Conformational Activation of a Transmembrane Proton Channel from Constant pH Molecular Dynamics

Wei Chen, Yandong Huang, and Jana Shen*

Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, MD

E-mail: jshen@rx.umaryland.edu

Computational details and protocols

Structure preparation. The coordinates of the crystal structure of M2TM (pdb 3LBW,[1] residues 25-46, G34A mutant, 1.65 Å resolution) was retrieved from the database Orientations of Proteins in Membranes (OPM).[2] The OPM database provides membrane protein structures with optimized spatial arrangements with respect to the lipid bilayer. The G34A mutant is functional, as demonstrated by electrophysiology studies.[3] The coordinates of 13 water molecules in the pore of the crystal structure were also kept. Hydrogen atoms were added using CHARMM-GUI.[4,5] Since it was suggested that two His37 residues are charged and the other two are neutral in the crystal structure,[1] we set two histidines along the diagonal to be doubly protonated and the other two single protonated at the ε site during equilibration (see below). The N-terminal 4-bromobenzoyl cap was replaced by the acetyl cap while the C-terminal amino cap was kept. Then the protein was inserted into a DMPC (1,2-Dimyristoyl-sn-glycero-3-phosphocholine) lipid bilayer with 58 lipids in the outer leaflet and 56 lipids in the inner leaflet. A 15 Å water layer was added to both sides of the bilayer. To account for the unresolved water in the pore between Val27 and Ala34 in the crystal structure,[1] CHARMM GUI[4,5] was used, which resulted in 3 water molecules. 10 sodium and 12 chloride ions were added to the bulk water to neutralize the system at pH 6.5 (assuming His37 tetrad is doubly charged and Asp44/Arg45 are charged) and introduce a 150 mM ionic strength. Our previous study showed that the pK_{a}'s from hybrid-solvent CpHMD are not sensitive to the number of explicit ions.[6]

System equilibration. In all simulations, CHARMM22/CMAP force field was used to represent the protein,[7,8] CHARMM36 force filed for lipids[9] and CHARMM modified TIP3P model[10] for water. van der Waals interactions were smoothly switched off between 8 and 12 Å using a force switching function. Particle-mesh Ewald (PME) method[11] was used for calculating electrostatics with a real space cutoff of 12 Å and reciprocal space has about

1 Å grid spacing and 6th-order spline interpolation. The bonds involving hydrogens were constrained using SHAKE[12] in CHARMM[10] or LINCS[13] in GROMACS.[14]

Following the protocol from CHARMM-GUI,[4,5] we first equilibrated the system using CHARMM (version c36b2).[10] Initially, the backbone and sidechain heavy atoms of the protein were restrained with respective force constants of 10 and 5 kcal/mol/Å^2. The 13 crystal water molecules in the pore of M2 were also restrained with a force constant of 5 kcal/mol/Å^2. With the MMFP module, a planar potential with a force constant of 2.5 kcal/mol/Å^2 was applied to keep water out of the hydrophobic region of the lipid bilayer. In addition, planar potentials with a force constant of 2.5 kcal/mol/Å^2 were used to keep the lipid head groups at the bilayer surface and the lipid tails in the center of the bilayer. Dihedral restraints with a force constant of 250 kcal/mol/Å^2 were applied to keep the lipids in the L-glyceride conformation. After energy minimization with steepest descent and adopted basis Newton-Raphson methods, the system was equilibrated in six steps. The force constants of the restraint potential on the protein and crystal water were gradually reduced to 1 kcal/mol/Å^2, while the force constants for other restraints were gradually reduced to zero. In the first two steps, the temperature was controlled at 310 K by Langevin dynamics. In the last four steps, the temperature was controlled at 310 K using Nosé-Hoover thermostat[15,16] while pressure was controlled at 1 atm using the Langevin piston pressure-coupling algorithm.[17] The time step in the first step was set to 1 fs while it was 2 fs in the remainder of the simulations. The first three steps lasted 50 ps each, while the last three steps lasted 100 ps each.

After the initial equilibration in CHARMM, we switched to GROMACS (version 5.0)[14] for further equilibration. We first carried out a 100-ps equilibration in the NVT ensemble. The temperature was controlled at 310 K by the Nosé-Hoover thermostat.[15,16] The heavy atoms of the protein were restrained with a force constant of 10 kJ/mol/Å^2. Then a 1-ns NPT equilibration was performed with the same restraints and temperature, while pressure was controlled at 1 bar by the Parrinello-Rahman approach.[18] After that, the force

constant of the restraints was reduced to 5, 2.5, 1, and 0.5 kJ/mol/Å^2 in four NPT equilibration steps. Each step was 250 ps long. Finally, without restraint, the system was equilibrated in the NPT ensemble for 60 ns. During the equilibration, His37 was fixed in the doubly charged state, while Asp44 and Arg45 were fixed in the charged state.

CpHMD simulations. Following equilibration runs, pH replica-exchange hybrid-solvent CpHMD simulations[6] were performed using CHARMM.[10] In the simulation, explicit water and bilayer were used for conformational dynamics, while the generalized Born (GB) membrane implicit-solvent model was used for propagating titration coordinates. Here, the GBSW model[19] was used with the refined atomic input radii[20,21] The implicit membrane was treated as an infinite planar low-dielectric slab,[22] with a thickness of 31 Å according to the average distance between the C2 atoms of the lipids in the outer and inner leaflets. The switching length for the transition from the implicit membrane to water was set to 5 Å. To exclude the implicit membrane from the protein and account for the continuous water wires in the pore of M2TM,[1,23] a "water" cylinder with a radius of 12 Å was added to the center of mass of the protein. The radius was chosen such that the cylinder includes the maximal number of protein atoms and minimal amount of lipids. The GB calculation and update of titration coordinates were carried out every 10 MD steps to allow relaxation of water molecules.[6]

Using the last snapshot from the equilibration in GROMACS, we added dummy hydrogen atoms to His/Asp residues for titration. Three short equilibration steps at pH 6.5 (the pH that M2 was crystallized) were carried out. The first step was a 100 ps NVT equilibration with temperature controlled at 310 K by the Langevin dynamics. Harmonic restraints with a force constant of 1 kcal/mol/Å^2 were applied on the heavy atoms of the protein. In the second step, a 100 ps NPT equilibration was run with the temperature controlled at 310 K by the Nosé-Hoover thermostat[15,16] and pressure controlled at 1 atm by the Langevin piston pressure-coupling algorithm.[17] The same restraints were maintained. In the last

step, another NPT equilibration run was performed. In place of the aforementioned atom-based restraints, a cylinder restraint was added to the center of mass of the protein heavy atoms by the MMFP module in CHARMM.

