# Supporting Information for: Multiscale kinetic modeling reveals ensemble of $Cl^-/H^+$ exchange pathways in ClC-ec1 antiporter

Heather B. Mayes*[,†,§], Sangyun Lee*[,†,⊥], Andrew D. White[†,∥], Gregory A. Voth**[,†,‡], and Jessica M. J. Swanson**[,†]

[†]*Department of Chemistry,* [‡]*James Franck Institute, and Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois, USA*

[§]*Department of Chemical Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA*

[⊥]*Computational Biology Center, IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, USA*

[∥] *Department of Chemical Engineering, University of Rochester, Rochester, New York 14627-0166, USA*

*Authors contributed equally
**Corresponding authors: gavoth@uchicago.edu, jmswanson@uchicago.edu

### *Additional Details on Simulation Methods*

#### *Initial system setup for MD simulations*

The simulation system consists of the ClC-ec1 dimer (PDB ID: 1OTS) with two $Cl^-$s at $S_{cen}$ and $S_{int}$ for each monomer as they are present in the crystal structure[1], then the $Cl^-$-bound protein structure is solvated with 163 POPE lipids, 17 $Cl^-$s, and ~11,000 water molecules in a 92 Å × 92 Å × 79 Å box with a periodic boundary condition. The lipid membrane was aligned on the xy plane of the simulation box. The CHARMM22 force field[2] was used for protein and ions, CHARMM27[3] for lipids, and the TIP3P water model.[4] For computational efficiency, the Drude oscillator force field (described below) was not implemented until after initial equilibration. Two systems were prepared: E148 in monomer A was set to be protonated in one system, and deprotonated in the other system. E148 in monomer B was set to be deprotonated in both systems. Previous experiments have shown that ClC functions as a monomer[5]; thus, all enhanced sampling used to calculate the potentials of mean force (PMFs) was applied only to monomer A. The protonation states of all other residues were determined based on previous p$K_a$ calculations on the same crystal structure.[6] Specifically, E113 in monomer B and D417 in monomer A were protonated, while the standard protonation states were chosen for all other residues. All other initial system setup and simulation details are described in our previous work.[7] The MD simulation was performed with Gromacs package.[8] The system was first equilibrated for 10 ns with 1 fs time step in the NVT ensemble with a temperature of 300 K. The temperature was controlled by Nose-Hoover thermostat.[9] $Cl^-$ at $S_{int}$ spontaneously diffused out to the intracellular solution, but $Cl^-$ at $S_{cen}$ (this $Cl^-$ referred to as Cl1) remained at the initial position during the equilibration steps.

From the equilibrated structure, a randomly chosen water molecule near the channel mouth on the extracellular solution side was switched to $Cl^-$ (this $Cl^-$ referred to as Cl2), and a randomly

chosen $Cl^-$ in the solution was switched to a water molecule. In order to generate the initial configurations for the 2D umbrella sampling windows in the next step, metadynamics (MetaD) simulation[10] was performed with PLUMED package,[11] for two different systems, with E148 either protonated or deprotonated. The collective variables (CVs) in the MetaD simulations were defined as the distances on the z axis of two $Cl^-$s, Cl1 and Cl2 from the center of mass of all alpha carbons of the protein. The z coordinate of the center of mass of alpha carbons of the initial configuration at time zero was set to be the origin (z = 0 Å) of the z axis. The hill height and the width of a MetaD Gaussian potential were set to be 0.1 kcal/mol and 0.35 Å, respectively. The Gaussian potential was added every 1000 time steps for Cl1 and Cl2, respectively. In the MetaD simulations, Cl1 was the "lower $Cl^-$", which was sampled from $S_{cen}$ to $S_{int}$, and the channel mouth at the intracellular solution, and Cl2 was the "upper $Cl^-$", which was sampled from the channel mouth at the extracellular solution to $S_{out}$, $S_{ext}$, and $S_{cen}$. $S_{out}$, $S_{ext}$, $S_{cen}$, and $S_{int}$ are located at around z = 6, 1, -4, and -7 Å, respectively (see Figure 1A and B in the main text for the positions of the binding sites in protein.) The upper and the lower boundaries for Cl1 were set to be 3 Å above and 20 Å below from $S_{cen}$ in the z axis, respectively, and for Cl2, 20 Å above and 2 Å below from $S_{cen}$ in the z axis, respectively. The half-sided harmonic potential with the force constant, 100 kcal/mol-Å, as added as a wall at each boundary. Additional wall potentials were added to the x and the y directions, in order to prevent Cl1 and Cl2 from escaping from the channel mouth horizontally, where the pore size of the channel mouth becomes larger. The wall potentials were placed at each side of the 35 Å × 30 Å rectangular box on the channel mouth at each side, which was large enough to cover the area of the pore. The MetaD simulation was run for ~40 ns, until the CV trajectories of Cl1 and Cl2 visited between the upper and the lower boundaries at least ~2-3 times. The system coordinates were written every 10 ps. The PMF can be obtained directly from the MetaD, but in

this work, MetaD was performed a bit more aggressively in order to obtain the initial configurations for the umbrella sampling in a short trajectory, where the positions of two $Cl^-$s were distributed along the pathway in the protein over all sampled configurations.

***2D PMF calculations for two $Cl^-$s with polarizable MD force field.***

The 2D umbrella sampling windows in the system with E148 protonated were constructed, where the range of Cl1 was from z = -16 to -2 Å, and for Cl2, from z = -7 to 12 Å. When E148 was deprotonated, Cl2 was repelled by the negatively charged E148, and was not sampled at the region near E148 in the MetaD simulation, which was between $S_{out}$ and $S_{cen}$. In the deprotonated-E148 case, the range of Cl1 was still from z = -16 to -2 Å, but for Cl2, it was from z = 0 to 12 Å. When Cl2 went from $S_{ext}$ to $S_{cen}$. with E148 deprotonated, Cl2 was blocked at E148 gate. The lower boundary for Cl2 in the PMF with E148 deprotonated was set to be z = -1 Å, where the PMF became greater than 15 kcal/mol. We assume that the channel pore is narrow enough and two $Cl^-$s repel each other, two $Cl^-$s cannot simultaneously exist at the same z coordinate. Later PMF calculation results showed that the system is at high energy region, when two $Cl^-$s get close to each other (the lower left corner of the PMFs in the main text Figure 4).

