THE UNIVERSITY OF CHICAGO

MULTISCALE MODELING OF ION EXCHANGE MECHANISM IN THE CLC ANTIPORTER

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

SANG YUN LEE

CHICAGO, ILLINOIS

JUNE 2016

To my family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Professor Gregory A. Voth for offering a precious opportunity for me to study many interesting problems in biological systems, and providing me a large computation resource for my research. His insight in computational chemistry was invaluable to me in the past years. Working with other great people in the group was another great experience for me.

First, I would like to thank Dr. Jessica M. J. Swanson, who has given me advice on any issues in my research and helped me finding the future career path. I appreciate, Professor Olaseni Sode, who worked with me in the early stage of the project on the transition path sampling for the ATP hydrolysis in actin, and Professor Andrew White, and Heather B. Mayes who worked with me in the development of the Python code for the Markov state model for ClC-ec1 antiporter. I also appreciate Prof. Chris Miller, who always gave me grateful advice on the ClC projects.

I would like to thank other former or current group members, who worked with me on different projects and had invaluable discussions, Dr. Christopher Knight, Dr. Yuxing Peng, Dr. Hui Li, Dr. Christopher Arntsen, Dr. Zhen Cao, Dr. Shule Liu, Dr. Rui Sun, Dr. Chen Chen, Dr. Rajib Biswas, Dr. Jesper Madsen, Ruibin Liang, Zhi Wang, and all others, especially people in the charge transport subgroup.

I appreciate Prof. Aaron Dinner and Prof. Suri Vaikuntanathan, becoming the committee members of my defense.

Above all, I would like to thank my family and friends who supported and encouraged me during all my Ph.D. years.

ABSTRACT

Ion channels or transporters are transmembrane proteins that conduct ions through a gated, water-filled pore between the intracellular and the extracellular environment. Such ion flux play a crucial role in controlling a wide range of biological functions, ranging from cell signaling, osmotic stress response to muscle contraction. Ion fluxes in some channel proteins are modulated by ligand binding, and the binding sites for the ligands are the attractive potential targets for diverse drugs, such as anesthetics, antiepileptics, and analgesics. Channels are distinguished from transporters depending on the need of energy input for transporting ions. In channels, ions diffuses down their electrochemical gradient through the membrane (passive transport), however in transporters, the uphill movement of ions against the gradient is driven by other energy sources from either the chemical energy, such as ATP hydrolysis (primary active transport), or the movement of other ions down the gradient (secondary active transport). ClC-ec1 is a bacterial transporter which mediates the uphill movement of H^+ using the concentration gradient of Cl^- through the membrane, and vice versa.

The physical structures of the proteins are the key source to understand their functions. Due to the recent progress in experimental techniques of recombinant expression, purification and crystallization, high-resolution crystal structures of ClC proteins have been revealed in several species. In addition to structural studies, the electrophysiological experiments along with site-directed mutation applied to some important residues provided some information for understanding how proteins work during the ion transport. However, despite the success accumulated over the past few decades, experimental approaches have some limitations, especially in elucidating all microscopic details. The system goes through multiple intermediate states which

accompany conformational changes of the protein, for example, the channel pore controls the ion conduction by changing between open and close structures. However, in many cases, the crystal structures have been reported for conformations of only a part of all possible intermediate states. Another limitation in site-directed mutagenesis is the difficulty to identify the function of the target residues when they are involved in multiple intermediate steps and the modification in each step leads to a loss of the function.

All-atom molecular dynamics (MD) simulation is a useful tool to override such difficulties because one can generate trajectories with realistic models (force field) of all components in the system, such as protein, lipid membrane, ions and solvents, and directly monitor the ion permeation process and associated conformational changes of the protein. Although MD provides microscopic details of the motions of all atoms, it generally has some limitations in the classical force field and the sampling time of the simulation. Due to its simplicity, classical force field cannot describe neither any chemical reactions nor the effect of electronic polarization. The chemical bonds in classical force field are generally modeled by harmonic potentials as a function of bond distance, which do not allow any bond dissociations. All atoms are treated as point charged particles to evaluate the electrostatic interaction in classical force field, however the interactions between higher order multipoles are not considered. These two limitations should be treated more explicitly in H^+ and Cl^- transport in ClC-ec1. The movement of H^+ through the hydrated cavity in the protein should be considered as a reactive process, where the excess proton migrates by Grotthuss shuttling mechanism between water molecules. The effect of induced polarization can be more significant for Cl^- in the protein than in the bulk solution, where the protein interior is generally in low dielectric environment.

The hybrid QM/MM simulation, which is the combination of quantum mechanics (QM) and empirical force field (MM, molecular mechanics) is a well-known method to simulate chemical reactions in a large biological system. The formation and the cleavage of chemical bonds are described at the core region by QM, and the surrounding region, which is chemically inert, is treated at the classical force field level. However, the simulation time accessible by the current computing power is typically on a time scale of picoseconds, when a DFT-level QM/MM method is applied in such large system. The time scale of the permeation of one ion through the protein is approximately in the order of milliseconds. In order to access the dynamics of long time scale, enhanced sampling techniques, such as umbrella sampling or metadynamics, are employed, which apply the bias potential on the pre-determined reaction coordinate and explore the free energy landscape in more efficient way. However, the success of the enhanced sampling depends on the converged sampling of other degrees of freedom orthogonal to the reaction coordinate. Although the reaction coordinate is well chosen for the enhanced sampling, the QM/MM MD may not be feasible if other degrees of freedom are on a time scale slower than picoseconds.

The MS-RMD (multi-scale reactive molecular dynamics) model has been developed in the group for a long time. The model explicitly describes the deprotonation reaction of amino acids. It enables more efficient sampling without much loss of accuracy, compared to the QM/MM MD. The model parameters were obtained by the force fitting algorithm (FitRMD), which finds the best fit to the reference data, which is the forces calculated by a DFT-level QM/MM using configurations sampled around the transition state. The resulting MS-RMD model was shown to faithfully reproduce the free energy profiles (potential of mean forces, PMF) of the reference QM/MM Hamiltonian for H^+ transport. The MS-RMD MD allowed us to calculate the two dimensional PMFs at an affordable computational cost. The two dimensional PMFs provided a

more detailed information about other important degrees of freedom, strongly coupled with H^+ transport, such as the increase of the hydration level in H^+ pathway and the changes of the orientation of the surrounding residues.

CHAPTER 1

DEVELOPMENT OF MS-RMD MODEL FOR DEPROTONATION REACTION OF GLUTAMIC ACID

1.1 Introduction

The hydrated excess "proton" is in fact a unit of net positive charge(1) due to a missing electron that can be passed between and among biomolecules(2) in a seemingly simple dance of charge neutralization. However, the molecular nature of this dance is complicated, involving a dynamic charge delocalization between molecules and a constant restructuring of covalent and hydrogen bond topology. Understanding and being able to characterize the migration of excess protons is important given the many roles that proton transport (PT) plays throughout biology. Virtually all biomolecules are sensitive to pH and many incorporate PT into their functional mechanisms, including transporters, proton pumps, proton channels, and enzymes. For example, the bacterial H^+/Cl^- antiporter ClC-ec1, a homolog of mammalian ClC antiporters, energetically couples the transmembrane transport of two Cl^- ions and one proton.(3, 4) The mammalian ClCs have many physiological functions, including maintenance of the membrane potential, regulation of transepithelial Cl^- transport, and control of pH in the cytoplasm and intracellular organelles.(5-9) Cytochrome *c* oxidase (C*c*O), a proton pump in the respiratory chain of mitochondria and bacteria, reduces oxygen to water and utilizes the released free energy to pump protons across the membrane, contributing to the transmembrane electrochemical potential gradient that is necessary for ATP

synthesis.(10-13) The influenza A M2 proton channel protein(14) transports the protons across the influenza virus membrane and triggers the dissociation of the viral matrix proteins, which is an essential step in the influenza virus replication cycle.(15) This is a short list of the many systems in which PT plays a role in a functional mechanism. For such systems the ionizable amino acids that influence in the PT pathways often play an active role via transient protonation and deprotonation. For example in the ClC-ec1 antiporter, an intracellular-facing glutamic acid, E203, shuttles an excess proton to an extracellular-facing glutamic acid, E148, during transmembrane PT.(16, 17) In C*c*O the highly conserved glutamic acid E286 is the key for both transmembrane proton pumping and PT for the chemical reaction.(18, 19) In the influenza A M2 channel, a highly conserved tetrad of histidine residues (H37) is responsible for the pH-dependent channel activation and proton selectivity.(20)

Computational approaches can play an important role in the investigation of PT mechanisms in proteins, adding molecular level insight as well as increased temporal and spatial resolution to experimental data. However, it is very challenging to *explicitly* model the PT process, even in simple bulk water solution, because it involves charge defect delocalization, Grotthuss shuttling (proton hopping), and solvent reorganization. Moreover, the migration of an excess proton in proteins and other confined spaces can be non-trivially coupled with changes in the hydration along the PT pathway.(21-23) As will be discussed later, water molecules move in/out of internal protein cavities in response to the excess positive charge as it moves between water molecules and ionizable amino acids. Therefore, to more completely describe PT in biological systems, a computational method should (1) explicitly treat charge defect delocalization and Grotthuss shuttling of the excess proton(s) between water molecules and ionizable amino acids undergoing protonation/deprotonation; (2) allow exchange of water molecules between different protein

internal cavities as well as between those cavities and bulk solution; and (3) be computationally efficient enough to achieve sufficient sampling of the charge translocation and protein and solvent configuration space for large-scale biomolecular systems, including protein, solvent, ions, and (where needed) membrane. Quantum mechanical approaches, such as *ab initio* molecular dynamics (AIMD) or quantum mechanics molecular mechanics (QM/MM) MD methods, can in principle explicitly treat the reactive nature of amino acid protonation/deprotonation and the Grotthuss hopping phenomenon. However, their computational expense limits the application of the former to small systems much smaller than proteins, and restricts the free energy sampling of the latter when applied to large biomolecular systems. A lack of sufficient free energy sampling generally leads to results with large systematic errors that depend on the initial state of system. In addition, QM/MM methods without adaptive partitioning(24, 25) prohibit exchange of MM and QM atoms across the QM/MM boundary, which is unphysical. Although adaptive QM/MM partitioning methods enable the exchange of MM and QM atoms across the boundary, they can suffer from both inaccuracy of forces on atoms near the boundary and a lack of sufficient energy conservation.(26) The QM/MM boundary may be especially problematic when the hydration along the PT pathway changes in response to the migration of the excess proton charge defect since this necessitates the exchange of water molecules between protein cavities and bulk solution, which often occurs beyond the nanosecond timescale.

An alternative approach that explicitly accounts for the reactive nature of the hydrated excess proton is multiscale reactive molecular dynamics (MS-RMD) method.(2, 27, 28) This approach describes the migration of an excess proton including explicit Grotthuss shuttling and charge defect delocalization by evolving the system on a reactive potential energy surface defined as a dynamical linear combination of diabatic basis states, as in the earlier multistate empirical valence bond (MS-

EVB) method.(2, 29, 30) As will be shown in detail later, the newer MS-RMD approach differs from the older MS-EVB approach primarily because the underlying reactive force field of MS-RMD is largely derived from AIMD or QM/MM data via force-matching and other means, rendering it far less empirical than the original MS-EVB approach. MS-RMD is also "multiscale" in the sense that quantum information on the electronic states is variationally bridged upward in scale to describe the forces felt by the system nuclei in the RMD model. As shown later as well, MS-RMD is several orders of magnitude more computationally efficient than MD from QM/MM, while still describing the charge delocalization and reactive nature of the PT. The MS-RMD method is also not complicated by the QM/MM boundary issues because the molecules that participate in the proton charge defect charge delocalization are dynamically redefined at each time step in such a way that the forces on the atoms are more continuous. Thus, as water molecules or residues move away from the excess proton, the forces acting on the atoms gradually transition to those described by the classical force field. These advantages make the MS-RMD method (and MS-EVB before it) a powerful tool for investigating PT in a variety of biological systems.

On the other hand, the primary challenge of the MS-RMD method is that it needs to be properly parameterized in order to faithfully simulate PT. The parameterization of MS-RMD models is the focus of this work. In ref (31), an approach for parameterizing MS-RMD models for ionizable amino acids in bulk water was developed. In this fitting approach, forces from QM/MM calculations are bridged via an iterative variational framework into the reactive MS-RMD model. In particular, the MS-RMD model parameters are obtained by sampling configurations with a guess MS-RMD model and then fitting the model parameters to best reproduce forces from QM/MM calculations that are as accurate as possible. Our fitting procedure was partially motivated by the work of Wang and co-workers, in which force fields for liquid water were

developed with an adaptive force matching method.(32, 33) It also bears similarity to the work of Zhou and Pu on reactive path force matching.(34) Due to the use of configurations from condensed phase MS-RMD trajectories, the resulting model takes into account the condensed phase environment, which is not captured in gas phase fitting. For a given MS-RMD potential energy functional form and with restrictions on the ranges of parameter values, this fitting approach was shown to work well for investigating glutamic acid (Glu) and aspartic acid (Asp) deprotonation in bulk water.(31) However, one cannot simply use these models to study deprotonation of amino acids such as Asp and Glu within proteins. The electrostatic and hydration environment affecting ionizable amino acid protonation/deprotonation inside biomolecules can be very different from that in bulk solution. For example, deprotonated Asp and Glu residues often form salt bridges with positively charged residues in proteins, and the surrounding hydration structure is rarely bulk-like. As a consequence, the pK_a values of amino acids inside proteins are often largely shifted from those measured in bulk solution. These differences require reparameterization of the MS-RMD models to fit the protein environment.

In this work, we demonstrated the use of FitRMD parameterization to generate MS-RMD models for ionizable amino acid in proteins. The paper is organized as follows. We first outline a framework for fitting MS-RMD models for amino acid residues in biomolecular systems. We then give the computational details for following this framework in two example systems: the ClC-ec1 antiporter and C*c*O. The ability of the generated MS-RMD models to faithfully reproduce the free energy profiles (potentials of mean force, PMFs) of the reference QM/MM Hamiltonian is demonstrated for PT inside the ClC-ec1 channel in two different states. The importance of sampling and related QM/MM limitations are also demonstrated in ClC-ec1. Finally, the advantages of developing efficient MS-RMD models are demonstrated with the calculation of 2-

dimensional free energy profiles (2D-PMFs) of PT coupled to hydration changes in the central hydrophobic region of C*c*O.

1.2 Methods

Multiscale reactive molecular dynamics (MS-RMD)

The MS-RMD (and MS-EVB) method describes electronic delocalization of the excess proton as a linear combination of distinct states with different chemical bonding topologies. In the central region of ClC-ec1 antiporter (See Figure 1 for the images of the whole protein structure and its central region), Figure 2 shows examples of a set of states in certain configurations, where the excess proton is delocalized into a glutamate residue and surrounding water molecules. The Hamiltonian is defined as:

$$\mathbf{H} = \sum_{ij} |i(\mathbf{r})\rangle h_{ij}(\mathbf{r}) \langle j(\mathbf{r})| \tag{1}$$

where **r** are the system nuclear coordinates, h_{ii} is the potential energy for state i described by a classical force field, and h_{ij} is the coupling between states $|i\rangle$ and $|j\rangle$. The diagonalization of the Hamiltonian matrix on the fly gives the energy and eigenvector of the ground state for every configuration of the nuclei **r**:

$$\mathbf{Hc} = \mathrm{E}_0 \mathbf{c} \tag{2}$$

The forces are evaluated by the Hellmann-Feynman theorem and are used to propagate the system in an MD simulation. It is important to note that the method explicitly treats the excess proton charge defect delocalization, Grotthuss shuttling, and the polarization effect associated with the

excess proton complex. The resulting MD trajectory is continuous and deterministic to within numerical error.

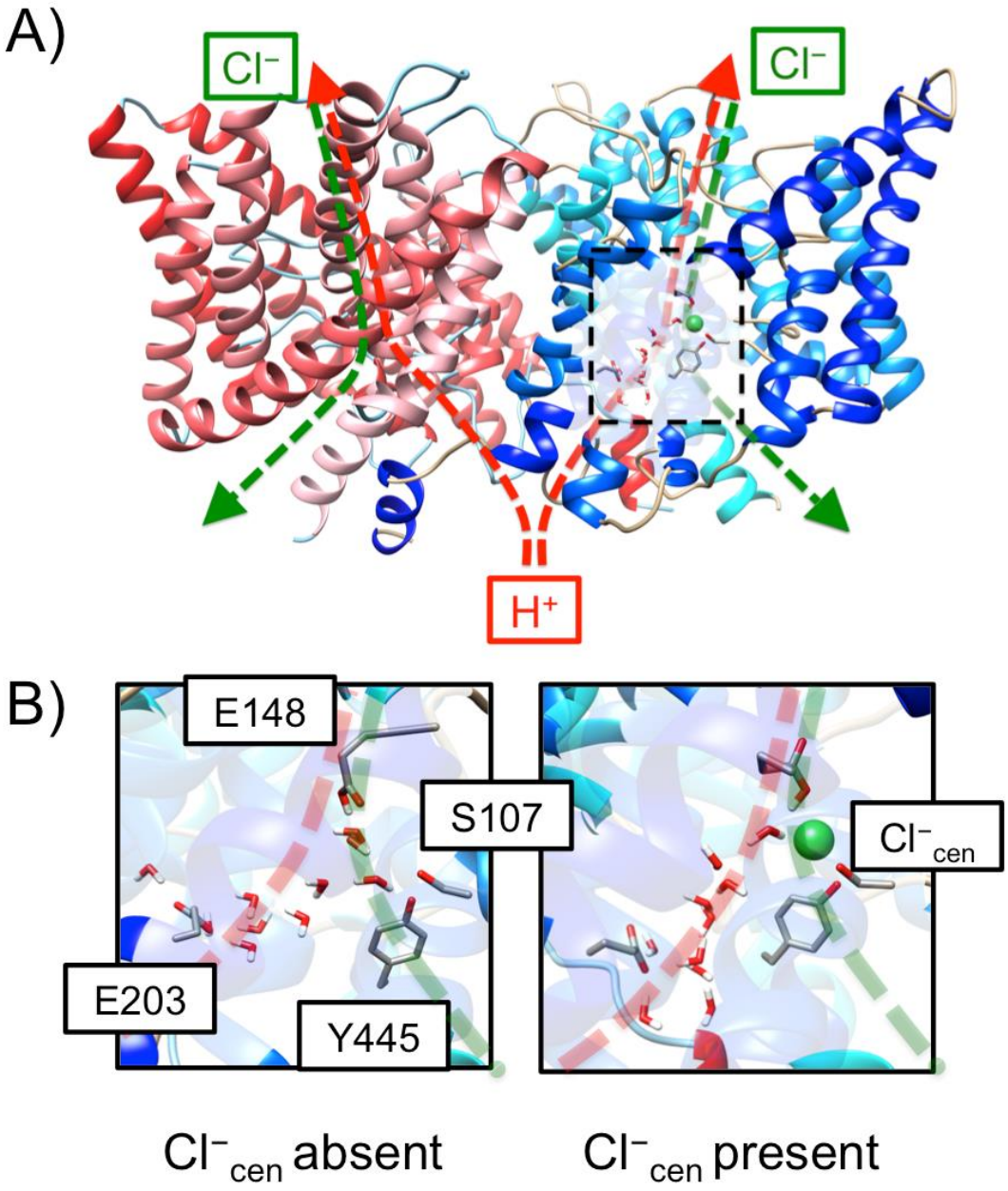


Figure 1.1. (A) Overview of the structure of the ClC-ec1 antiporter and transport pathways for Cl^- (green dashed) and H^+ (red dashed). The central region is highlighted in the black box. (B) Representative configurations of the central region with (left) Cl^-_{cen} absent and (right) present for the "P" state in the PMFs in Fig 3. The protein and water molecules (in sticks) and Cl^-_{cen} (in a sphere) are included in the QM region in the QM/MM calculation of the reference force.

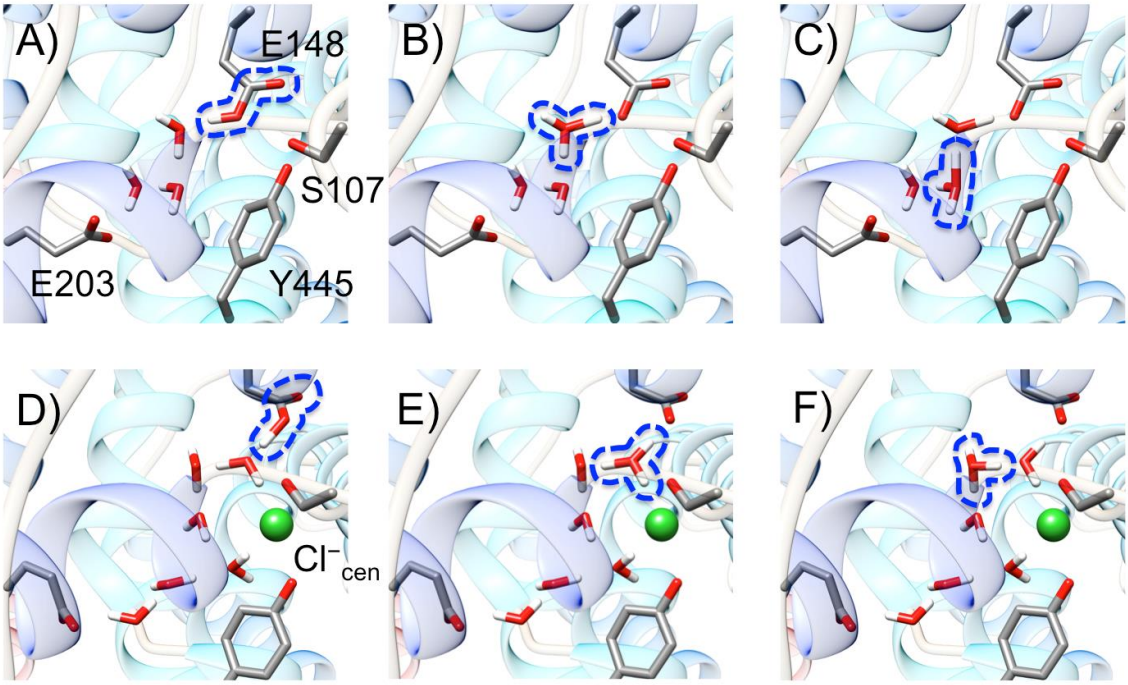


Figure 1.2. Representative configurations of a set of MS-RMD states with different bonding topologies for the same atomic coordinates in the ClC-ec1 antiporter. The configurations represent the three mostly populated diabatic states (with the highest values of $c_i^2(\mathbf{r})$), captured at the windows for $\zeta_R = 0.8$, when Cl^-_{cen} is either absent (A-C) or present (D-F). The excess proton is attached to E148 (A and D), the first water from E148 (B and E), or the second water (C and F), respectively. The protonated species in each diabatic state is shown inside the blue dashed line.

The diagonal elements h_{ii} of the MS-RMD Hamiltonian are given by the potential energy function of each basis state i. Note that there is a single excess proton in the system, and either glutamate or water is protonated at each state. The h_{ii} corresponding to the state with protonated glutamate (GLUH) is described as,

$$h_{ii}^{GLUH} = V_{surrounding}^{intra} + V_{GLUH}^{intra} + V_{surrounding,GLUH}^{inter} + V_{ii} \quad (3)$$

where the first three terms are the inter- and the intra-molecular potentials of protonated glutamate and all other surrounding molecules, such as waters, other protein residues, lipids, and ions in the system. They are computed with CHARMM22 force field,(35) with the exception of the O-H bond in the carboxyl (-COOH) group of glutamate. To properly represent its dissociation as a proton transfers from the carboxyl oxygen to the water oxygen, the harmonic bond stretch potential is replaced with a standard Morse potential, $U^{Morse}(r)$:

$$U^{Morse}(r) = D\left[1 - \exp\left(-\alpha(r - r_0)\right)\right]^2 \quad (4)$$

where r is the O-H bond length, and D, α, and r_0 are parameters, which are taken from our previous work.(31) Since the classical force fields between two protonated forms of glutamate and water do not share a common energy origin, V_{ii} is introduced to compensate the constant energy shift between the two states.

In order to correct overestimated electrostatic interaction at a short distance(36) between opposite charges on a hydronium and deprotonated glutamate two repulsive terms, $V_{OO_k}^{rep}$ and $V_{H_jO_k}^{rep}$, are introduced in h_{ii} corresponding to the state with deprotonated glutamate:

$$V_{OO_\epsilon}^{rep}\left(R_{OO_\epsilon}, R_{H_jO_\epsilon}; j = 1,2,3\right) = B\,\exp\left[-b\left(R_{OO_\epsilon} - d_{OO}^0\right)\right] \cdot \sum_j^3 \exp\left[-b'\left(R_{H_jO_\epsilon}\right)^2\right] \quad (5)$$

$$V_{H_jO_\epsilon}^{rep}\left(R_{HO_\epsilon}\right) = C\,\exp\left[-c\left(R_{H_jO_\epsilon} - d_{OH}^0\right)\right] \quad (6)$$

where R_{OO_ϵ} is the distance between the hydronium oxygen, O, and the carboxyl oxygen of glutamate, O_ϵ (OE1 and OE2 in the PDB), and $R_{H_jO_\epsilon}$ is the distance between each of three hydronium hydrogen atoms, H_j, and the carboxyl oxygen of glutamate. The functional forms for the repulsive terms are the same as those used in the MS-EVB3.1 model. (31) B, b, b', C, and c

are fitted parameters, and d_{OO}^0 and d_{OH}^0 are fixed the same value used in MS-EVB3.1, which are 2.4 and 1.0 Å, respectively.

The off-diagonal element h_{ij} for the coupling between protonated glutamate and water is given by,

$$h_{ij}^{GLUH} = c_1 \exp[-c_2(r_{OH} - c_3)^2] \quad (7)$$

where r_{OH} is the distance between the donor oxygen of the carboxyl group of glutamate and the acceptor hydrogen of the adjacent hydronium molecule. c_1, c_2, and c_3 are fitted parameters. h_{ij} for the coupling between hydronium and water is the same as the one used in the MS-EVB3.1 model.

The MS-RMD (and MS-EVB) formalism also provides a convenient and physically intuitive description of the excess proton center of the excess charge (CEC), defined as:(2)

$$\mathbf{r}_{\mathrm{CEC}} = \sum_{i}^{N_{\mathrm{EVB}}} c_i^2(\mathbf{r})\mathbf{r}_i^{\mathrm{COC}} \quad (8)$$

where the N_{EVB} is the total number of EVB states, $c_i^2(\mathbf{r})$ is the population of state i contributing to the MS-RMD ground state, and $\mathbf{r}_i^{\mathrm{COC}}$ is the geometric center-of-charge for state i. This CEC definition allows the use of a continuous reaction coordinate (further discussed below) for the PMF calculation of PT in biological systems. [2] The protonated moiety in the state with the largest coefficient, c_1 (Equation 8), possesses the majority of the excess positive charge. When in bulk water this state is the most hydronium-like species(29, 30, 37, 38) (or the so-called "pivot" hydronium, a technical term used below).

FitRMD parameterization scheme

One possible fitting procedure of an MS-RMD potential energy function to QM/MM data was first described in ref (31). In such an approach, configurations along the PT reaction coordinate are sampled by MS-RMD umbrella sampling simulations with initial guess MS-RMD model parameter set. In the present paper, the initial guess amino acid models were taken from the previous work done in bulk water,(31) except that the constant energy shift between protonated/deprotonated states in the model (V_{ii}) was determined by the difference in the Coulomb energy of the RMD (EVB) complex between the most favorable hydronium state and the protonated state. The range of the PT reaction coordinate was set to sample configurations for both protonated/deprotonated states (further defined below). Next, for each configuration a QM/MM calculation was performed to collect the reference forces on each atom in the MS-RMD reactive complex. Then the MS-RMD model parameter set was optimized by minimizing the variational residual:

$$\chi^2 = \frac{1}{3N_C N_A} \sum_{j=1}^{N_C} \sum_{i=1}^{N_A} w(\mathbf{r}_{ij}) \left| \mathbf{F}_{ij} - \mathbf{F}_{ij}^{\mathrm{QM/MM}} \right|^2 \tag{9}$$

where N_C and N_A are the number of configurations and the number atoms in each configuration, respectively. The weight of each atom $w(\mathbf{r}_{ij})$ is set to unity here, but it should be noted other choices are possible. The atomic force $\mathbf{F}_{ij}$ is the one obtained from the current MS-RMD model parameter set and $\mathbf{F}_{ij}^{\mathrm{QM/MM}}$ is the reference force from the QM/MM calculations. The whole set of the model parameters were then divided into two groups: 1) the diagonal terms V_{ii}, and 2) the off-diagonal and repulsive terms. The model optimization was done individually for each group. First,

the off-diagonal and repulsive terms were fit with the value of V_{ii} fixed, and then V_{ii} was refit with new values of the off-diagonal and repulsive terms held fixed.

