyl groups of glycerol make hydrogen bonds with exposed carbonyl oxygen atoms, polar hydrogen atoms, and water, whereas the aliphatic backbone of glycerol faces the opposite hydrophobic side of the channel (6). The amphipathic channel lining in GlpF ensures the selectivity for linear polyalcohols (5). The amphipathicity of the channel's interior also seems to be important for water transport and is found, although to a less extent, in pure water-conducting aquaporins, such as aquaporin-1 (11–13). Water transport through completely hydrophobic channels, carbon nanotubes, also has been simulated recently (14).

An earlier equilibrium molecular dynamics (MD) study of glycerol-saturated GlpF suggested the mechanism of glycerol conduction (6). The displacement of glycerol was found to be

nonuniform through the constriction region and to occur in steps of 2–2.5 Å, reflecting an exchange of one set of channel-glycerol hydrogen bonds for another. This exchange was thermally driven, mediated by a competition between water and glycerol for the same hydrogen-bonding sites in the channel; glycerol and water translocations were accordingly correlated. Glycerol was found to be least mobile at the SF and the NPA motifs, regions that were identified as binding sites in the 100 K crystal structure (5). The conduction pathway was identified by piecing together the glycerol displacements (about 1 Å over 1 ns, per glycerol) occurring in four monomers (6) and agrees completely with the one found for water transport along the channel in other studies (10, 15).

The present work seeks to investigate complete, continuous conduction events of glycerol through GlpF. Such events are extremely rare on the nanosecond time scale of MD simulations. To realize such events nevertheless, we employ the technique of steered MD (SMD; ref. 16) that accelerates the conduction by applying external forces on glycerol in a manner that retains the intrinsic conduction mechanism. The resulting MD trajectories reveal the mechanism of the conduction and selectivity. Through a time series analysis of MD trajectories, we reconstruct a potential of mean force (PMF), namely the free-energy profile along the conduction pathway, which succinctly characterizes the conduction process.

There have been earlier attempts to reconstruct PMFs from SMD simulations (17–19); the suggested methods, however, either neglected the nonequilibrium character of the simulations or required knowledge of the friction coefficient, which is generally unknown. Recently, Hummer and Szabo (20) used Jarzynski's identity between free energy and irreversible work (21, 22) and reconstructed PMFs in the quasi-equilibrium regime of atomic force microscopy, without needing a friction coefficient. We adapt their method to suit the strong nonequilibrium regime of SMD and reconstruct a PMF for the glycerol conduction through GlpF from actual MD simulations.

Theory and Methods

We model and simulate the entire tetrameric protein in a fully hydrated lipid bilayer. Two complementary SMD schemes are used to induce continuous glycerol conduction through the channel: (i) application of a *constant force* to the center-of-mass (c.o.m.) of glycerol (cf-SMD), and (ii) attachment of the glycerol c.o.m. to a harmonic constraint moving with a *constant velocity* (cv-SMD). This section describes the procedures chosen and

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: GlpF, *E. coli* glycerol facilitator; NPA, Asn-Pro-Ala; SF, selectivity filter; MD, molecular dynamics; SMD, steered MD; PMF, potential of mean force; c.o.m., center-of-mass; cf-SMD, constant force SMD; cv-SMD, constant velocity SMD.

[‡]Present address: Quantum Protein Centre, Department of Physics, Technical University of Denmark, DK-2800 Lyngby, Denmark.

[§]M.Ø.J., S.P., and E.T. contributed equally to this work.

[†]To whom reprint requests should be addressed. E-mail: kschulte@ks.uiuc.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

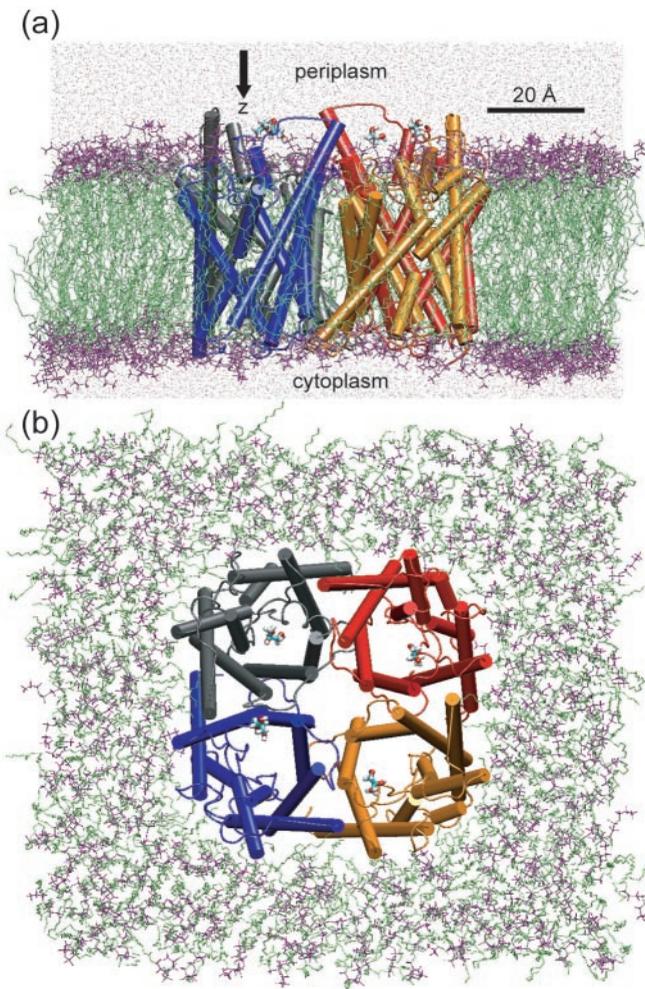


Fig. 1. The GlpF tetramer embedded in a hydrated POPE membrane. (a) Side view. (b) Top view from the periplasmic side. The system contains 106,245 atoms. One glycerol molecule (licorice representation) is positioned at the periplasmic side above each monomer and is pulled along the z-axis during the simulations. The four monomers are colored individually. The lipid head groups and fatty acid part are shown in purple and green, respectively. The water molecules appear as red points (a only).

presents the theory underlying PMF reconstruction from cv-SMD trajectories.