The distances between the neighboring centers in the umbrella sampling were set to be 0.5 Å in both dimensions. The total number of windows in the PMF with E148 deprotonated was ~600 windows, and ~1000 in the PMF with E148 protonated. The initial configurations for the 2D umbrella sampling were taken from the MD snapshots of the system, written in the MetaD. The snapshot was chosen at each window, when the sum of each distance from the positions of Cl1 and Cl2 in the snapshot to each window center was the smallest among all snapshots. For the umbrella sampling MD simulations, the CHARMM36 force field was employed for protein[12] and

lipid.[13] The Drude polarizable force field[14-21] was employed for all atoms, which can account for the redistribution of the electron densities, induced by the change of the surrounding electrostatic field. Previous DFT-level QM calculation studies[22, 23] showed that the polarization effect of $Cl^-$ is significant in such a low dielectric environment in the protein interior. For this system size, the Drude force field is a computationally reasonable choice to properly describe the polarization effect. A subset of Drude model parameters for the interaction between ion and protein residues were updated based on a recent study.[24] The coordinates and the topology files for the CHARMM/Drude MD simulation were generated at CHARMM-GUI website.[25] The MD simulations were performed with NAMD package,[26] where a highly parallelized module for treating the Drude oscillator model was implemented.[27] The system was first equilibrated for 2-300 ps before the production runs with umbrella potentials applied, with 1 fs time step in the NVT ensemble with a temperature of 300 K. The temperature was controlled by Nose-Hoover thermostat.[9] The Drude oscillators were kept at a temperature of 1 K, and the damping coefficient was set to be 20 $ps^{-1}$. An additional quadratic potential was applied to the Drude oscillator when its length exceeds 0.25 Å. The screened Coulomb correction of Thole[28, 29] was calculated for the Drude oscillators that are not excluded from nonbonded interactions within 5 Å. The force constants for Cl1 and Cl2 were 15 kcal-$mol^{-1}$-$Å^{-2}$ at all windows. Replica exchange umbrella sampling (REUS)[30] was performed to improve the convergence in the PMF. During REUS, the conventional Metropolis Monte Carlo exchange criterion was applied every 10 ps to determine swapping the neighboring windows. The CV trajectories of the z coordinates of Cl1 and Cl2 were written every time step during the production runs for 2-3 ns at each window, and the PMFs were calculated from the CV trajectories using the Weighted Histogram Analysis Method (WHAM).[31, 32]

***Calculation of the rate coefficients for $Cl^-$ transport from the PMF***

The rate coefficients were estimated using transition state theory as follows,[7, 33]

$$k_{rxn} = \frac{\omega_0}{2\pi} \exp\left( -\frac{\Delta F^{\ddagger}}{k_B T} \right) \tag{1}$$

where $k_B$ is Boltzmann's constant, $T$ is the simulation temperature (300 K), and $\Delta F^{\ddagger}$ is the free energy barrier height in the PMF. The fundamental frequency $\omega_0$ is that of the reactant state oscillations around its minimum, which is defined as

$$\omega_0 = \sqrt{\frac{\partial^2 PMF(r) / \partial r^2}{m_{eff}}}\Bigg|_{r=r_0} \tag{2}$$

where $r_0$ is the local minimum in the PMFs, and the effective mass of $Cl^-$, $m_{eff}\left\langle v^2 \right\rangle = k_B T / 2$, where the value of $\left\langle v^2 \right\rangle$ was calculated at $r_0$.

***Calculation of the rate coefficients for the in/out diffusions of $Cl^-$***

$Cl^-$ in the solution diffuses into either $S_{out}$ or $S_{int}$ in the protein, depending on its starting position in the solution. The average time for $Cl^-$ in the solution binding to either $S_{out}$ or $S_{int}$, which is the inverse of the rate coefficient, was measured in the Brownian Dynamics (BD) simulations. The BD method was chosen, instead of a regular MD, because the time scale of this step is typically microseconds. The initial configurations for the BD simulation were taken from the equilibrated MD trajectory with classical force field described in the former section. Solvents and ions were removed from the MD configurations, and protein and lipids were fixed at their initial positions. The $Cl^-$ binding time were measured in eight different cases, depending on the starting position of

$Cl^-$, the protonation state of E148, and $Cl^-$ occupancies at $S_{out}$, $S_{cen}$, and $S_{int}$. The concentration of $Cl^-$ in the solution was set to be 300mM. The initial positions of $Cl^-$s in the solution are randomly chosen in the simulation box at least 20 Å away from any protein and lipids atoms. Non-bonded interactions between $Cl^-$ and protein, or lipids in implicit solvent were treated by Generalized Born formalism. Lennard-Jones parameters and atomic charges followed classical CHARMM force field; the polarizable Drude force field was not used for this simulation as it requires 4-5 times more computational time, and was thus prohibitively expensive for these simulations. The BD simulation was performed with Gromacs package.[8] Time step was 5 fs, and temperature was 300 K. The friction coefficient $\gamma$ for $Cl^-$ was 35 $ps^{-1}$, which was determined by using the diffusion coefficient $D$ of $Cl^-$ in the bulk ($2.03 \times 10^{-9}$ $m^2/s$), and Einstein's relation, $D=k_BT/m\gamma$.[34] The reaction zone was defined as 5 Å radius spheres placed at $S_{out}$ and $S_{int}$. Time for any $Cl^-$ in the solution first entering the reaction zone was measured in the BD simulation. To get the average, ten trajectories with randomized initial positions for $Cl^-$s were repeated for each case.

When the external $Cl^-$ transport occurs outward, $Cl^-$ at either $S_{out}$ or $S_{int}$ diffuses out to the solution. The $Cl^-$ escape time was also measured in eight cases, like when measuring the $Cl^-$ binding time. Initial configurations were taken from the 2D umbrella sampling windows described in former section, where $Cl^-$ is bound to either $S_{out}$ or $S_{int}$. Then, umbrella sampling potential was removed, and unbiased MD simulation with the same setups (using CHARMM/Drude force field) was initiated from the configurations. The time from $Cl^-$ originally bound at either $S_{out}$ or $S_{int}$ to first moving 20 Å away from their binding sites, was measured ten times to determine an average "first escape" time.

***Additional MD Results***

Figure S1 shows the results of two-dimensional umbrella sampling to determine how additional residues move as chloride ions transit the ClC-ec1 pore.