The production CpHMD simulation was performed with the pH-based replica-exchange protocol.[6] 12 pH replicas were used in the pH range of 3.5-9.0 with an interval of 0.5. The exchange between neighboring replicas was attempted every 1 ps (500 MD steps). The average exchange acceptance rate was above 45%. The same cylinder restraint was applied to prevent drifting of the protein. The production run lasted 80 ns per replica, resulting in a total sampling time of 960 ns.

pK_{a} calculations. For individual titration sites, we counted the numbers of the protonated ($\lambda < 0.1$) and deprotonated states ($\lambda > 0.9$). Intermediate λ values were discarded as unphysical states (less than 15%). Then, we calculated the deprotonated fraction S at all pH conditions. The site-specific microscopic pK_{a} values were obtained by fitting S to the generalized Henderson-Hasselbalch equation (Hill equation):

$$S = \frac{1}{1 + 10^{n(\mathrm{p}K_{\mathrm{a}} - \mathrm{pH})}} \tag{1}$$

where n is the Hill coefficient. We also calculated the pK_{a}'s of the stepwise titration of His37 tetrad.

$$\begin{aligned}
[\mathrm{His}_4\mathrm{H}^+] &\Leftrightarrow [\mathrm{His}_4^0] + [\mathrm{H}^+] \\
[\mathrm{His}_4\mathrm{H}_2^{2+}] &\Leftrightarrow [\mathrm{His}_4\mathrm{H}^+] + [\mathrm{H}^+] \\
[\mathrm{His}_4\mathrm{H}_3^{3+}] &\Leftrightarrow [\mathrm{His}_4\mathrm{H}_2^2+] + [\mathrm{H}^+] \\
[\mathrm{His}_4\mathrm{H}_4^{4+}] &\Leftrightarrow [\mathrm{His}_4\mathrm{H}_3^3+] + [\mathrm{H}^+]
\end{aligned} \tag{2}$$

The fraction of unprotonated state was calculated as follows.

$$
\begin{aligned}
S_{0\text{-}1} &= \frac{[\mathrm{His}_4^0]}{[\mathrm{His}_4^0] + [\mathrm{His}_4\mathrm{H}^+]} \\
S_{1\text{-}2} &= \frac{[\mathrm{His}_4\mathrm{H}^+]}{[\mathrm{His}_4\mathrm{H}^+] + [\mathrm{His}_4\mathrm{H}_2^{2+}]} \\
S_{2\text{-}3} &= \frac{[\mathrm{His}_4\mathrm{H}_2^{2+}]}{[\mathrm{His}_4\mathrm{H}_2^{2+}] + [\mathrm{His}_4\mathrm{H}_3^{3+}]} \\
S_{3\text{-}4} &= \frac{[\mathrm{His}_4\mathrm{H}_3^{3+}]}{[\mathrm{His}_4\mathrm{H}_3^{3+}] + [\mathrm{His}_4\mathrm{H}_4^{4+}]}
\end{aligned}
\tag{3}
$$

The stepwise pK_a's were obtained by fitting $S_{0\text{-}1}$, $S_{1\text{-}2}$, $S_{2\text{-}3}$, and $S_{3\text{-}4}$ to the Hill equation (Eqn. 1).

Experimental structures of M2TM

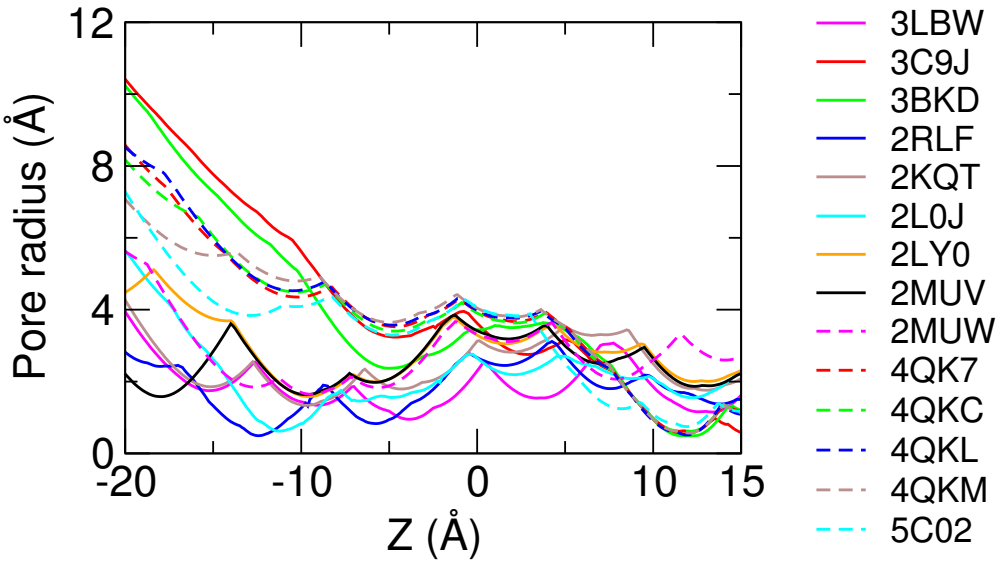

Figure S1: **Differences among the experimental structures of M2TM.** Pore radius along the channel z axis of M2TM for different structures. PDB IDs are indicated in legends. PDB 3LBW, 3C9J, 3BDK, 4QK7/C/L/M and 5C02 are crystal structures. The rest of them are solution or solid-state NMR models. Pore radius were calculated using the HOLE program.[24]