Developing MS-RMD models using FitRMD for ClC-ec1

MS-RMD simulations of the PT process in the bacterial ClC homolog, ClC-ec1, are extensively described in ref (39). The FitRMD method was used to parameterize the MS-RMD models from QM/MM data for the two glutamate residues in the central region, namely the E148 (Glu_{ex}) and the E203 (Glu_{in}), which have been identified as intermediate sites for proton binding along the transport pathway (see Figure 1). Depending on the presence of Cl^- at the central site (Cl^-_{cen}), two different systems were setup. For each system setup, two sets of umbrella sampling simulations were performed (deprotonating either E148 or E203) to sample configurations for generation of the QM/MM force data. In order to obtain the initial configurations for the umbrella sampling, 50 ns of unconstrained classical MD was run both with Cl^-_{cen} present and absent. In this time water penetrated the central region from the intracellular side of the membrane, forming a continuous hydrogen bonded network between the carboxyl oxygens of E148 and E203. A harmonic potential with a force constant of 20 $\text{kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-2}$ was applied to a collective variable (CV), defined as $|\mathbf{r}_{CEC} - \mathbf{r}_X|$, where "*X*" is the carboxyl center of mass for either E148 or E203. The centers of the windows ranged from 2.0 to 4.0 Å and were separated by 0.25 Å. At CV = 2.0 Å all configurations have the amino acid fully protonated, while at CV=4.0 Å all were deprotonated with protonated state contributing less than 0.01% ($c_i^2(\mathbf{r}) \leq 10^{-4}$).

The configurations were selected with a 2 ps interval from each window for ~500 configurations for each protonation site. For each configuration, a single point QM/MM

calculation was performed to evaluate the forces acting on the QM atoms. As shown in Figure 1B, the QM region included Cl^-_{cen}, if present, the water molecules in the central region, and the side chain of the pore-lining residues, including E113, E148, E202, E203, Y445, and S107. The waters in the central region within the third solvation shell from E148 and E203 were also included in the QM region. The QM box size was set to be 20 – 30 Å in each dimension to ensure it was 6-8 Å larger than the size of the QM atoms in each dimension. The Gaussian Expansion of the Electrostatic Potential (GEEP) scheme was used to treat the QM/MM electrostatic coupling with periodic boundary conditions (PBCs),(40, 41) and the spurious QM/QM periodic image interactions were decoupled as described in ref (42). The $C_\alpha - C_\beta$ chemical bonds that crossed the QM/MM boundary were capped with hydrogen atoms, the forces on which were calculated following the IMOMM scheme(43) with a scaling factor of 1.50. The QM atoms were treated with density function theory (DFT) using the BLYP functional(44, 45) with empirical dispersion corrections,(46) under the Gaussian and plane wave (GPW) scheme.(47) Goedecker-Teter-Hutter (GTH) pseudopotentials(48) were used and the Kohn-Sham orbitals were expanded in the Gaussian TZV2P basis set.

After the QM/MM forces were obtained, the MS-RMD model parameters were optimized by minimizing the residual given in eq 9 using a genetic algorithm.(49) All atoms in the reactive complex, as defined by the MS-RMD state-selection algorithm,(30) were included in the fitting (i.e., the glutamic acid side chains and solvent atoms in the central region). The V_{ii} and the off-diagonal terms were iteratively optimized for 3 – 6 rounds, depending on the system, until they changed less than 1 %. The models for E203 and E148 were developed independently because they are separated by 10-14 Å, depending on the presence of Cl^-_{cen}, and therefore never participate in coupled delocalization of the excess charge. The MS-RMD simulations were performed with

the RAPTOR software(28) interfaced with the LAMMPS MD package (http://lammps.sandia.gov),(50) and the umbrella potentials were controlled by the PLUMED package.(51) All single point QM/MM calculations are performed by CP2K package.(52) The FitRMD calculation was performed by an in-house code.(31) Parameters for E148 and E203 for ClC-ec1 with Cl^-_{cen} either absent or present are given in Table 1.1.

	ClC-ec1			
	Cl^-_{cen} absent		Cl^-_{cen} present	
	E148	E203	E148	E203
B	0.063153	0.118175	0.012536	0.003282
b	1.571751	0.558680	0.232384	0.424087
b'	1.320947	1.311301	1.469007	0.480047
d^0_{OO}	2.4	2.4	2.4	2.4
C	0.363648	0.076533	0.014044	0.026472
c	1.117167	1.063753	1.152912	1.309021
d^0_{OH}	1.0	1.0	1.0	1.0
r_s^l	3.5	3.5	3.5	3.5
r_s^h	4.0	4.0	4.0	4.0
V_{ii}	-147.095673	-144.522565	-151.11473	-147.01392
c_1	-36.090543	-40.406344	-30.414842	-35.686750
c_2	1.879933	3.826462	3.331769	1.291340
c_3	1.193253	1.439519	1.422240	1.598104
D	143.003	143.003	143.003	143.003
α	1.8	1.8	1.8	1.8
r_0	0.975	0.975	0.975	0.975

Table 1.1 The MS-RMD model parameters of E148 and E203 in ClC-ec1.

1D-PMF calculations of PT from E203 to E148 in ClC-ec1

In order to directly compare MS-RMD and QM/MM free energy profiles, umbrella sampling was performed with a CV defined as ζ_R which is a function of the excess proton CEC ($\mathbf{r}_{\mathrm{CEC}}$) and the center of mass of the carboxyl groups of the E203 ($\mathbf{r}_{203}$) and the E148 ($\mathbf{r}_{148}$) residues:

$$\zeta_R = \frac{|\mathbf{r}_{\mathrm{CEC}} - \mathbf{r}_{203}|}{|\mathbf{r}_{\mathrm{CEC}} - \mathbf{r}_{203}| + |\mathbf{r}_{148} - \mathbf{r}_{\mathrm{CEC}}|} \tag{10}$$

The excess proton CEC coordinate, however, was defined in this care as(53)

$$\mathbf{r}_{\mathrm{CEC}} = \sum_{i=1}^{N_H} \mathbf{r}^{H_i} - \sum_{j=1}^{N_X} w^{X_j} \mathbf{r}^{X_j} - \sum_{i=1}^{N_H} \sum_{j=1}^{N_X} f_{sw}\left(d_{X_j H_i}\right)\left(\mathbf{r}^{H_i} - \mathbf{r}^{\mathrm{X}_j}\right) \tag{11}$$

where the $\mathbf{r}^{X_j}$ is the position of the *j*th heavy atom in the QM region, which is either the water oxygen atoms or the carboxyl oxygen atoms in the E203 or the E148, and the $\mathbf{r}^{H_i}$ is the position of the *i*th hydrogen atom bound to those heavy atoms. (It should be noted that this CEC definition was used for both the QM/MM and MS-RMD simulations, as opposed to eq 8 for the latter, in order to have a common definition of the CV in the two PMFs for comparison.) The weighting factor, w^{X_j}, was set to be two for all the water oxygen atoms and zero for carboxyl oxygen atoms in the E203 and E148, which reflect the hydrogen coordination number in the deprotonated state of the heavy atom. The term $d_{X_j H_i}$ denotes the distance between X_j and H_i atoms, and the $f_{sw}\left(d_{X_j H_i}\right) = 1/(1 + \exp[(d_{X_j H_i} - r_{sw})/d_{sw}])$ is the switching function describing the coordination number of H_i to X_j, with the parameters chosen as $d_{sw} = 0.04$ Å, $r_{sw} = 1.25$ Å.(54)

The CV ζ_R varies from 0 to 1 as the excess proton CEC moves from E203 to E148. The centers of the harmonic umbrella potentials were separated by 0.1 – 0.2 Å between adjacent windows. The

umbrella sampling force constant for ζ_R was chosen to be in the range of 3000 – 7000 $kcal \cdot mol^{-1}$, depending on the sampling overlap between umbrella windows. Given that the denominator of ζ_R, $|\mathbf{r}_{\mathrm{CEC}} - \mathbf{r}_{203}| + |\mathbf{r}_{\mathrm{CEC}} - \mathbf{r}_{148}|$, is ~ 15 Å, the effective force constant acting on $|\mathbf{r}_{\mathrm{CEC}} - \mathbf{r}_{203}|$ was 15 – 30 kcal $\cdot$ mol^{-1} $\cdot$ Å^{-2}. Because the central region is accessible by the water molecules from the intracellular bulk, the CEC can locate outside the central region at the windows near E203 when a restraint is only placed on ζ_R. To avoid sampling irrelevant positions (leading to the intracellular bulk phase), an additional restraint defining the upper limit for $|\mathbf{r}_{\mathrm{CEC}} - \mathbf{r}_{148}|$ was applied such that the CEC was always situated between E148 and E203. The upper limit was chosen to be 2 σ above the average of $|\mathbf{r}_{\mathrm{CEC}} - \mathbf{r}_{148}|$ in each window.

The details of the preparation of the initial configuration and the classical MD equilibration for the ClC-ec1 antiporter are described in ref (39). The initial configurations used for the FitRMD umbrella sampling were used to initiate the MS-RMD umbrella sampling simulations. Near the middle of the reaction path, when E203 and E148 were both deprotonated, a water molecule close to the center of each umbrella window was replaced with a hydronium cation. All windows were first equilibrated for 100 ps, followed by production runs of ~ 2 ns. The integration time step was 1 fs. The CV ζ_R was collected every ten time steps (10 fs) and the PMF was constructed by the weighted histogram analysis method (WHAM).(55) Statistical error bars in the PMFs were estimated using the block averaging method by dividing each trajectory into four consecutive blocks.

1D-PMF calculation with QM/MM

The initial configurations used in the MS-RMD simulations were also used for the QM/MM umbrella sampling simulations. The simulation details for the QM/MM MD were kept consistent with the single point force calculations used in FitRMD. The window spacing and force constants for the umbrella windows were similar to those used in the MS-RMD simulations. All windows were equilibrated with QM/MM MD for another ~ 5 ps, followed by production runs of ~ 20 ps. (It should be noted that in our experience if a FitRMD model differs significantly from the underlying QM/MM forces due to a bad fit, then subsequent QM/MM MD configurations taken starting with the MS-RMD initial conditions will diverge quickly from the MS-RMD ones.) The integration time step was 0.5 fs and the CV ζ_R was collected every time step. The PMF was again constructed by WHAM. All QM/MM MD simulations were performed by CP2K package.(52)

Upon finding that the PMFs calculated with QM/MM and MS-RMD MD disagreed for the Cl^-_{cen} absent system in the CV range $0.18 < \zeta_R < 0.26$, a second set of QM/MM simulations was run in order to investigate the origin of the PMF discrepancy. The second set of QM/MM umbrella sampling simulations were initiated from the last MD snapshots of the MS-RMD umbrella sampling simulations, which had an increased hydration level compared to the original QM/MM simulations. All windows in the CV range $0.18 < \zeta_R < 0.26$ were then sampled for 5 ps. All other simulation details were unchanged from the original QM/MM umbrella sampling.

Comparative structural analysis in ClC-ec1

Structural elements of the ClC-ec1 protein are shown in Figures 1-6. To obtain these structures the production run simulation conformations were averaged in each umbrella window following three steps. First, configurations were aligned based on alpha carbon atoms at least 15 Å away from the central region. Second, the positions of the side chains of the four residues in the central region (S107, E148, E203, and Y445) and Cl^-_{cen}, if present, were averaged and full moieties representative of these averages were depicted. Third, the positions of the water molecules and the excess proton were taken from the last MD frame.

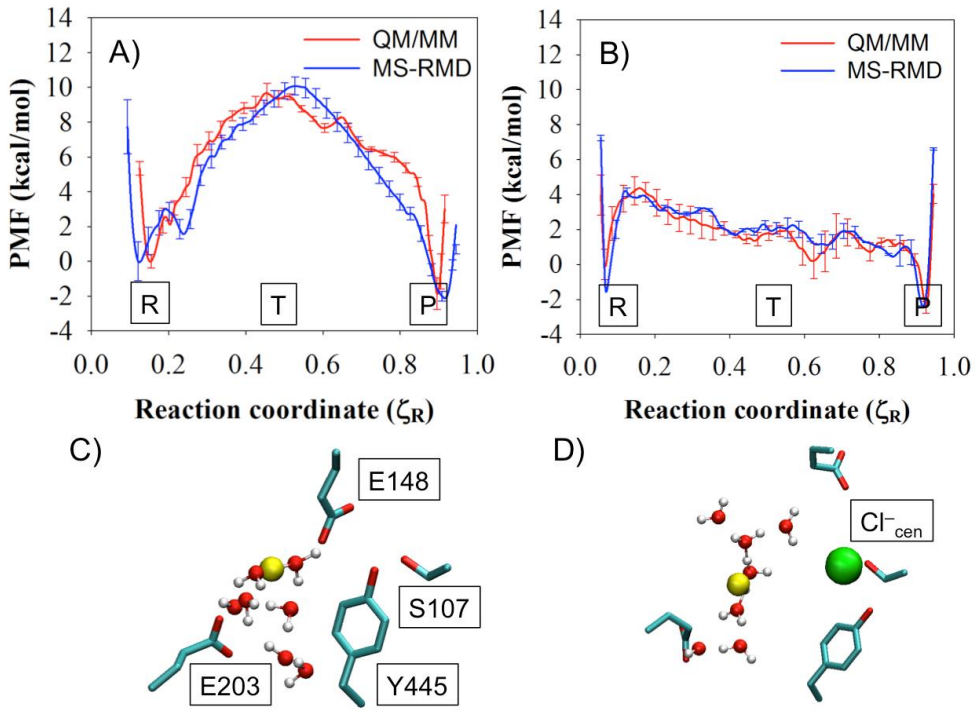


Figure 1.3. PMFs of PT through central region of ClC-ec1 antiporter with (A) Cl^-_{cen} absent and (B) present, as calculated with MS-RMD (blue) and QM/MM (red). Labels "R", "T", and "P" indicate

the reactant state, the center of the CV range, and the product state, respectively. The CV ζ_R varies from 0 to 1 as the excess proton CEC moves from E203 to E148. The average side chain positions of four important residues are shown for state "T" in the MS-RMD systems with (C) Cl^-_{cen} absent or (D) present. The solvent positions and proton center of the excess charge (CEC, yellow sphere) are taken from the last snapshot.

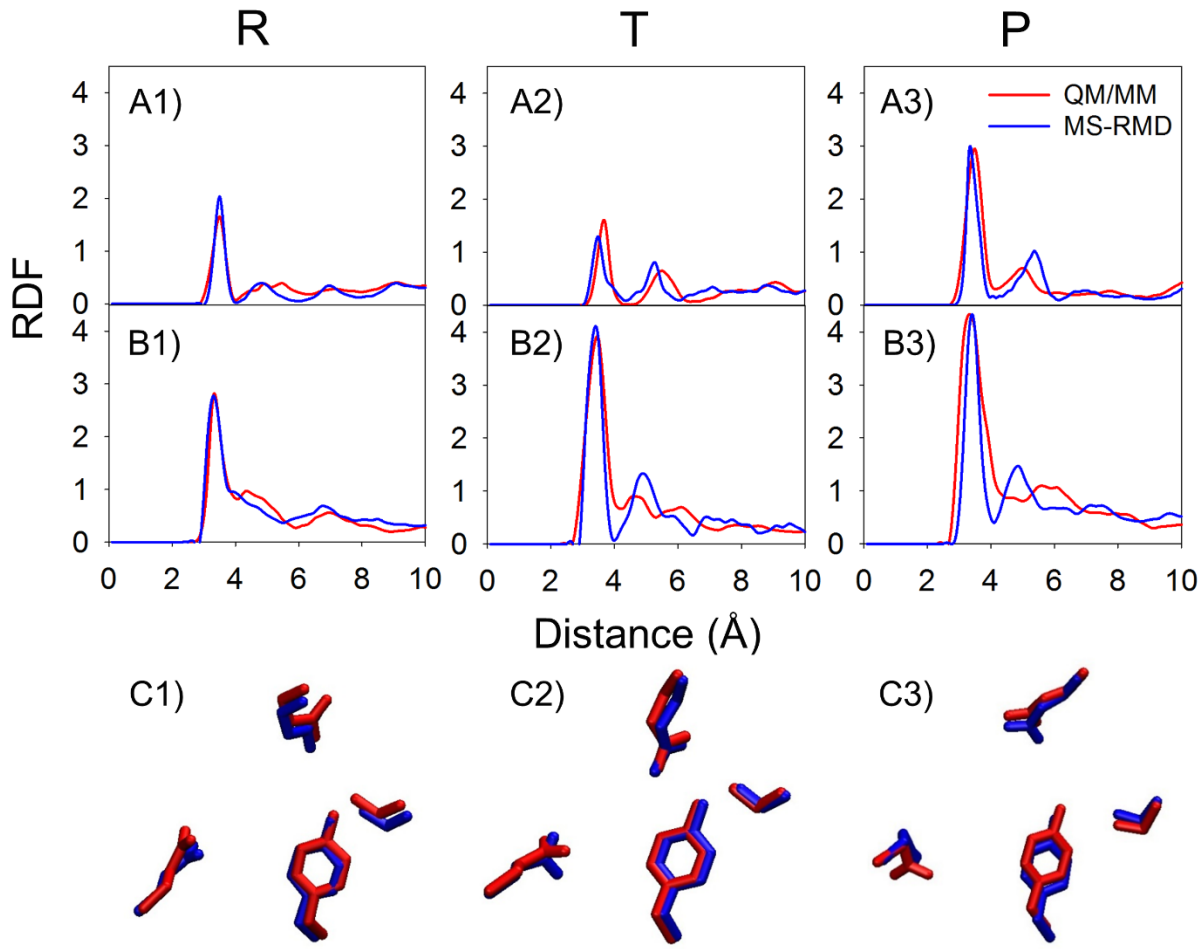


Figure 1.4. In the system with Cl^-_{cen} absent, RDFs from E148 carboxyl carbon (CD in the PDB) to water oxygens are shown in A1 to A3, for the "R", "T", and "P" states (Fig 3), respectively. The RDFs from E203 carboxyl carbon to water oxygens are shown in B1 to B3 in the same order. The averaged positions of the four side chains are shown in C1 to C3 in the same order, with the QM/MM structures in red and the MS-RMD structures in blue.

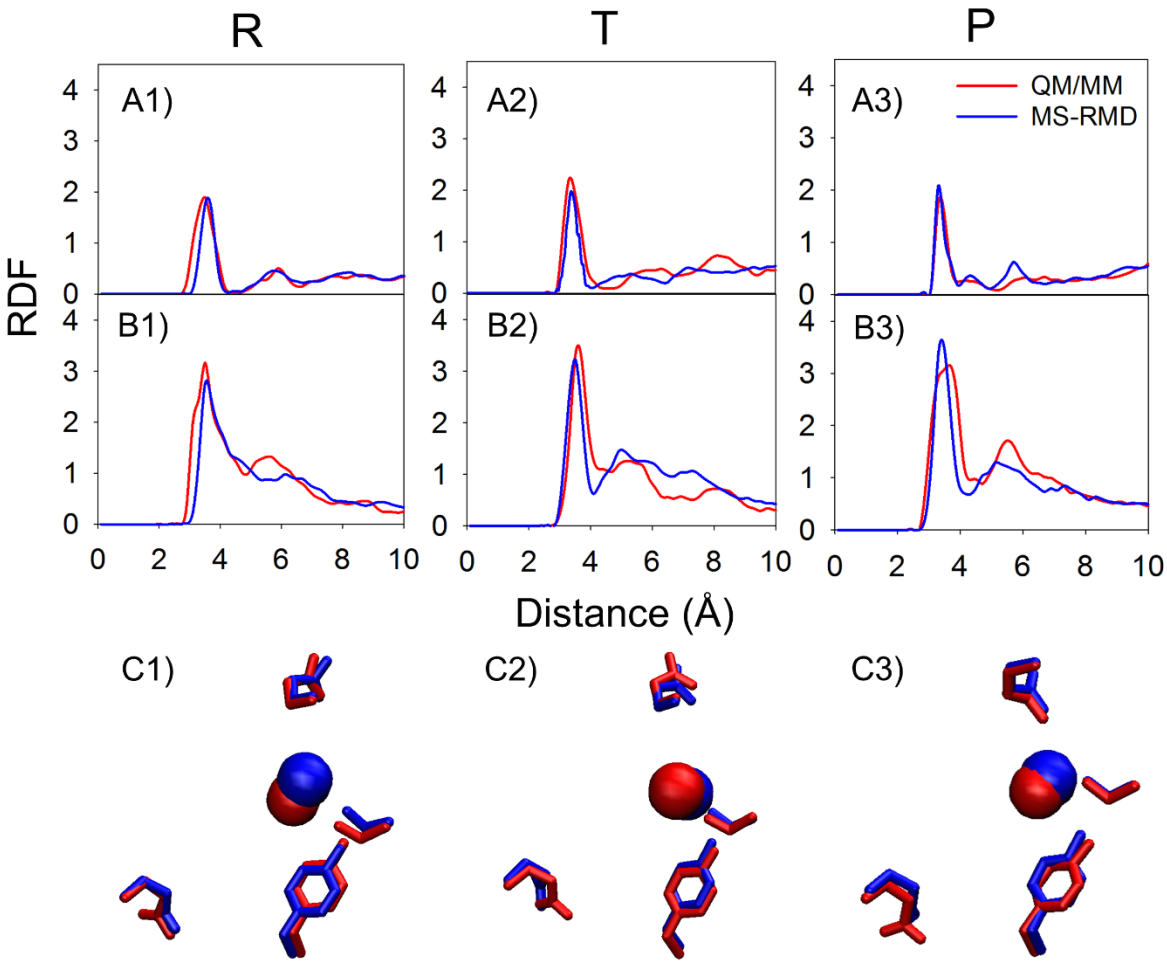
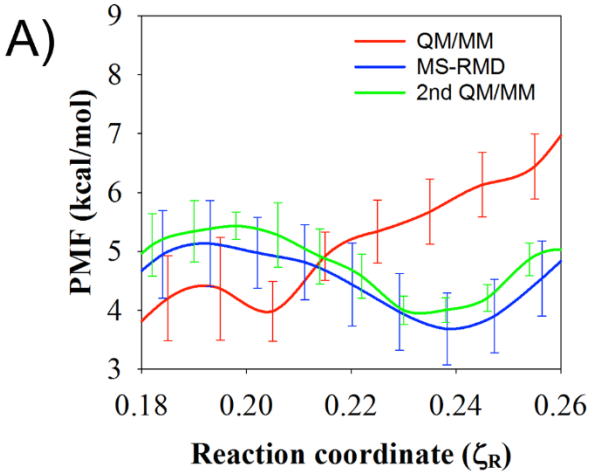
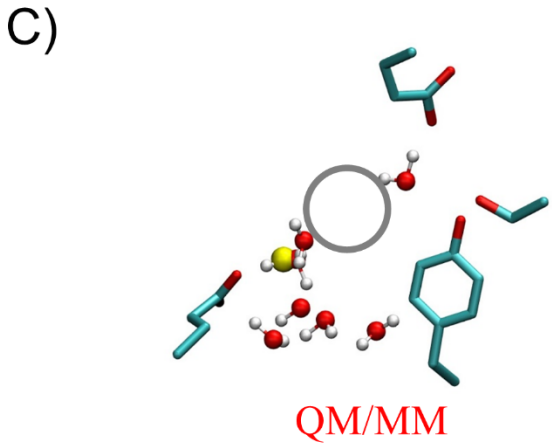


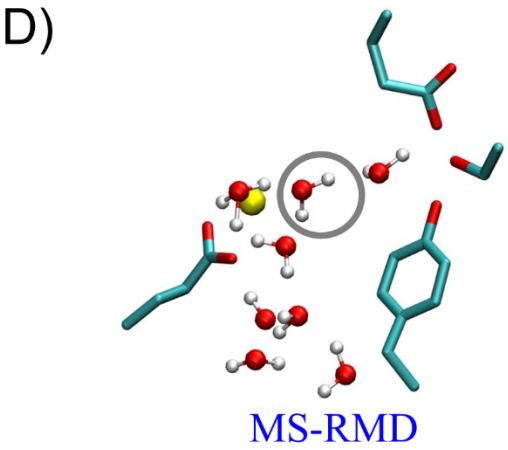
Figure 1.5. In the CLC system as in Fig. 4, but with Cl^{-}_{cen} present. All notation is consistent with Fig 4.

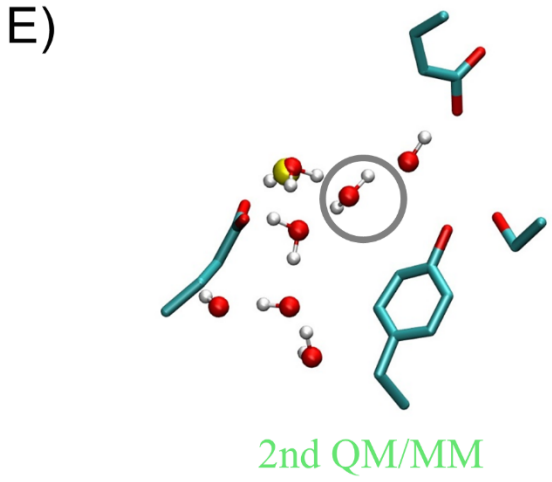


B)

System	$\langle c_1^2(\mathbf{r})\rangle$
QM/MM	0.55 ± 0.06
MS-RMD	0.70 ± 0.04
2nd QM/MM	0.69 ± 0.05







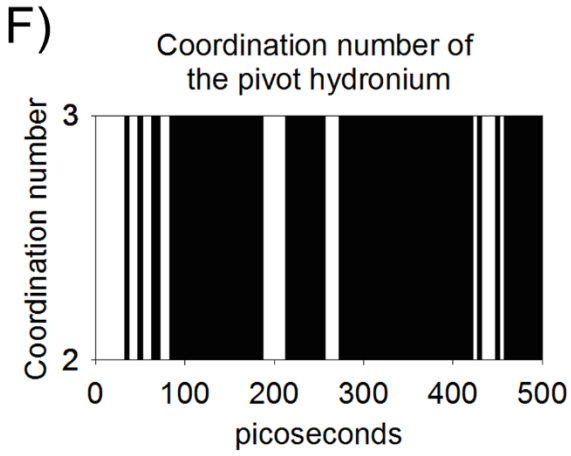


Figure 1.6. Panel (A): The PMF from Fig 3A for $\zeta_R = 0.18 - 0.26$. The second QM/MM PMF (green) was initiated from the final MS-RMD configuration in each window. (B) The average and the standard error of $c_1^2(\mathbf{r})$, reflecting the delocalization of the hydrated excess proton (see text),

from the original QM/MM, the MS-RMD, and the second QM/MM umbrella sampling at window $\zeta_R = 0.24$. Panels (C-E): The averaged structure of the four residues and the water structure in the last snapshot from the (C) the QM/MM, (D) the MS-RMD, and (E) the second QM/MM umbrella sampling at the window $\zeta_R = 0.24$. The region circled by a gray line represents the position of the second water. (See further description in the main text.) Panel (F): The coordination number of the pivot hydronium to oxygen atoms of surrounding water molecules or the carboxyl group of E203 in the MS-RMD simulation, initiated from the configuration at low hydration level, which mimics the configuration of the QM/MM in Fig. 6C.