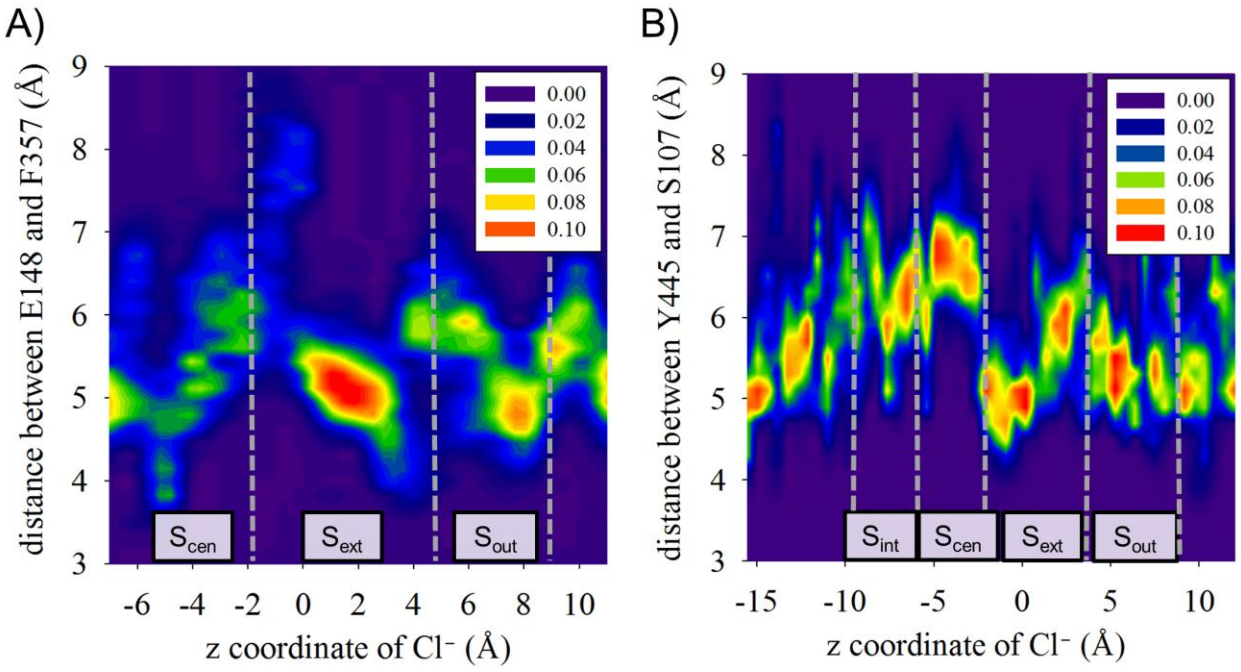


Figure S1. The probability distribution of the opening of the external gate (A) and the internal gate (B). The distance between the center of mass of the carboxyl group of E148 and the phenyl group of F357 for plot A, and between the center of mass of the phenyl group of Y445 and the hydroxyl group of S107 for plot B, are chosen to be the collective variables for the opening of the external and the internal gates, respectively. The distance between two residues is changed by ~ 2 Å, when $Cl^-$ migrates from $S_{ext}$ to $S_{cen}$ (A), when $Cl^-$ is present at $S_{ext}$ and $S_{cen}$, and $S_{cen}$ and beyond $S_{int}$ (B).

### *Adjusting Rate Coefficients from p$K_a$ estimates*

Previous work determined the effect of the presence of $Cl^-$ in the $S_{cen}$ position on the proton transport rate coefficients, but did not specifically look at how the transport rate coefficients are affected by presence of the $Cl^-$ in the $S_{out}$ or $S_{int}$ positions. The change in rate coefficients can be estimated by the determining the change in p$K_a$, which relates to the change in free energy through the relationship:

$$\Delta G = -RT \ln K_a \quad (3)$$

We used PropKa 3.1[35, 36] to determine the p$K_a$s of the key E148 and E203 residues with and without chloride ions present at the three locations used in the MKM. Protein structures were taken from the 2D umbrella sampling window in which chloride ions were present at $S_{cen}$ and $S_{out}$ and E148 was in the up orientation. Chloride ions were then relocated to the three sites for each p$K_a$ calculation. Although the pKa of E148 is certainly influenced by its orientation, all protein coordinates were held fixed in these calculations in order to measure the relative influence of chloride ions at $S_{out}$ or $S_{int}$. The effective dielectric constant for the p$K_a$ calculation was set to be 80 in solvent, and 4 in protein. The results are presented in Table S1. The p$K_a$ of E148 is increased when $Cl^-$ is present at $S_{out}$ or $S_{cen}$, but was not found to be sensitive to the presence at $S_{int}$. The p$K_a$ of E203 is not sensitive to the $Cl^-$ presence at any of the sites.

Table S1. The calculated p$K_a$ of E148 and E203 using PropKa 3.1[35] with E148 in the "up" conformation. Three digit numbers on the first row of the table represent the presence of $Cl^-$ at each binding site, $S_{out}$, $S_{cen}$, and $S_{int}$. For each binding state, the same protein coordinates were used.

| $Cl^-$ binding | 000 | 001 | 010 | 100 | 011 | 101 | 110 | 111 |
|---|---|---|---|---|---|---|---|---|
| p$K_a$(148) | 6.36 | 6.36 | 6.68 | 6.88 | 6.68 | 6.88 | 7.21 | 7.21 |
| p$K_a$(203) | 3.43 | 3.42 | 3.42 | 3.43 | 3.40 | 3.42 | 3.42 | 3.40 |

### *Initial Rate Coefficients Estimated from Simulations*

Table S2 shows the calculated energy barriers and rate coefficients for $Cl^-$ transport in ClC-ec1. In this table, four chloride ion binding sites are designated. However, the MKM treats $S_{out}$ and $S_{ext}$ as one site, a simplification justified by the small barrier between them. The initial rate coefficients used in the MKM are listed in Table S3.