Modeling. The modeling of the GlpF tetramer in a hydrated 16:0/18:1c9-palmitoyloleoyl-phosphatidylethanolamine (POPE) lipid bilayer was carried out as in ref. 6. The POPE membrane accounts well for the properties of *E. coli*'s cell membrane (6, 23, 24) and ensures hydrophobic matching at the protein–lipid interface (6), as indicated by the decrease of the membrane thickness toward the protein–lipid interface noticeable in Fig. 1a.

Crystallographically resolved glycerol molecules located inside the channel were replaced by water molecules (6), and the system was equilibrated for 1 ns at constant temperature and pressure (*NPT* ensemble) through the Langevin piston method (25). The equilibrated structure of GlpF was very close to the crystal structure, the rms deviation of C_{α} atoms being only about 1 Å. For each monomer, one glycerol molecule then was placed at a position lying along the channel axis on either the periplasmic (system 1) or the cytoplasmic (system 2) side of the protein (see Fig. 1). A separation of at least 5 Å between glycerol and the protein was initially imposed. Glycerol and water then were

minimized with the protein fixed and, subsequently, the two systems were each subjected to 100 ps of *NPT* equilibration as described in ref. 6, with all glycerol-oxygen atoms constrained harmonically (with a spring constant of 350 pN/Å) to positions optimal for subsequent pulling. Because the *NPT*-equilibrated volume is not expected to change during the pulling of glycerol, the SMD simulations were conducted at constant volume and temperature, which is computationally less demanding. Langevin dynamics was used to keep the temperature constant, but the glycerol molecules were not coupled to the temperature bath.

All simulations were carried out at 310 K (physiological temperature and above the gel-liquid crystalline phase-transition temperature of POPE), and a pressure of 1 atm (1 atm = 101.3 kPa) was assumed in the *NPT* simulations. With periodic boundary conditions, the Particle Mesh Ewald method (26) was used for computation of electrostatic forces without cutoff. A time step of 1 fs was used throughout. The size of the system counted 106,245 atoms. The MD simulation program NAMD (27) was used with the CHARMM 27 parameter set (28, 29). The simulations were carried out on a 32-processor Linux cluster, and the required time for the 1-ns simulation was about 10 days.

SMD Simulations. The cf-SMD scheme was implemented by applying a constant force (250 pN) along the z-axis (the channel axis) to each glycerol c.o.m. In the cv-SMD simulations used for PMF reconstruction, a harmonic constraint (with a spring constant $k = 150$ pN/Å) was attached to each glycerol c.o.m. and was pulled along the z-axis with a constant velocity (3×10^{-2} Å/ps) (see Eq. 3). The present spring constant is typically chosen two orders of magnitude larger than those used in atomic-force microscopy experiments to realize a finer spatial resolution; the thermal fluctuation of the constrained coordinate is $(k_B T/k)^{1/2} \approx 0.5$ Å, and the corresponding force fluctuation is $(k_B T k)^{1/2} \approx 80$ pN. As indicated in Fig. 2a, the cv-SMD simulations were performed independently through 12 sections of the channel. Pulling through each section started from a configuration prepared by a 100-ps equilibration with the glycerol c.o.m. constrained along the z-axis to the respective initial position. The total simulation times for cf-SMD and cv-SMD schemes were 1 ns and 5 ns, respectively.

Reconstruction of the PMF. The SMD simulation describes a classical mechanical system that is subject to an external perturbation and is governed by a Hamiltonian $\mathcal{H}(\mathbf{r}, t)$ where \mathbf{r} specifies a configuration of the whole system. Jarzynski (22) showed that because such a system evolves according to a Markov process satisfying detailed balance, the free energy $G(t') \equiv -\beta^{-1} \log \int d\mathbf{r} e^{-\beta \mathcal{H}(\mathbf{r}, t')}$ and external work $\mathcal{W}(t') \equiv \int_0^{t'} dt \partial_t \mathcal{H}[\mathbf{r}(t), t]$ are related by the identity

$$e^{-\beta[G(t') - G(0)]} = \langle e^{-\beta \mathcal{W}(t')} \rangle \quad [1]$$

where the average $\langle \cdot \rangle$ is taken over the ensemble of trajectories evolving from the equilibrium distribution for $\mathcal{H}(\mathbf{r}, 0)$. Going one step backward in Jarzynski's proof, one obtains a more informative formulation of the identity (Eq. 20 in ref. 22):

$$e^{-\beta[\mathcal{H}(\mathbf{r}', t') - G(0)]} = \langle \delta[\mathbf{r}(t') - \mathbf{r}'] e^{-\beta \mathcal{W}(t')} \rangle, \quad [2]$$

which is exploited here.