Hydration dynamics in the central region of ClC-ec1

To calculate the timescale for the transition from a low to high hydration state in the central region of ClC-ec1, the low hydration state first needed to be created in the MS-RMD structure. To do this, a high hydration configuration was taken from the last MS-RMD snapshot at umbrella window $\zeta_R = 0.24$ (Fig 6D). An additional harmonic potential was applied to the water density CV (N_w) in the pre-defined box (see ref (23) and (39) and for definition). The force constant of the harmonic potential was set to be 10 $kcal \cdot mol^{-1}$, and the center of the potential was at $N_w = 0$. The center of the box was chosen to be the midpoint between the center of mass of each carboxyl group of E148 and E203, and the box size was set to be 2 x 2 x 2 $Å^3$, which covered the position of the second water. The second water present in the initial configuration was expelled from the central region by the harmonic potential centered at $N_w = 0$. The MS-RMD simulation was run for 200 ps, until the system was equilibrated in the absence of the second water.

Ten independent trajectories were initiated from the last MD frame of the simulation above, after the velocities of all atoms were randomized and the harmonic potential centered at $N_w = 0$ was released. The coordination number of the pivot hydronium was defined as the number of oxygen atoms (from water or the E203 carboxyl group) within 3 Å from the pivot hydronium

oxygen. The time for the second water entering the central region was estimated by averaging the ten simulations time to when there was a productive (not transient) transition of the coordination number from 2 to 3.

Developing MS-RMD models using FitRMD for CcO

The FitRMD method was also used to parameterize the MS-RMD models from QM/MM data for three protonatable sites in the hydrophobic cavity of C*c*O, including the E286, PRDa_3, PRAa_3 (Figure 7). Umbrella sampling simulations were first carried out along the PT pathway identified by the MS-RMD metadynamics (MTD) simulation.(56, 57) Then for each protonatable site ~ 100 configurations were selected from the trajectories from windows within 4 Å of the transition state of proton dissociation (defined as the windows with ~ 50% of the configurations having the largest amplitude on the amino acid and the other 50% on the first water molecule) for each protonatable site. Single point QM/MM calculations were then performed for each configuration using the B3LYP level density functional theory.(58) The MM models were the CHARMM22 and CHARMM36(59) force fields for the protein and lipids, respectively. The QM region included the side chain of each protonatable amino acid, the hydrated excess proton, and water molecules within 3 solvation shells of the carboxyl group (Figure 7B). In all calculations, the QM box size was chosen to be 6-8 Å larger than the actual size of the QM atoms in each dimension. The GEEP scheme was used to treat the QM/MM electrostatic coupling with periodic boundary conditions (PBCs), and the spurious QM/QM periodic image interactions were decoupled as described in ref (42). The $C_\alpha - C_\beta$ chemical bonds that cross the QM/MM boundary were capped with hydrogen atoms, the forces on which were calculated following the IMOMM scheme with a scaling factor

of 1.50. The forces generated by the QM/MM calculations were used to parameterize the MS-RMD parameters for the protonatable sites using FitRMD approach. The MS-RMD simulations were performed with the RAPTOR software interfaced with the LAMMPS MD package, as described earlier. The QM/MM calculation was performed with the CP2K package, and FitRMD was carried out with in-house software, again as described earlier. Parameters for E286, PRDa_3 and PRAa_3 are given in Table 1.2.

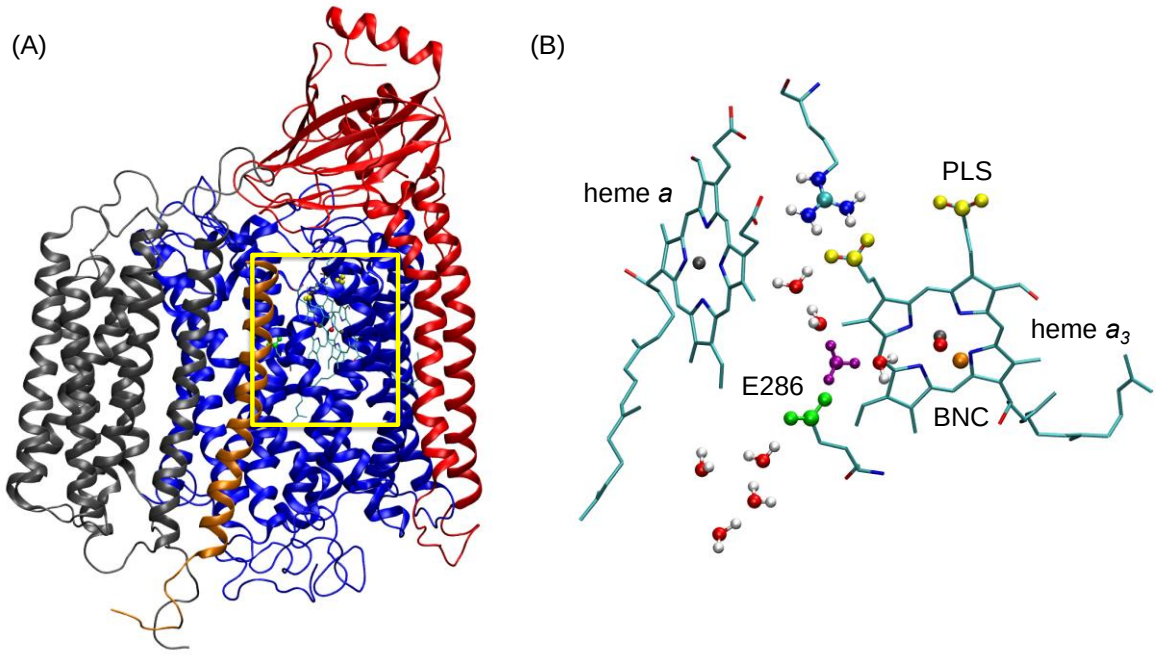


Figure 1.7. (A) Cytochrome c oxidase from *Rhodobacter sphaeroides*. The subunit I, II, III and IV are colored in blue, red, gray and orange, respectively. The yellow rectangular box highlights the E286 residue and the heme-copper groups that are essential for proton pumping and reaction. (B) Enlargement of the region where the E286, heme *a*, heme a_3 and BNC are located. The E286 is shown in green, the propionate groups of the PLS in yellow, the hydrated excess proton in purple, the iron atom of the heme groups in gray and the copper atom in the BNC in orange. The heme groups are shown as sticks. In the QM/MM calculation, the QM atoms include the E286 side chain,

the hydronium and the 3 solvation shells of water molecules around the E286 residue (shown in VDW representation).

	E286	PRD$\underline{a}_3$	PRAa_3
B	0.037588	0.995945	0.994509
b	0.208388	1.999090	1.993038
b'	0.533589	0.000009	0.006308
d_{OO}^0	2.4	2.4	2.4
C	4.925621	0.988369	0.322227
c	1.975947	1.990330	1.981954
d_{OH}^0	1.0	1.0	1.0
r_s^l	3.5	3.5	3.5
r_s^h	4.0	4.0	4.0
V_{ii}	-135.809617	-149.99445	-146.174
c_1	-21.659933	-31.931186	-38.677822
c_2	2.785857	2.689357	1.903222
c_3	1.299987	1.147862	1.245548
D	143.003	143.003	143.003
α	1.8	1.8	1.8
r_0	0.975	0.975	0.975

Table 1.2 The MS-RMD model parameters of E286, PRD$\underline{a}_3$ and PRAa_3 in C*c*O.

The QM/MM calculations for C*c*O were performed at the B3LYP DFT level and as such were much more expensive than the BLYP level DFT QM/MM calculations carried out for ClC-ec1. As a result, fewer configurations were used for C*c*O than for ClC-ec1 in the FitRMD, and for the same reason no explicit PMF was calculated with the B3LYP level QM/MM for C*c*O (see below). In addition, the C*c*O system is too complex to converge any reasonable accurate QM/MM PMF and this fact serves to highlight the strength of the MS-RMD approach.

2D-PMF for PT in the C*c*O hydrophobic cavity with MS-RMD

Full details of the C*c*O simulations and PMF calculations are presented in ref (60). Some of the relevant details are given here, including that the MS-RMD umbrella sampling simulations in the hydrophobic cavity (HC) were carried out by restraining the excess proton CEC position (eq 8) along the PT pathway defined from the MTD simulations and the water density in a predefined box that encompasses the HC (see ref (23) for definition). The force constant for the harmonic umbrella sampling restraint potential was 10 kcal/mol/Å^2 on the proton migration CV and 20 kcal/mol on the water density. The window spacing was ~ 0.5 Å for the CEC and ~ 0.5 for the water density. For each umbrella window the MS-RMD simulation length was ~500 ps. The integration time step was 1 fs. The CV ζ_R was collected every ten time steps (10 fs). The 2D PMF was constructed by the WHAM.

Comparing the computational efficiency of different methods

The computational speed for the MS-RMD, self-consistent density functional tight binding (SCC-DFTB)-based QM/MM(61, 62) and BLYP-based QM/MM methods were compared for the MD simulation of the C*c*O system. (The QM/MM with B3LYP is much too slow for viable QM/MM MD in these systems.) For MS-RMD, the setup was the same as described in the above section. For the BLYP QM/MM method, the QM/MM setup was the same as described in the above section for FitRMD, except that the QM atoms were treated by the BLYP functional with empirical dispersion corrections. More details on the BLYP-based QM/MM MD simulation setup are presented in ref (60). For the SCC-DFTB-based QM/MM setup, the QM atoms were the same as in the BLYP-based QM/MM setup. The point charge based Ewald summation was used to treat

QM/MM electrostatic coupling under PBCs.(63) More details on the SCC-DFTB-based QM/MM setup are described in ref (64). The MS-RMD simulation and BLYP-based QM/MM simulations were performed as described previously. The SCC-DFTB-based QM/MM simulation was performed with CHARMM package.(65)

1.3 Results and Discussion

Comparing QM/MM and MS-RMD free energy profiles in the ClC-ec1 antiporter

An important measure of success of the FitRMD method is its ability to reproduce results of the reference Hamiltonian for properties other than those that were fit (atomistic forces), and there is arguably no property more important than the PT free energy profile (PMF). Unlike our previous work demonstrating the FitRMD method for amino acid deprotonation in bulk water,(31) calculating a PMF with QM/MM is possible in a protein environment because the water molecules are more confined. (In the bulk environment, the QM/MM boundary issues introduce such large errors that a direct comparison between MS-RMD and QM/MM PMFs is highly problematic at best.) However, for many protein cavities the waters involved in PT between two residues are largely surrounded by protein and they all fit into the QM region, making exchange across the QM/MM boundary less of a complication. Although many protein systems will still require extensive sampling (beyond the limits of QM/MM) as well as exhibit water exchange across the QM/MM boundary on longer time scales, the level of confinement in the ClC-ec1 antiporter system enables a direct comparison of the QM/MM and MS-RMD PMFs.

PT through the central region of ClC antiporters is one of the essential intermediate steps in Cl^-/H^+ exchange. The migration of an excess proton from the internal E203 to the external E148 through water molecules in the central region is coupled with migration of approximately 2 Cl^- ions in the opposite direction (Fig 1). Therefore, the PMFs for PT in the central region of the ClC-ec1 antiporter were calculated both in the presence and absence of Cl^- at the central binding site, Cl^-_{cen}.

The MS-RMD and QM/MM PMFs (Figures 3A and 3B) show excellent agreement for most parts of the reaction coordinate and for both states of the system (with and without Cl^-_{cen}), although there are some differences (discussed below). The free energy barrier of the PT process is significantly decreased when Cl^-_{cen} is present, mainly due to the electrostatic interaction between the excess proton and the Cl^-_{cen} ion. These results demonstrate that the FitRMD approach is capable of generating MS-RMD models that reproduce the free energy profile of the reference QM/MM Hamiltonian, using only forces on the atomic nuclei from a relatively small set of configurations as input for the fitting. Moreover, the FitRMD approach is robust enough to quantitatively capture the effect of the Cl^-_{cen} ion on the free energy surface of the reference QM/MM Hamiltonian. This suggests that the FitRMD approach is also capable of describing the shift in proton affinity in different protein environments, which has significant value for simulating PT in different protein systems with the MS-RMD method.

The hydration structure surrounding the reactive protein residues was also compared. The radial distribution functions (RDFs) from the carboxyl carbon of either E148 or E203 to the water and hydronium oxygens were calculated with the excess proton CEC restrained in the reactant, transition, and product states of the PMF (the positions are labeled as “R”, “T”, and “P” in Figure

3A and 3B respectively). The RDFs (Figs 4 and 5 with and without Cl^-_{cen}, respectively) demonstrate that the solvation structure around E203 and E148 is quite similar between the QM/MM and MS-RMD methods. This result provides additional evidence that the model generated by the FitRMD approach faithfully reproduces the underlying free energy landscape of the reference QM/MM Hamiltonian, even though only QM/MM forces on the atom nuclei were used as input for the fitting. The discrepancies in the RDFs shown in figs 4 and 5 are likely due to the different water behavior of QM versus MM water as well as short QM/MM sampling where the water dynamics are slow in the confined space of the central region, which will be discussed below. We note, however, that this level of agreement may not be expected for more bulk-like water environments since the QM and MM water will have such different structural properties.

Limitations of the QM/MM free energy profiles

Although the PMFs in Fig 3A and 3B show that the MS-RMD and QM/MM PMFs generally agree well, some discrepancies appear. Focusing on the most significant, Figure 6A highlights the region $0.18 < \zeta_R < 0.26$ from Figure 3A, where the MS-RMD PMF (blue) dips to a modest metastable minimum but the QM/MM PMF (red) shows an uphill rise with only a slight dip at $\zeta_R \cong 0.21$. This discrepancy between the two PMFs is caused by the change of the local environment for the excess proton. The structures of the central region captured from the $\zeta_R = 0.24$ umbrella windows, where the two PMFs disagree most, are shown in Figure 6C (QM/MM) and 6D (MS-RMD). The MS-RMD structure shows one additional water (grey circle) compared to the QM/MM structure. The center of excess charge (yellow sphere) is close to the first water coordinated to E203 in this window for both the QM/MM and MS-RMD configurations, meaning

that the first water is the most hydronium-like species (the "pivot" hydronium, which has the largest MS-RMD state coefficient, c_1, in Equation 8). To better understand the difference in solvation structures and charge delocalization, we calculated the average value of c_1^2 for all configurations in the $\zeta_R = 0.24$ umbrella windows from the QM/MM and MS-RMD simulations (Fig 6B). It should be kept in mind that the c_1^2 value for an Eigen cation ($H_9O_4^+$) is ~ 0.65, which indicates that the pivot hydronium holds ~ 65% of the positive charge (hydronium-like) defect, while that for a Zundel cation ($H_5O_2^+$) is ~ 0.5 since in that case excess charge is more equally shared between two water molecules. The MS-RMD simulations yield larger $\langle c_1^2(\mathbf{r})\rangle$ values (0.7) because the pivot hydronium is coordinated to two waters in the central region, in addition to the carboxyl group, leading to a stable Eigen-like ($CO\text{-}H_7O_3^+$) complex. However, one of these waters is missing in the QM/MM simulations (Fig 6C - gray circle), shifting $\langle c_1^2(\mathbf{r})\rangle$ to a lower value (0.55) and the delocalization to more of a Zundel-like complex between the carboxyl group and first water molecule ($CO\text{-}H_3O^+$). The MS-RMD Eigen-like complex forms a contact ion pair (CIP) with the carboxyl group of deprotonated E203 and the stabilizing electrostatic attraction between oppositely charged ions causes a small energy well (~1 kcal/mol) in the MS-RMD PMF. However, in the QM/MM configurations, no hydrogen bond acceptor is found near one of hydrogen atoms of the pivot hydronium, causing the PMF to continue its uphill climb.

To see if the PMF discrepancy was caused by sampling different conformational phase space, another set of QM/MM umbrella sampling simulations was run for 5 ps (as described earlier), initiated from the last snapshots from the MS-RMD umbrella sampling simulations in the CV range $0.18 < \zeta_R < 0.26$. The resulting QM/MM PMF (green in Fig 6A) shows much better agreement with the MS-RMD PMF. Figure 6B and E show that for this second QM/MM

simulation the value of $\langle c_1^2(\mathbf{r})\rangle$ is close to the MS-RMD value and the missing second water is present. Thus, the PMF discrepancy is indeed caused by sampling different conformational phase space with that sampled by MS-RMD having one additional water molecule to stabilize the excess proton.

To estimate the time scale required for the additional water molecule (missing in the original QM/MM simulations) to enter the central region, an MS-RMD configuration was prepared (see Methods) at $\zeta_R = 0.24$ in which this water was removed (mimicking the QM/MM solvation structure shown in Figure 6C). The system was equilibrated in this low solvation state, and then ten independent MS-RMD simulations were run for 500 ps. Figure 6F shows the coordination number of the pivot hydronium to the oxygen atoms of the surrounding water molecules or the carboxyl group of E203 in one of the trajectories. The coordination number at $t = 0$ ps is 2, when the pivot hydronium is coordinated to one water oxygen and E203's carboxyl oxygen. Although there are some transient transitions to a coordination of 3 around $t \simeq 30$-70 ps, due to an additional water oxygen, the second water is not stable for another 100 ps. The second water then enters the central region from the intracellular bulk. The coordination number remains 3 for most of the last 400 ps of the trajectory (~ 89 % of the time). This trajectory is representative of the other nine trajectories, in which the second water did not enter the central region until t = 60 - 185 ps ($\langle t \rangle$ = 99 ps), after which the coordination number remained 3 for 88 ± 5 % of the rest of the trajectories. These results suggest that the entrance of the second water is energetically favored in this umbrella window, but that the time scale of this event exceeds the sampling time possible in the QM/MM simulations.

Although the discrepancy that motivated this detailed investigation was small, it highlights another important limitation of QM/MM MD PMFs. In addition to the errors and artificial dynamics introduced by QM/MM boundary issues, limited sampling, especially of slow degrees of freedom such as changing hydration, can hide insidious PMF errors. This reinforces the importance of being able to map accurate *ab initio* forces onto an efficient method that is capable of extensive sampling. Free energy profiles in complicated condensed phase environments are generally a balance of enthalpic and entropic contributions, requiring both accurate potential energy representations and extensive sampling. This is further explored below.

MS-RMD can capture coupling between hydration and PT in CcO

C*c*O offers another example of PT through an interior protein region being coupled to hydration changes. During the reaction cycle of aa_3-type C*c*O, as found in mitochondria, protons from the intracellular side of the membrane are transported through the so-called D-channel to the glutamic acid E286 in the middle of the membrane. The protons are then transported through water molecules in a hydrophobic cavity (the HC) to either the pump loading site (PLS) to be further released to the periplasmic side of the membrane, or to the binuclear center (BNC) to react with oxygen and form water. A particularly interesting aspect of the proton pumping mechanism in C*c*O is the role of water molecules in the HC during PT from amino acid E286 to the PLS or BNC (Figure 7B). The number of water molecules in the HC and their role in PT has been the focus of much debate.(21, 22, 66-70) To further complicate this issue, the water molecules can move in and out of the HC in C*c*O in response to the migration of the excess proton, and the two processes can be intrinsically coupled to each other. Capturing this type of cooperativity often requires

computationally demanding enhanced sampling of multiple degrees of freedom. Here we show that the computationally efficient MS-RMD method parameterized by FitRMD approach allowed us to address this challenge.

We calculated 2D PMFs for the PT from E286 to the PLS in different redox states during the A→ F transition of C*c*O. The collective variables used to define these 2D PMFs are (1) the progress of the excess proton CEC through the HC (horizontal axis) and (2) the degree of hydration of the HC (vertical axis) (Figure 8, see Methods for more discussion). The 2D PMF shown in Figure 8 clearly reveals the cooperativity between the PT and dynamic hydration in the HC. The minimum free energy pathway (black line) along the 2D PMF (Figure 8) verifies that as the excess proton migrates from E286 to the PRDa_3, the hydration level of HC gradually increases. The non-horizontal feature of the minimum free energy pathway indicates that the HC hydration is intrinsically coupled to the proton charge defect translocation in this activated process.

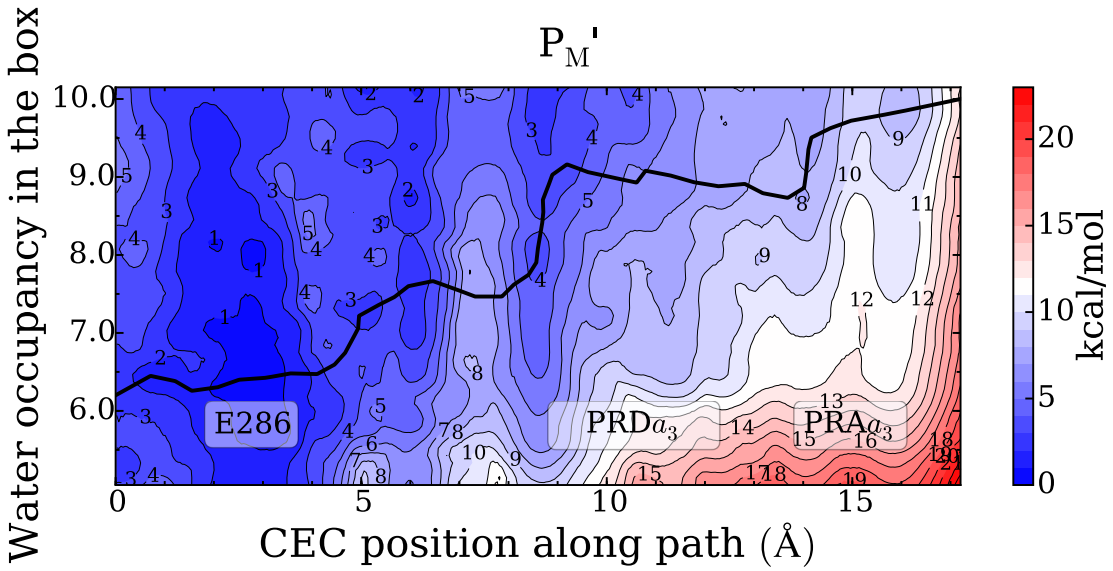


Figure 1.8. Two-dimensional free energy profiles (2D-PMFs) for PT from the E286 to the PLS in the P_M' state, as a function of the CEC coordinate through the hydrophobic cavity (HC) as the horizontal axis and the water hydration in the HC as the vertical axis. The minimum free energy pathways (black lines) are diagonal in nature, indicating the two processes are coupled. The statistic errors of the 2D-PMFs is ~ 0-3 kcal/mol.

It is also important to emphasize how the computational efficiency of the MS-RMD method makes it feasible to calculate such 2D PMFs. Figure 9 shows a relative speed and scaling over processors plot of the MS-RMD, SCC-DFTB-based QM/MM and BLYP-based QM/MM MD methods for C*c*O. It is seen that MS-RMD is at least three orders of magnitude faster than the BLYP-based QM/MM simulation, and two orders of magnitude faster than the SCC-DFTB-based QM/MM simulation. MS-RMD also scales better over processors than the other two methods (the SCC-DFTB CHARMM code is not scalable at all.)

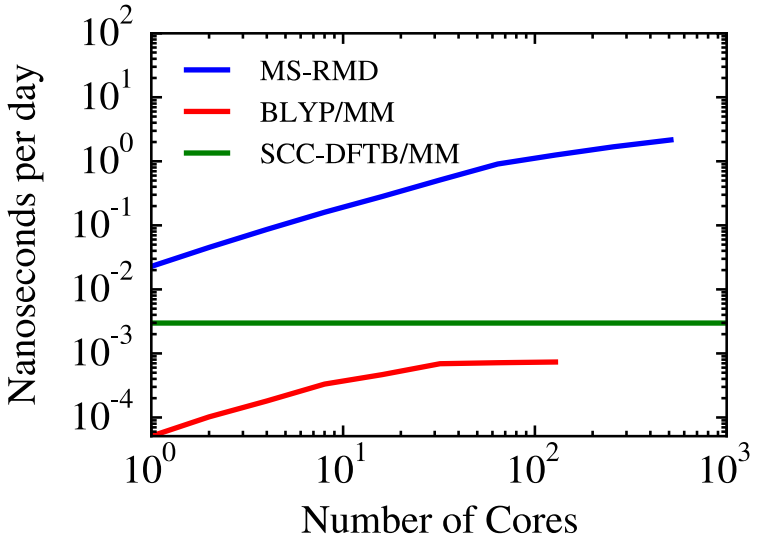


Figure 1.9. Scaling plot for the MS-RMD, SCC-DFTB-based QM/MM and BLYP-based QM/MM simulations of C*c*O. The blue curve is for MS-RMD, the green is for SCC-DFTB, and the red curve is for BLYP-based QM/MM. The MS-RMD method is at least 3 orders of magnitude faster than BLYP-based QM/MM method and 2 orders of magnitude faster than SCC-DFTB-based QM/MM, while more favorable scaling properties over processors (the SCC-DFTB CHARMM code is not scalable).

1.4 Conclusions

Describing reactive processes in biomolecular systems remains a challenging domain for molecular simulations. The *ab initio* methods that are capable of describing chemical reactions generally do not scale to the size of most biomolecules or to the timescales needed to converge multidimensional free energy profiles of rare events in condensed phase environments. Thus, bridging a quantum description of the reactive processes with more computationally efficient approaches (e.g., classical dynamics and enhanced sampling methods) that are capable of extensive phase space sampling is of great utility and fundamental importance. The work presented herein

contributes a multiscale framework aimed at accomplishing this multiscale bridging, specifically for the purpose of simulating the protonation/deprotonation of ionizable moieties such as amino acids in biomolecules. The core of our framework is the MS-RMD method, which describes reactive processes in classical MD and explicitly treats the charge defect delocalization and Grotthuss shuttling of the hydrated excess proton in the PT process. However, MS-RMD must be carefully parameterized in order to simulate PT faithfully and the approach is not yet a "black box". Considerable effort has gone into the parameterization of the MS-RMD model for PT in water(27, 29, 30, 37, 38) and amino acids.(31, 71, 72) Herein we have extended to proteins the parameterization scheme from QM/MM data to MS-RMD models (FitRMD), and demonstrated its use on protonatable amino acids in biomolecular systems.

As presented here, the FitRMD approach variationally maps quantum data (DFT-level QM/MM forces in this case) onto the MS-RMD nonlinear reactive force field. We have demonstrated how FitRMD can be used to parameterize MS-RMD models for amino acids in proteins with two example systems, ClC-ec1 and C*c*O. In ClC-ec1 the MS-RMD models were shown to faithfully reproduce the PT PMFs of the reference QM/MM MD, both in the presence and absence of a Cl^- ion in the central region. Thus, the FitRMD parameterization was robust enough to capture quite different free energy profiles due to the presence or absence of a single ion. Moreover the local structure of the protein and confined water was shown to be quite similar between the MS-RMD and QM/MM simulations. In C*c*O the developed MS-RMD models were shown to be efficient enough to capture the coupling between PT and hydration changes in the HC region. Two-dimensional free energy surfaces in which both PT and hydration levels were explicitly sampled are required for this analysis. Given that individual umbrella windows often required over 500ps to converge, these calculations would not be possible with a QM/MM MD

approach. Moreover QM/MM boundary issues, such as the lack of exchange of water molecules, can lead to large systematic errors in such PT PMFs. Even for the ClC-ec1 system where exchange of water across the QM/MM boundary is minimal due to protein confinement, lack of water exchange was argued to introduce errors in the calculated QM/MM free energy profile.

It is worth noting that in this paper the original parameterization method of Nelson *et. al.*(31) was extended and shown to work in much more complex molecular situations. For example, in ClC-ec1 the MS-RMD models produce a PMF that is consistent with the QM/MM Hamiltonian even though the CV used for the configuration sampling (absolute distance) is different from the CV used for the PMF calculation (ratio of distances) (see Methods for a full description). Moreover, different models for the same Glu residues were parameterized in different states, with and without the central Cl^- ion, resulting in quite different PMFs that were independently consistent with the QM/MM PMFs. It should also be emphasized that although a low-level QM method (BLYP-D) was used here in the QM/MM calculations, our goal was to demonstrate that a reference *ab initio* Hamiltonian (and free energy surface) can be reproduced. Thus, a MS-RMD model obtained from FitRMD using a higher level reference *ab initio* method can be expected to also faithfully reproduce the free energy surface from the latter, if only the latter could be calculated, which it currently cannot be. In principal, therefore, our multiscale FitRMD with MS-RMD framework can be used in the future to estimate high-level (e.g., MP2) free energy profiles for reactive processes, although the accuracy of the surrounding environment will remain dependent on the chosen classical force field.