Simulation analysis

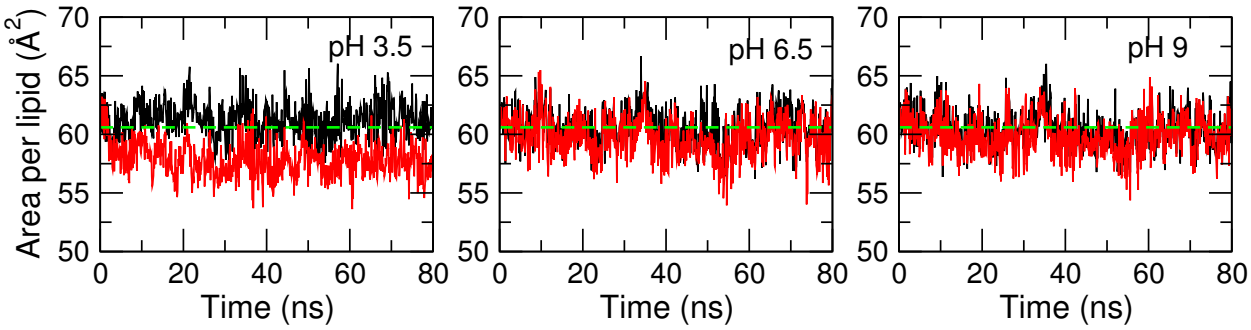

Figure S2: **Area per lipid remains stable in the production CpHMD run.** Three pH replicas are shown. Black and red lines are the values for the outer and inner leaflets, respectively. Green dashed line is the experimental value for the DMPC bilayer.[25] The area per lipid was calculated using the VTMC program.[26] Only the lipids that are more than 5 Å away from the protein were included for average. This is because in the region closer than 5 Å from the protein, lipid and protein atoms overlap resulting in an "artificial" decrease in the calculated area per lipid. This phenomenon was also noted by the developers of the VTMC program.[26] The calculated area per lipid is similar to the published result in the work of the CHARMM36 force field for lipids.[9]

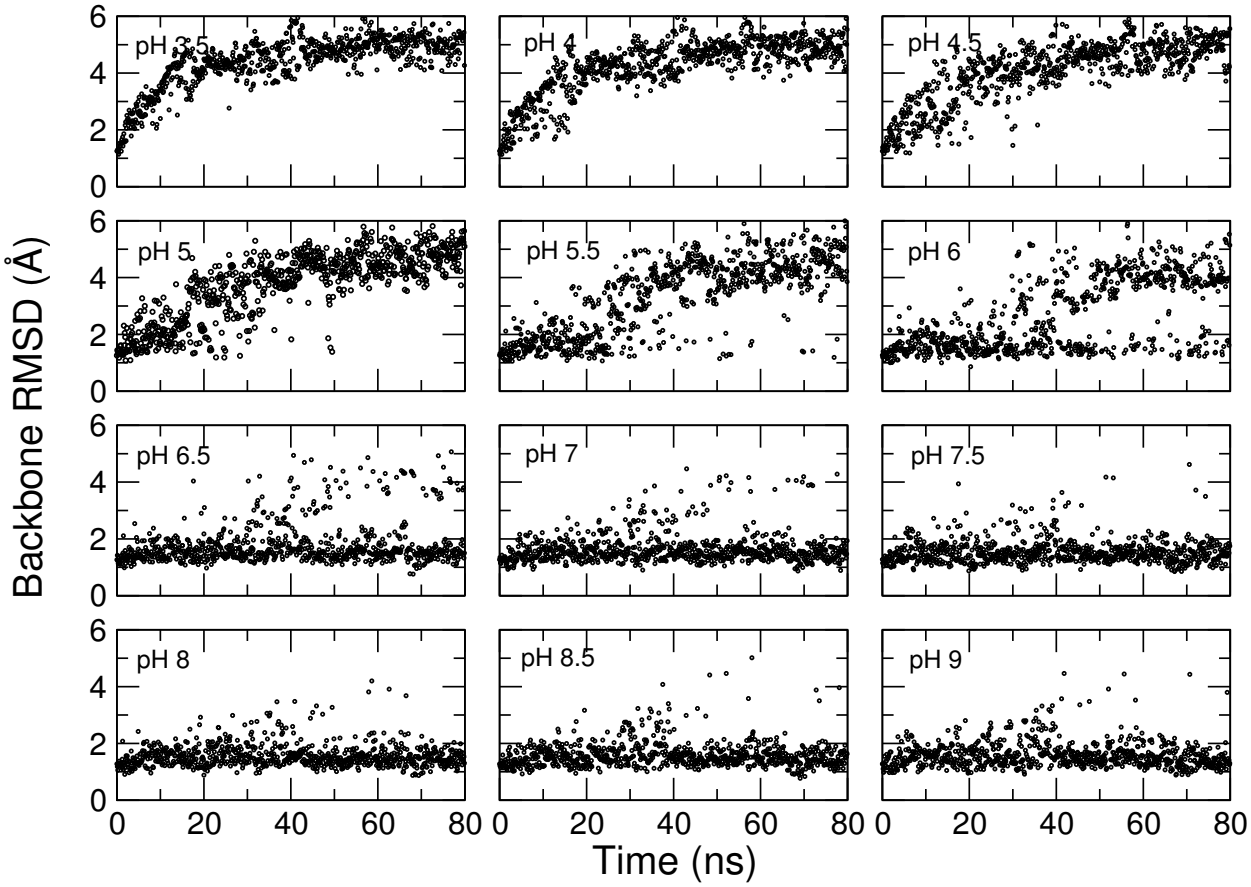

Figure S3: **Convergence of the backbone RMSD.** Time series of backbone RMSD relative to the starting structure (pdb 3LBW) under simulated pH conditions.