Table S2. Calculated energy barriers and rates for $Cl^-$ transport. The $Cl^-$ binding states at $S_{out}$, $S_{ext}$, $S_{cen}$, and $S_{int}$ are defined in the 2D PMFs with E148 deprotonated and protonated (see main text Figures 1 and 4). The free energy barrier (ΔF) and the rate coefficients (for transition from the initial to the final states) were calculated between two adjacent states. "0" and "1" at each column represent $Cl^-$ is either "absent" or "present" at the binding site, respectively. The units for the free energy barrier and the rate coefficients are kcal/mol and $ms^{-1}$, respectively. The estimated error for each of these coefficients was 2 orders of magnitude.

| initial state | | | | final state | | | | ΔF | rate coefficient |
|---|---|---|---|---|---|---|---|---|---|
| $S_{out}$ | $S_{ext}$ | $S_{cen}$ | $S_{int}$ | $S_{out}$ | $S_{ext}$ | $S_{cen}$ | $S_{int}$ | | |
| E148 deprotonated | | | | | | | | | |
| 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0.4 | 6.8E+05 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 2.9 | 1.0E+04 |
| 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0.8 | 3.6E+05 |
| 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 3.8 | 2.4E+03 |
| E148 protonated | | | | | | | | | |
| 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0.2 | 1.0E+06 |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 2.4 | 2.2E+04 |
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 4.2 | 1.1E+03 |
| 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 4.3 | 1.0E+03 |
| 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 4.8 | 4.4E+02 |
| 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 4.4 | 8.5E+02 |
| 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 3.3 | 5.2E+03 |
| 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 2.3 | 2.7E+04 |
| 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 4.2 | 1.2E+03 |
| 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1.9 | 5.9E+04 |
| 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0.9 | 3.1E+05 |
| 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 4.2 | 1.2E+03 |

Table S3. Initial rate coefficients for transitions between positions tracked with the MKM. Unless designated with a star, all units are $ms^{-1}$. Those with a star correspond to second order elementary steps and have units of $ms^{-1}mM^{-1}$, for ion transport from solution to a protein binding site. As noted in the main text, values of "0" and "1" at each column represent "down" and "up" states (respectively) for the rotation of E148 side-chain; "deprotonated" and "protonated" states (respectively) for E148 and E203 (columns 2 and 3); and "$Cl^-$ absent" and "present" at each of the three tracked $Cl^-$ binding sites. "x" represents that the rate is independent of the position for that descriptor (the rate coefficient is not changed with either "0" or "1" for that column). The rate coefficients listed under "external $Cl^-$" transport are calculated from BD and unbiased MD simulations; those for "internal $Cl^-$" are from the presented 2D PMFs for two $Cl^-$s; and those for the "external $H^+$", "internal $H^+$", and "E148 rotation" are taken from our previous work, and are shown with a light blue background.[7, 37]