It also seems clear that the MS-RMD model parameters will depend on the given system under study and hence not likely to be "transferable" to some other system. However, this is precisely

the point of the MS-RMD approach (and also the QM/MM approach which it is fitting). It is very unlikely that it is generally possible to have transferable potential parameters for reactive processes. On the other hand, we note that the MS-RMD fitting methodology is itself transferable to different systems and its use is practical, because the total computational cost for both the model fit and the MD sampling with MS-RMD will be still much cheaper than directly performing QM/MM MD simulations on the same system.

Our future efforts will focus on improving the FitRMD protocol to make it more robust to choice of the trial MS-RMD model, as free as possible from co-evolution of parameters, and insensitive to discrepancies between the QM and MM descriptions of non-reactive atoms in the system. We anticipate that reactive processes in many biomolecular and other systems can eventually be studied with the MS-RMD approach.

1.5 References

1. Knight C & Voth GA (2012) The Curious Case of the Hydrated Proton. *Acc. Chem. Res.* 45(1):101-109.

2. Swanson JMJ, *et al.* (2007) Proton solvation and transport in aqueous and biomolecular systems: Insights from computer simulations. *J. Phys. Chem. B* 111(17):4300-4314.

3. Engh AM & Maduke M (2005) Cysteine Accessibility in ClC-0 Supports Conservation of the ClC Intracellular Vestibule. *J. Gen. Physiol.* 125(6):601-617.

4. Zhang X-D, Li Y, Yu W-P, & Chen T-Y (2006) Roles of K149, G352, and H401 in the Channel Functions of ClC-0: Testing the Predictions from Theoretical Calculations. *J. Gen. Physiol.* 127(4):435-447.

5. Chen TY (2005) Structure and function of ClC channels. *Annu. Rev. Physiol.* 67(1):809-839.

6. Estevez R & Jentsch TJ (2002) CLC chloride channels: correlating structure with function. *Curr. Opin. Struct. Biol.* 12(4):531-539.

7. Jentsch TJ (2008) CLC chloride channels and transporters: from genes to protein structure, pathology and physiology. *Crit. Rev. Biochem. Mol. Biol.* 43(1):3-36.

8. Accardi A & Miller C (2004) Secondary active transport mediated by a prokaryotic homologue of ClC Cl^- channels. *Nature* 427(6977):803-807.

9. Miller C & Nguitragool W (2009) A provisional transport mechanism for a chloride channel-type Cl^-/H^+ exchanger. *Phil. Trans. R. Soc. B* 364(1514):175-180.

10. Wikstrom MKF (1977) Proton Pump Coupled to Cytochrome-C Oxidase in Mitochondria. *Nature* 266(5599):271-273.

11. FergusonMiller S & Babcock GT (1996) Heme/copper terminal oxidases. *Chem. Rev.* 96(7):2889-2907.

12. Kaila VR, Verkhovsky MI, & Wikstrom M (2010) Proton-Coupled Electron Transfer in Cytochrome Oxidase. *Chem. Rev.* 110:7062-7081.

13. Wikström M, Sharma V, Kaila VRI, Hosler JP, & Hummer G (2015) New Perspectives on Proton Pumping in Cellular Respiration. *Chem. Rev.*

14. Pringle CR (1999) Virus Taxonomy *Arch. Virol.* 144(2):421-429.

15. Wang C, Holsinger LJ, Lamb RA, & Pinto LH (1994) The M2 Protein of Influenza-a Virus Forms an Ion-Channel in Mammalian-Cells. *Biophys. J.* 66(2):A350-A350.

16. Wang D & Voth GA (2009) Proton Transport Pathway in the CIC Cl^-/H^+ Antiporter. *Biophys. J.* 97(1):121-131.

17. Zhang Y & Voth GA (2011) The Coupled Proton Transport in the ClC-ec1 Cl^-/H^+ Antiporter. *Biophys. J.* 101(10):L47-L49.

18. Yamashita T & Voth GA (2010) Properties of Hydrated Excess Protons near Phospholipid Bilayers. *J. Phys. Chem. B* 114:592-603.

19. Lee HJ, *et al.* (2010) Intricate Role of Water in Proton Transport through Cytochrome *c* Oxidase. *J. Am. Chem. Soc.* 132:16225-16239.

20. Liang R, Li H, Swanson JMJ, & Voth GA (2014) Multiscale simulation reveals a multifaceted mechanism of proton permeation through the influenza A M2 proton channel. *Proc. Natl. Acad. Sci. U. S. A.* 111(26):9396-9401.

21. Yamashita T & Voth GA (2012) Insights into the mechanism of proton transport in cytochrome *c* oxidase. *J. Am. Chem. Soc.* 134(2):1147-1152.

22. Goyal P, Lu J, Yang S, Gunner MR, & Cui Q (2013) Changing hydration level in an internal cavity modulates the proton affinity of a key glutamate in cytochrome c oxidase. *Proc. Natl. Acad. Sci. U. S. A.* 110(47):18886-18891.

23. Peng Y, Swanson JMJ, Kang S-g, Zhou R, & Voth GA (2015) Hydrated Excess Protons Can Create Their Own Water Wires. *J. Phys. Chem. B* 119(29):9212-9218.

24. Bulo RE, Ensing B, Sikkema J, & Visscher L (2009) Toward a Practical Method for Adaptive QM/MM Simulations. *J. Chem. Theory Comput.* 5(9):2212-2221.

25. Park K, Gotz AW, Walker RC, & Paesani F (2012) Application of Adaptive QM/MM Methods to Molecular Dynamics Simulations of Aqueous Systems. *J. Chem. Theory Comput.* 8(8):2868-2877.

26. Mones L, *et al.* (2015) The Adaptive Buffered Force QM/MM Method in the CP2K and AMBER Software Packages. *J. Comput. Chem.* 36(9):633-648.

27. Knight C, Lindberg GE, & Voth GA (2012) Multiscale reactive molecular dynamics. *J. Chem. Phys.* 137(22).

28. Yamashita T, Peng Y, Knight C, & Voth GA (2012) Computationally Efficient Multiconfigurational Reactive Molecular Dynamics. *J. Chem. Theory Comput.* 8(12):4863-4875.

29. Schmitt UW & Voth GA (1999) The computer simulation of proton transport in water. *J. Chem. Phys.* 111(20):9361-9381.

30. Wu YJ, Chen HN, Wang F, Paesani F, & Voth GA (2008) An improved multistate empirical valence bond model for aqueous proton solvation and transport. *J. Phys. Chem. B* 112(2):467-482.

31. Nelson JG, Peng Y, Silverstein DW, & Swanson JMJ (2014) Multiscale Reactive Molecular Dynamics for Absolute pK(a) Predictions and Amino Acid Deprotonation. *J. Chem. Theory Comput.* 10(7):2729-2737.

32. Akin-Ojo O, Song Y, & Wang F (2008) Developing ab initio quality force fields from condensed phase quantum-mechanics/molecular-mechanics calculations through the adaptive force matching method. *J. Chem. Phys.* 129(6):064108.

33. Wang F, Akin-Ojo O, Pinnick E, & Song Y (2011) Approaching post-Hartree–Fock quality potential energy surfaces with simple pair-wise expressions: parameterising point-charge-based force fields for liquid water using the adaptive force matching method. *Mol. Simul.* 37(7):591-605.

34. Zhou Y & Pu J (2014) Reaction Path Force Matching: A New Strategy of Fitting Specific Reaction Parameters for Semiempirical Methods in Combined QM/MM Simulations. *J. Chem. Theory Comput.* 10(8):3038-3054.

35. MacKerell AD, *et al.* (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* 102(18):3586-3616.

36. Akin-Ojo O & Wang F (2009) Improving the Point-Charge Description of Hydrogen Bonds by Adaptive Force Matching. *J. Phys. Chem. B* 113(5):1237-1240.

37. Day TJF, Soudackov AV, Čuma M, Schmitt UW, & Voth GA (2002) A second generation multistate empirical valence bond model for proton transport in aqueous systems. *J. Chem. Phys.* 117(12):5839-5849.

38. Biswas R, Tse Y-LS, Tokmakoff A, & Voth GA (2015) Role of Pre-Solvation and Anharmonicity in Aqueous Phase Hydrated Proton Solvation and Transport. *J. Phys. Chem. B*:(In Press).

39. Lee S, Swanson JMJ, & Voth GA (2015) Multiscale simulations reveal the proton transport mechanism in the ClC-ec1 antiporter. *Biophys. J.*:(submitted).

40. Laino T, Mohamed F, Laio A, & Parrinello M (2006) An efficient linear-scaling electrostatic coupling for treating periodic boundary conditions in QM/MM simulations. *J. Chem. Theory Comput.* 2(5):1370-1378.

41. Laino T, Mohamed F, Laio A, & Parrinello M (2005) An efficient real space multigrid OM/MM electrostatic coupling. *J. Chem. Theory Comput.* 1(6):1176-1184.

42. Blochl PE (1995) Electrostatic Decoupling of Periodic Images of Plane-Wave-Expanded Densities and Derived Atomic Point Charges. *J. Chem. Phys.* 103(17):7422-7428.

43. Maseras F & Morokuma K (1995) Imomm - a New Integrated Ab-Initio Plus Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition-States. *J. Comput. Chem.* 16(9):1170-1179.

44. Becke AD (1988) Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys. Rev. A* 38(6):3098-3100.

45. Lee CT, Yang WT, & Parr RG (1988) Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys. Rev. B* 37(2):785-789.

46. Grimme S, Antony J, Ehrlich S, & Krieg H (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* 132(15).

47. Lippert G, Hutter J, & Parrinello M (1997) A hybrid Gaussian and plane wave density functional scheme. *Mol. Phys.* 92(3):477-487.

48. Hartwigsen C, Goedecker S, & Hutter J (1998) Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* 58(7):3641-3662.

49. Muhlenbein H & Schlierkamp-Voosen D (1993) Predictive Models for the Breeder Genetic Algorithm I. Continuous Parameter Optimization. *Evol. Comput.* 1(1):25-49.

50. Plimpton S (1995) Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* 117(1):1-19.

51. Tribello GA, Bonomi M, Branduardi D, Camilloni C, & Bussi G (2014) PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* 185(2):604-613.

52. VandeVondele J, *et al.* (2005) QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* 167(2):103-128.

53. Konig PH, *et al.* (2006) Toward theoretical analyis of long-range proton transfer kinetics in biomolecular pumps. *J. Phys. Chem. A* 110(2):548-563.

54. Riccardi D, *et al.* (2006) "Proton holes" in long-range proton transfer reactions in solution and enzymes: A theoretical analysis. *J. Am. Chem. Soc.* 128(50):16302-16311.

55. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, & Kollman PA (1992) THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13(8):1011-1021.

56. Laio A & Parrinello M (2002) Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* 99(20):12562-12566.

57. Zhang Y & Voth GA (2011) Combined Metadynamics and Umbrella Sampling Method for the Calculation of Ion Permeation Free Energy Profiles. *J. Chem. Theory Comput.* 7(7):2277-2283.

58. Becke AD (1993) Density-Functional Thermochemistry .3. The Role of Exact Exchange. *J. Chem. Phys.* 98(7):5648-5652.

59. Klauda JB, *et al.* (2010) Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* 114(23):7830-7843.

60. Liang R, *et al.* (2015) Multiscale simulations reveal the proton pumping mechanism in cytochrome *c* oxidase. *Proc. Natl. Acad. Sci. U. S. A.*:(submitted).

61. Elstner M, *et al.* (1998) Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* 58(11):7260-7268.

62. Cui Q, Elstner M, Kaxiras E, Frauenheim T, & Karplus M (2001) A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J. Phys. Chem. B* 105(2):569-585.

63. Riccardi D, Schaefer P, & Cui Q (2005) pK(a) calculations in solution and proteins with QM/MM free energy perturbation simulations: A quantitative test of QM/MM protocols. *J. Phys. Chem. B* 109(37):17715-17733.

64. Liang R, Swanson JMJ, & Voth GA (2014) Benchmark Study of the SCC-DFTB Approach for a Biomolecular Proton Channel. *J. Chem. Theory Comput.* 10(1):451-462.

65. Brooks BR, *et al.* (2009) CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* 30(10):1545-1614.

66. Riistama S, *et al.* (1997) Bound water in the proton translocation mechanism of the haem-copper oxidases. *Febs Lett* 414(2):275-280.

67. Wikstrom M, Verkhovsky MI, & Hummer G (2003) Water-gated mechanism of proton translocation by cytochrome *c* oxidase. *Bba-Bioenergetics* 1604(2):61-65.

68. Tuukkanen A, Kaila VRI, Laakkonen L, Hummer G, & Wikstrom M (2007) Dynamics of the glutamic acid 242 side chain in cytochrome *c* oxidase. *Bba-Bioenergetics* 1767(9):1102-1106.

69. Sharma V, Enkavi G, Vattulainen I, Róg T, & Wikström M (2015) Proton-coupled electron transfer and the role of water molecules in proton pumping by cytochrome *c* oxidase. *Proc. Natl. Acad. Sci. U. S. A.* 112(7):2040-2045.

70. Ghosh N, Prat-Resina X, Gunner MR, & Cui Q (2009) Microscopic pKa Analysis of Glu286 in Cytochrome *c* Oxidase (Rhodobacter sphaeroides): Toward a Calibrated Molecular Model. *Biochemistry* 48:2468-2485.

71. Maupin CM, Wong KF, Soudackov AV, Kim S, & Voth GA (2006) A Multistate Empirical Valence Bond Description of Protonatable Amino Acids. *J. Phys. Chem. A* 110(2):631-639.

72. Maupin CM, McKenna R, Silverman DN, & Voth GA (2009) Elucidation of the Proton Transport Mechanism in Human Carbonic Anhydrase II. *J. Am. Chem. Soc.* 131(22):7598-7608.

CHAPTER 2

THE FREE ENERGY SAMPLING OF PROTON TRANSPORT IN THE CENTRAL REGION OF CLC-ec1

2.1 Introduction

The ClC proteins are a family of chloride channels and transporters that are widely distributed in species ranging from bacteria to human beings (1). They have a broad range of physiological functions, including maintenance of the membrane potential, regulation of transepithelial Cl^- transport, and control of pH in the cytoplasm and intracellular organelles (2, 3). In addition to their physiological importance, two aspects of ClC proteins have attracted interest in recent years. First, despite having very similar structures the ClC proteins can be either passive ion channels, facilitating selective anion transport down an electrochemical gradient, or secondary active transporters, coupling the uphill movement of protons (H^+) to the downhill transport of Cl^- ions (and vice versa) (4-6). Second, the mechanism of ClC transporters seems to be unique. Most exchangers transfer ions across the membrane in opposite directions in separate half reactions via large conformational changes in what is called a ping-pong or alternating access mechanism. However, ClC transporters can bind both ions simultaneously and transport them across the membrane through partially congruent pathways, resulting in an intriguing mechanism that has yet to be fully explained (7). Although recent experimental studies (8-11) have suggested that there are some pH- or Cl^- dependent conformational changes in regions outside of the ion pathways, their magnitude and connection to ion transport remain unknown.

ClC-ec1, a bacterial ClC transporter homologue from *Escherichia coli*, controls the stoichiometric exchange of 2 Cl^- for 1 H^+ (12, 13). Following the determination of a high-

resolution crystal structures for ClC-ec1 (14, 15), a series of electrophysiological experiments (12, 16, 17) and computational studies (18-21) revealed the transport channels for the Cl^- ions. Like all ClC transporters, ClC-ec1 forms a dimer in vivo and in vitro, and can function as a monomer in vitro (22). Each monomer has three intermediate binding sites for Cl^- (S_{ext}, S_{cen}, and S_{int}) along the Cl^- pathway (see Fig. 1) (15). The protonation of glutamic acid residue E148 (Glu_{ex}) at the center of the conducting pore allows Cl^- ions to enter S_{cen} from the extracellular side (23). The H^+ pathway is thought to separate from the Cl^- pathway on the intracellular side of Glu_{ex}, but the two pathways overlap on the extracellular side (16, 24-26). S_{ext} and S_{cen} are located along the H^+ pathway, and S_{int} is separated from S_{cen} by only 6.5 Å.

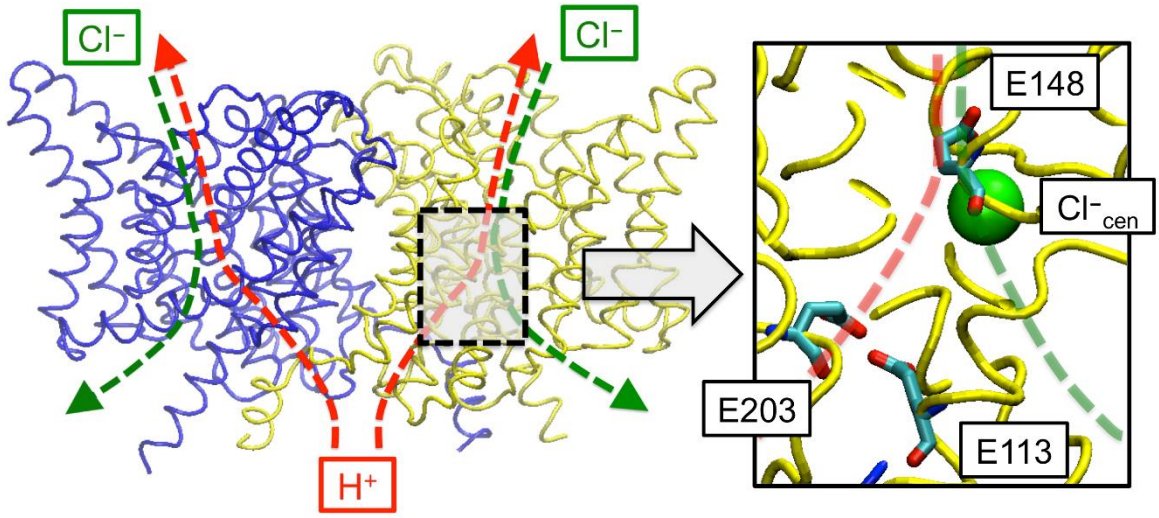


Figure 2.1. Overview of the structure of the ClC-ec1 antiporter and transport pathways for Cl^- (green dashed) and H^+ (red dashed). ClC-ec1 consists of two monomers, where monomer A colored in yellow, and monomer B in blue. The central region is highlighted in the black box. Two pathways for Cl^- and H^+ meet at E148. In the figure, one Cl^- is bound to S_{cen} site, which is very close to the H^+ pathway. The other Cl^- binding sites (S_{ext} and S_{int}) are not shown here.

Site-directed mutagenesis studies (16, 27) revealed that two glutamate residues, E148 (Glu_{ex}) and E203 (Glu_{in}), which are located near the external and internal sides of the membrane respectively, are essential to proton transport (PT). Thus, it is assumed that Glu_{in} accepts a proton from the intracellular side and transfers it to Glu_{ex}, but how it does so in concert with Cl^- transport is unknown. Since the two residues are separated by about 15 Å, water molecules or protonatable residues must occupy this space in order to facilitate PT. However, the residues surrounding this region (I109, L186, F199, F357, I402, and T407) are mostly hydrophobic prior to any possible large conformational changes (8-11), and none of the crystal structures from *E. coli* (15) or other homologs (28) exhibit any crystallographic water molecules in this region. Moreover, it is unclear what influences the presence or absence of Cl^- at the binding sites or the protonation states of the surrounding residues have on the PT from Glu_{in} to Glu_{ex}. Thus, the molecular mechanism by which protons traverse the ClC-ec1 core, and the impact PT has on H^+:Cl^- antiporting, remains unclear.

In the current study, PT through the central region is investigated by calculating the free energy profiles (potentials of mean force, PMFs) for *explicit* proton migration from Glu_{in} to Glu_{ex}. Multiple PMFs are calculated to determine the effect of various factors on the migration of the excess proton: 1) the binding of Cl^- at the central site (Cl^-_{cen}), 2) the dynamic hydration of the central region, and 3) the rotation of Glu_{in} and Glu_{ex}. A combination of quantum mechanics/molecular mechanics (QM/MM) and multiscale reactive molecular dynamics (MS-RMD) (29, 30) is used to enable realistic description of the excess proton migration. The MS-RMD model, which is force-matched to *ab initio* data, also provides the required computational efficiency to enable extensive statistical sampling of the slow hydration dynamics in the central region along with the excess proton migration through the water and involved residues via Grotthuss shuttling. By virtue of MS-RMD's accuracy and efficiency, this coupled behavior is

quantitatively expressed, for the first time, in terms of *two-dimensional* (2D) PMFs calculated as a function of the excess proton coordinate and the water occupancy in the central region. The presented quantitative simulations build significantly upon our earlier and more preliminary work on this problem (24, 25), while at the same time being quite different from MD simulations that have studied the water occupancy alone in the absence of explicit proton transport (the latter then being inferred from the water hydration patterns rather than explicitly simulated) (26, 31).

Our results provide insight the elementary mechanism of PT in the ClC-ec1 central region. First, the excess proton, which likely resides on E113 or is shared between E113 and E203 in the crystal structure as previously predicted (32), is shown to transfer readily from E113 to E203 (Glu_{in}) prior to PT to E148 (Glu_{ex}). Second, PT from Glu_{in} to Glu_{ex} is shown to be dynamically coupled to the hydration of the central region (and vice versa). Third, the presence of Cl^-_{cen} is found to facilitate PT between Glu_{in} and Glu_{ex}, despite blocking the downward rotation of Glu_{ex}. Surprisingly, however, we find that proton transport is also kinetically feasible in the absence of Cl^-_{cen}, when side chain rotation facilitates PT, mimicking a classical heavy-light-heavy reaction (33). The calculated PT rate constants suggest that PT through the central region is not rate determining, as suggested by experiments (34). Although experimentalists have proposed two mechanisms for wild-type (WT) ClC-ec1 (13, 35), differing in the presence or absence of Cl^-_{cen} during PT through the central region, our simulations reveal that PT can happen in either case, though the presence of the anion does lower the PT free energy barrier. This result further supports the notion that the requisite coupling between proton and Cl^- transport occurs elsewhere in the overall PT process (e.g., during proton uptake to, or release from, residues E203 or E148), as previously suggested based on calculated pKa shifts of E148 depending on ion occupancy (36).

Finally, PT through the central region of the interesting E203K mutant is also explored. Despite the replacement of Glu_{in} with a basic residue, experiments show that H^+ permeation is not blocked and that the Cl^-/H^+ coupling remains intact (though slightly weaker) in this mutant (27). Our simulations suggest that PT from K203 to Glu_{ex} is indeed kinetically and thermodynamically possible, but only in the presence of Cl^-_{cen}. Similar to WT, the calculated rate for PT through the central region is not rate limiting.

2.2 Methods

System setup

The initial configuration of the ClC-ec1 dimer (PDB ID: 1OTS), embedded in 163 POPE lipids, 17 Cl^-s, and ~11,000 water molecules in a 92 Å x 92 Å x 79 Å simulation box, was taken from our previous study (24). In this structure, Cl^- ions were present at S_{cen} and S_{int} of each monomer, as also given in the PDB structure. Consistent with pKa calculations (32), standard protonation states were chosen for all residues other than E113 in monomer B and D417 in monomer A, which were both protonated. Previous experiments have shown that ClC functions as a monomer (22); thus, all enhanced sampling used to calculate the PMFs was applied only to monomer A. Similarly, all variations in the core region (e.g., the presence or absence of Cl^-, different protonation states, and the E203K substitution) were only made in monomer A. Removing all restraints used in previous studies, the system was equilibrated for 10 ns.

Classical molecular dynamics simulations

The classical molecular dynamics (MD) simulations were performed with the Gromacs MD package (37), the CHARMM27 force field (38, 39), and the TIP3P water model (40) under periodic boundary conditions in the constant NVT ensemble with a temperature of 300 K. Long-ranged electrostatic interactions were treated with the Particle Mesh Ewald (PME) method (41). The cutoff distance for both the Lennard-Jones and the real space Coulomb interaction was set to be 12 Å. The systems in the presence of Cl^-_{cen} and Cl^-_{int} (the same configurations with PDB structure) were 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 (42). Cl^-_{int} spontaneously diffused out to the bulk on the intracellular side during the equilibration steps. In order to generate the configurations without Cl^-_{cen}, well-tempered metadynamics (43) was performed with the PLUMED package (44) until Cl^-_{cen} is gradually displaced from S_{cen} to the intracellular side. A Gaussian potential was added for the z coordinate of Cl^-_{cen} every 1 ps. The initial Gaussian height was set to be 0.1 kcal/mol and damping parameter ΔT was 300 K.

For the E203K mutant, E203 was substituted for Lys from the equilibrated structures of the wild type in the presence and the absence of Cl^-_{cen}, respectively, and the protonation state of E202 was set to be either protonated or deprotonated. Then, all systems were equilibrated again for another 10 ns with the same simulation setups.

Classical MD of the hydration of the central region

Starting from the equilibrated structure for each system, the classical MD simulation was performed for 50 ns. The protonation states of the residues in the central region were set to be

E113(0)E148(0)E203(H) in the WT, and E113(0)E148(0)K203(H) in the E203K mutant. Note that "0" in the parenthesis denotes the side chain is in the deprotonated state, and "H" for the protonated state. Because the hydration process does not require proton transfer, E148 is set to be deprotonated. Previous experimental (15) and simulation (18) studies showed that Cl^- is destabilized at S_{ext} when E148 is deprotonated, so the system with Cl^- at S_{ext} is not considered. Similarly, the possibility of Cl^- at S_{int} is not examined because despite the presence of the anion density at S_{int} in the crystal structure, previous MD studies showed that when E148 is deprotonated, Cl^- is not stabilized at S_{int}, and easily diffuses out to the intracellular side (20, 24-26). All atomic coordinates were recorded every 100 ps. The number of water molecules in the central region was calculated, and the presence of a continuous hydrogen bond network was determined at each frame. A continuous hydrogen bond network is defined by the presence of at least one pathway through which the carboxyl oxygen atoms of Glu_{ex} and Glu_{in} are continuously connected by hydrogen bonds through water molecules between them, without any breaks. A hydrogen bond is designated when the distance between two oxygen atoms is smaller than 3.5 Å, and the angle between two oxygen and a shared hydrogen is smaller than 30°.

QM/MM free energy profiles for PT from E113 to E203

The initial configurations for umbrella sampling (45) were prepared after 10 ns of equilibration MD runs with either Cl^-_{cen} present or absent. E113, which was protonated during equilibration, was deprotonated and a water molecule close to the center of each umbrella sampling window was replaced with a protonated water molecule (hydronium cation). The effective spatial position of the excess proton is defined as the center of excess charge (CEC), $\vec{r}_{CEC}$ (46):

$$\vec{r}_{CEC} = \sum_{i=1}^{N_H} \vec{r}^{H_i} - \sum_{j=1}^{N_X} w^{X_j} \vec{r}^{X_j} - \sum_{i=1}^{N_H} \sum_{j=1}^{N_X} f_{sw}(d_{X_j H_i}) \left(\vec{r}^{H_i} - \vec{r}^{X_j} \right) \quad (1)$$

where the $\vec{r}^{X_j}$ is the position of the *j*th heavy atom in the QM region (one of the water oxygen atoms or glutamic acid carboxyl oxygen atoms), and the $\vec{r}^{H_i}$ is the position of the *i*th hydrogen atom bound to those heavy atoms. The weighting factor, w^{X_j}, was set to be 2 for the water oxygen and 0 for the glutamic acid OE1 and OE2 atoms, which reflects the hydrogen coordination number in the deprotonated state of the heavy atom. The term $d_{X_j H_i}$ here denotes the distance between X_j and H_i atoms, and the switching function $f_{sw}(d_{X_j H_i})$ is given as follows,

$$f_{sw}(d_{X_j H_i}) = \frac{1}{1 + \exp[(d_{X_j H_i} - r_{sw}) / d_{sw}]} \quad (2)$$

where d_{sw} is 0.04 Å and r_{sw} is 1.25 Å (47). The transfer of the CEC is mapped onto a one-dimensional collective variable (CV), denoted as ζ_R, as a function of the distances between $\vec{r}_{CEC}$ and the center of mass of the carboxyl group atoms of two Glu residues. For example, for the PMF of PT from E113 ($\vec{r}_{113}$) to E203 ($\vec{r}_{203}$), ζ_R is defined as

$$\zeta_R = \frac{\left| \vec{r}_{CEC} - \vec{r}_{113} \right|}{\left| \vec{r}_{CEC} - \vec{r}_{113} \right| + \left| \vec{r}_{CEC} - \vec{r}_{203} \right|} \quad (3)$$

such that ζ_R varies from 0 to 1 as the CEC moves from E113 to E203. The centers of the harmonic potential were separated by 0.1 – 0.2 Å between adjacent windows. The force constants for ζ_R were chosen to be in the range of 1500 – 3000 $kcal \cdot mol^{-1}$, depending on the windows. Considering that the denominator of ζ_R, $|\vec{r}_{CEC} - \vec{r}_{113}| + |\vec{r}_{CEC} - \vec{r}_{203}|$, varies from 7 – 8 Å, the effective force constants acting on $|\vec{r}_{CEC} - \vec{r}_{113}|$ were 30 – 40 $kcal \cdot mol^{-1} \cdot \text{Å}^{-2}$. The initial

configurations were taken from the classical MD simulations, which were described in the SI, and both E113 and E203 were set to be deprotonated. A water molecule located between E113 and E203 was converted to a hydronium, by adding a hydrogen atom to water oxygen. Then, the system was further equilibrated for 1 ns using the multistate empirical valence bond (MS-EVB) model (48, 49) version 3.1 (29). The simulation was performed with the RAPTOR software (50), interfaced with the LAMMPS MD package (http://lammps.sandia.gov) (51), and the harmonic potential was controlled by the PLUMED package (44).