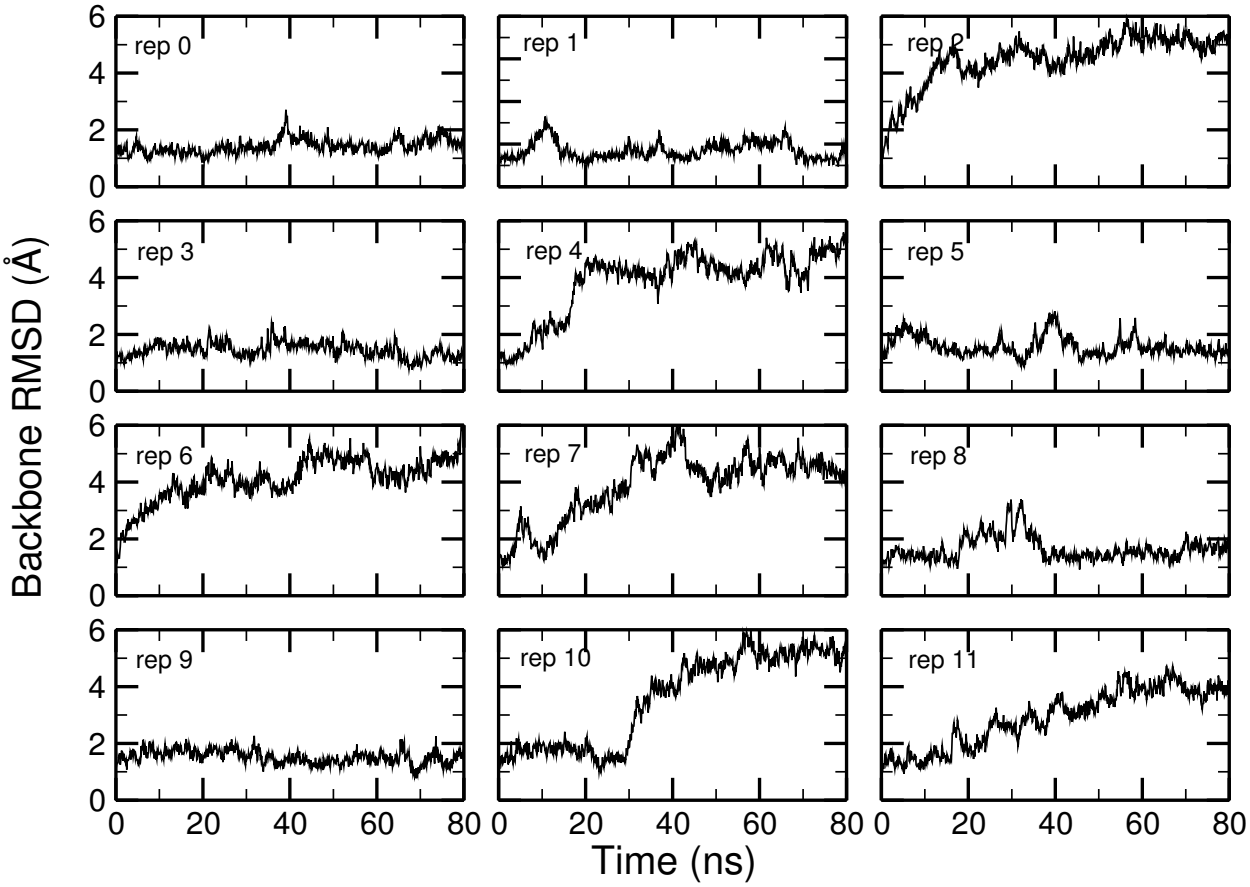

Figure S4: **Conformational change of individual replicas.** Time series of backbone RMSD relative to the starting structure (pdb 3LBW) for all replicas.

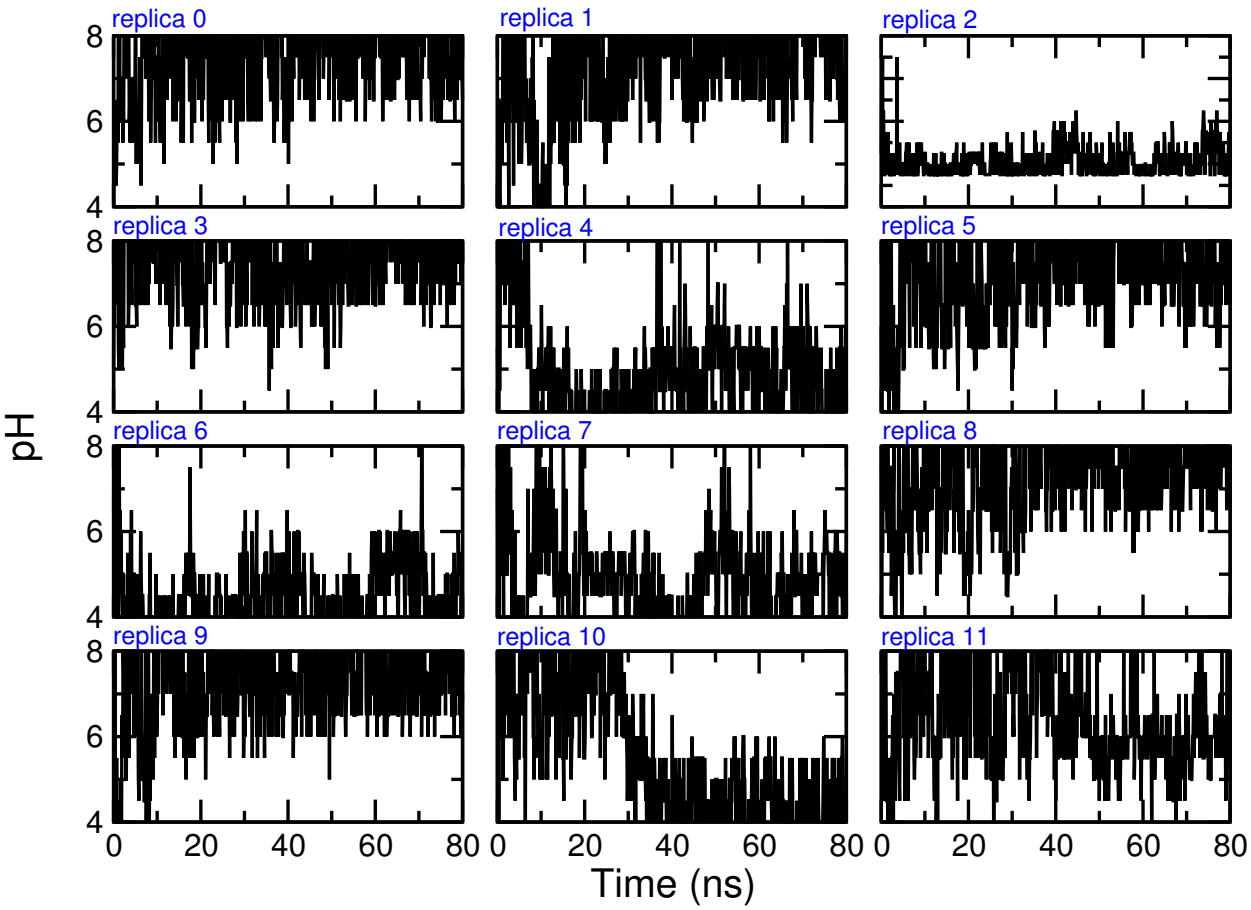

Figure S5: **Replica walk across the entire pH ladder.** Replica walk in the pH space as a function of simulation time.