| initial state | | | | | | final state | | | | | | rate coefficient |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rot | E148 | E203 | $S_{out}$ | $S_{cen}$ | $S_{int}$ | rot | E148 | E203 | $S_{out}$ | $S_{cen}$ | $S_{int}$ | (1/ms)* |
| $Cl^-$ transitions outside the protein pore from BD simulations | | | | | | | | | | | | |
| x | 0 | x | 1 | 0 | 0 | x | 0 | x | 0 | 0 | 0 | 5.2E+05 |
| x | 1 | x | 1 | 0 | 0 | x | 1 | x | 0 | 0 | 0 | 2.6E+05 |
| x | 0 | x | 0 | 0 | 0 | x | 0 | x | 0 | 0 | 1 | 8.8E+01* |
| x | 1 | x | 0 | 0 | 0 | x | 1 | x | 0 | 0 | 1 | 1.23E+02* |
| x | 0 | x | 0 | 0 | 0 | x | 0 | x | 1 | 0 | 0 | 2.79E+02* |
| x | 1 | x | 0 | 0 | 0 | x | 1 | x | 1 | 0 | 0 | 5.96E+02* |
| x | 0 | x | 0 | 0 | 1 | x | 0 | x | 0 | 0 | 0 | 3.7E+05 |
| x | 1 | x | 0 | 0 | 1 | x | 1 | x | 0 | 0 | 0 | 7.2E+05 |
| x | 0 | x | 0 | 0 | 1 | x | 0 | x | 1 | 0 | 1 | 1.74E+02* |
| x | 1 | x | 0 | 0 | 1 | x | 1 | x | 1 | 0 | 1 | 4.15E+02* |
| x | 0 | x | 1 | 0 | 0 | x | 0 | x | 1 | 0 | 1 | 6.2E+01* |
| x | 1 | x | 1 | 0 | 0 | x | 1 | x | 1 | 0 | 1 | 8.8E+01* |
| x | 0 | x | 1 | 0 | 1 | x | 0 | x | 0 | 0 | 1 | 5.4E+05 |
| x | 1 | x | 1 | 0 | 1 | x | 1 | x | 0 | 0 | 1 | 2.7E+05 |
| x | 0 | x | 1 | 0 | 1 | x | 0 | x | 1 | 0 | 0 | 3.9E+05 |
| x | 1 | x | 1 | 0 | 1 | x | 1 | x | 1 | 0 | 0 | 7.8E+05 |
| x | 0 | x | 0 | 1 | 0 | x | 0 | x | 0 | 1 | 1 | 8.3E+01* |
| x | 1 | x | 0 | 1 | 0 | x | 1 | x | 0 | 1 | 1 | 9.1E+01* |
| x | 0 | x | 0 | 1 | 0 | x | 0 | x | 1 | 1 | 0 | 1.11E+02* |
| x | 1 | x | 0 | 1 | 0 | x | 1 | x | 1 | 1 | 0 | 2.11E+02* |
| x | 0 | x | 1 | 1 | 0 | x | 0 | x | 0 | 1 | 0 | 5.6E+05 |
| x | 1 | x | 1 | 1 | 0 | x | 1 | x | 0 | 1 | 0 | 3.5E+05 |
| x | 0 | x | 0 | 1 | 1 | x | 0 | x | 0 | 1 | 0 | 9.8E+05 |
| x | 1 | x | 0 | 1 | 1 | x | 1 | x | 0 | 1 | 0 | 1.9E+06 |
| x | 0 | x | 0 | 1 | 1 | x | 0 | x | 1 | 1 | 1 | 1.79E+02* |
| x | 1 | x | 0 | 1 | 1 | x | 1 | x | 1 | 1 | 1 | 2.1E+02* |
| x | 0 | x | 1 | 1 | 1 | x | 0 | x | 0 | 1 | 1 | 5.8E+05 |
| x | 1 | x | 1 | 1 | 1 | x | 1 | x | 0 | 1 | 1 | 2.9E+05 |
| x | 0 | x | 1 | 1 | 0 | x | 0 | x | 1 | 1 | 1 | 7.7E+01* |
| x | 1 | x | 1 | 1 | 0 | x | 1 | x | 1 | 1 | 1 | 8.0E+01* |
| x | 0 | x | 1 | 1 | 1 | x | 0 | x | 1 | 1 | 0 | 1.1E+06 |
| x | 1 | x | 1 | 1 | 1 | x | 1 | x | 1 | 1 | 0 | 2.0E+06 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Cl^-$ transitions within the pore from MD simulations with a polarizable FF | | | | | | | | | | | | |
| x | 0 | x | 0 | 0 | 1 | x | 0 | x | 0 | 1 | 0 | 6.8E+05 |
| x | 1 | x | 0 | 0 | 1 | x | 1 | x | 0 | 1 | 0 | 1.0E+06 |
| x | 0 | x | 0 | 1 | 0 | x | 0 | x | 0 | 0 | 1 | 1.0E+04 |
| x | 1 | x | 0 | 1 | 0 | x | 1 | x | 0 | 0 | 1 | 2.2E+04 |
| x | 1 | x | 0 | 1 | 0 | x | 1 | x | 1 | 0 | 0 | 7.3E+02 |
| x | 1 | x | 0 | 1 | 1 | x | 1 | x | 1 | 0 | 1 | 8.7E+02 |
| x | 1 | x | 1 | 0 | 0 | x | 1 | x | 0 | 1 | 0 | 4.1E+02 |
| x | 1 | x | 1 | 0 | 1 | x | 1 | x | 0 | 1 | 1 | 9.5E+03 |
| x | 0 | x | 1 | 0 | 1 | x | 0 | x | 1 | 1 | 0 | 3.6E+05 |
| x | 1 | x | 1 | 0 | 1 | x | 1 | x | 1 | 1 | 0 | 3.1E+05 |
| x | 0 | x | 1 | 1 | 0 | x | 0 | x | 1 | 0 | 1 | 2.4E+03 |
| x | 1 | x | 1 | 1 | 0 | x | 1 | x | 1 | 0 | 1 | 1.2E+03 |
| Adjusting rate coefficients for in/out $H^+$ transitions from E203 | | | | | | | | | | | | |
| x | x | 0 | x | 0 | x | x | x | 1 | x | 0 | x | 1.0E+04* |
| x | x | 0 | x | 1 | x | x | x | 1 | x | 1 | x | 1.0E+04* |
| x | x | 1 | x | 0 | x | x | x | 0 | x | 0 | x | 1.0E+03 |
| x | x | 1 | x | 1 | x | x | x | 0 | x | 1 | x | 1.0E+03 |
| In/out $H^+$ transitions from E148 estimated from previous simulations[37] | | | | | | | | | | | | |
| 1 | 1 | x | 0 | 0 | x | 1 | 0 | x | 0 | 0 | x | 1.1E+03 |
| 1 | 1 | x | 0 | 1 | x | 1 | 0 | x | 0 | 1 | x | 1.8E+00 |
| 1 | 0 | x | 0 | 0 | x | 1 | 1 | x | 0 | 0 | x | 4.3E+02* |
| 1 | 0 | x | 0 | 1 | x | 1 | 1 | x | 0 | 1 | x | 3.0E+03* |
| In/out $H^+$ transitions from E148 from pKa calculations described above | | | | | | | | | | | | |
| 1 | 1 | x | 1 | 0 | x | 1 | 0 | x | 1 | 0 | x | 1.4E+00 |
| 1 | 1 | x | 1 | 1 | x | 1 | 0 | x | 1 | 1 | x | 2.3E-01 |
| 1 | 0 | x | 1 | 0 | x | 1 | 1 | x | 1 | 0 | x | 4.3E+02* |
| 1 | 0 | x | 1 | 1 | x | 1 | 1 | x | 1 | 1 | x | 3.0E+03* |
| $H^+$ transitions between E148 and E203 from previous simulations[7] | | | | | | | | | | | | |
| 0 | 0 | 1 | 0 | 0 | x | 0 | 1 | 0 | 0 | 0 | x | 2.4E+02 |
| 0 | 1 | 0 | 0 | 0 | x | 0 | 0 | 1 | 0 | 0 | x | 2.9E-04 |
| 1 | 0 | 1 | 0 | 1 | x | 1 | 1 | 0 | 0 | 1 | x | 6.4E+05 |
| 1 | 1 | 0 | 0 | 1 | x | 1 | 0 | 1 | 0 | 1 | x | 7.7E-01 |
| $H^+$ transitions between E148 and E203 from p$K_a$ calculations described above | | | | | | | | | | | | |
| 0 | 0 | 1 | 1 | 0 | x | 0 | 1 | 0 | 1 | 0 | x | 7.9E+02 |
| 0 | 1 | 0 | 1 | 0 | x | 0 | 0 | 1 | 1 | 0 | x | 8.8E-05 |
| 1 | 0 | 1 | 1 | 1 | x | 1 | 1 | 0 | 1 | 1 | x | 2.2E+06 |
| 1 | 1 | 0 | 1 | 1 | x | 1 | 0 | 1 | 1 | 1 | x | 2.3E-01 |
| E148 rotation from previous simulations[37] | | | | | | | | | | | | |
| 0 | 1 | 0 | x | 0 | x | 1 | 1 | 0 | x | 0 | x | 1.8E+01 |
| 1 | 1 | 0 | x | 0 | x | 0 | 1 | 0 | x | 0 | x | 2.5E+06 |
| 0 | 0 | 0 | x | 0 | x | 1 | 0 | 0 | x | 0 | x | 4.9E+01 |
| 1 | 0 | 0 | x | 0 | x | 0 | 0 | 0 | x | 0 | x | 4.9E+08 |

***Discrete States in the MKM***

Based on the MD simulations shown in the manuscript Results, we found that it was appropriate to discretize the E148 rotation into two positions, rather than the three positions described based on crystal structures.[38] While more than three binding locations for chloride have been suggested based on crystal structures, we found it sufficient to explicitly include three of the four sites in the MKM ($S_{out}$, $S_{cen}$, and $S_{int}$, also further discussed in the Results). We did not find that explicit description of Y445 position was required to describe ion transport. Some movement of Y445 was observed during movement of chloride ions through ClC-ec1 (Figure S1 in the SI). However, these movements were sufficiently coupled that we did not find it appropriate to separate out transitions into separate steps and model distinct protein states based on Y445 position. As noted in the introduction, there is some evidence that additional protein movements further from the binding sites are involved in ion transport, although their magnitude and involvement in antiport is unclear. As described further in the Results and Discussion, we did not find sufficient data to support including such movements in the MKM at this time.