For the present cv-SMD simulations, the Hamiltonian is

$$\mathcal{H}(\mathbf{r}, t) = H(\mathbf{r}) + \frac{k}{2} [z(\mathbf{r}) - z_0 - vt]^2 \quad [3]$$

where $z(\mathbf{r})$ is the position of the glycerol c.o.m. along the channel axis. The first term describes the unperturbed system, and the

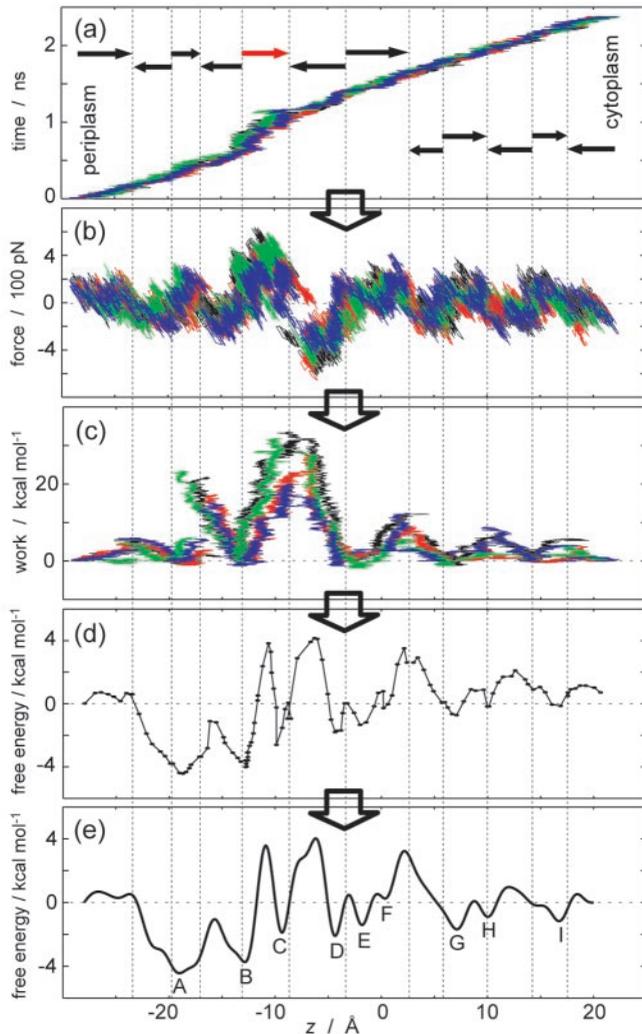


Fig. 2. PMF reconstruction procedure. (a) Trajectory $z(t)$ of the glycerol c.o.m. (b) External force vs. z . (c) External work vs. z . (d) PMFs in each section. (e) Combined PMF. In a, b, and c, different colors (black, red, green, and blue) represent data from different monomers. The position z along the channel axis is measured relative to the midpoint between the N_{δ} atoms of Asn-68 and Asn-203 of the NPA motifs. Arrows in a denote pulling directions, and the red arrow indicates pulling in the SF region with reduced velocity (see text); for reverse pullings, $z(-t)$ are plotted. The PMFs, reconstructed independently in each section, are plotted in d such that they match at the boundaries.

second term describes the harmonic constraint (coupled to glycerol c.o.m. through a spring constant k and initially located at z_0) moving with constant velocity v . The external work applied to glycerol is

$$\mathcal{W}(t') = - \int_0^{t'} dt kv\{z[\mathbf{r}(t)] - z_0 - vt\} = \int_0^{t'} dt v f(t) \quad [4]$$

where $f(t)$ is the force due to the harmonic constraint. We seek to reconstruct the PMF along the coordinate z , which is defined through the equation $G(z') \equiv -\beta^{-1} \log \int d\mathbf{r} \delta[z(\mathbf{r}) - z'] e^{-\beta H(\mathbf{r})}$. By substituting the Hamiltonian (Eq. 3) into Eq. 2 one finds (20)

$$\begin{aligned} G(z') &= -\frac{1}{\beta} \log \langle \delta[z[\mathbf{r}(t')] - z'] e^{-\beta W(t')} \rangle \\ &= -\frac{1}{\beta} \log \langle \delta[z[\mathbf{r}(t')] - z'] \rangle \langle e^{-\beta W(t')} \rangle_{z'} \end{aligned} \quad [5]$$

where we define $W(t') \equiv \mathcal{W}(t') - (k/2)(z' - z_0 - vt')^2$ and where $\langle \cdot \rangle_{z'}$ denotes an ensemble average restricted to trajectories satisfying $z[\mathbf{r}(t')] = z'$. Direct application of Eq. 5 requires a large number of trajectories because it involves evaluating an exponential average (19). We therefore resort to the second order cumulant expansion

$$\begin{aligned} G(z') &\approx -\frac{1}{\beta} \log \langle \delta[z[\mathbf{r}(t')] - z'] \rangle \\ &\quad + \langle W(t') \rangle_{z'} - \frac{\beta}{2} [\langle W(t')^2 \rangle_{z'} - \langle W(t') \rangle_{z'}^2] \end{aligned} \quad [6]$$

since the first and second order moments, $\langle W(t') \rangle_{z'}$ and $\langle W(t')^2 \rangle_{z'}$, can be calculated quite accurately from relatively few trajectories and still produce satisfactory PMFs (30).