The initial configurations for each window in the QM/MM umbrella sampling were taken from the final configurations of the 1 ns MS-EVB equilibration. The QM atoms were treated with the density functional theory (DFT) BLYP-D3/TZV2P level of theory (52-54) including the Grimme dispersion correction (55). The QM region included Cl^-_{cen} if present, the water molecules in the central region, and the side chains of residues E113, E148, E203, Y445, and S107 for the WT system. Y445 and S107 were included in the QM region because their side chain OH groups are coordinated to Cl^-_{cen}. The waters within two solvation shells of E113 and E203 were also included in the QM region. The QM box size was set to be 20 – 30 Å in each dimension to ensure it was 6-8 Å larger than the size of the QM atoms. A quadratic confining wall potential was used to restrain the QM atoms within the QM box, with the wall skin thickness being 1 Å × 1 Å × 2 Å. The QM box size was chosen such that the quadratic wall potential did not affect the QM atoms for the vast majority of the simulation. The Gaussian Expansion of the Electrostatic Potential (GEEP) scheme was used to treat the QM/MM electrostatic coupling with periodic boundary conditions (PBCs) (56, 57), and the spurious QM/QM periodic image interactions were decoupled as described in ref (58). The chemical bonds that crossed the QM/MM boundary were capped with hydrogen atoms, the forces on which were calculated following the IMOMM scheme with a scaling factor of 1.50

(59). Otherwise, the QM setup was consistent with previous QM/MM umbrella sampling of PT in the influenza A M2 channel (60). The temperature was set to 300 K and controlled by a Nose-Hoover thermostat (42). The QM/MM umbrella sampling was performed with the CP2K software package (61). All windows were equilibrated for 5 ps, followed by 20 ps production runs. The value of ζ_R was collected every time step (0.5 fs). The PMF was constructed with the WHAM method (62).

The MS-RMD model configuration

The MS-RMD models were constructed for Glu_{in} and Glu_{ex} in WT ClC, and for K203 and Glu_{ex} in the E203K mutant. Detailed procedures of the MS-RMD force-matching of QM/MM data have been described in a previous studies, where the models were developed for describing the deprotonation reaction of Glu and Asp in aqueous solution (29) as well as in protein environments (30). The model parameters from that work were refit in this study to the QM/MM data calculated in protein environment. There were 2 different system setups in total for the WT, as well as depending on the presence of Cl^-_{cen}, 4 setups for E203K mutant, depending on the presence of Cl^-_{cen} and the protonation state of E202. The E202 residue was set to be deprotonated in the WT. For each system setup, two sets of umbrella sampling were performed from either E148 or E203 (K203 in E203K) to sample configurations for generating QM/MM force data. The harmonic potential with a force constant of 20 $kcal \cdot mol^{-1} \cdot \text{Å}^{-2}$ was applied to the CV, $|\vec{r}_{CEC} - \vec{r}_X|$. The "*X*" in the previous expression is either E148 or E203 (or K203) for each set. The vector $\vec{r}_{203}$ in E203K was defined as the center of mass of the nitrogen atom in the terminal zeta position (NZ in the PDB) and the epsilon carbon atom (CE in the PDB) in Lys. For each set, the centers of the windows ranged from 2.0 to 4.0 Å, and were separated by 0.25 Å between adjacent windows. ~1000

configurations were evenly sampled for the CV in all range. Each configuration was selected with a 2 ps interval. For each configuration, a single point energy calculation was performed with the CP2K software to calculate the QM/MM forces on the atoms in the QM region. To avoid the boundary issues, more waters were included in the QM region, which were selected within the third solvation shell from E148 and E203 (or K203). The side chains of S107, E113, E148, E202, E203 (or K203), and Y445 were included in the QM region. Other simulation details were kept as the same with the QM/MM umbrella sampling. The MS-RMD force field was fit to the QM/MM reference forces by using the previously described force-fitting algorithm (29). Parameters for E148 and E203 in WT, and E148 and K203 in the E203K mutant are given in Table S2.

The MS-RMD umbrella sampling

Four sets of the 2D PMFs were calculated in the WT, two for each of the Cl^-_{cen} present and absent cases, and two for each of two different sets of CVs. The two sets of CVs were defined as follows. In the first set, CV1 is the difference in the two distances from the CEC to the center of mass of the carboxyl group of either of E148 or E203, $|\vec{r}_{CEC} - \vec{r}_{203}| - |\vec{r}_{CEC} - \vec{r}_{148}|$, and CV2 is the distance between the center of mass of the carboxyl groups of two residues, $|\vec{r}_{148} - \vec{r}_{203}|$. The vector $\vec{r}_{CEC}$ in the MS-RMD umbrella sampling was defined by Eq. 1, using all water hydrogens for $\vec{r}^{H_i}$, and water oxygens and glutamic acid carboxyl oxygens for $\vec{r}^{X_j}$. The collective variable CV1 represents the migration of the excess proton from E203 to E148. The collective variable CV2 represents the rotation of E148 and E203, where it decreases as the side chains of the two residues rotate and face toward each other. In the second set of 2D PMFs, CV1 is the distance of

the CEC along the pathway sampled from metadynamics, (described below) which represents PT, and CV2 is the water density, N_w, for hydration of the central region.

The CV for water density (the effective number of water molecules in the central region) N_w is defined as (63)

$$N_w = \sum_{i}^{N_{H_2O}} N_i \tag{4}$$

where N_{H_2O} is total number of the water molecules and N_w is the occupancy of the *i*th water molecule in a defined rectangular box, given by

$$N_i = \prod_{\alpha}^{x,y,z} \left(\frac{1-R_{i,\alpha}{}^{6}}{1-R_{i,\alpha}{}^{12}} \right) \tag{5}$$

and $R_{i,\alpha}$ is defined as,

$$R_{i,\alpha} = \begin{cases} 0 & \left(\left| r_{i,\alpha} - r_{0,\alpha} \right| \le b_\alpha \right) \\ \dfrac{\left| r_{i,\alpha} - r_{0,\alpha} \right| - b_\alpha}{d} & \left(\left| r_{i,\alpha} - r_{0,\alpha} \right| > b_\alpha \right) \end{cases} \tag{6}$$

where r_0 is the position of the center of the box, which is chosen to be the midpoint between the center of mass of each of the carboxyl groups of E148 and E203, and r_i are the position of the oxygen atom of *i*th water molecule. The central axis of the box is aligned with the vector connecting the center of mass of the carboxyl group of E148 to that of E203. The distance *d* was chosen to be 5 Å. The parameters b_x, b_y, and b_z were set to be 3, 3, and 6 Å in the system for Cl^-_{cen} present, and 2, 2, and 5 Å in the system with Cl^-_{cen} absent, respectively. The size of the box is bigger in the case of Cl^-_{cen} present, because the side chain of E148 rotates up due to the hindrance

by Cl^-_{cen} located between E148 and Y445 OH group, and the distance between E148 and E203 becomes longer. The boundary of the box was determined to avoid counting the waters outside the central region.

Metadynamics (MetaD) (64) simulations were used to determine the curvilinear CEC pathway (65), where CV1 in the second set of PMF calculations is defined as the distance along the pathway. The initial configurations for the MetaD simulations were taken from the classical MD simulations, where a continuous hydrogen bond network formed in each system setup. The protonation states of the residues in the central region in the initial configurations are E148(0)E203(H) in the WT, and E148(0)K203(H) in E203K mutant. Note that "0" in the parenthesis denotes the side chain is in the deprotonated state, and "H" for the protonated state. A Gaussian MetaD potential was added for the z coordinate of the CEC every 1 ps. The Gaussian height was set to be 0.2 kcal/mol. The wall potentials (half-sided harmonic potential) were added to the boundary of the central region, acting on the CEC when its z coordinate is located 2 Å above from the center of mass of the side chain of E148, or 2 Å below from E203, in order to prevent sampling the CEC outside of the central region. The MetaD was performed for each system setup, until the CEC moves between E148 and E203 (or K203) two to three times. The coordinate of the CEC was recorded every time step. The *x* and the *y* coordinates of the CEC were averaged with each 0.25 Å interval in the *z* coordinate. The CEC pathway (CV1 in the second set) is defined as a 3D curvilinear path connecting the averaged points. In the umbrella sampling, the harmonic potential was applied to the tangent vector defined at the window center on the CEC pathway.

The 2D PMFs from the MS-RMD simulations were calculated by umbrella sampling. The force constants of the harmonic potentials in the umbrella sampling were set to be 30 and 5 $kcal \cdot mol^{-1} \cdot Å^{-2}$ for CV1 and CV2 of the first set, and 15 $kcal \cdot mol^{-1} \cdot Å^{-2}$ and 15 $kcal \cdot mol^{-1}$ for

CV1 and CV2 of the second set, respectively. All windows had an equilibrium time of 100 ps after the harmonic potential is applied, and production sampling of ~1 ns. For the E203K mutant, the one dimensional PMF was calculated for the proton pathway sampled from the MetaD for each system setup. The PMF was constructed by the WHAM method. Statistical errors in the PMFs were estimated using the block average method by dividing each trajectory into four consecutive blocks. The average error in the PMF calculated in this way is about 0.8 $kcal \cdot mol^{-1}$.

Committor analysis (66) was performed to check the appropriateness of the chosen CVs as reaction coordinates. A total of 10 configurations were captured from the windows at the top of the free energy barrier in the PMFs. Each configuration was spaced by 20 ps. The velocities of all atoms were randomized again for 300 K, following the Maxwell-Boltzmann distribution, and the bias potential from the umbrella sampling was released. The unconstrained committor simulations were performed from each configuration, until the system reached either the E148(0)E203(H) or E148(H)E203(0) state.

Proton transport rate constants

The minimum free energy path in the 2D PMF was calculated by using the string method (67). The rate for PT from E203 to E148 was then estimated by using transition state theory between the two states as follows (68)

$$k_{rxn} = \frac{\omega_0}{2\pi} \exp\left(-\frac{\Delta F^{\ddagger}}{k_B T} \right) \tag{7}$$

where k_B is Boltzmann's constant, T is the temperature (300 K), and $\Delta F^{\ddagger}$ is the free energy barrier height along the minimum free energy path. The fundamental frequency ω_0 is that of the reactant state oscillations around its minimum, which is defined as

$$\omega_0 = \sqrt{\frac{\partial^2 PMF(r)/\partial r^2}{m_{eff}}} \quad (8)$$

where r is the local minimum in the PMFs for the protonated E203 (or K203). The effective mass of the excess proton CEC, m_{eff}, was determined using the equipartition theorem, $m_{eff}\langle v^2\rangle = k_B T/2$, where the value of $\langle v^2\rangle$ was calculated from the MS-RMD trajectory sampled at r. When calculating $\langle v^2\rangle$ in the 2D PMFs, v is defined as $\delta\xi/\delta t$, where $\delta\xi$ is the displacement of the trajectory (CV1, CV2), projected onto the tangent vector of the lowest free energy path at r, and δt is the time step in the trajectory.

2.3 Results and Discussions

Proton transport from E113 to Glu_{in} (E203)

The side chains of E113 and E203 are close together in the crystal structure, with their respective side-chain O atoms only 2.5 Å apart. Faraldo-Gómez and Roux calculated the pK_a of these residues using the Poisson-Boltzmann equation (32). Their results suggested that in the crystal structure (PDB 1OTS) E113 is protonated, but shares a hydrogen bond with E203 such that the two residues share the excess proton. Additionally, experiments on the E113Q mutant provide indirect evidence of a coupling between the protonation states of E113 and E203, as the H^+ permeation rate of this mutant is 7 times lower than that of the WT (27). This presumably occurs because the electrically neutral Q113 suppresses the proton affinity of E203 (34). Importantly, the possibility that the excess proton is shared between E113 and E203 has not been explored in

previous studies using non-reactive classical MD (24, 26, 69, 70) since the residues were simulated in fixed protonation states.

In order to assess the preferred location of the excess proton prior to PT from Glu_{in} to Glu_{ex}, we calculated the PMF for the migration of the excess proton from E113 to E203 with DFT-based QM/MM umbrella sampling. The CV used in this PMF is defined in Eq. 3. As shown in Fig. 2, a shared excess proton between E113 and E203 is not the dominant state; it is ~5 kcal/mol less stable than either protonated E113 ($\zeta_R = 0.15$) or protonated E203 ($\zeta_R = 0.85$), regardless of the presence or absence of $Cl^-_{cen.}$. The PMF also shows that PT from E113 to E203 is slightly favored (by about 1-2 kcal/mol). Given their similar proton affinities and the low barrier for PT between the two residues, it is possible that E113 plays a direct role in proton uptake (e.g., delivering the proton to E203 in the WT protein or in place of it in E203 mutants). Figure S1 shows representative structures for when the excess proton resides on E113 or E203, and compares them with the crystal structure. When E203 is deprotonated, it forms a salt bridge with the positively charged R28 of the other monomer. This interaction is present in the crystal structure, indicating that E203 is likely deprotonated, which is consistent with previous calculations (32). As E203 is protonated, R28 breaks its salt bridge with E203 and moves closer to E113. The movement of R28 due to the protonation of E203 also has been reported in other computational work (69). These combined results suggest that ClC-ec1 evolves from E113(H)E203(0), as captured in the crystal structure, to E113(0)E203(H) prior to PT from E203 to E148. Thus, the proton was localized on E203, as opposed to being shared between E113 and E203, in the starting structure for our analysis of PT between E203 and E148.

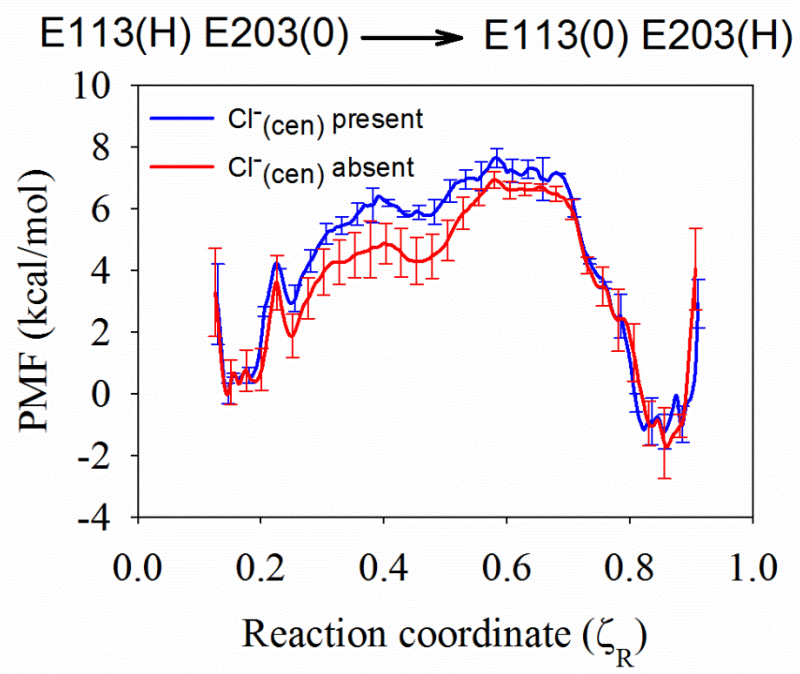


Figure 2.2. The PMF for PT from E113 to E203, calculated with the QM/MM umbrella sampling, when Cl^-_{cen} is present (blue) and absent (red).

Hydration of the central region

The hydration state of the central region has been previously examined in a series of computational studies. Kuang and colleagues (31) predicted positions of three water molecules using the DOWSER program (71), which places waters in empty cavities of static structures given a favorable interaction energy between the water and the surrounding protein. E203 and E148 were not directly connected by the hydrogen bonded network of these waters, but through the hydroxyl group of Y445 and Cl^-_{cen}. However, experiments demonstrated that proton permeation was conserved in Y445F mutant, suggesting that proton permeation does not necessitate the hydroxyl group (72). In a later study (73), four water positions connecting E203 and E148 were predicted by DOWSER, though their stability was never validated with MD simulations.

Several studies have used MD to evaluate the hydration of the central region, though none have focused on the protonation state most relevant to PT that is studied herein - namely E113(0)E203(H)E148(0). In our previous work, an MD simulation of E203(0)E148(0) demonstrated that some number of waters remained stable for 1.4 – 3.7 ns (24). Ko and Jo (69) simulated E203(H)E148(H) and found that water spontaneously entered the central region from the intracellular bulk when Cl^-_{cen} was absent and remained stable for ~ 8 ns. Han et al. (26) observed the formation of transient water wire in the central region in their MD simulations, but for the E203(0)E148(0) state with Cl^-_{cen} present.

Herein, classical MD was performed for the protonation states most relevant to PT: E113(0)E203(H)E148(0) in WT and E113(0)K203(H)E148(0) in the E203K mutant. We found that a continuous hydrogen bonded network can form in both the presence and absence of Cl^-_{cen}, but that it forms infrequently and its stability is transient (Fig. S2). As discussed below, hydration of the central region is thermodynamically uphill when E203 (or K203) is protonated and E148 is deprotonated, which is consistent with the rare hydration observed in previous MD simulations (26, 69). However, as discussed below (in 'Coupling to Hydration Changes') water molecules in the central region are stabilized by the entry of the excess proton to the first solvation shell above E148. Thus, the hydration stability changes significantly as the excess proton enters this region because the two processes are coupled.

Proton transport from Glu_{in} (E203) to Glu_{ex} (E148)

Mechanistic Predictions from Experiment The mechanism of the PT from Glu_{in} to Glu_{ex} has been hypothesized based on experimental findings. MacKinnon and co-workers (28) observed that the crystal structures from *E. coli* (ClC-ec1) and its eukaryotic homolog (cmClC) are different around

the central region. They hypothesized that the two conformations capture different intermediate states, and proposed a mechanism in which Glu_{ex} moves from S_{ext} to S_{cen} in association with the movement of Cl^- ions from S_{cen} and S_{int} to the intracellular side. Thereafter, Glu_{ex} (still at S_{cen}) receives a proton from Glu_{in}. Hence, PT from Glu_{in} to Glu_{ex} occurs with Cl^-_{cen} absent. A later study (35) showed that, although the E148A mutant is itself not H^+ conducting, its proton conductivity recovers upon addition of a carboxylate-containing substrate as an analog for E148, even in the absence of Cl^- ions. In an E148A crystal structure containing this substrate, the substrate's carboxyl group is bound to S_{cen}, displacing Cl^-_{cen}. These findings further suggest that PT in the central region from Glu_{in} to Glu_{ex} can occur in the absence of Cl^-_{cen}.

Nguitragool and Miller (13) proposed a different mechanism wherein PT from Glu_{in} to Glu_{ex} occurs with Cl^-_{cen} present and Glu_{ex} in position between S_{ext} and S_{cen}. They showed that the Cl^-/H^+ coupling of the WT and various Y445 mutants are correlated with the presence of an anion at S_{cen} in the crystal structures (72). When Br^-, which is a substitute for Cl^- in the preparation of the crystal structure, is replaced by $SeCN^-$, the protein could transport $SeCN^-$, but H^+ permeation was blocked. The corresponding crystal structures displayed no $SeCN^-$ density at S_{cen}, suggesting the importance of an anion in this position for H^+ permeation (74). It is important to recognize, however, that the observed coupling between H^+ transport and an anion at S_{cen} could occur somewhere other than the PT between Glu_{in} and Glu_{ex} (e.g., during proton uptake or release). Indeed, Miller and co-workers have also suggested that it is proton uptake that is rate limiting for proton permeation (34). Nevertheless, these two proposed mechanisms (13, 35) differ in whether Cl^-_{cen} is present during PT from Glu_{in} to Glu_{ex}. Interestingly, our MS-RMD simulations reported below show that PT can happen in either the presence or absence of Cl^-_{cen}, although the former has a kinetically faster rate.

MS-RMD Simulation Results To examine the Cl^-/H^+ coupling using a computational approach, Zhang and Voth (25) calculated the PMF for PT from Glu_{in} to Glu_{ex} using an older MS-EVB model, and showed that the free energy barrier for PT is decreased by Cl^-_{cen}. In the present work, the free energy profile for PT from Glu_{in} to Glu_{ex} is recalculated with more realistic initial configurations and more accurate and extensive MS-RMD simulations, such that the coupling between PT and the conformational changes of Glu_{in} and Glu_{ex} as well as the hydration of the central region, can be explicitly characterized. The new initial configurations include the full dimer of the protein in the membrane, explicit waters, ions, and an equilibrium distribution of hydrogen bonded water networks in the central region. The MS-RMD model (50, 75) was developed based on the MS-EVB framework, which has been employed to model proton dynamics in aqueous systems with high efficiency and accuracy (48, 49) and fit to high level QM/MM data (see Methods). With this MS-RMD approach, mechanistically revealing 2D PMFs using two different sets of CVs (described in Methods) could be calculated both in the presence and the absence of Cl^-_{cen}. In turn, the minimum free energy path along the 2D PMFs could be obtained with the string method, which finds the most likely collective path (and the lowest free energy barrier) between the protonated Glu_{in} and Glu_{ex} minima in the 2D PMF.

In addition, the committor probability (66) could be calculated in order to verify that the chosen CVs are representative of the correct reaction coordinate for PT from Glu_{in} to Glu_{ex}. As described earlier in Methods, by starting from the configurations captured at the top of the free energy barrier in each PMF, unconstrained simulations were initiated with randomized velocities. Demonstrating that half of these trajectories proceed to reactant (Glu_{in} protonated) and half to product (Glu_{ex} protonated) (Table S1) verifies that the free energy barriers in the presented PMFs represent a good

transition state of the reaction. This analysis is often overlooked and is critical because relative distance coordinates (e.g. Eq. 3 for PT from Glu_{in} to Glu_{ex}) can force the system to sample a more limited configuration space that is not reflective of the unconstrained reaction.

Coupling to Conformational Changes Figures 3A and 4A show the coupling between PT from Glu_{in} to Glu_{ex} and the inter-residue distance. When Cl^-_{cen} is present this inter-residue distance (CV1) does not change significantly, primarily because Cl^-_{cen} blocks the downward rotation of E148. However, when Cl^-_{cen} is absent the distance between the donor (E203) and acceptor (E148) changes significantly as the proton is transferred; similar to a heavy-light-heavy reaction, the inter-residue distance (CV1, the horizontal axis coordinate in Fig. 4A) decreases in the transition region when the excess proton is in between the residues, but lengthens in the reactant and product minima when either residue is protonated. The 1D PMF along the minimum free energy path in each 2D PMF in the bottom of Figs. 3 and 4 show that PT from Glu_{in} to Glu_{ex} is downhill process, in both the presence and the absence of Cl^-_{cen}. This may seem to contradict previous calculations of the pKa value of E148 in the crystal structure (32). However, the solvation environment that enables protonation of E148 changes significantly during the PT process (further described below) from that found in the crystal structure. Figure 5 shows representative structures for the excess proton on E203 (R for 'reactant'), in between E203 and E148 (T for 'transition'), and on E148 (P for 'product'). They show how the downward rotation of E148 is hindered by Cl^-_{cen} (Fig 5T top), but possible when Cl^-_{cen} is absent, such that E148 rotates down and the E203 and E148 side chains face each other (Fig. 5T bottom). This downward orientation is also stable for state 'P', when E148 is protonated. We do not observe a stable, direct interaction between the excess proton and Cl^- (forming HCl), as suggested in a previous experimental study (74). Instead the proton passes

through water molecules. Since the protonation of Cl^- was not included in our MS-RMD model, this conclusion was verified by the same finding in our DFT-level QM/MM PMF of same system (30), where QM atoms in the central region, including Cl^-_{cen}, are all reactive. Nevertheless, the barrier for PT is decreased by ~ 4 kcal/mol in the presence of Cl^-_{cen}, indicating that Cl^-_{cen} does, as expected, facilitate the deprotonation of E203 by stabilizing the excess proton in the central region.

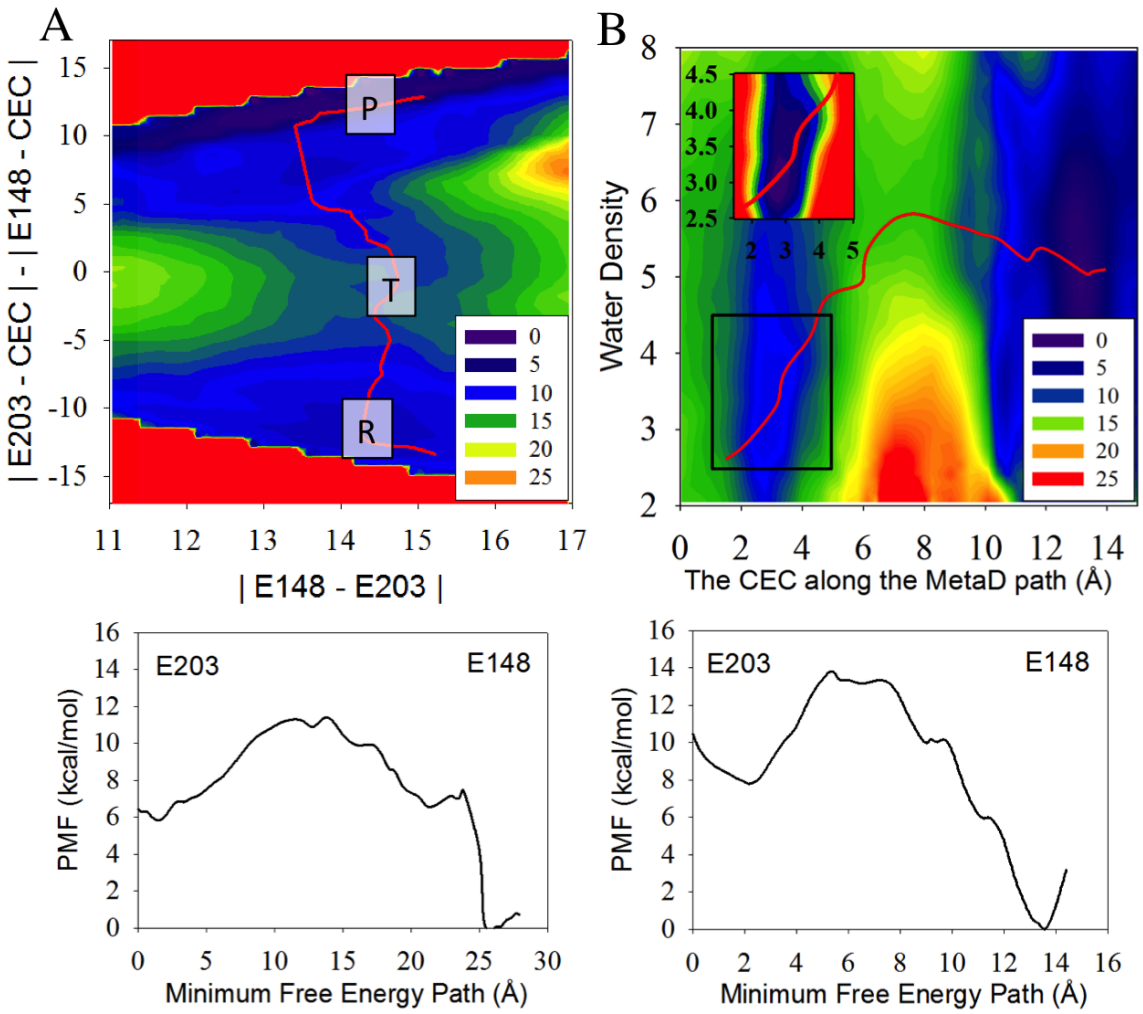


Figure 2.3. The 2D PMFs of WT, with Cl^-_{cen} present, calculated with MS-RMD. For panel A, CV1 (the horizontal axis) is defined as the distance between E148 and E203, and CV2 (the vertical axis) is the difference in the two distances from the excess proton CEC to either E148 or E203. For panel B, CV1 is the distance of the CEC along the proton pathway sampled by

the MetaD, and CV2 is the water density in the central region. The red line on each 2D PMF represents the minimum free energy path for the transition from E203(H) to E148(H). The plot below each 2D PMF is the PMF calculated along that minimum free energy path. The boxes "R", "T", and "P" in panel A represent the positions of the reactant, transition state, and product state, respectively. The inset in panel B highlights the free energy in the reactant minimum more clearly with a finer color scale of purple through red corresponding to the PMF range of 7.5 – 10.5 kcal/mol.