Since each descriptor has two possible positions, the total number of possible protein states in the MKM is $2^6 = 64$. The associated transition matrix is $64 \times 64 = 4096$; each of these options had to be analyzed to determine if they were possible, and if so, a rate coefficient determined. The full values for the full transition matrix are included in the SI. Briefly, transitions that were found not to be possible, and thus given a rate coefficient of zero in the transition matrix, include movement of a chloride ion directly from $S_{int}$ to $S_{ext}$, without passing through $S_{cen}$. The simulations showed some descriptors changed in a coupled fashion (e.g. movement of a chloride ion from $S_{int}$ to $S_{cen}$ was only found to happen simultaneously with E148 rotating from the "down" to "up" position) while other position changes were independent of each other (e.g. a chloride ion could

bind at $S_{int}$ from the bulk fluid independent of whether E148 was protonated). Very few simultaneous position changes were found (e.g. chloride binding was not found to happen in a concerted transition with E148 protonation). As a result, the transmission matrix is sparse; the rate coefficient is zero (representing a disallowed transition) for 3768 of the 4096 entries. As the MKM would be used to determine the most likely populations of each state and the transitions between them at steady state, the 64 matrix diagonal values were determined from the mass balance requirement, leaving 264 transition values needed. The simulations suggested that only 68 unique transition rate coefficients were needed, as some remained the same independent of one or more of the size descriptors. For example, the E148 protonation from the bulk fluid required that the side chain be located in the "up" position, required that the $S_{out}$ position be unoccupied, did not depend on whether E203 was protonated, and did not depend on whether $S_{int}$ was occupied. E148 could be protonated whether or not either $S_{out}$ or $S_{cen}$ was occupied, but the rate of protonation did depend on their occupancies. Thus, two E148 protonation rate coefficients from the bulk fluid had to be determined from simulation (with and without $S_{cen}$ occupied, which had been determined in a previous study[37]) that could then be used to populate 8 entries in the transition matrix (since each rate coefficient was independent of the position of two descriptors). As shown in Table S3, of the 68 required transition rate coefficients, 32 were obtained from BD simulations reported in this study, 12 from the MD simulations for $Cl^-$ transit with a polarizable force field (also reported in this study), 4 adjusting parameters for E203 protonation/deprotonation, 12 were taken from previously reported studies[7, 37] of proton transport (PT) in ClC-ec1 obtained via multiscale reactive molecular dynamics (MS-RMD) simulations, and 8 were determined using previously reported PT rate coefficients and the p*K*a shift of E148 for different $Cl^-$ binding states. The uncertainties were

estimated to be approximately 2 kcal/mol which corresponds to approximately two orders of magnitude in the resulting rate coefficients.

All rate coefficients associated with residue movements and ion movements from internal protein positions have units of 1/ms, corresponding to first-order elementary steps. Ion binding from bulk fluid, whether protons or chloride ions, corresponds to a second-order elementary step. In order to convert all rate coefficients to rates in the same manner, second-order rate coefficients are converted in the MKM to pseudo-first order coefficient by multiplying them by the appropriate concentration value. For consistency with experimental data, the chloride concentration was specified at 300 mM for the external bulk and 1 mM for the internal bulk. The proton concentration was calculated according the pH specified.

***MKM Rate Coefficient Fitting and Filtering***

As noted in the main text, the individual rate coefficients transitions between states, shown in Table S3, each calculated rate coefficient has associated error. Propagating this error when calculating pathways combining many individual steps would result in uncertainty so large as to obscure valuable information that could be gleaned from the model. As noted in the main text, a variety of macroscopic properties of ClC-ec1 that directly relate to its mechanism have been measured and reported in the literature, from ion exchange rates to p$K_a$ values.[39] We used six experimental data points in a Monte Carlo (MC) fitting procedure, as described below, screened the results to match additional data points, and then left some data points for use to validate the model.

The six experimental values used to refine the MKM by adjusting rate coefficients to fit these values were the $Cl^-$:$H^+$ transport ratios and $Cl^-$ transport rates at pH 4.5, 6.0, and 7.5. The

fitting algorithm used a particle swarm optimization procedure,[40] with multiple independent fitting runs performed in parallel (12000 runs). For each run, the initial parameters (rate coefficients) were those shown in Table S3, and maximum and minimum values were determined based on the estimated uncertainty. First, a pseudorandom number was generated using the standard python library "random.py" to determine which parameter would be adjusted. A maximum change in value was set to one order of magnitude above or below the value, with a second pseudorandom number chosen from a uniform distribution within this maximum adjustment range. If a step of this size would cause the parameter value to fall outside the allowable range based on estimated uncertainty, the magnitude of the step changed was adjusted to the edge of the allowable range. The eigenvectors corresponding to eigenvalues near zero (within $10^{-8}$, to account for floating point precision) were then determined. These eigenvalues represent the population of each state at steady state. To account for half of the proteins being oriented in an opposite orientation, the transition matrix for the "opposite" (as opposed to biological) orientation was determined by recalculating the pseudo-first order rate coefficients, and then eigenvectors for the opposite orientation were also determined. For each orientation, net ion flow through the protein was calculated, and each orientation given a weight of 0.5. A log-likelihood score was then assigned based on difference between the calculated ion transport rates (assuming each orientation should contribute half of the total flow) and those specified by experiment at pH 4.5, 6.0, and 7.5 (pseudo first-order rate coefficients for proton binding from bulk differed between the pH values). Acceptance of the parameter perturbation was determined by a Monte Carlo criterion. After a minimum of 6000 parameter perturbation steps were repeated, solutions which met the following conditions were selected: 1) for pH 4.5 and pH 6.0, the overall chloride transport rate fell within 0.1 ions/ms of the experimental value, and each orientation separately contributed no more than 60% of the

experimental value; and 2) the $Cl^-$:$H^+$ ratio fell within 0.1 of the experimental ratio of 2.2:1 for pH 4.5, 6.0, and 7.5. These parameter sets were then tested to determine if they allowed ion transport in the absence of an ion gradient at pH 4.5 and a $Cl^-$ concentration of 1 mM or 300 mM. For the gradient-free test, only the biological orientation was used; if half of the proteins were in either orientation, any flow would cancel and the results would not be meaningful. This combined process of fitting and filtering resulted in ten unique parameter sets which were then used to generate the results shown in the main text. Pathways for ion transport were determined by following all possible paths through the protein for each ion.