For N trajectories $\mathbf{r}_n(t)$ ($n = 1, \dots, N$) obtained from cv-SMD simulations, we calculate the corresponding work $\mathcal{W}_n(t)$ using Eq. 4. The trajectories then are divided into time windows of width Δt ($= 20$ ps), and the mean position in each window, $\bar{z}_n(j) \equiv \Delta t^{-1} \int_j dt z[\mathbf{r}_n(t)]$, is evaluated for each trajectory where $\int_j dt$ denotes the integral over the j -th window. In general, the mean position in a given time window fluctuates among trajectories, and one needs to combine data in different time windows through a weighting scheme seeking to minimize the variance of the quantity being estimated. In contrast to the weighted histogram method (31), in which variances can be deduced from means, the present approach requires one to obtain variances from higher order moments. This could induce an additional statistical inaccuracy. To get around this difficulty, we minimized the fluctuation of $\bar{z}_n(j)$ among trajectories (see Fig. 2a) by using a stiff spring ($k = 150$ pN/Å), which permits us to use all N trajectories within the j -th time window for evaluating the PMF at $\langle \bar{z}(j) \rangle \equiv N^{-1} \sum_n \bar{z}_n(j)$ according to the scheme.

$$G(\langle \bar{z}(j) \rangle) = \langle \bar{W}(j) \rangle - \frac{\beta}{2} [\langle \bar{W}(j)^2 \rangle - \langle \bar{W}(j) \rangle^2] \quad [7]$$

$$\langle \bar{W}(j)^q \rangle \equiv \frac{1}{N} \sum_n \left\{ \int_j dt \left[\mathcal{W}_n(t) - \frac{k}{2} (\langle \bar{z}(j) \rangle - z_0 - v\bar{t}(j))^2 \right]^q \right\} \quad [8]$$

where $\bar{t}(j) \equiv \Delta t^{-1} \int_j dt t$. Note that under this scheme the first term in Eq. 6 vanishes.

By dividing the channel into 12 sections and independently reconstructing a PMF for each section, we minimize the statistical inaccuracy in the external work. The PMF for the full channel (between the periplasmic outlet at $z_p = -28$ Å and the cytoplasmic outlet at $z_c = 20$ Å) is then obtained by combining the sectional PMFs through a linear combination of M ($= 40$) sine functions, $\tilde{G}(z) \equiv \sum_{m=1}^M a_m \sin[m\pi(z - z_p)/(z_c - z_p)]$. The smallest wave length included is 2.4 Å, which indicates the resolution we seek for the PMF. We exclude cosine functions because glycerol is almost in bulk water at both channel outlets, implying $\tilde{G}(z_p) \approx \tilde{G}(z_c)$ such that cosine functions should not contribute. An error is defined for each section by $E \equiv \sum_j [G(\langle \bar{z}(j) \rangle) - \tilde{G}(\langle \bar{z}(j) \rangle) + \tilde{G}(z_{eq})]^2$ where z_{eq} is the mean position of the glycerol c.o.m. in each initially equilibrated configuration. The coefficients a_m are then determined such that they minimize the total error, i.e., the sum of the sectional errors. In each section, we used four trajectories ($n = 4$). The overall procedure is illustrated in Fig. 2.

Results and Discussion

The intrinsic conduction mechanism, namely rupture and formation of channel-glycerol and glycerol-water hydrogen bonds driven by a competition between glycerol and water for the same

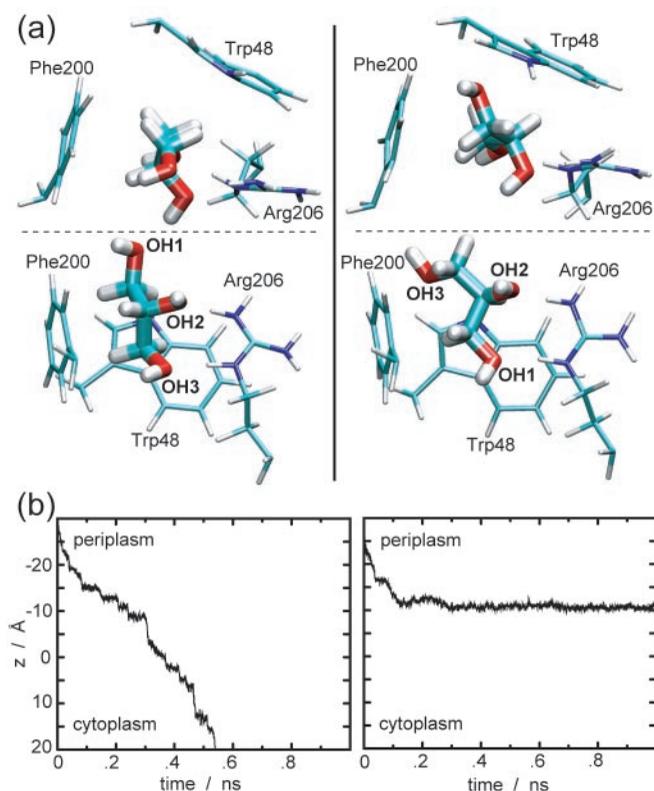


Fig. 3. Stereoselectivity of GlpF for glycerol. (a) Representative snapshots, from cv-SMD simulations, demonstrating favorable (*Left*) and unfavorable (*Right*) orientations of glycerol at the SF region, seen from the top (*Upper*) and the side (*Lower*). The entrance of glycerol with OH3 first (favorable orientation) into the SF leads to the formation of hydrogen bonds between all three hydroxyl groups of glycerol and Arg-206, whereas the aliphatic backbone of glycerol is facing the hydrophobic corner of the SF. This feature is not optimally established when OH1 enters first (unfavorable orientation). (b) Representative cf-SMD trajectories reflecting the stereoselectivity of the channel. With the same magnitude (250 pN) of external forces applied, a favorable orientation of glycerol resulted in a stepwise translocation of glycerol inside the channel (*Left*), whereas an unfavorable orientation resulted in an immobilization of glycerol in the SF region (*Right*) during a 1-ns simulation.