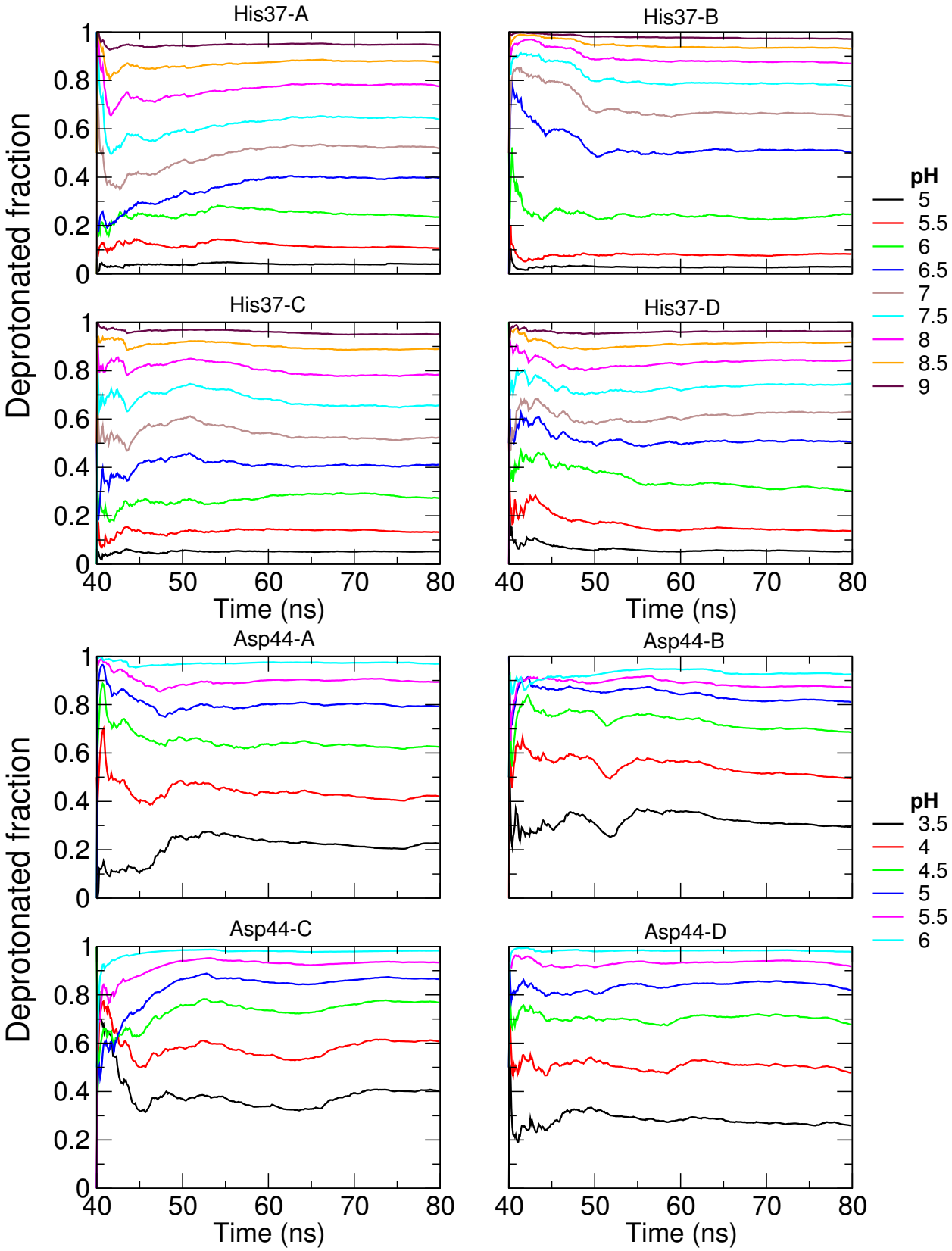

Figure S6: **Convergence of unprotonated fractions.** Cumulative values of the unprotonated fraction of His37 and Asp44 as a function of simulation time.