***Discussion of Experimental Data used in the MKM Rate Coefficient Fitting***

Multiple groups have reported a $Cl^-$:$H^+$ exchange ratio of 2.2:1 for $Cl^-/H^+$ over a wide range of relevant pH values.[41-44] Similarly, the chloride ion transport rate has been consistently reported as 2.337 chloride ions per millisecond for an external and internal $Cl^-$ concentration of 300 mM and 1 mM, respectively, at pH 4.5 on both sides of the membrane bilayer.[44, 45] However, two sets of data have been published showing relative rates of chloride ion transport rate at multiple pH values.[39, 46] Lim and Miller reported relative chloride ion transport rates at pH 3.9, 4.9, 5.9, and 7.4.[46] Since the only absolute rate published is at pH 4.5, a relative rate at pH 4.5 must be estimated to convert these relative rates to absolute ones. Figure S2 shows such a conversion to absolute rates by obtaining a relative rate at pH 4.5 from linear interpolation between values for pH 3.9 and 4.9, using an absolute rate of 2.3 ions/ms, published by the same group.[45] However, we did not use these values for training our model, since it was later found that ions were likely leaking through the lipid bilayer in addition to exchanging through the antiporter in this study, complicating the interpretation of these results.[39] In the same work that carefully found the leak through the lipid

bilayer, Picollo et al. determined that chloride ion uptake could be described by a curve of the form $K = A/(1 + K_a/[H^+])$, where A is the maximal rate and $K_a$ is the proton binding constant for E148, which, in the same study, they determined to be 6.2. We set the value for A = 2.38 $Cl^-$ ions/ms to reproduce the groups measured transport rate of 2.337 $Cl^-$ ions/ms at pH 4.5.[44] The MKM results after fitting to two points on this curve from Picollo et al. (and to $Cl^-$:$H^+$ ratio at 3 pH values) are shown in the main text, Figure 6. When the MKM was fit to two points based on the data from Lim and Miller (pH 4.5 and pH 5.9), predicted chloride rates at other pH values (3.9, 4.9, and 7.4) did not match rates for Lim and Miller's data (converted to absolute rates as described above). Interestingly, these points do fall along a curve of the form proposed by Picollo et al. when the values for A and $K_a$ is are adjusted to also reproduce the two chloride transport values that were used to adjust the MKM rate coefficients (Figure S2). The robustness of the equation in describing overall chloride transport rates indicates that the individual transitions in the MKM do indeed correspond to a macroscopically observable chloride transport rate coupled to proton transport rate through a pH-dependent gate.

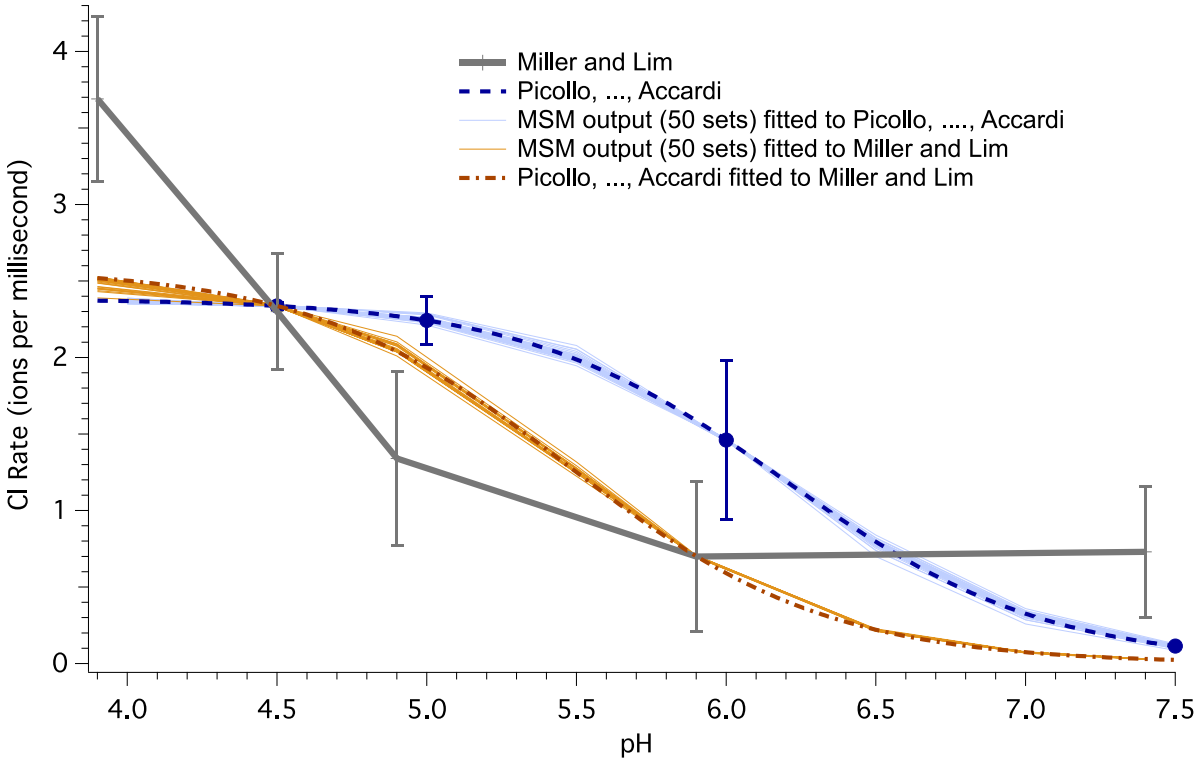


Figure S2. Curves depicting $Cl^-$ transport rate through ClC-ec1 at various pH values when the external $Cl^-$ concentration is 300 mM and internal is 1 mM, and the pH is the same on both sides of the lipid bilayer. Points on the thick, solid gray curve represent experimentally reported values from Lim and Miller,[45, 46] as discussed above. The dashed blue curve represents the formula published by Accardi and coworkers with the parameters fit to their own data.[44] The multiple light blue lines represent MKM output from 50 runs fit to two points (pH 4.5 and 6.0; based on overall flux and not filtered as described above) on the Accardi et al. curve (values output for pH 4.0—7.5 in 0.5 increments), and the orange lines represent MKM output from 50 runs fit to two points from Lim and Miller (fit at pH 4.5 and 5.9; MKM values also output for pH 3.9, 5.5, 6.5, 7.0, and 7.4). The red dashed-and-dotted line shows the formula from Accardi and coworkers with the parameters fit to the same two points from Lim and Miller used in the fitting the MKM output shown in orange.