hydrogen bonding sites in the channel (6), was preserved throughout our SMD simulations. This result is evident from the peaks of the distribution of the distances, $d(\text{H}_{\text{glycerol}}-\text{O}_{\text{water}})$ and $d(\text{H}_{\text{glycerol}}-\text{O}_{\text{GlpF}})$, at 1.9 Å and 1.8 Å, respectively [in the calculation of $d(\text{H}_{\text{glycerol}}-\text{O}_{\text{GlpF}})$, only the relevant carbonyl oxygen atoms of residues 64–66 and 198–201 were included; data not shown]. The mechanistic details described in the following can, therefore, be considered intrinsic.

Stereoselectivity of the Channel. Stereoselectivity of GlpF was originally suggested based on different fluxes into liposomes measured for stereoisomers of chiral polyalcohols; the favorable enantiomeric forms were conducted at higher rates (3). Although glycerol is not chiral and the stereoselectivity for its conduction cannot be experimentally measured in terms of conduction rates, the crystal structure of the glycerol-saturated GlpF suggests the channel's preference for one of the prochiral forms of glycerol (5).

GlpF's stereoselectivity for glycerol can be inferred from our SMD simulations. Fig. 3a, taken from cv-SMD simulations, shows two representative orientations of glycerol at the SF. When the hydroxyl group labeled OH3 enters the SF first [favorable orientation, in accord with the glycerol orientation at

the SF in the crystal structure (5); Fig. 3a *Left*], the three hydroxyl groups of glycerol make a series of hydrogen bonds with carbonyl oxygen atoms and polar hydrogen atoms of the protein, while the aliphatic backbone is optimally facing the hydrophobic corner of the SF (amphipathic matching). For the upside-down entry (unfavorable orientation; Fig. 3a *Right*) the amphipathic matching is not optimally established. Consequently, the favorably oriented glycerol could be pulled through the SF by a force of about 400 pN, whereas the unfavorable orientation required a larger force of 450–600 pN.

The stereoselectivity is also reflected in cf-SMD simulations, as illustrated for two representative trajectories in Fig. 3b. For a favorable orientation of glycerol, the applied force of 250 pN led to a stepwise translocation inside the channel (Fig. 3b *Left*). On the other hand, the same magnitude of force was not enough to squeeze unfavorably oriented glycerol through the SF within 1 ns, but rather, glycerol remained immobile at the entrance of the SF (Fig. 3b *Right*). For both orientations, the first channel-glycerol hydrogen bond at the SF region involves Arg-206:H_η, which agrees with the suggestion in ref. 5 that Arg-206:H_η serves as a catalyst for glycerol passage through the SF.

We note that because of the short length of glycerol (as compared with other linear polyalcohols) and its conformational flexibility, not all trajectories showed significant differences between the favorable and unfavorable orientations.

Energetics of the Conduction. A PMF was successfully reconstructed from cv-SMD trajectories, as illustrated in Fig. 2. The most troublesome region in this regard was the SF, functionally the most important part of the channel. For a pulling speed of 0.03 Å/ps, glycerol was occasionally jerked by a large force resulting from a distant constraint point rather than being smoothly translocated by formation of proper contacts with the SF residues. Such events induce strong nonequilibrium effects onto the PMF reconstruction and need to be compensated for by several additional trajectories. Lowering the pulling speed to 0.015 Å/ps in this region, indicated by the red arrow in Fig. 2a, resolved this technical difficulty and ensured smooth translocations through the SF.

Identical free energies can be assumed for a solvated glycerol molecule in the cytoplasmic and periplasmic aqueous media. Nevertheless, the obtained PMF displays a clear asymmetry in the two entrance vestibules of the channel. At the periplasmic vestibule, we note a deep well (A in Fig. 4b), whereas the corresponding site at the cytoplasmic vestibule exhibits only a shallow minimum (I in Fig. 4b). This asymmetry seems related to the structure of GlpF. The periplasmic vestibule is bigger and more protruding from the membrane than the cytoplasmic one (Figs. 1a and 4a); the observed free-energy well in the periplasmic vestibule corresponds to a major structural feature in GlpF formed by the loops connecting helices M4 and M5 as well as M7 and M8. It is noteworthy that a similar structural feature also is present and even more pronounced in maltoporin (32–34) but is absent in such aquaporins as aquaporin-1 that only transport water (11–13).