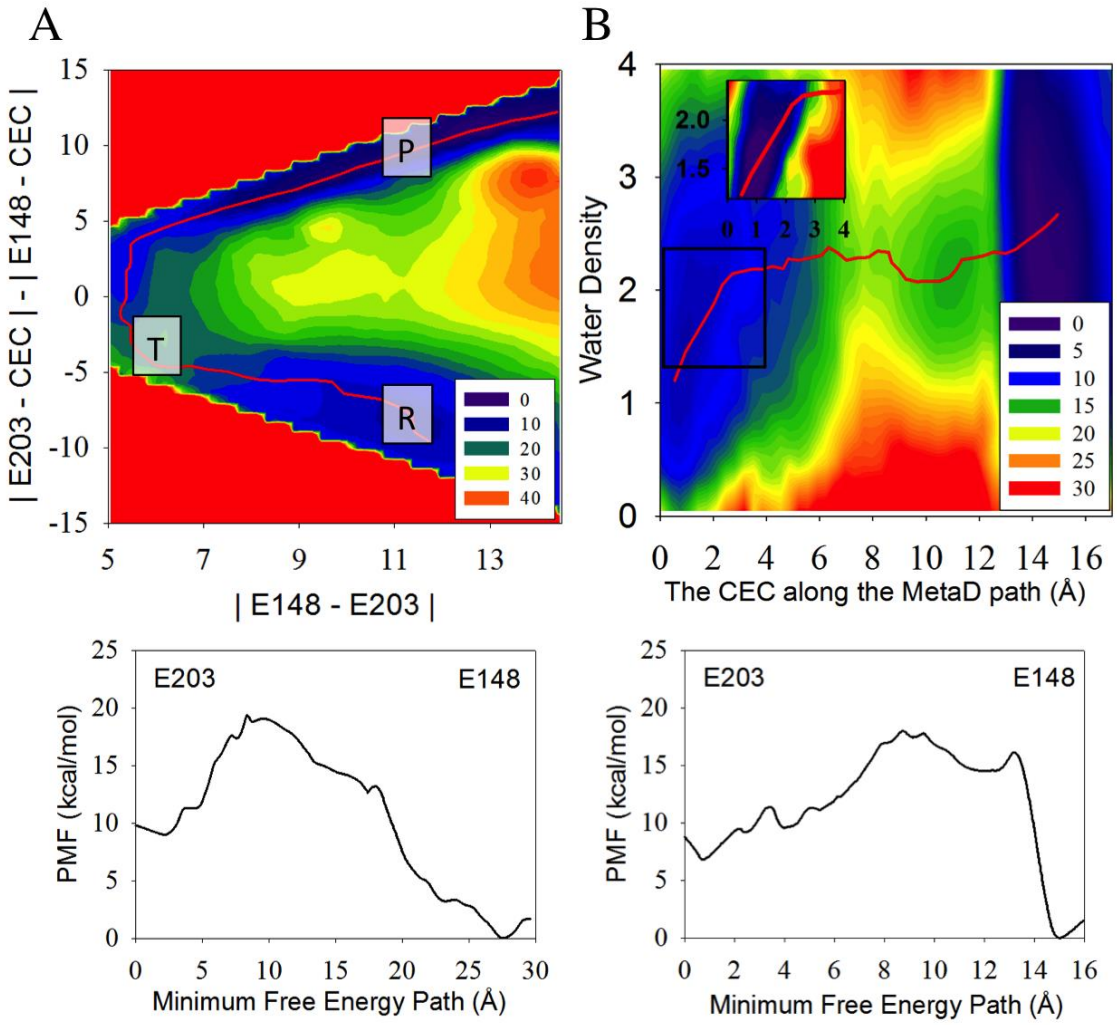


Figure 2.4. The 2D PMFs of WT, with Cl^-_{cen} absent and with all other quantities as otherwise described in the caption for Fig. 3. Note the higher water occupancy and degree of dynamical coupling to the proton CEC translocation evident from panel B in this case of the Cl^-_{cen} present.

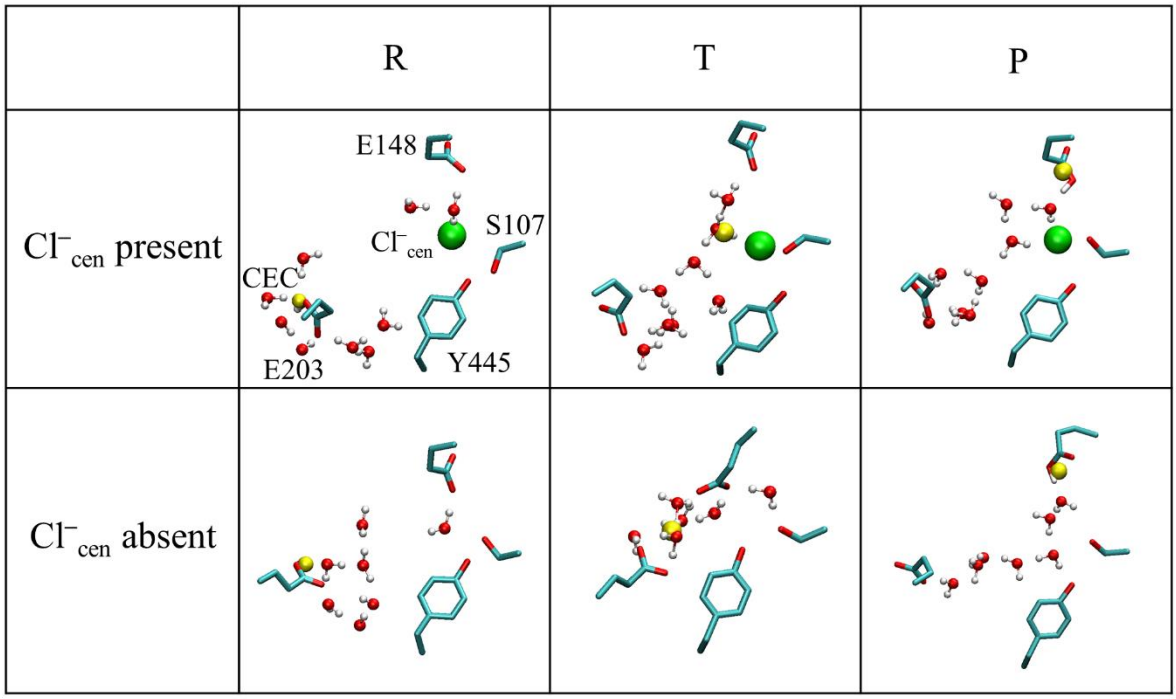


Figure 2.5. Representative configurations of the central region for each of the boxes R, T, and P in the 2D PMFs in Figs. 3A and 4A. (Top) All configurations from 1 ns of MD simulation are first aligned based on alpha carbons more than 15 Å from the central region. Then intact residues and water molecules are overlaid over the average positions of the highlighted protein and water molecules. The positions of the intact residues and are averaged over the trajectories, and the water molecules are taken from the last MD frame of each trajectory. (Bottom) The excess proton CEC is represented by a yellow sphere.

Coupling to Hydration Changes The 2D PMFs in Figures 3B and 4B show that the hydration of the central region is coupled with the migration of the excess proton from Glu_{in} to Glu_{ex}. In the reactant state (with the proton on Glu_{in}) water molecules can transiently penetrate the central region, but there is an energetic cost to this process (Figs 3B and 4B), making it infrequent (see Fig S1). However, once the excess proton enters the first solvation shell of Glu_{in} more water molecules are pulled in from the intracellular bulk. Thus, the hydration of the central region is stabilized as the proton migrates to E148. Importantly, the presence of a (meta-)stable full water

chain between Glu_{in} to Glu_{ex} before the proton transfer event of Glu_{in} is *not* a prerequisite for proton transport. Thus, we find an interesting phenomenon - the dynamic hydration of a hydrophobic space induced by an incoming excess proton, which was recently predicted in an idealized carbon nanotube system (63), but is here, for the first time, observed in a biomolecular system. The coupling between the excess proton motion and the hydration of this region is shown more clearly when Cl^-_{cen} is present (Fig. 3 B), due to the increased size of the central region accommodating more water.

Calculated Rate Constants The rate constants of PT from E203 to E148 in the presence and absence of Cl^-_{cen} were calculated with transition state theory using the minimum free energy paths in Fig 3 and 4 and reported in Table 1. The PT rate constant is roughly 3 orders of magnitude greater with Cl^-_{cen} present. However, in both cases the calculated rate constants are on the order of $0.7 - 380\ \mu s^{-1}$, which is 3–4 orders of magnitude greater rate than the experimentally measured values for the H^+ permeation (~$0.2 - 1\ ms^{-1}$depending on the Cl concentration gradient and for pH values between 4 and 7) (9, 34). It is critical to note, however, that these experimental values correspond to the *complete* permeation process, which not only includes the migration of an excess proton from one side of membrane to the other side, but also the associated Cl^- migration. Thus, our results suggest that PT through the central region is facilitated by Cl^-_{cen}, but that it is kinetically possible in the presence or absence of Cl^-_{cen} (i.e., it is not rate limiting for H^+ permeation in either case). This has indeed been proposed in previous experimental work (34). These results clearly suggest that the requisite Cl^-/H^+ coupling occurs during proton uptake to, or release from, E203 or E148. This is consistent with an early prediction from Yin et al. (36) who used continuum pK_a calculations to show that the pKa of E148 is quite low in the absence of nearby anions.

	Cl^-_{cen}	PMF figure	k (μs^{-1})	$\Delta F^{\ddagger}$ (kcal/mol)
WT	present	3 A	$(3.8 \pm 1.8) \times 10^{2}$	5.1
		3 B	$(6.4 \pm 1.5) \times 10^{2}$	5.9
	absent	4 A	0.71 ± 0.25	10.2
		4 B	0.24 ± 0.09	10.9
E203K	present	6 A	$(2.3 \pm 0.3) \times 10^{-2}$	11.4
		6 B	1.2 ± 0.2	9.4
	absent	6 C	$(7.0 \pm 0.7) \times 10^{-9}$	20.4
		6 D	$(1.9 \pm 0.9) \times 10^{-2}$	12.0

Table 2.1 The rate constants for PT in the WT ClC-ec1 and E203K mutant. The rate constants were calculated with transition state theory. Error in the rate constant was estimated using the block average method by calculating the rate constant in four consecutive blocks in the trajectories for each window. The experimental value for proton permeation rate in the WT protein is $(1.0 \pm 0.1) \times 10^{-3}$ ms^{-1}, (34) while that for E203K is ~ 3 times slower (27).

Proton transport in the E203K mutant

Glu_{in} and Glu_{ex} are deemed essential to PT because replacement of either residue with a nonprotonatable side chain eliminates H^+ transport, but retains Cl^- transport (16). However, the chemical identity of Glu_{in} (E203) is more flexible than one might guess. Most protonatable side chains (D, H, K, and R) retain Cl^-/H^+ exchange while others (C and Y) block H^+ transport and significantly inhibit Cl^- flow (27).

To explore how the replacement of E203 with a basic residue retains Cl^-/H^+ exchange capacity, we calculated the PMF for PT both in the presence and the absence of Cl^-_{cen} in the E203K mutant. The curvilinear pathway for PT from K203 to Glu_{ex} was determined with MetaD and the PMFs were calculated with umbrella sampling and force integration of the excess proton CEC along the

determined pathways. The resulting PMFs are shown in Fig. 6. The free energy barrier for PT in the WT is decreased in the presence of Cl^-_{cen}. However, for the mutant this decrease is more significant both in magnitude and consequence. In the absence of Cl^-_{cen} the calculated rate constant (Table 1) is six orders of magnitude lower than that of the WT, yet experimentally the rate for H^+ permeation is only seven times slower (27). Thus, Fig. 6 C indicates that H^+ permeation is not allowed in E203K in the absence of Cl^-_{cen}, which is consistent with the decreased coupling observed for the E203K mutant (27). (Note that decreased coupling means fewer H^+ ions are exchanged per Cl^- ion). Since the E203K mutant has the additional constraint to proton permeation that Cl^-_{cen} must be present, it follows that fewer H^+ ions would be transferred from one side of the membrane to the other. This is not the case in the presence of the Cl^-_{cen} (Fig. 6 A) since PT is only ten times lower than for the WT, which is still two orders of magnitude greater than the mutant's lowest rate constant for H^+ permeation. Although PT from K203 to E148 is slightly uphill, the thermodynamic cost could be compensated by coupled downhill steps (e.g., proton release going with a proton gradient or Cl^- transport). Thus, excluding the participation of other residues (as discussed below), PT from K203 to Glu_{ex} is kinetically and thermodynamically possible, but only in the presence of Cl^-_{cen}. This may also speak to the overall mechanism of coupled ClC transport, suggesting that in WT the PT typically occurs with Cl^-_{cen} present, albeit with the PT not being the rate limiting step(s) in the transport cycle as stated earlier.

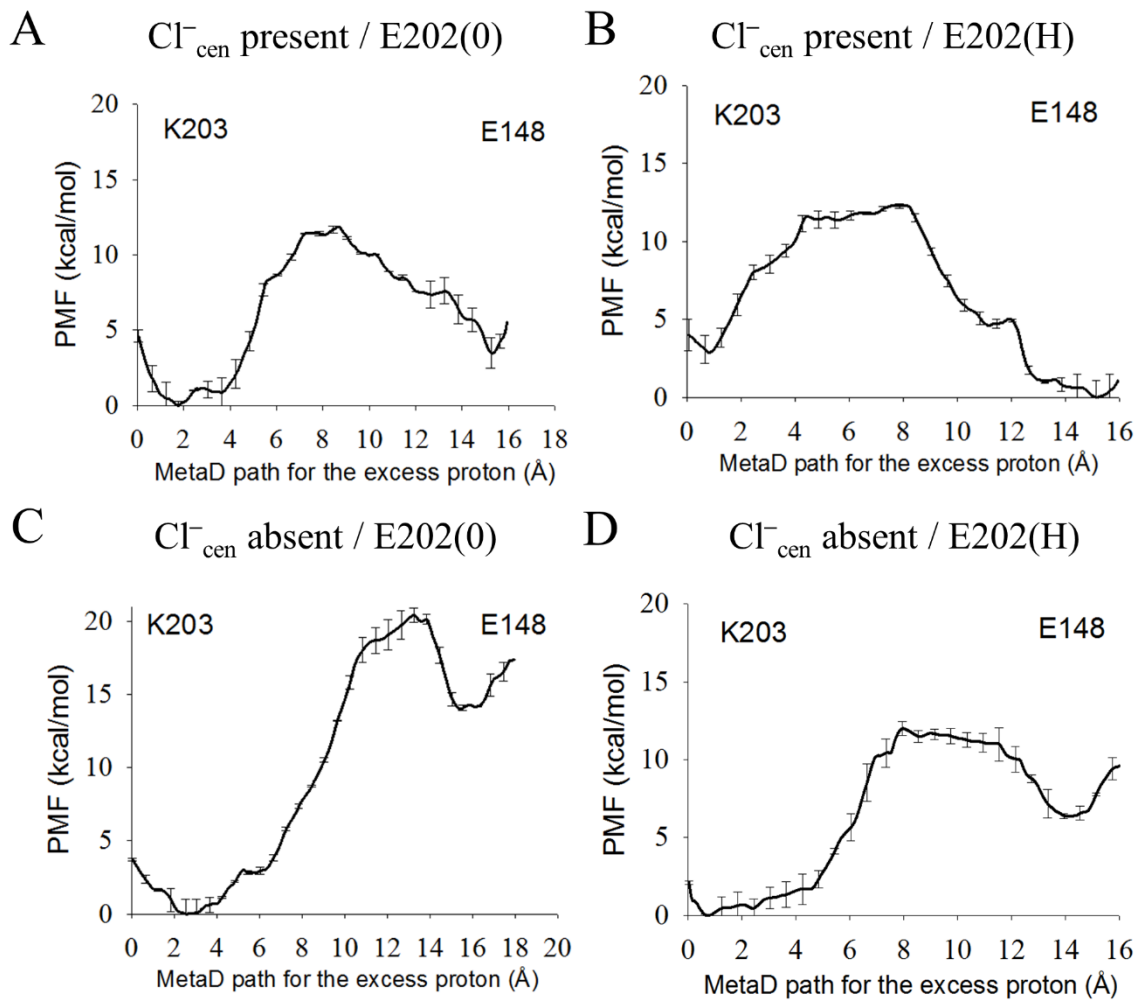


Figure 2.6. The PMFs for PT in E203K mutant with Cl^-_{cen} present (panels A and B) or absent (panels C and D), and with deprotonated E202 (panels A and C) or protonated E202 (panels B and D). The reaction CV is defined as the distance of the excess proton CEC along the proton pathway sampled by the MetaD.

One interesting possibility is that other residues may also play a role in the PT process in the E203K mutant. Following a long equilibration (see Methods), we found that the side chain of E202, which is deprotonated in the WT, reorients toward K203, decreasing the average distance between the E202 carboxyl group and the K203 zeta nitrogen (NZ in the PDB) to 4.5 Å (see Fig. S3). However, E202 is more likely to be protonated in E203K than in the WT, because a basic residue

(K203) is located nearby. When E202 is protonated, the average distance between the two residues is increased to 10.2 Å. In the WT protein E202 points away from the PT pathway regardless of its protonation state, such that the average distance between the carboxyl groups of E202 and E203 is 7.0 Å with E202 deprotonated and 6.5 Å with E202 protonated, respectively. Thus, the proton affinity of K203 is likely more influenced by the protonation state of E202 than E203 is in the WT protein.

To further explore the role of E202 in the E203K mutant, PMFs for the protonated E202 system were also calculated. The PMFs in Fig. 6 B and D show that the free energy barriers for PT are significantly reduced both in the presence and the absence of Cl^-_{cen}. The free energy of the final states (protonated Glu_{ex}, deprotonated K203) are also decreased, making the PT process thermodynamically favorable in the presence of Cl^-_{cen}. The computed rate constants for PT with protonated E202 are both around 10 μs^{-1} (Table 1), indicating that, as in the WT, PT through the central region would not be rate-determining for the overall PT process. Thus, it is tempting to conclude that E202 plays either a direct or indirect role in the central PT process in this mutant. However, these results do not seem to explain the 3-fold decreased rate of H^+ permeation for E203K (27), nor do our calculations account for the cost of protonating two residues (E202 and K203) in the ClC core.

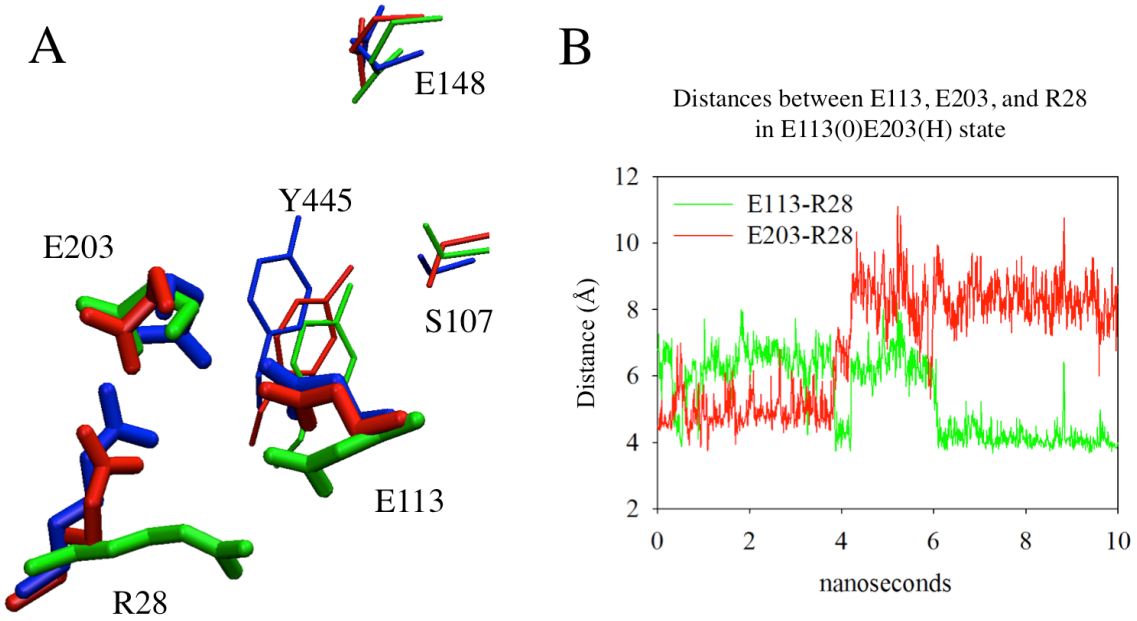


Figure 2.7 (A) Representative configurations of several residues located around the central region, when $\zeta_R = 0.15$ (colored in red) and $\zeta_R = 0.85$ (in green) sampled in the PMF for PT from E113 to E203 as shown in Fig. 2. The X-ray crystal structure (PDB 1OTS) is colored in blue. Corresponding protonation states for the configurations at $\zeta_R = 0.15$ and 0.85 are E113(H)E203(0) and E113(0)E203(H), respectively. E113, E203, and R28 are represented by thick stick, and S107, E148, and Y445 by thin sticks. (B) The distance between the center of mass (COM) of the carboxyl group of E113 and the zeta-nitrogen of R28 (green) and between E203 and R28 (red) in the 10 ns classical MD run. The initial configuration was obtained from the equilibrated structure in the last snapshot from previous classical MD run with the protonation states E113(H)E203(0), then the simulation was initiated after switching the protonation states from E113(H)E203(0) to E113(0)E203(H).

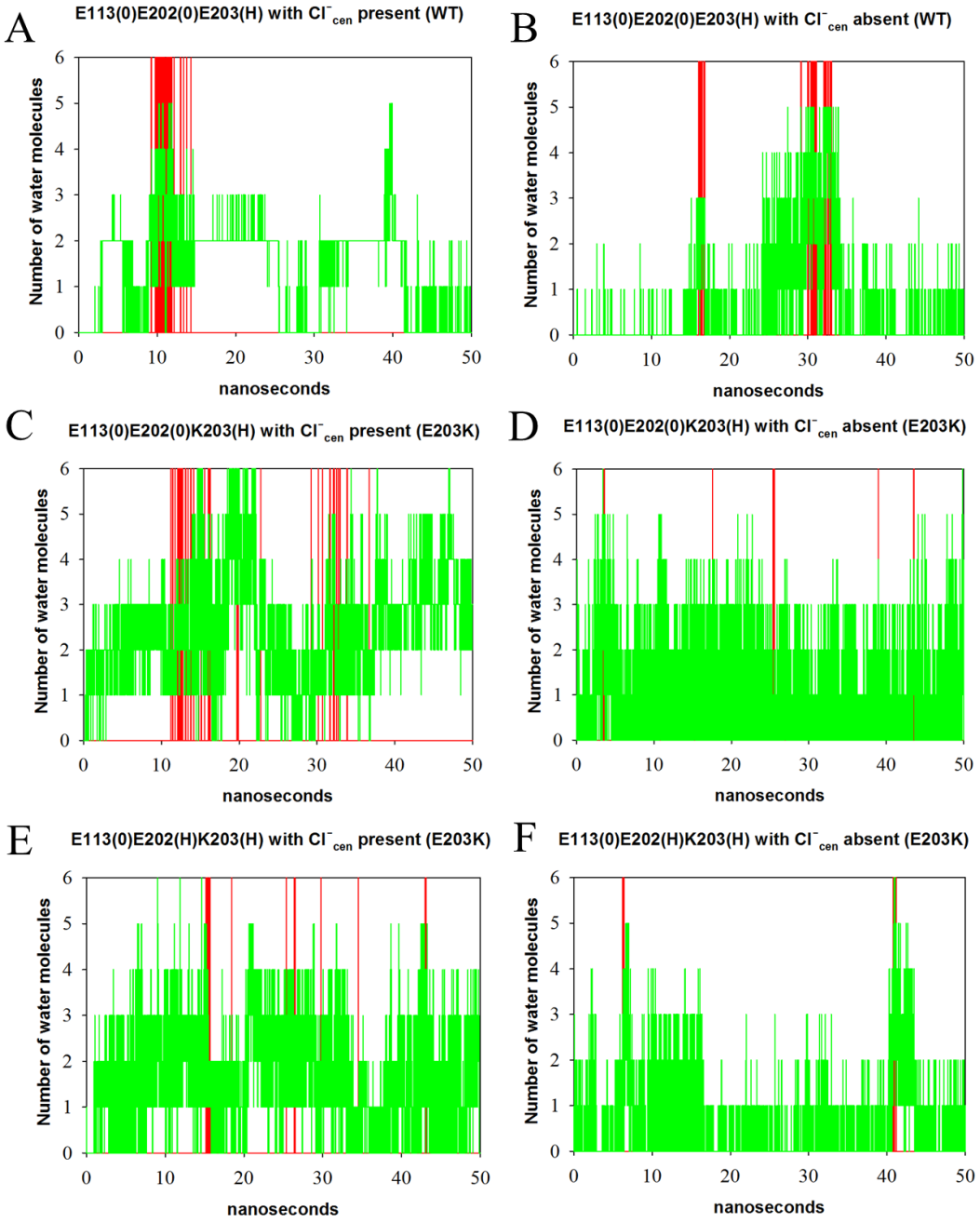


Figure 2.8 The classical MD results for the hydration of the central region. The simulations were performed for 6 different system setups in WT (panels A and B), and E203K mutant (panels C - F), depending on the protonation states of the residues around the central region and the presence of Cl^-_{cen}. The green line represents the number of water molecules in the region, and the red line represents the presence of a continuous hydrogen bond network connecting E148 and E203 (or K203 in K203 mutant). The central region is not fully hydrated most of time, and a formation of a continuous hydrogen bond network is a rare event in all cases.