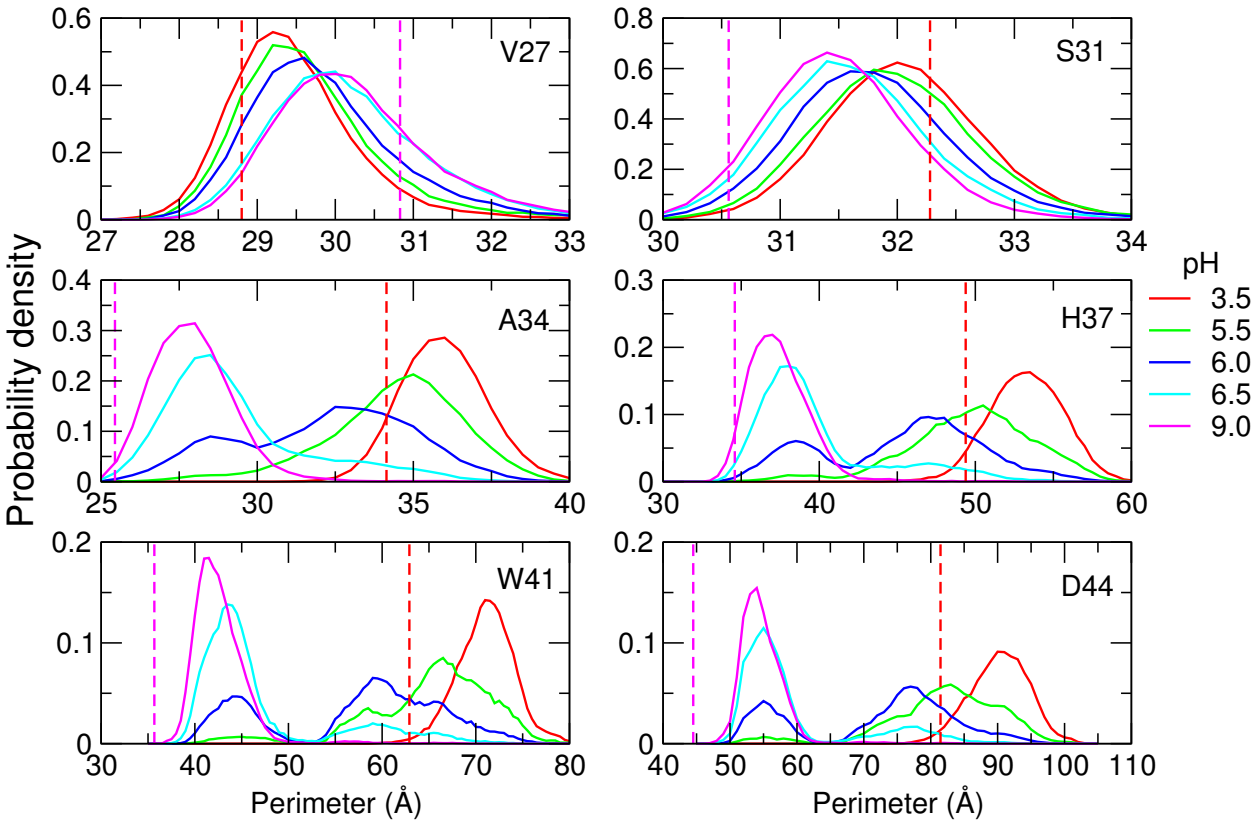

Figure S7: Probability distribution of perimeters at different locations along the pore of M2 under different pH conditions. At the given residue as labeled in each panel, a tetragon was defined using the Cα atoms of four symmetric residues. Then the perimeter of the tetragon was measured. The magenta and red dashed lines are the values in the closed (pdb 3LBW) and open (pdb 3C9J) crystal structures, respectively.

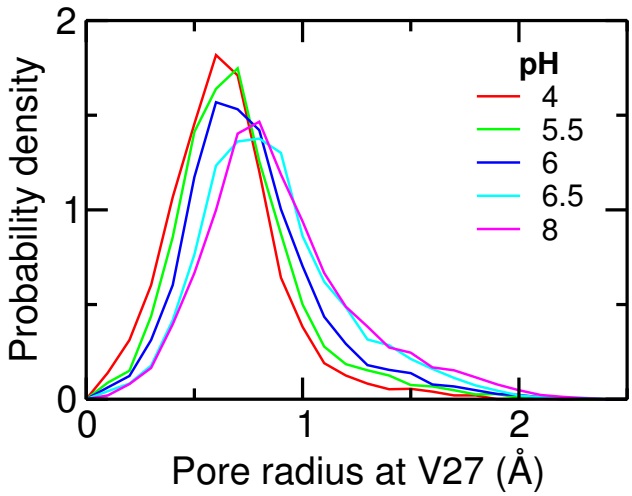

Figure S8: Probability distribution of the pore radius at Val27 at different pH conditions. Pore radius was calculated as the minimum radius near Val27 (10 Å $\leq$ Z $\leq$ 15 Å) using the program HOLE.[24]

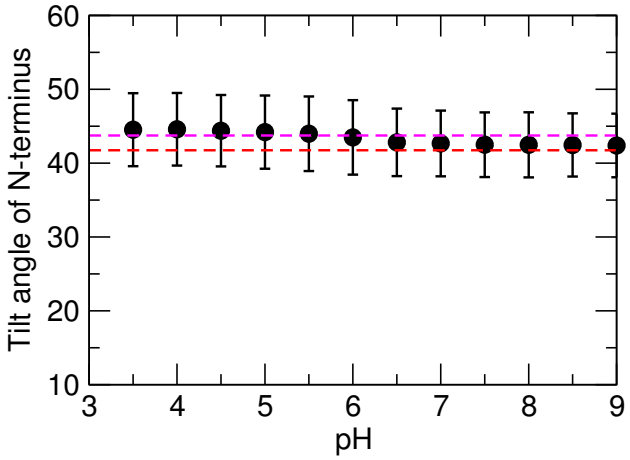

Figure S9: Tilt angle of the N-terminal portion of helices relative to the channel axis. Tilt angle is defined as the principal axis of the N-terminal portion (residues 25-32) of the helix relative to the channel axis. Channel axis is defined as the principal axis of the N-terminal portion (residues 25-32) of the helix bundle. Principal axis was calculated using Cα atoms. Angles were averaged over all four helices and simulation time. Error bars indicate the standard deviations. Magenta and red dashed lines indicate the values in the closed (pdb 3LBW) and open (pdb 3C9J) crystal structures, respectively.