GlpF is a passive channel, for which, under equilibrium conditions, the inward and outward transport rates are the same. However, the presence of such a significant structural domain on the periplasmic side, and not inside the cell, may be indicative of some functional implications. At physiological conditions, the channel facilitates almost exclusively inward transport of glycerol, because glycerol becomes phosphorylated immediately after entering the cytoplasmic region. Trapping glycerol from the periplasm is likely to be the most important, and probably the rate limiting, step of the transport, especially under low-concentration conditions; the channel must snatch glycerol while it is nearby. Furthermore, in a tetrameric structure, the periplasmic vestibules of the four channels form an even larger attractive

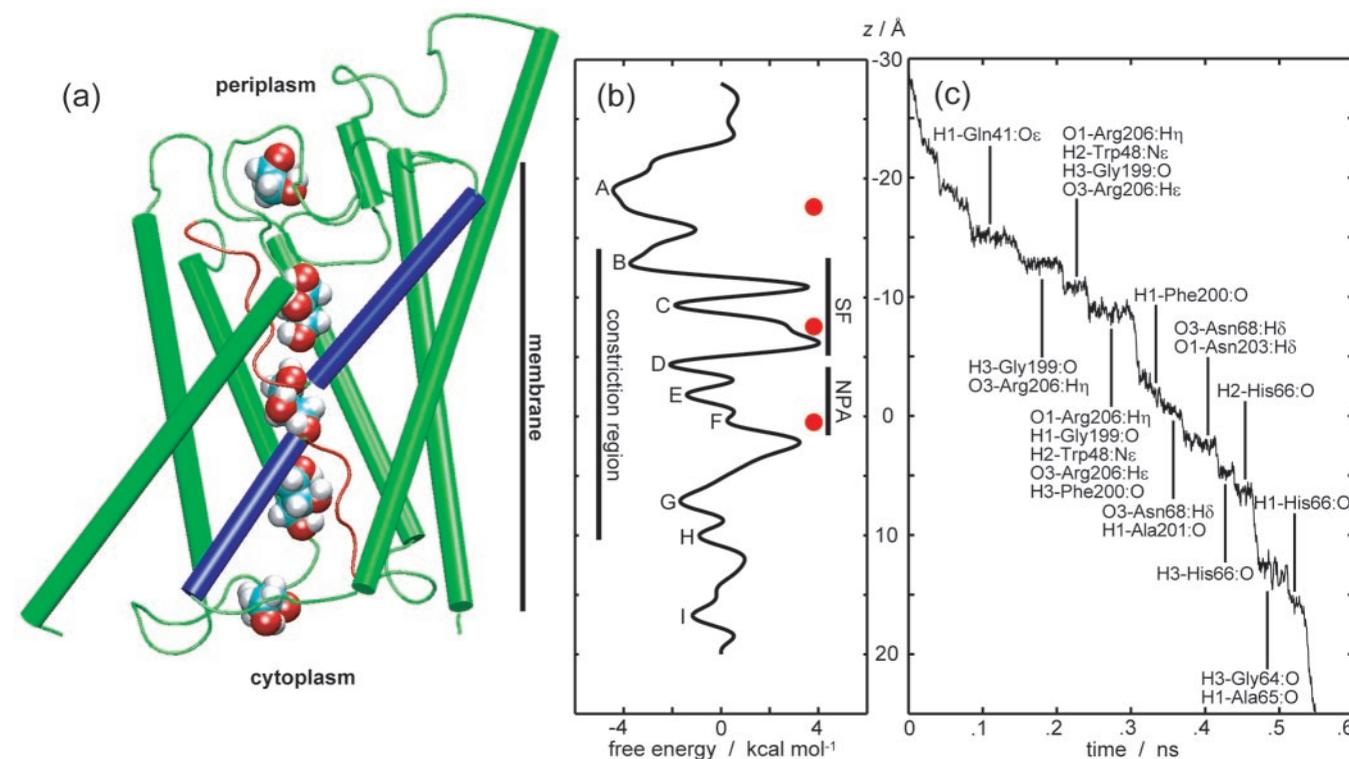


Fig. 4. Energetics of glycerol conduction through GlpF. (a) Monomeric channel drawn on the same scale as *b* and *c*. Five snapshots of glycerol are depicted at positions corresponding to minima, from top to bottom, A, C, E, G, and I in *b*. The nonhelical parts of the half-membrane spanning repeats (red) constitute the conduction pathway. The helical parts of the repeats are shown in blue. (b) The PMF reconstructed from cv-SMD simulations (same as Fig. 2e). Free-energy minima are labeled A–I. Red dots mark the c.o.m. positions of the glycerol molecules in the x-ray structure (5). Vertical bars indicate the SF, the NPA motifs, and the constriction region. The largest barriers are encountered during passage through the SF. The PMF displays a significant difference between the periplasmic vestibule (minimum A) and the cytoplasmic vestibule (minimum I). (c) A trajectory of the glycerol c.o.m. being pulled by a constant force of 250 pN through the channel (same as Fig. 3*b Left*). Residues forming hydrogen bonds with glycerol are listed for each quasi-stationary region of the trajectory. Atoms of the hydroxyl groups of glycerol are labeled 1–3 as in Fig. 3*a*.

region. After being attracted to this region, a glycerol molecule has a higher chance to be transported through one of the four channels available to it. Even if some of the channels are already occupied by other glycerol molecules because of a slow diffusion of glycerol in the channel (as estimated later in this paper, the glycerol transport through GlpF is three orders of magnitude slower than the glycerol diffusion in water), the glycerol molecule can explore other pores of the tetrameric complex before being lost in the bulk.

The PMF reveals further structural and functional features (Fig. 4 *a* and *b*). The highest free-energy barrier against the passage of glycerol is observed at the SF, which is the narrowest part of the channel (5) and is designed to exclude large molecules, thereby acting as a filter. Glycerol and other small linear polyalcohols, however, can compensate the repulsive substrate–protein steric interactions by establishing amphipathic matching with the SF. The PMF shows two minima (B and C in Fig. 4*b*) in the SF region, corresponding to two differently hydrogen-bonded configurations of glycerol during its sliding movement through this region. Displacement of glycerol from these minima toward the center of the channel disrupts the channel–glycerol hydrogen bonds and requires turning of the glycerol backbone in the SF region. Therefore, the passage is accompanied by transient increases of free energy as characterized by high barriers observed in the SF region of the PMF.