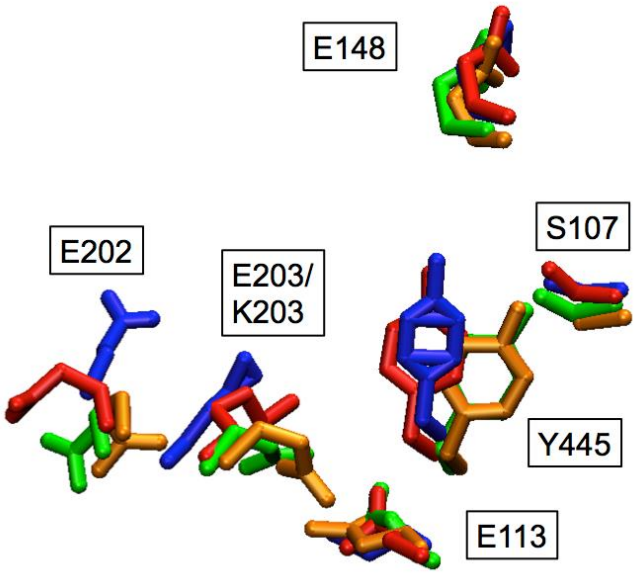


Figure 2.9 The rotation of E202 in the E203K mutant. The structure of the WT protein was obtained either with E202 deprotonated (green) or E202 protonated (orange), and the structure of E203K was obtained either with E202 deprotonated (blue) or E202 protonated (red), respectively. The structure was averaged from the MD simulations for each system. The alpha carbons located more than 15 Å from the central region were aligned at each MD frame, and the positions of all heavy atoms were averaged over the trajectory.

The number of trajectories from unconstrained simulations starting from the top of the free energy barrier in the PMF are shown in Table S1, depending on whether it reached either E148(H)E203(0) or E148(0)E203(H) state, respectively. The commitment probability P_B is calculated as the number of trajectories reaching E148(H)E203(0) state over the total number of trajectories. Assuming that all trajectories are statistically independent, the standard deviation in P_B can be calculated as $\sigma = \sqrt{P_B(1-P_B)/N}$, where N is the total number of trajectories. Although P_B was not exactly 0.5 in Cl^-_{cen} absent, the difference is smaller that the standard deviation σ, 0.16 at $P_B = 0.5$.

PMF	E148(H)E203(0)	E148(0)E203(H)	P_B
Cl^-_{cen} absent	4	6	0.4
Cl^-_{cen} present	5	5	0.5

Table 2.2. The committor analysis. The MS-RMD parameters for E148 and E203 in WT (A) and E148 and K203 in E203K mutant, when E202 is either deprotonated (B) or protonated (C), are shown in Table S2. The definitions of the parameters are described in our previous work (29, 30). The O-H bond stretches in the carboxyl (-COOH) group of glutamic acid residues are treated by standard Morse potentials (defined by D, α, and r_0). The N-H bonds in the ε-amino group (-NH3) group of lysine residues are treated by the original harmonic potentials in CHARMM force field.

Table 2.3. The MS-RMD model parameters.

A)

	WT			
	Cl^-_{cen} absent		Cl^-_{cen} present	
	E148	E203	E148	E203
B	0.063153	0.118175	0.012536	0.003282
b	1.571751	0.558680	0.232384	0.424087
b'	1.320947	1.311301	1.469007	0.480047
d^0_{OO}	2.4	2.4	2.4	2.4
C	0.363648	0.076533	0.014044	0.026472
c	1.117167	1.063753	1.152912	1.309021
d^0_{OH}	1.0	1.0	1.0	1.0
r_s^l	3.5	3.5	3.5	3.5
r_s^h	4.0	4.0	4.0	4.0
V_{ii}	-147.095673	-144.522565	-151.11473	-147.01392
c_1	-36.090543	-40.406344	-30.414842	-35.686750
c_2	1.879933	3.826462	3.331769	1.291340
c_3	1.193253	1.439519	1.422240	1.598104
D	143.003	143.003	143.003	143.003
α	1.8	1.8	1.8	1.8
r_0	0.975	0.975	0.975	0.975

(Table 2.3, continued)

B)

	E203K mutant (E202 deprotonated)			
	Cl^-_{cen} absent		Cl^-_{cen} present	
	E148	K203	E148	K203
B	0.475545	0.395946	0.368462	0.015901
b	0.057427	0.062900	0.033645	0.786306
b'	0.016288	0.006991	0.007927	1.083358
d^0_{OO}	2.4	2.4	2.4	2.4
C	0.385755	0.193260	0.449148	0.020434
c	1.884732	1.524637	1.754929	1.120238
d^0_{OH}	1.0	1.0	1.0	1.0
r_s^l	3.5	3.5	3.5	3.5
r_s^h	4.0	4.0	4.0	4.0
V_{ii}	-149.939246	-46.932166	-151.977503	-44.572462
c_1	-31.867844	-31.580329	-29.863163	-35.609380
c_2	2.535868	2.444616	2.361152	1.760981
c_3	1.568009	1.489707	1.487692	1.781791
D	143.003		143.003	
α	1.8		1.8	
r_0	0.975		0.975	

C)

	E203K mutant (E202 protonated)			
	Cl^-_{cen} absent		Cl^-_{cen} present	
	E148	K203	E148	K203
B	0.200350	0.008170	0.491293	0.006962
b	0.020911	1.486142	0.143266	0.553671
b'	1.343726	0.721547	0.004143	1.062345
d^0_{OO}	2.4	2.4	2.4	2.4
C	0.478908	0.011863	0.353012	0.015041
c	1.735206	0.344983	1.691584	1.837319
d^0_{OH}	1.0	1.0	1.0	1.0
r_s^l	3.5	3.5	3.5	3.5
r_s^h	4.0	4.0	4.0	4.0
V_{ii}	-152.990826	-45.646055	-153.443147	-36.700277

(Table 2.3, continued)

c_1	-28.703903	-32.586354	-25.418764	-34.914226
c_2	2.819715	2.989506	2.993419	1.143951
c_3	1.489503	1.427786	1.576890	1.573202
D	143.003		143.003	
α	1.8		1.8	
r_0	0.975		0.975	

2.4 Conclusions

We have performed extensive MS-RMD and QM/MM simulation studies of the PT process through the central region of the WT ClC-ec1 antiporter and its E203K mutant. MS-RMD was shown to offer an accurate and efficient way of sampling the reactive dynamics of an excess proton in this large biomolecular system while accounting for other slow dynamics, such as water movement in the central region, and by capturing longer time scales than are accessible to *ab initio* MD or QM/MM MD methods.

We first focused on the most likely starting protonation state for PT through the central region, which is an essential step in proton permeation through the protein. A free energy profile of PT between E113 and E203 shows that the proton more likely resides on E203 than being shared between E113 and E203 prior to transfer to E148. We then turned to the proper starting solvation state for the central region. Although our standard MD simulations suggested that water molecules can transiently enter the region between E203 to E148, their stability is limited. Thus, we turned to a combination of MS-RMD and QM/MM simulations to calculate 2D PMFs that describe not only the stability of hydration in the central region, but also its coupling to PT between E203 (Glu_{in}) and E148 (Glu_{ex}). These 2D free energy profiles show that the hydration of the central region is

indeed coupled with the proton translocation, such that waters enter this space as the excess proton migrates through it.

Using the Glu_{in} to Glu_{ex} inter-residue distance as an additional CV, we then calculated 2D free energy profiles to show that during PT from Glu_{in} to Glu_{ex} the presence of a Cl^- ion in the central site (Cl^-_{cen}) both blocks the downward rotation of Glu_{ex}, and reduces the free energy barrier for PT. However, the calculated rate constants for both the Cl^-_{cen} present and absent cases are significantly greater than the experimental rate of proton permeation, suggesting that PT through the central region is kinetically possible, and not rate limiting, in the presence or absence of Cl^-_{cen}. This suggests that the experimentally observed coupling between Cl^- and H^+ transport occurs outside of the central region. For additional comparisons to experiment, future simulations must account for PT in other regions (i.e., from the intracellular side to Glu_{in} and from Glu_{ex} to the extracellular side), and the kinetic coupling to Cl^- transport must be more explicitly determined. Attention should also be given to the potential role of the recently observed protein conformational changes (9-11). This work is in progress in our group. Nevertheless, the results presented herein provide an interesting new perspective on the story of coupling in the ClC antiporters; that PT from Glu_{in} to Glu_{ex} can occur in either the presence or absence of Cl^-_{cen} (albeit the fomer is faster), thus supporting the notion that the requisite coupling between Cl^- and H^+ transport occurs outside of the central region. Our results neither support nor negate the experimentally suggested mechanisms for this particular PT step in the central region, which differ in the presence of Cl^-_{cen} (13, 35).

Lastly, we explored the PT process in the E203K mutant, for which H^+ permeation remains intact experimentally despite the substitution of E203 with a basic residue. Our results suggest that PT from K203 to E148 is both kinetically and thermodynamically possible, but only when Cl^-_{cen} is present. Additionally, we find that E202 deviates from its position in the WT system, rotating

such that it interacts more closely with K203. Although the PMFs and rate calculations show that PT is most favorable when E202 is protonated, this calculation does not account for the cost of protonating two residues in the ClC-ec1 core. Once again, PT through the central region is facilitated by Cl^-_{cen} but not expected to be rate determining in E203K mutant. These findings add to our understanding of how the ClC-ec1 protein exchanges ions through partially overlapping pathways. Our future efforts will aim to add further insight into our mechanistic understanding of this important antiporter.

2.5 References

1. Chen TY (2005) Structure and function of ClC channels. *Annu. Rev. Physiol.* 67(1):809-839.

2. Estevez R & Jentsch TJ (2002) CLC chloride channels: correlating structure with function. *Curr. Opin. Struct. Biol.* 12(4):531-539.

3. Jentsch TJ (2008) CLC chloride channels and transporters: from genes to protein structure, pathology and physiology. *Crit. Rev. Biochem. Mol. Biol.* 43(1):3-36.

4. Maduke M, Miller C, & Mindell JA (2000) A decade of CLC chloride channels: structure, mechanism, and many unsettled questions. *Annu. Rev. Biophys. Biomol. Struct.* 29(1):411-438.

5. Chen TY & Hwang TC (2008) CLC-0 and CFTR: chloride channels evolved from transporters. *Physiol. Rev.* 88(2):351-387.

6. Miller C (2006) ClC chloride channels viewed through a transporter lens. *Nature* 440(7083):484-489.

7. Picollo A, Xu Y, Johner N, Berneche S, & Accardi A (2012) Synergistic substrate binding determines the stoichiometry of transport of a prokaryotic H(+)/Cl(-) exchanger. *Nat. Struct. Mol. Biol.* 19(5):525-531, S521.

8. Khantwal CM, *et al.* (2016) Revealing an outward-facing open conformational state in a CLC Cl/H exchange transporter. *Elife* 10.7554/eLife.11189.

9. Basilio D, Noack K, Picollo A, & Accardi A (2014) Conformational changes required for H(+)/Cl(-) exchange mediated by a CLC transporter. *Nat. Struct. Mol. Biol.* 21(5):456-463.

10. Bell SP, Curran PK, Choi S, & Mindell JA (2006) Site-Directed Fluorescence Studies of a Prokaryotic ClC Antiporter. *Biochemistry* 45(22):6773-6782.

11. Elvington SM, Liu CW, & Maduke MC (2009) Substrate-driven conformational changes in ClC-ec1 observed by fluorine NMR. *EMBO J.* 28(20):3090-3102.

12. Accardi A & Miller C (2004) Secondary active transport mediated by a prokaryotic homologue of ClC Cl^- channels. *Nature* 427(6977):803-807.

13. Miller C & Nguitragool W (2009) A provisional transport mechanism for a chloride channel-type Cl^-/H^+ exchanger. *Phil. Trans. R. Soc. B* 364(1514):175-180.

14. Dutzler R, Campbell EB, Cadene M, Chait BT, & MacKinnon R (2002) X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* 415(6869):287-294.

15. Dutzler R, Campbell EB, & MacKinnon R (2003) Gating the selectivity filter in ClC chloride channels. *Science* 300(5616):108-112.

16. Accardi A, *et al.* (2005) Separate ion pathways in a Cl^-/H^+ exchanger. *J. Gen. Physiol.* 126(6):563-570.

17. Jayaram H, Accardi A, Wu F, Williams C, & Miller C (2008) Ion permeation through a Cl^--selective channel designed from a CLC Cl^-/H^+ exchanger. *Proc. Natl. Acad. Sci. U. S. A.* 105(32):11194-11199.

18. Bostick DL & Berkowitz ML (2004) Exterior site occupancy infers chloride-induced proton gating in a prokaryotic homolog of the ClC chloride channel. *Biophys. J.* 87(3):1686-1696.

19. Gervasio FL, Parrinello M, Ceccarelli M, & Klein ML (2006) Exploring the gating mechanism in the ClC chloride channel via metadynamics. *J. Mol. Biol.* 361(2):390-398.

20. Cheng MH & Coalson RD (2012) Molecular dynamics investigation of Cl^- and water transport through a eukaryotic CLC transporter. *Biophys. J.* 102(6):1363-1371.

21. Ko YJ & Jo WH (2010) Chloride ion conduction without water coordination in the pore of ClC protein. *J. Comput. Chem.* 31(3):603-611.

22. Robertson JL, Kolmakova-Partensky L, & Miller C (2010) Design, function and structure of a monomeric ClC transporter. *Nature* 468(7325):844-847.

23. Accardi A (2015) Structure and gating of CLC channels and exchangers. *J. Physiol.* 593(18):4129-4138.

24. Wang D & Voth GA (2009) Proton transport pathway in the ClC Cl^-/H^+ antiporter. *Biophys. J.* 97(1):121-131.

25. Zhang Y & Voth GA (2011) The coupled proton transport in the ClC-ec1 Cl(-)/H(+) antiporter. *Biophys. J.* 101(10):L47-49.

26. Han W, Cheng RC, Maduke MC, & Tajkhorshid E (2014) Water access points and hydration pathways in CLC H^+/Cl^- transporters. *Proc. Natl. Acad. Sci. U. S. A.* 111(5):1819-1824.

27. Lim HH & Miller C (2009) Intracellular proton-transfer mutants in a CLC Cl^-/H^+ exchanger. *J. Gen. Physiol.* 133(2):131-138.

28. Feng L, Campbell EB, Hsiung Y, & MacKinnon R (2010) Structure of a eukaryotic CLC transporter defines an intermediate state in the transport cycle. *Science* 330(6004):635-641.

29. Nelson JG, Peng Y, Silverstein DW, & Swanson JM (2014) Multiscale Reactive Molecular Dynamics for Absolute pKa Predictions and Amino Acid Deprotonation. *J. Chem. Theory Comput.* 10(7):2729-2737.

30. Lee S, Liang R, Swanson JM, & Voth GA (2015) Computationally efficient multiscale reactive molecular dynamics to describe amino acid deprotonation in proteins. *J. Chem. Theory Comput.* DOI: 10.1021/acs.jctc.5b01109.

31. Kuang Z, Mahankali U, & Beck TL (2007) Proton pathways and H^+/Cl^- stoichiometry in bacterial chloride transporters. *Proteins* 68(1):26-33.

32. Faraldo-Gomez JD & Roux B (2004) Electrostatics of ion stabilization in a ClC chloride channel homologue from Escherichia coli. *J. Mol. Biol.* 339(4):981-1000.

33. Skodje RT (1993) The Adiabatic Theory of Heavy-Light-Heavy Chemical-Reactions. *Annu. Rev. Phys. Chem.* 44(1):145-172.

34. Lim HH, Shane T, & Miller C (2012) Intracellular proton access in a Cl(-)/H(+) antiporter. *PLoS Biol.* 10(12):e1001441.

35. Feng L, Campbell EB, & MacKinnon R (2012) Molecular mechanism of proton transport in CLC Cl^-/H^+ exchange transporters. *Proc. Natl. Acad. Sci. U. S. A.* 109(29):11699-11704.

36. Yin J, Kuang Z, Mahankali U, & Beck TL (2004) Ion transit pathways and gating in ClC chloride channels. *Proteins: Struct. Funct. Bioinf.* 57(2):414-421.

37. Hess B, Kutzner C, van der Spoel D, & Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* 4(3):435-447.

38. MacKerell AD, *et al.* (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* 102(18):3586-3616.

39. Feller SE & MacKerell AD (2000) An Improved Empirical Potential Energy Function for Molecular Simulations of Phospholipids. *J. Phys. Chem. B* 104(31):7510-7515.

40. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79(2):926-935.

41. Darden T, York D, & Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98(12):10089-10092.

42. Nosé S (1984) A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* 81(1):511-519.

43. Barducci A, Bussi G, & Parrinello M (2008) Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* 100(2):020603.

44. Tribello GA, Bonomi M, Branduardi D, Camilloni C, & Bussi G (2014) PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* 185(2):604-613.

45. Torrie GM & Valleau JP (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* 23(2):187-199.

46. König PH, *et al.* (2005) Toward Theoretical Analyis of Long-Range Proton Transfer Kinetics in Biomolecular Pumps. *J. Phys. Chem. A* 110(2):548-563.

47. Riccardi D, *et al.* (2006) "Proton Holes" in Long-Range Proton Transfer Reactions in Solution and Enzymes: A Theoretical Analysis. *J. Am. Chem. Soc.* 128(50):16302-16311.

48. Swanson JM, *et al.* (2007) Proton solvation and transport in aqueous and biomolecular systems: insights from computer simulations. *J. Phys. Chem. B* 111(17):4300-4314.

49. Wu Y, Chen H, Wang F, Paesani F, & Voth GA (2008) An improved multistate empirical valence bond model for aqueous proton solvation and transport. *J. Phys. Chem. B* 112(2):467-482.

50. Yamashita T, Peng Y, Knight C, & Voth GA (2012) Computationally Efficient Multiconfigurational Reactive Molecular Dynamics. *J. Chem. Theory Comput.* 8(12):4863-4875.

51. Plimpton S (1995) Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* 117(1):1-19.

52. Becke AD (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* 38(6):3098-3100.

53. Lee C, Yang W, & Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* 37(2):785-789.

54. Schäfer A, Huber C, & Ahlrichs R (1994) Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.* 100(8):5829-5835.

55. Grimme S, Antony J, Ehrlich S, & Krieg H (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* 132(15):154104.

56. Laino T, Mohamed F, Laio A, & Parrinello M (2006) An efficient linear-scaling electrostatic coupling for treating periodic boundary conditions in QM/MM simulations. *J. Chem. Theory Comput.* 2(5):1370-1378.

57. Laino T, Mohamed F, Laio A, & Parrinello M (2005) An efficient real space multigrid OM/MM electrostatic coupling. *J. Chem. Theory Comput.* 1(6):1176-1184.

58. Blochl PE (1995) Electrostatic Decoupling of Periodic Images of Plane-Wave-Expanded Densities and Derived Atomic Point Charges. *J. Chem. Phys.* 103(17):7422-7428.

59. Maseras F & Morokuma K (1995) Imomm - a New Integrated Ab-Initio Plus Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition-States. *J. Comput. Chem.* 16(9):1170-1179.

60. Liang RB, Li H, Swanson JMJ, & Voth GA (2014) Multiscale simulation reveals a multifaceted mechanism of proton permeation through the influenza A M2 proton channel. *Proc. Natl. Acad. Sci. U. S. A.* 111(26):9396-9401.

61. VandeVondele J, *et al.* (2005) Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* 167(2):103-128.

62. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, & Kollman PA (1992) THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13(8):1011-1021.

63. Peng Y, Swanson JMJ, Kang S-g, Zhou R, & Voth GA (2014) Hydrated Excess Protons Can Create Their Own Water Wires. *J. Phys. Chem. B* 119(29):9212-9218.

64. Laio A & Parrinello M (2002) Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* 99(20):12562-12566.

65. Zhang Y & Voth GA (2011) Combined Metadynamics and Umbrella Sampling Method for the Calculation of Ion Permeation Free Energy Profiles. *J. Chem. Theory Comput.* 7(7):2277-2283.

66. Dellago C & Bolhuis P (2007) Transition Path Sampling Simulations of Biological Systems. *Atomistic Approaches in Modern Biology,* Topics in Current Chemistry, ed Reiher M (Springer Berlin Heidelberg), Vol 268, pp 291-317.

67. E W, Ren W, & Vanden-Eijnden E (2007) Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.* 126(16):164103.

68. Chandler D (1987) Introduction to modern statistical mechanics. (Oxford University Press, New York), pp 234-270.

69. Ko YJ & Jo WH (2010) Secondary water pore formation for proton transport in a ClC exchanger revealed by an atomistic molecular-dynamics simulation. *Biophys. J.* 98(10):2163-2169.

70. Miloshevsky GV, Hassanein A, & Jordan PC (2010) Antiport mechanism for Cl(-)/H(+) in ClC-ec1 from normal-mode analysis. *Biophys. J.* 98(6):999-1008.

71. Zhang L & Hermans J (1996) Hydrophilicity of cavities in proteins. *Proteins* 24(4):433-438.

72. Accardi A, Lobet S, Williams C, Miller C, & Dutzler R (2006) Synergism between halide binding and proton transport in a CLC-type exchanger. *J. Mol. Biol.* 362(4):691-699.

73. Kieseritzky G & Knapp EW (2011) Charge transport in the ClC-type chloride-proton anti-porter from Escherichia coli. *J. Biol. Chem.* 286(4):2976-2986.

74. Nguitragool W & Miller C (2006) Uncoupling of a CLC Cl^-/H^+ Exchange Transporter by Polyatomic Anions. *J. Mol. Biol.* 362(4):682-690.

75. Knight C, Lindberg GE, & Voth GA (2012) Multiscale reactive molecular dynamics. *J. Chem. Phys.* 137(22):22A525.

Chapter 3

Multiscale simulations reveal the key mechanism of the Cl^-/H^+ coupling beyond the central region in ClC-ec1 antiporter.

3.1 Introduction

The ClC channels and transporters constitute a large membrane protein family, where their isoforms found in a wide range of organisms, ranging from many prokaryotes to nearly all eukaryotic cells. (1, 2) They are involved in a broad range of physiological functions in a cell, such as stabilization of the membrane potential (ClC-1), regulation of transepithelial Cl^- transport (ClC-2, Ka, and Kb), ion homeostasis of endosomes (ClC-3,4,5, and 6), acidification of lysosome (ClC-7), acid resistance in a bacterial cell (ClC-ec1). (3-6) Defects of the proteins are known to cause many hereditary diseases, such as myotonia congenita, Dent's disease, Bartter's syndrome, osteopetrosis, and idiopathic epilepsy. (2, 6)

ClC-ec1, a bacterial ClC transporter homologue from *Escherichia coli*, mediates the ion exchange (antiport) mechanism of Cl^- and H^+ through the membrane. (See Fig. 1 for the structural overview.) It can pump H^+ through the membrane thermodynamically uphill in one direction, by using a downhill gradient of Cl^- concentration in the opposite direction as a driving force. It can also pump Cl^- in one direction through the membrane, using H^+ concentration gradient in the opposite direction. (7) Transmembrane ionic current was measured in various conditions of different concentration gradients for Cl^- and H^+. From experiment, the reversal potential, which is the voltage at thermodynamic equilibrium (when the net ionic current is zero), was shown to be the sum of the Nernst potentials of the concentrations of Cl^- and H^+ respectively, weighted by a certain ratio. This ratio was independent from the change of

concentration gradients of both Cl^- and H^+, showing that ClC-ec1 mediates a robust coupled exchange of Cl^- and H^+ through the membrane, with a fixed stoichiometric ratio, nearly two Cl^-s for one H^+. (7, 8) Later experiments (9, 10) directly measured the Cl^- efflux out of the liposome and the H^+ influx against the pH gradient, and confirmed the Cl^-:H^+ exchange ratio again to be 2.2 ± 0.1:1. (11)

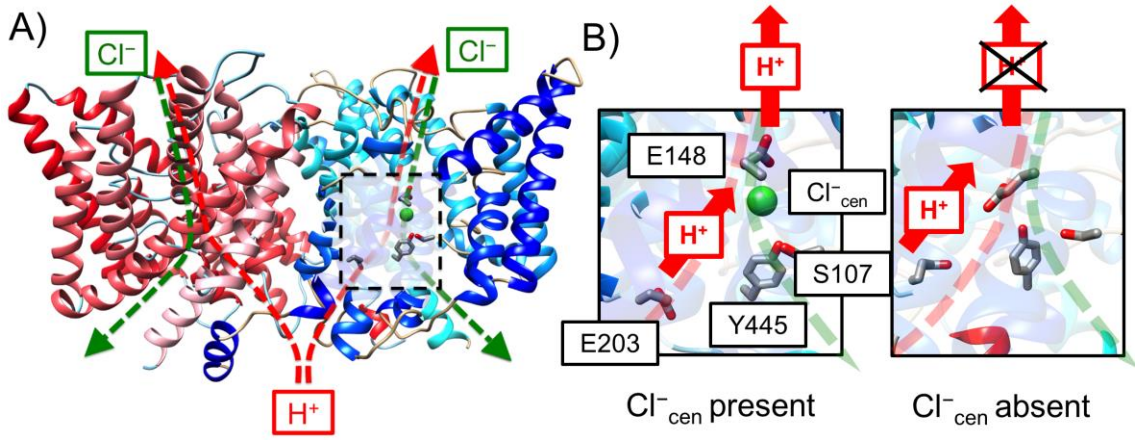


Figure 3.1. (A) Overview of the structure of the ClC-ec1 antiporter and transport pathways for Cl^- (green dashed) and H^+ (red dashed). ClC-ec1 consists of two monomers, where monomer A colored in yellow, and monomer B in blue. The central region is highlighted in the black box. (B) Schematic picture of the PT mechanism of this study. PT from E203 to E148 occurs, regardless of the presence of Cl^-_{cen}, however, Cl^-_{cen} facilitates the second step, deprotonation of E148 to the extracellular region.

A mutagenesis scan was performed at several ionizable residues in the protein, such as Glu or Asp, which can be protonated and deprotonated while H^+ (proton) migrates through its pathway in the protein structure. In E148A and E203Q mutants, where they were mutated to non-protonatable side chains, the H^+ flux through the membrane was blocked. The Cl^- flux was still intact in both mutants, but had the pH dependence in E203Q mutant, but not in E148A mutant. Here, a molecular-level mechanism of the Cl^-/H^+ exchange process in ClC-ec1 was proposed from experimental findings, where 1) E148 (Glu_{ex}) and E203 (Glu_{in}) participate in proton

transport (PT) in the protein, 2) Cl^- transport is controlled by the protonation of E148, and 3) the transport pathways for Cl^- and H^+ are overlapped at the region between E148 and the extracellular side of the membrane, and the H^+ pathway diverges to E203. (12)

Although E148 and E203 were known to be essential for PT, the two side chains were separated by ~ 15 Å in the X-ray crystal structure (13), and it was an unsolved question for a long time how the proton can transfer over this distance from E203 to E148, which is called as the central region. Based on previous experimental findings, the points listed above in the Cl^- and H^+ transport mechanism commonly agree with different experimental groups, however there were some differences in a more detailed picture from different groups, especially about the PT mechanism in the central region. Miller *et. al.* (14, 15) proposed that PT from E203 to E148 occurs in the presence of Cl^- at the central binding site (Cl^-_{cen}), then while E148 is protonated, Cl^- is transported through the E148 gate, and a new Cl^- coming from the extracellular region is bound to the central site. Y445 is a key residue in the central region, where its hydroxyl group is coordinated to the Cl^-_{cen} site (S_{cen}), and stabilizes Cl^- binding. Their experiment showed that, the loss of the Cl^-/H^+ coupling in several cases of Y445 mutants (the loss of the H^+ flux with the pH gradient) is coupled with the loss of anion signal at the Cl^-_{cen} site in the crystal structures. (16) However, MacKinnon *et. al.* suggested a different PT mechanism in the central region, where the side chain of E148 changes to the down conformation, and occupies S_{cen} in the absence of Cl^-. Then, PT from E203 to E148 occurs while Cl^- is absent at S_{cen}. A new Cl^- is transported through the protonated E148 gate, and E148 returns to the up conformation when a new Cl^- is present at S_{cen}. This conformational coupling of E148 and the presence of Cl^-_{cen} was found between the crystal structures from *E. coli* (13), and its eukaryotic homologue (cmClC) (17). In later experiment, (18) they showed that, the proton conductivity of E148A mutant was recovered by

addition of a free glutamate ion in the solution, which can substitute the function of E148 in WT. The carboxyl group of this substrate was bound at S_{cen} in the crystal structure of E148A mutant, as E148 in WT does in the absence of Cl^-_{cen}.