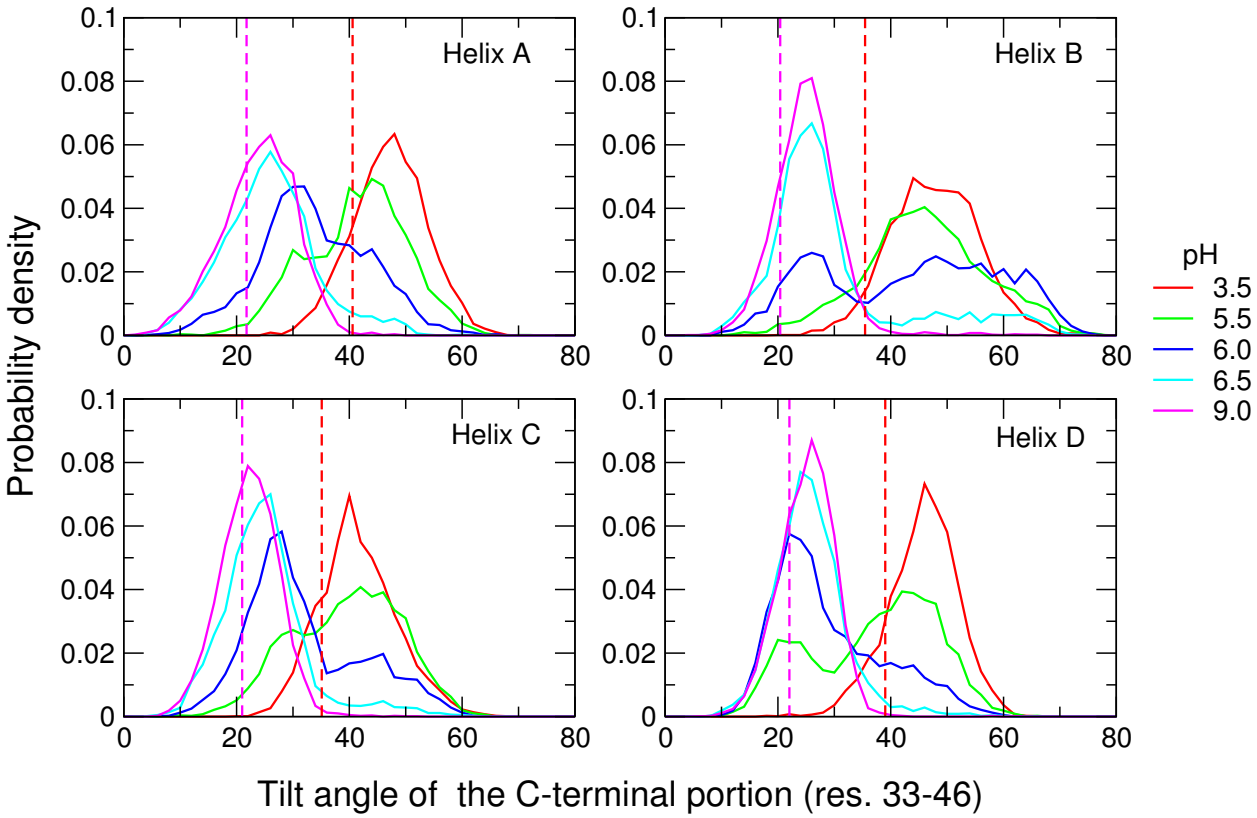

Figure S10: Probability distribution of the tilt angle of C-terminal portion of the helices (residues 33-46) relative to the channel axis. Each panel includes the data for one helix as labeled. The magenta and black dashed lines represent the values in the closed (pdb 3LBW) and open (pdb 3C9J) crystal structures, respectively.

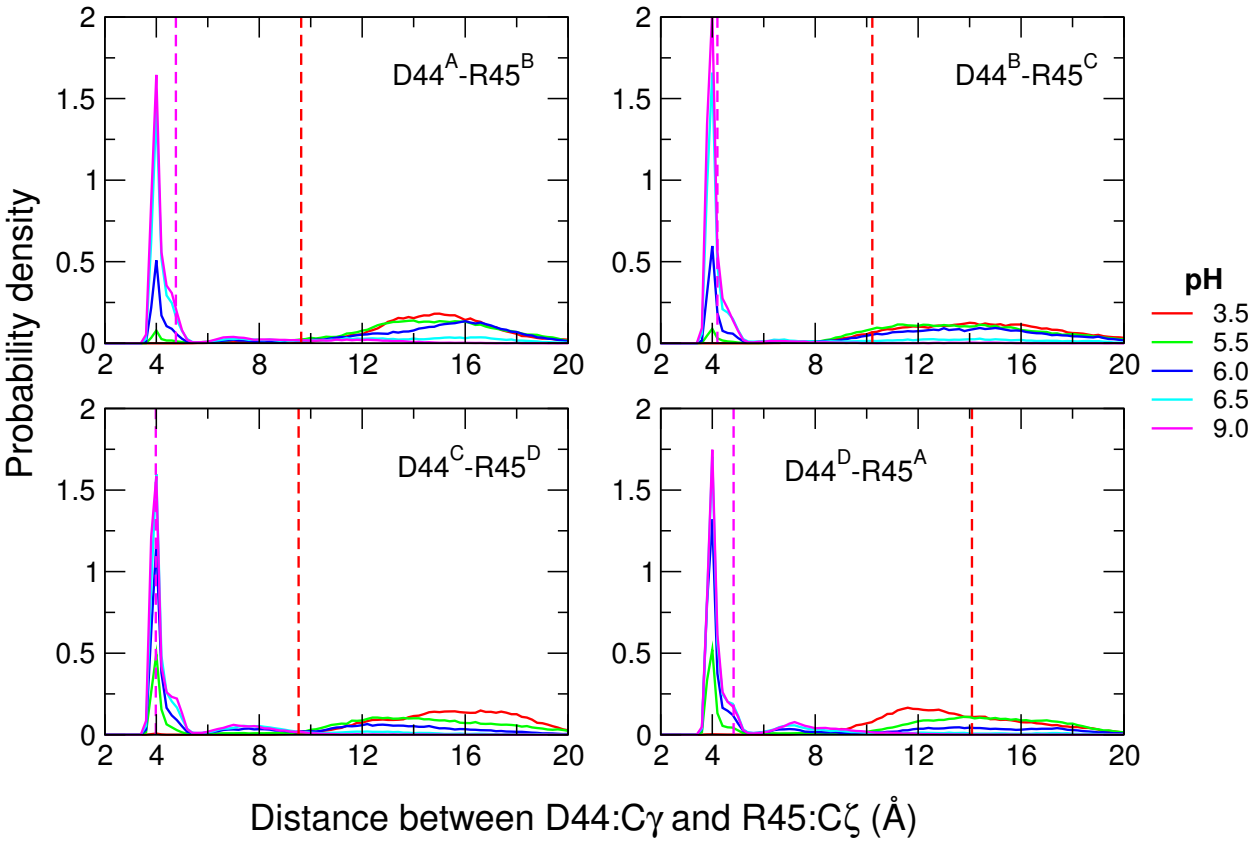

Figure S11: Probability distribution of the distance between the Cγ atom of Asp44 of one helix and the Cζ atom of Arg45 of its neighboring helix. Each panel is for one D44-R45 pair. The magenta and black dashed lines are the values in the closed (pdb 3LBW) and open (pdb 3C9J) crystal structures, respectively.

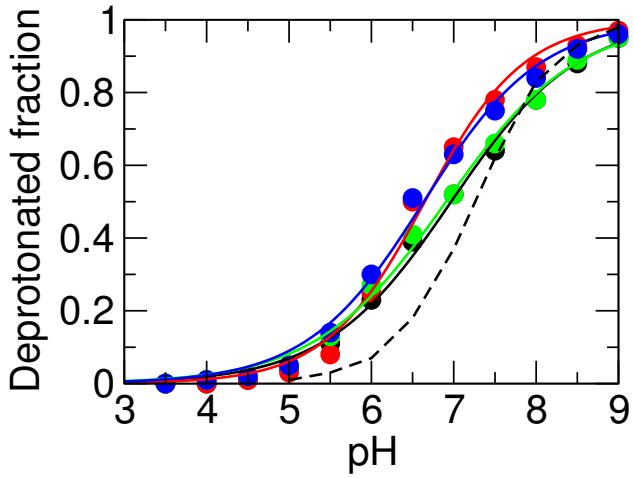

Figure S12: **Microscopic titration of His37.** Deprotonated fraction of individual His37 residues at different pH. Four colors represent four His37 residues. The solid lines are fitting to the Hill equation, which gave four microscopic pK_{a}'s: 7.0, 6.6, 6.9, and 6.6. The dashed line represents the titration curve of model histidine.