The NPA motifs also contribute a binding site. Three free-energy minima (D–F in Fig. 4*b*) are related to the formation of different sets of hydrogen bonds between glycerol and hydrogen atoms of the two asparagines of the NPA motifs. Another

binding site (minima G and H in Fig. 4*b*) is located close to the cytoplasmic side. Therefore, we conclude that, in accord with our earlier suggestion (6), there are three major binding sites for glycerol inside the channel. We recall here that two of these, namely the SF and the NPA motifs, are the only channel segments with polar hydrogen atoms exposed toward the channel interior and that three glycerols are crystallographically observed in these segments (5). Differences between the observed positions of the rather closely spaced glycerol molecules and the PMF minima can be attributed to interactions between the glycerols.

Overall, the largest free-energy barrier is about 7.3 kcal/mol, which compares well with the Arrhenius activation energy 9.6 ± 1.5 kcal/mol inferred from a measurement of glycerol permeability of GlpF-reconstituted liposomes (4).

Analysis of the substrate–channel interactions in the cf-SMD simulations complements the information obtained from the PMF, which was calculated from cv-SMD trajectories. One representative cf-SMD trajectory is shown in Fig. 4*c*. Hydrogen bonding interaction between glycerol and the channel is a major factor in stabilizing quasi-stationary regions characterized by long residence times in the trajectory. The residence time is longest at the SF, where glycerol engages in hydrophobic dispersion interactions with the aromatic rings of Trp-48 and Phe-200 while making hydrogen bonds with Arg-206. The complete passage through the SF is found, for the present trajectory, to involve three quasi-stationary regions separated by 2–2.5 Å along *z*. At the NPA motifs, we find two quasi-stationary regions where glycerol forms hydrogen bonds with the H_δ atoms of

Asn-68 and Asn-203. These hydrogen bonds resemble those between glycerol and the H_γ and H_ε atoms of Arg-206 in the SF region, whereas in the remaining part of the channel, all channel-glycerol hydrogen bonds are between the hydrogen atoms of glycerol and exposed carbonyl groups. The identified hydrogen acceptors are, with the exception of Gln-41:O_ε and Trp-48:N_ε, the exposed carbonyl oxygen atoms of residues 64–66 and 196–201. These findings confirm our earlier identification of the complete pathway and its constituents (6). The long residence at the SF and NPA regions is in accord with the glycerol positions in the crystal structure (5) and with the reduced mobility of glycerol in these regions observed in equilibrium MD simulations (6).

Kinetics of the Conduction. Based on the reconstructed PMF $\tilde{G}(z)$, kinetic quantities can be calculated. We estimate the time scale for glycerol transport through GlpF as follows. We assume that because of the attractive periplasmic vestibule, when a glycerol molecule exits the channel through the periplasmic side, another one immediately enters the channel. On the cytoplasmic side, we assume that glycerol is converted into glycerol-phosphate as soon as it exits the channel. These assumptions are consistent with a reflecting boundary at the periplasmic outlet ($z_p = -28 \text{ \AA}$) and an absorbing boundary at the cytoplasmic outlet ($z_c = 20 \text{ \AA}$). Under these boundary conditions, the mean first passage time from the periplasmic outlet to the cytoplasmic outlet is given by $\tau = \beta \gamma \int_{z_p}^{z_c} dz \int_{z_p}^{z_c} dz' e^{\beta[G(z) - G(z')]} (35)$, where the friction coefficient γ is assumed to be independent of the position of the substrate. By using the friction coefficient for glycerol in bulk water as an approximate value for the intra-channel friction coefficient [$\gamma = k_B T / D$; $D = 1.1 \times 10^{-5} \text{ cm}^2/\text{s}$ (36)], we find $\tau \approx 3.2 \times 10^{-5} \text{ s}$. This provides an estimate for the *single-channel* transport time of glycerol through GlpF, for which there is currently no experimental data available.

Most kinetic quantities, including mean first passage times, depend exponentially on PMFs. More accurate calculations of kinetic quantities would require, therefore, together with a

calculation of the intra-channel friction coefficient, improvement of the PMF by using slower pulling speeds and/or generating more trajectories.

Conclusion

Within the framework of SMD, we have induced complete, continuous conduction events of glycerol through GlpF and confirmed the conduction mechanism and pathway suggested earlier (6). The simulations also reflect the channel's stereoselectivity between two prochiral forms of glycerol. Based on Jarzynski's identity, we have reconstructed a PMF along the conduction pathway. The obtained PMF locates binding sites and barriers inside the channel, capturing structural features and elucidating the selection mechanism. The PMF is asymmetric and reveals a low-energy periplasmic vestibule of the channel, which is suited for efficient uptake of glycerol from the periplasm. The PMF permits calculations of kinetic quantities such as the transport time. SMD simulations and PMF reconstructions, therefore, seem promising for studies of transmembrane channels in general.

The present results may guide future research on GlpF and other transmembrane channels. The selection mechanism for/against various substrates should be investigated with focus on the selective regions identified as key barriers in the PMF. Simulating conduction of longer polyalcohols will give further insight into the stereoselectivity of GlpF. The asymmetry found in the PMF is an intriguing attribute that should be explored further in terms of its physiological implications and also be addressed in studies of other transmembrane channels.