## Supplemental References


1. Dutzler, R.; Campbell, E. B.; MacKinnon, R., *Science* **2003,** *300* (5616), 108-112.
2. MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S., *et al.*, *J. Phys. Chem. B* **1998,** *102* (18), 3586-3616.
3. Feller, S. E.; MacKerell, A. D., Jr., *J. Phys. Chem. B* **2000,** *104* (31), 7510-7515.
4. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., *J. Chem. Phys.* **1983,** *79* (2), 926-935.
5. Robertson, J. L.; Kolmakova-Partensky, L.; Miller, C., *Nature* **2010,** *468* (7325), 844-847.
6. Faraldo-Gomez, J. D.; Roux, B., *J. Mol. Biol.* **2004,** *339* (4), 981-1000.
7. Lee, S.; Swanson, J. M.; Voth, G. A., *Biophys. J.* **2016,** *110* (6), 1334-1345.
8. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., *J. Chem. Theory Comput.* **2008,** *4* (3), 435-447.
9. Nosé, S., *J. Chem. Phys.* **1984,** *81* (1), 511-519.
10. Laio, A.; Parrinello, M., *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (20), 12562-12566.

11. Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G., *Comput. Phys. Commun.* **2014,** *185* (2), 604-613.
12. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr., *J. Chem. Theory Comput.* **2012,** *8* (9), 3257-3273.
13. Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D., Jr.; Pastor, R. W., *J. Phys. Chem. B* **2010,** *114* (23), 7830-7843.
14. Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D., Jr., *J. Chem. Theory Comput.* **2005,** *1* (1), 153-168.
15. Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D., Jr.; Roux, B., *J. Chem. Theory Comput.* **2006,** *2* (6), 1587-1597.
16. Chowdhary, J.; Harder, E.; Lopes, P. E. M.; Huang, L.; MacKerell, A. D., Jr.; Roux, B., *J. Phys. Chem. B* **2013,** *117* (31), 9142-9160.
17. Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D., Jr., *Chem. Phys. Lett.* **2006,** *418* (1–3), 245-249.
18. Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D., Jr.; Roux, B., *J. Chem. Theory Comput.* **2010,** *6* (3), 774-786.
19. Luo, Y.; Jiang, W.; Yu, H.; MacKerell, A. D., jr.; Roux, B., *Faraday Discuss.* **2013,** *160* (0), 135-149.
20. Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D., Jr., *J. Chem. Theory Comput.* **2013,** *9* (12), 5430-5449.
21. Li, H.; Chowdhary, J.; Huang, L.; He, X.; MacKerell, A. D., Jr.; Roux, B., *J. Chem. Theory Comput.* **2017,** *13* (9), 4535-4552.
22. Church, J.; Pezeshki, S.; Davis, C.; Lin, H., *J. Phys. Chem. B* **2013,** *117* (50), 16029-16043.
23. Smith, M.; Lin, H., *Chem. Phys. Lett.* **2011,** *502* (1–3), 112-117.
24. Li, H.; Ngo, V.; Da Silva, M. C.; Salahub, D. R.; Callahan, K.; Roux, B.; Noskov, S. Y., *J. Phys. Chem. B* **2015,** *119* (29), 9401-9416.
25. Jo, S.; Kim, T.; Iyer, V. G.; Im, W., *J. Comput. Chem.* **2008,** *29* (11), 1859-1865.
26. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K., *J. Comput. Chem.* **2005,** *26* (16), 1781-1802.
27. Jiang, W.; Hardy, D. J.; Phillips, J. C.; MacKerell, A. D., Jr.; Schulten, K.; Roux, B., *J. Phys. Chem. Lett.* **2011,** *2* (2), 87-92.
28. Thole, B. T., *Chem. Phys.* **1981,** *59* (3), 341-350.
29. van Duijnen, P. T.; Swart, M., *J. Phys. Chem. A* **1998,** *102* (14), 2399-2407.
30. Sugita, Y.; Kitao, A.; Okamoto, Y., *J. Chem. Phys.* **2000,** *113* (15), 6042-6051.
31. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., *J. Comput. Chem.* **1992,** *13* (8), 1011-1021.
32. Grossfield, A. *WHAM: an implementation of the weighted histogram analysis method*, ver. 2.0.9.
33. Chandler, D., *Introduction to Modern Statistical Mechanics*. Oxford University Press: New York, 1987; p 234-270.
34. Corry, B.; Allen, T. W.; Kuyucak, S.; Chung, S.-H., *Biophys. J.* **2001,** *80* (1), 195-214.
35. Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H., *J. Chem. Theory Comput.* **2011,** *7*, 525-537.
36. Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H., *J. Chem. Theory Comput.* **2011,** *7*, 2284-2295.
37. Lee, S.; Mayes, H. B.; Swanson, J. M. J.; Voth, G. A., *J. Amer. Chem. Soc.* **2016,** *138* (45), 14923-14930.
38. Vien, M.; Basilio, D.; Leisle, L.; Accardi, A., *J. Gen. Physiol.* **2017,** *149* (4), 523-529.
39. Picollo, A.; Xu, Y.; Johner, N.; Berneche, S.; Accardi, A., *Nat. Struct. Mol. Biol.* **2012,** *19* (5), 525-531.
40. Kennedy, J.; Eberhart, R. C.; Shi, Y., The Particle Swarm. In *Swarm Intelligence*, Morgan Kaufmann: San Francisco, 2001; pp 287-325.
41. Accardi, A.; Miller, C., *Nature* **2004,** *427* (6977), 803-807.
42. Miller, C.; Nguitragool, W., *Phil. Trans. R. Soc. B* **2009,** *364* (1514), 175-180.
43. Nguitragool, W.; Miller, C., *J. Mol. Biol.* **2006,** *362* (4), 682-690.
44. Basilio, D.; Noack, K.; Picollo, A.; Accardi, A., *Nat. Struct. Mol. Biol.* **2014,** *21* (5), 456-463.
45. Lim, H. H.; Shane, T.; Miller, C., *PLoS Biol.* **2012,** *10*, e1001441.
46. Lim, H. H.; Miller, C., *J. Gen. Physiol.* **2009,** *133* (2), 131-138.