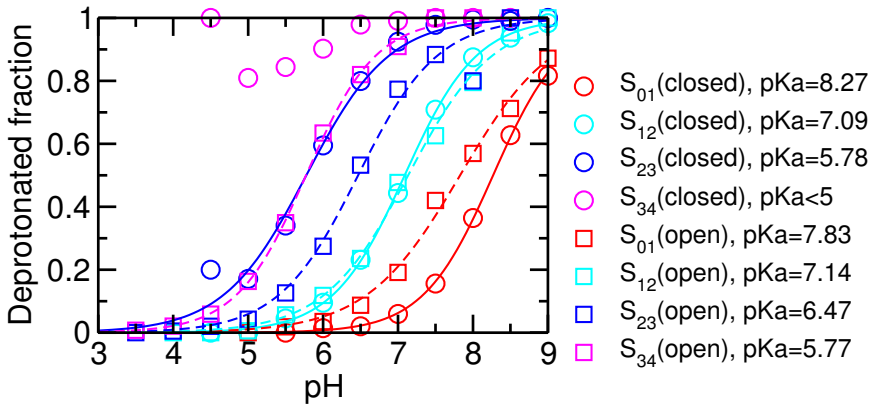

Figure S13: **Stepwise titration of His37 tetrad in the closed and open states.** Deprotonated fraction of His37 tetrad in the closed or open state for each titration step as a function of pH. Four colors represent the four titration steps. Circle and square represent the closed and open states, respectively. The solid lines represent fitting to the Hill equation for the closed state, while dash lines represent fitting for the open state. The corresponding pK_{a} values are listed. The pK_{a} for the last protonation step (+3 to +4 charge) of the closed state could not be determined because the population is too small at low pH.

References

(1) Acharya, R.; Carnevale, V.; Fiorin, G.; Levine, B. G.; Polishchuk, A. L.; Balannik, V.; Samish, I.; Lamb, R. A.; Pinto, L. H.; DeGrado, W. F. et al. Structure and mechanism of proton transport through the transmembrane tetrameric M2 protein bundle of the influenza A virus. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15075–15080.

(2) Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. OPM: orientations of proteins in membranes database. *Bioinformatics* **2006**, *22*, 623–625.

(3) Balannik, V.; Carnevale, V.; Fiorin, G.; Levine, B. G.; Lamb, R. A.; Klein, M. L.; DeGrado, W. F.; Pinto, L. H. Functional Studies and Modeling of Pore-Lining Residue Mutants of the Influenza A Virus M2 Ion Channel. *Biochemistry* **2010**, *49*, 696–708.

(4) Jo, S.; Kim, T.; Im, W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS ONE* **2007**, *2*, e880.

(5) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M. et al. CHARMM-GUI *Membrane Builder* toward realistic biological membrane simulations. *J. Comput. Chem.* **2014**, *35*, 1997–2004.

(6) Wallace, J. A.; Shen, J. K. Continuous constant pH molecular dynamics in explicit solvent with pH-based replica exchange. *J. Chem. Theory Comput.* **2011**, *7*, 2617–2629.

(7) MacKerell Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr., R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(8) MacKerell Jr., A. D.; Feig, M.; Brooks III, C. L. 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.* **2004**, *25*, 1400–1415.

(9) Klauda, J. B.; Venable, R. M.; Freites, J. A.; OâĂŹConnor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; Alexander D. MacKerell, Jr.,; Pastor, R. W. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **2010**, *114*, 7830–7843.

(10) Brooks, B. R.; Brooks III, C. L.; Mackerell Jr., A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartles, C.; Boresch, S. et al. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(11) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(12) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(13) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(14) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D. et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(15) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255–268.

(16) Hoover, W. G. Canonical dynamics: Equilibration phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(17) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613–4621.

(18) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(19) Im, W.; Lee, M. S.; Brooks III, C. L. Generalized Born model with a simple smoothing function. *J. Comput. Chem.* **2003**, *24*, 1691–1702.

(20) Chen, J.; Im, W.; Brooks III, C. L. Balancing solvation and intramolecular interactions: toward a consistent generalized Born force field. *J. Am. Chem. Soc.* **2006**, *128*, 3728–3736.

(21) Nina, M.; Beglov, D.; Roux, B. Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations. *J. Phys. Chem. B* **1997**, *101*, 5239–5248.

(22) Im, W.; Feig, M.; Brooks III, C. L. An implicit membrane generalized Born theory for the study of structure, stability, and interactions of membrane proteins. *Biophys. J.* **2003**, *85*, 2900–2918.

(23) Thomaston, J. L.; Alfonso-Prieto, M.; Woldeyes, R. A.; Fraser, J. S.; Klein, M. L.; Fiorin, G.; DeGrado, W. F. High-resolution structures of the M2 channel from influenza A virus reveal dynamic pathways for proton stabilization and transduction. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 14260–14265.

(24) Smart, O. S.; Goodfellow, J. M.; Wallace, B. A. The pore dimensions of gramicidin A. *Biophys. J.* **1993**, *65*, 2455–2460.

(25) Kučerka, N.; Liu, Y.; Chu, N.; Petrache, H. I.; Tristram-Nagle, S.; Nagle, J. F. Structure of fully hydrated fluid phase DMPC and DLPC lipid bilayer using X-ray scattering from oriented multiamellar arrays and from unilamellar vesicles. *Biophys. J.* **2005**, *88*, 2626–2637.

(26) Mori, T.; Ogushi, F.; Sugita, Y. Analysis of lipid surface area in protein-membrane systems combining voronoi tessellation and monte carlo integration methods. *J. Comput. Chem.* **2012**, *33*, 286–293.