This work was supported by the Roy J. Carver Charitable Trust, National Institutes of Health Grant PHS 5 P41RR05969, and the Danish National Research Foundation via a grant to MEMPHYS Center for Biomembrane Physics. M.Ø.J. acknowledges financial support from the Danish Natural Science Council. Molecular images in the paper were created with the software program VMD (37).

- Preston, G. M., Piazza-Carroll, P., Guggino, W. B. & Agre, P. (1992) *Science* **256**, 385–387.
- Borgnia, M., Nielsen, S., Engel, A. & Agre, P. (1999) *Annu. Rev. Biochem.* **68**, 425–458.
- Heller, K. B., Lin, E. C. & Wilson, T. H. (1980) *J. Bacteriol.* **144**, 274–278.
- Borgnia, M. J. & Agre, P. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 2888–2893.
- Fu, D., Libson, A., Miercke, L. J. W., Weitzman, C., Nollert, P., Krucinski, J. & Stroud, R. M. (2000) *Science* **290**, 481–486.
- Jensen, M. Ø., Tajkhorshid, E. & Schulten, K. (2001) *Structure (London)* **9**, 1083–1093.
- Zhu, F., Tajkhorshid, E. & Schulten, K. (2001) *FEBS Lett.* **504**, 212–218.
- de Groot, B. L., Engel, A. & Grubmüller, H. (2001) *FEBS Lett.* **504**, 206–211.
- Nollert, P., Harries, W. E. C., Fu, D., Miercke, L. J. W. & Stroud, R. M. (2001) *FEBS Lett.* **504**, 112–117.
- Tajkhorshid, E., Nollert, P., Jensen, M. Ø., Miercke, L. J. W., O'Connell, J., Stroud, R. M. & Schulten, K. (2002) *Science* **296**, 525–530.
- Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J. B., Engel, A. & Fujiyoshi, Y. (2000) *Nature (London)* **407**, 599–605.
- Ren, G., Reddy, V. S., Cheng, A., Melnyk, P. & Mitra, A. K. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 1398–1403.
- Sui, H., Han, B.-G., Lee, J. K., Walian, P. & Jap, B. K. (2001) *Nature (London)* **414**, 872–878.
- Hummer, G., Rasaiah, J. C. & Noworyta, J. P. (2001) *Nature (London)* **414**, 188–190.
- de Groot, B. L. & Grubmüller, H. (2001) *Science* **294**, 2353–2357.
- Israelowitz, B., Gao, M. & Schulten, K. (2001) *Curr. Opin. Struct. Biol.* **11**, 224–230.
- Balsera, M., Stepaniants, S., Izrailev, S., Oono, Y. & Schulten, K. (1997) *Biophys. J.* **73**, 1281–1287.
- Evans, E. & Ritchie, K. (1997) *Biophys. J.* **72**, 1541–1555.
- Gullingsrud, J., Braun, R. & Schulten, K. (1999) *J. Comp. Phys.* **151**, 190–211.
- Hummer, G. & Szabo, A. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 3658–3661.
- Jarzynski, C. (1997) *Phys. Rev. Lett.* **78**, 2690–2693.
- Jarzynski, C. (1997) *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat.* **56**, 5018–5035.
- Tieleman, D. P. & Berendsen, H. J. C. (1998) *Biophys. J.* **74**, 2786–2801.
- Tieleman, D. P., Forrest, L. R., Sansom, M. S. P. & Berendsen, H. J. C. (1998) *Biochemistry* **37**, 17554–17561.
- Feller, S. E., Zhang, Y. H., Pastor, R. W. & Brooks, B. R. (1995) *J. Chem. Phys.* **103**, 4613–4621.
- Darden, T., York, D. & Pedersen, L. (1993) *J. Chem. Phys.* **98**, 10089–10092.
- Kalé, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K. & Schulten, K. (1999) *J. Comp. Phys.* **151**, 283–312.
- Schlenkerich, M., Brickmann, J., MacKerell, A. D., Jr., & Karplus, M. (1996) in *Biological Membranes: A Molecular Perspective from Computation and Experiment*, eds. Merz, K. M. & Roux, B. (Birkhauser, Boston), pp. 31–81.
- MacKerell, A. D., Jr., Bashford, D., Bellott, M., Dunbrack, R. L., Jr., Evanseck, J., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. (1998) *J. Phys. Chem. B* **102**, 3586–3616.
- Hummer, G. (2001) *J. Chem. Phys.* **114**, 7330–7337.
- Ferrenberg, A. M. & Swendsen, R. H. (1989) *Phys. Rev. Lett.* **63**, 1195–1198.
- Schirmer, T., Keller, T. A., Wang, Y.-F. & Rosenbusch, J. P. (1995) *Science* **267**, 512–514.
- Dutzler, R., Wang, Y.-F., Rizkallah, P. J., Rosenbusch, J. P. & Schirmer, T. (1996) *Structure (London)* **4**, 127–134.
- Hilty, C. & Winterhalter, M. (2001) *Phys. Rev. Lett.* **86**, 5624–5627.
- Szabo, A., Schulten, K. & Schulten, Z. (1980) *J. Chem. Phys.* **72**, 4350–4357.
- Lide, D. R., ed. (1994) *CRC Handbook of Chemistry and Physics* (CRC, Boca Raton, FL), 75th Ed.
- Humphrey, W., Dalke, A. & Schulten, K. (1996) *J. Mol. Graphics* **14**, 33–38.