Another issue in the PT mechanism in the central region was that, some number of water molecules are needed in the central region, which can connect E148 and E203 with a hydrogen-bond network, then a proton can transfer across the ~15 Å distance through the water structure via the Grothuss mechanism. However, the central region is surrounded by many hydrophobic residues (I109, L186, F199, F357, I402, and T407), and none of the crystal structures resolved so far showed any water molecules in the central region. Although there were some computational approaches showing some water molecules spontaneously filling in the central region with classical MD simulations, the water structure was unstable, lasting only in the nanoseconds time scale. This may be too short that E203 is deprotonated, then the excess proton departed from E203 is transferred to E148. (19, 20)

In order to understand a more detailed mechanism of PT from E203 to E148 and the other coupled changes in the system listed above, in our previous study (21), some of the two dimensional free energy profiles (the potential of mean forces; PMF) were calculated in both the presence and the absence of Cl^-_{cen}, where one dimension in the PMF was defined by PT, and the other dimension was chosen to be either the conformational change of E148 and E203, or the hydration of the central region. The free energy calculation results were able to confirm the previous experimental findings and provide some answers on those unsolved issues in the PT mechanism as mentioned above: 1) PT from E203 to E148 is facilitated by the presence of Cl^-_{cen}, 2) the deprotonation of E203 induces the increase of the hydration level in the central region, and 3) Protonated E148 goes to the down conformation only in the absence of Cl^-_{cen}. However, the

PT rate crossing the free energy barrier from E203 to E148 in the PMF was much faster in both the presence and the absence of Cl^-_{cen}, than the rate measured in experiment for the overall PT process. Based on our previous results, PT from E203 to E148 was not a rate-limiting step regardless of the presence of Cl^-_{cen}, and it still remained unclear how PT is coupled with Cl^- transport in the overall mechanism, and which intermediate step is rate-limiting for the overall PT process.

There are two other intermediate steps for PT occurring out of the central region: 1) from the intracellular side of the membrane to E203, and 2) from E148 to the extracellular region. PT in the latter step is considered to be more strongly coupled with Cl^- transport than the former step, because the Cl^- and H^+ transport pathways are overlapped in the latter step, but in the former step, the crystal structures show that the Cl^- pathway is departed from E203 by ~10 Å. In this work, the PT mechanism from E148 to the extracellular region was studied by calculating the free energy profile for PT in both the presence and the absence of Cl^-_{cen} using the H^+ transport pathway obtained from the metadynamics simulations, where it can describe a two-step process with a single continuous collective variable (CV): the rotation of E148 side chain from down to up conformation, followed by the deprotonation of E148 to the extracellular region. The reactive nature of the proton dynamics and the protonation/deprotonation reaction of the Glu residue was described by the MS-RMD model (22), following the procedure of previous work (21). The simulation results showed several important points to explain the Cl^-/H^+ coupling mechanism in a molecular level, where the deprotonation of E148 is a rate-limiting step, and coupled with the presence of Cl^-_{cen}. In the absence of Cl^-_{cen}, E148 favors the down conformation at the S_{cen}, resulting that the free energy barrier for PT to the extracellular region becomes high enough to prevent the proton leakage below the experimentally measurable range.

Not only with the rotation of E148, but PT can be coupled with other conformational changes of the protein outside of the central region. Although the crystal structures of the ClC proteins did not show any large-scale conformational change among different structures, unlike other transporters, several studies (11, 23-25) reported that the conformational changes of some residues or the helical structures outside of the central region are coupled with transport of Cl^- and H^+. There was also a computational approach characterizing the conformational change with normal mode analysis. (25, 26) However, it is worth to note that the results do not provide the information about how large the protein structure is changed, nor in which intermediate step the conformational change occurs.

In addition to the study of the PT mechanism in WT, the free energy profile was calculated for PT in E148A mutant starting from E203, through the central region, then going to the extracellular region. In E148A mutant, due to the introduction of hydrophobic environment around A148, the water molecules approaching A148 from the extracellular region are disconnected with the water molecules in the central region. This causes the increase of the free energy barrier, when the excess proton penetrates through the dehydrated region at around A148. The PT rate constants calculated from the PMFs are below the experimentally measurable range in both the presence and the absence of Cl^-_{cen}. Experimentally, E148A mutant becomes like a Cl^- channel, where the H^+ flux is blocked, but the Cl^- flux is still present in the direction of its concentration gradient. Our results agree with the experimental results, where PT is effectively blocked in both the presence and the absence of Cl^-_{cen}.

3.2 Methods

System setup

The initial system setup is described in our previous work (21). Briefly speaking, the simulation system consists of the ClC-ec1 dimer (PDB ID: 1OTS), 163 POPE lipids, 17 Cl^-s, and ~11,000 water molecules in a 92 Å x 92 Å x 79 Å box with a periodic boundary condition. The CHARMM22 force field (27) was used for protein and ions, and CHARMM27 (28) for lipids. The excess proton and water molecules were described by the MS-EVB model (29) ver. 3.1 (30). Except the Glu residues in the central region of monomer A, the protonation states of all other residues were determined based on previous pKa calculation results on the same crystal structure (31). E113 in monomer B and D417 in monomer A were set to be protonated, and standard protonation states were chosen for all other residues.

The metadynamics simulation for sampling the rotation of protonated E148

The initial configurations for the MD simulation with MS-RMD model were taken from the last MD snapshot of the windows corresponding to the product state in the PMFs for PT from E203 to E148 in both the presence and the absence of Cl^-_{cen} from previous calculations (21). (The PMF in Fig. 3B for the case with Cl^-_{cen} present, and the PMF in Fig. 4B with Cl^-_{cen} absent) The protonation states for E148 and E203 of the initial configurations are E148(H)E203(0). With the same MS-RMD models for E148 and E203 from previous work, the metadynamics (MetaD) simulation (32) was performed from the initial configurations for both cases with Cl^-_{cen} present and absent, where the CV was defined as the z coordinate of the center of the excess charge (CEC),

which represents the effective position of the delocalized excess proton in the system. The CEC, $\mathbf{r}_{CEC}$, is defined as follows,

$$\mathbf{r}_{CEC} = \sum_{i}^{N} c_i^2(\mathbf{r})\mathbf{r}_i^{COC} \tag{1}$$

where $c_i^2(\mathbf{r})$ is the population of *i* th MS-RMD state, and $\mathbf{r}_i^{COC}$ is the center of charge (COC) of the *i* th state given by,

$$\mathbf{r}_i^{COC} = \sum_{k}^{\{i\}} |q_k| \mathbf{r}_k \Big/ \sum_{k}^{\{i\}} |q_k| \tag{2}$$

where the coordinate of the *k* th atom, compromising the MS-RMD complex in the *i* th state, $\mathbf{r}_k$, is averaged, weighted by the absolute value of its atomic partial charge, q_k. The CEC is located near the carboxyl group of E148 in the initial configurations. The hill height and the width of a MetaD Gaussian potential is 0.1 kcal/mol and 0.35 Å, respectively. The Gaussian hill was added every 1000 time steps in the trajectory. The half-sided harmonic potential with the force constant, 100 $kcal \cdot mol^{-1} \cdot Å^{-2}$, was added as a wall at the lower boundary of the CV, which is 2 Å below in the z axis from the position of E148 side chain, in order to prevent a proton from transferring back to E203. Another wall was added at the upper boundary, which is 6 Å above from the E148 side chain, in order to prevent a proton from deprotonating to the extracellular region. The MetaD simulation was run for ~3 ns, until the CV trajectory visited between the upper and the lower boundaries 2~3 times. The other MD simulation details, such as thermostat, the cut-off distances

for the non-bonded interactions, and so on, followed the same procedure used in previous work. The simulation was performed with the RAPTOR software (33), interfaced with the LAMMPS MD package (http://lammps.sandia.gov) (34), and the bias potential was controlled by the PLUMED package (35).

In the whole MetaD trajectory, E148 was kept protonated, and the CEC stays at the E148 carboxyl group. As the z coordinate of the CEC varies between the boundaries at both sides, E148 rotates its side chain from the down conformation (when the E148 carboxyl group contacts with the water molecules in the central region) to the up conformation (when it contacts with the water molecules approaching from the extracellular region to E148.) The x, y, and z coordinates of the CEC was recorded every time step, then the x and the y coordinates were averaged with each 0.25 Å interval in the z coordinate. The MetaD pathway for the CEC was obtained by interpolating the averaged x and y coordinates at every z coordinate with 0.25 Å interval, using Mathematica ver. 9.0.

The MS-RMD umbrella sampling for the rotation of E148

The two sets of the PMFs for the rotation of protonated E148 in both the presence and the absence of Cl^-_{cen} were calculated with umbrella sampling on the MetaD pathway for the CEC (36), combined with the replica exchange method (REUS) (37) to more efficiently sample other slow degrees of the freedom in the system. The same MS-RMD models for E148 and E203 from previous work were used. Each window was separated by every 0.25 Å in the z direction. The force constant of the harmonic potential was set to be 30 $kcal \cdot mol^{-1} \cdot Å^{-2}$. The direction of the harmonic potential was defined as the tangent vector of the MetaD pathway calculated at the

window center. A MD snapshot from previous MetaD simulation, where the position of the CEC is in the closest distance from the center of each window, was chosen to be the initial configuration at the window. The conventional Metropolis Monte Carlo exchange criterion was applied every 5000 time steps to determine swapping the neighboring windows. The average acceptance ratio for the successful exchange was about 20 % over all windows. Each window was sampled for 1~2 ns. The CV values were recorded every steps, and the PMF was calculated by the WHAM method (38).

The MS-RMD modeling for E148 in the up conformation

The MS-RMD model for E148 was re-parameterized, where E148 is in the up conformation and deprotonates to the water molecules from the extracellular region. The initial configuration was taken from the last MD snapshot at the window, corresponding to the free energy basin for the up conformation of E148, from previous umbrella sampling simulation. Another umbrella sampling simulation was initiated using a different CV definition, which was the distance from the CEC to the center of mass of the carboxyl group of E148, in order to sample configurations for generating the reference force data in QM/MM. The MS-RMD models for E148 and E203 in the umbrella sampling were taken from previous work, where E148 is in the down conformation, and transfers the excess proton to the water molecule in the central region. The centers of the windows ranged from 2 to 4 Å, and separated by 0.25 Å between neighboring windows. The force constant of the harmonic potential was set to be 30 $kcal \cdot mol^{-1} \cdot Å^{-2}$. ~500 configurations for each case with Cl^{-}_{cen} present and absent (5~60 configurations from each window) were evenly sampled for the CV in all range. Each configuration was selected with a 5 ps interval. For

each configuration, a single point energy calculation was performed with the CP2K software (39) to calculate the QM/MM forces on the atoms in the QM region.

In the QM/MM simulation, the QM atoms were treated with the density functional theory (DFT) BLYP-D3/TZV2P level of theory (40-42) including the Grimme dispersion correction (43). The QM region included Cl^-_{cen} if present, the side chains of residues E148, Y445, and S107, the water molecules in the central region within 3 Å distance from E148, and the water molecules from the extracellular region within 8 Å distance from E148. The total number of the QM atoms is ~130 in average among all configurations. The QM box size was set to be 20 – 30 Å in each dimension to ensure it was 6-8 Å larger than the size of the QM atoms. The Gaussian Expansion of the Electrostatic Potential (GEEP) scheme was used to treat the QM/MM electrostatic coupling with periodic boundary conditions (PBCs) (44, 45), and the spurious QM/QM periodic image interactions were decoupled as described in ref (46). The chemical bonds that crossed the QM/MM boundary between the alpha carbon (MM) and the beta carbon (QM) of the residue, were capped with hydrogen atoms, and the forces on which were calculated following the IMOMM scheme with a scaling factor of 1.50 (47). Otherwise, the QM setup was consistent with previous works (21, 22, 48), where the QM/MM simulation was performed in other biological systems as well as ClC-ec1.

Once the QM/MM force was calculated for all configurations sampled from the umbrella sampling, the MS-RMD model parameters for E148 in the up conformation were optimized by using the FitRMD method (22, 30), which minimizes the residuals between the forces calculated by the QM/MM and the MS-RMD. The new MS-RMD parameters for E148 in the up conformation for both cases with Cl^-_{cen} present and absent are given in Table 1.

	Cl^-_{cen} absent		Cl^-_{cen} present	
	E148 (down)	E148 (up)	E148 (down)	E148 (up)
B	0.063153	0.454014	0.012536	0.413895
b	1.571751	0.075642	0.232384	0.025128
b'	1.320947	0.010841	1.469007	0.024147
d^0_{OO}	2.4	2.4	2.4	2.4
C	0.363648	0.371124	0.014044	0.356169
c	1.117167	0.709177	1.152912	0.665787
d^0_{OH}	1.0	1.0	1.0	1.0
r_s^l	3.5	3.5	3.5	3.5
r_s^h	4.0	4.0	4.0	4.0
V_{ii}	-147.095673	-157.105468	-151.11473	-154.325345
c_1	-36.090543	-42.974750	-30.414842	-30.319574
c_2	1.879933	2.552372	3.331769	3.436706
c_3	1.193253	1.556170	1.422240	1.456306
D	143.003	143.003	143.003	143.003
α	1.8	1.8	1.8	1.8
r_0	0.975	0.975	0.975	0.975

Table 3.1 The MS-RMD model parameters for E148 in the down/up conformations. The model parameters for E148 down conformations are taken from our previous work (21).

The MS-RMD MetaD and umbrella sampling for the two-step process: the rotation of E148 and the PT from E148 to the extracellular region.

A new two sets of the MetaD simulation were initiated, using a new MS-RMD model for E148, with both Cl^-_{cen} present and absent, where the CV was still defined as the z coordinate of the CEC, but sampled for a wider range, including the rotation of E148 from down to up conformation, followed by the deprotonation of E148 to the solution at the extracellular region. The MetaD simulation was set up consistently with a previous run, except the wall potential for the upper boundary at the entrance for the both Cl^- and H^+ transport pathways at the extracellular

region, which is located at 15 Å above in the z direction from E148 in the down conformation. Additional wall potentials were added to the x and the y directions, in order to prevent the CEC from escaping from the entrance horizontally, where the pore size of the protein channel becomes larger. The wall potentials were placed at each side of the 35 Å × 30 Å rectangular box on the xy plane, which is large enough to cover the whole area of the pore at the channel entrance. The MetaD simulation was run for ~5 ns, until the CV trajectory visited between the upper and the lower boundaries 2~3 times. The MetaD pathway for the CEC was calculated from the MetaD trajectory with the same fashion as a previous run.

The two sets of the REUS simulation for the CEC on the MetaD pathway were initiated with both Cl^{-}_{cen} present and absent. Each window was separated by every 0.25 Å in the z direction. The force constant of the harmonic potential was set to be 30 $kcal \cdot mol^{-1} \cdot Å^{-2}$. The direction of the harmonic potential at each window was set to be a tangent vector of the MetaD pathway at the window center. The cylindrical wall potential with a 5 Å radius was added to the direction perpendicular to the MetaD pathway. The cylindrical potential is a flat-bottom harmonic potential, where the bias is zero, when the perpendicular distance to the MetaD pathway is smaller than 5 Å, then when it is greater than 5 Å, the half-sided harmonic potential was applied to force the CEC toward the MetaD pathway. The cylindrical potential is widely used in other studies, when the free energy profile of ion transport in the protein channel is described by the one-dimensional PMF along the channel axis. The PMF is more efficiently converged by limiting the sampling space around the channel axis. The radius of the cylindrical potential is taken into account in later step, when the pKa of E148 was calculated from the PMF, which corresponds to the free energy change when the proton migrates from E148 to the solution in the extracellular region.

The MS-RMD MetaD and umbrella sampling for PT in E148A mutant.

The calculation of the PMF for PT in E148A mutant was done in a similar fashion with that for WT, but the CV was sampled for a wider range, which includes PT from E203, through the central region, through by A148, then to the extracellular region. The initial configurations for the MetaD simulation were chosen from WT, with both Cl^-_{cen} present and absent, when E203 is protonated, and E148 is deprotonated. Then, E148 is mutated to Alanine residue, and the system is equilibrated for another 1 ns. The MS-RMD for E203 was taken from previous work. The MetaD simulation was initiated from the configuration with E203 protonated, and the MetaD potential was added to the z coordinate of the CEC to explore sampling from E203 to the extracellular region. The wall potentials were added in the same fashion with a previous MetaD run. Once the MetaD pathway for the CEC was obtained from the MetaD simulations, the REUS simulations were performed to calculate the PMFs for PT from E203 to the extracellular region with both Cl^-_{cen} present and absent. All other REUS setups were consistent with a previous REUS run for WT.

The calculation of PT rate constants from the PMF

The PT rate constant was estimated by using transition state theory between the two states as follows (49)

$$k_{rxn} = \frac{\omega_0}{2\pi} \exp\left(-\frac{\Delta F^{\ddagger}}{k_B T} \right) \tag{3}$$

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 ω_0 is that of the reactant state oscillations around its minimum, which is defined as

$$\omega_0 = \sqrt{\frac{\partial^2 PMF(r) / \partial r^2}{m_{eff}}} \tag{4}$$

where r is the local minimum in the PMFs. The effective mass of the excess proton CEC, m_{eff}, was determined using the equipartition theorem, $m_{eff}\langle v^2 \rangle = k_B T/2$, where the value of $\langle v^2 \rangle$ was calculated from the MS-RMD trajectory sampled at r.

The calculation of pKa of E148

The pKa of E148 was estimated using the equation for calculating the equilibrium constant of binding of the substrate at the binding site of the protein, using the one-dimensional PMF for the substrate moving along the channel axis with the cylindrical potential applied at the channel entrance, as follow (50)

$$K_a^{-1} = C^0 \cdot e^{\Delta G_{site}/k_B T} \cdot \pi r^2 \cdot \int_{site} dz\, e^{-\left[w(z) - w_{ref}\right]/k_B T} \tag{5}$$

where the substrate is the excess proton and the binding site is E148 in our PMFs. K_a is the acid dissociation constant of E148. C^0 is the standard state concentration (1 M = 1/1660 $Å^{-3}$). Remember that K_a is unitless, and the other terms except C^0 on the right hand side of Eq. 5 are in the unit of $Å^3$. ΔG_{site} is the free energy cost introduced by the cylindrical potential, when the excess proton is at E148, which is zero in all PMFs, because the pore size at E148 is so narrow that the cylindrical potential is not felt by the CEC at this region. r is the radius of the cylindrical potential, which is set to be 5 Å in all PMFs. $w(z)$ is the one-dimensional PMF as a function of the CV, z, and w_{ref} is the asymptotic value of $w(z)$, when the excess proton is at a long distance at the extracellular region. In the last term, the Boltzmann factor of the free energy change is integrated over the region for the binding site (protonated E148) in the PMF. The lower boundary of the range of the binding site is set to be at the position of E148. Although E148 gradually changes to the deprotonated state in the PMF, the pKa of E148 from Eq. 3 was insensitive to the choice of the upper boundary of the binding site. The pKa of E148 changes only 0.001, when the upper boundary varies between the extracellular region and 2 Å above from the position of E148.

3.3. Result and Discussion

The first step of PT in WT: the rotation of E148

Two sets of the PMF for PT (See Fig. 2A and B) were calculated for each case with Cl^-_{cen} either present or absent: 1) one PMF of the rotation of E148 calculated with the MS-RMD model for E148 when it is in the down conformation (with a blue line in Fig. 2A and B), and 2) another PMF of a continuous two-step process, including both the rotation of E148 and the deprotonation

of E148 to the extracellular region with the MS-RMD model for E148 in the up conformation (with a red line in Fig. 2A and B). As shown in our earlier work (22), the MS-RMD model parameters depend on the change of the protein environment surrounding the Glu residue, therefore the model for E148 in this study was parameterized, depending on its conformational change. The MS-RMD model for E148 down conformation was taken from previous work (21), where E148 accepts/donates a proton from/to the water molecules in the central region, and the model for E148 up conformation was newly parameterized in this study, where E148 accepts/donates a proton from/to the water molecules at the extracellular region. (See Fig. 3)

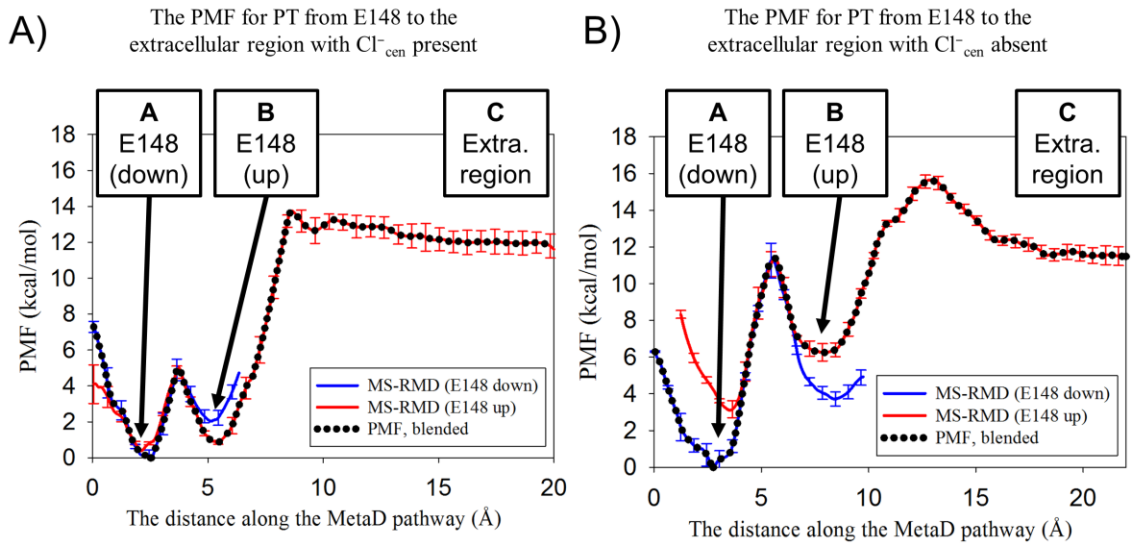


Figure 3.2. (A) The PMF for a two-step PT process with Cl^-_{cen} present, including the rotation of E148 from down (A) to up (B) conformations, followed by the deprotonation of E148 to the extracellular region (C). (B) The same PMF with Cl^-_{cen} absent.

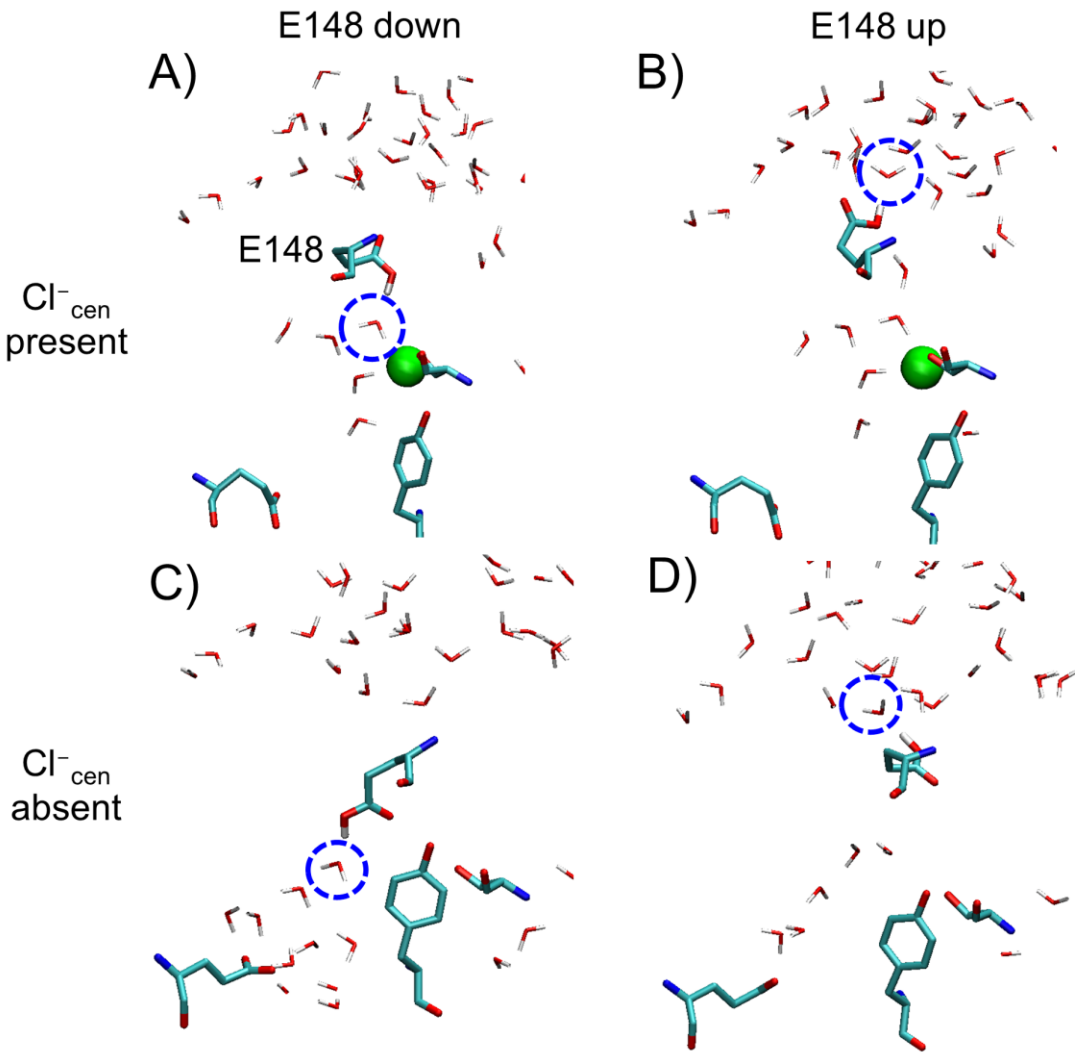


Figure 3.3. The representative configuration captured at the windows for the local energy minima of the PMF in Fig. 2, with Cl^-_{cen} present (A, B) and absent (C, D), or with E148 in the down conformation (A, C) and the up conformation (B, D).

The two PMFs, calculated with different models for E148, are well overlapped at the region ~3 Å long around the top of the free energy barrier for the first step in the PMFs, where the protonated E148 changes from the down to the up conformation (the transition between the states A and B in Fig. 2A and B). Then, by setting the position of the barrier top as a boundary, the two PMFs were merged into a single PMF (with a dotted line in Fig. 2A and B), where the region for

E148 in the down conformation was calculated by using the MS-RMD model parameterized for E148 in the down conformation, and the other region in the PMF using the model for E148 up conformation.

The two models give different results on the PMFs, where the difference in the free energy between the local minima, as well as the height of the free energy barrier in the transition between the E148 up and down conformations are changed. The differences between the two PMFs are larger with Cl^-_{cen} absent, which is ~3 kcal/mol at the both local energy minima for E148 up and down conformations, compared to those with Cl^-_{cen} present, ~1 kcal/mol at the same positions. This difference was caused by the difference in the model. Although the MS-RMD state with protonated E148 was always the most dominant state during the down/up conformational change, other energetically feasible states with deprotonated E148 still contributed some populations to the total system, and the contribution from different states was calculated differently with two models. In order to effectively describe the polarization of the charge density on E148, in response to the local electrostatic environmental change by surrounding water molecules and some of the protein residues, the populations of some of deprotonated states with the energetically feasible range are included in the total state in the MS-RMD model.

Fig 4 shows how the population of the state with protonated E148 is changed between the two models along the same umbrella sampling window. The difference in the population at the same window between the two models was greater with Cl^-_{cen} absent than with Cl^-_{cen} present, causing that the two PMFs more diverge with Cl^-_{cen} absent. The population of protonated E148 state with Cl^-_{cen} absent was decreased by using the model for E148 down conformation, causing that the excess proton on E148 is more delocalized into the surrounding water molecules, and the

system becomes more stabilized, compared to that using the model for E148 up conformation. The space around E148 is so narrow that the water molecule in the central region, which makes a hydrogen bond with E148, does not move along with the carboxyl group of E148 rotating from the down to the up conformation. The carboxyl group of E148 breaks a hydrogen bond with the water molecule in the central region, and makes a new bond with another water at the extracellular region during the rotation. (See Fig. 3) When the excess proton is more delocalized with the model for E148 down conformation, the free energy cost required for breaking a hydrogen bond between E148 and the adjacent water molecule would be more increased, resulting that the free energy barrier in the PMF is increased, compared to that with the model for E148 down conformation.

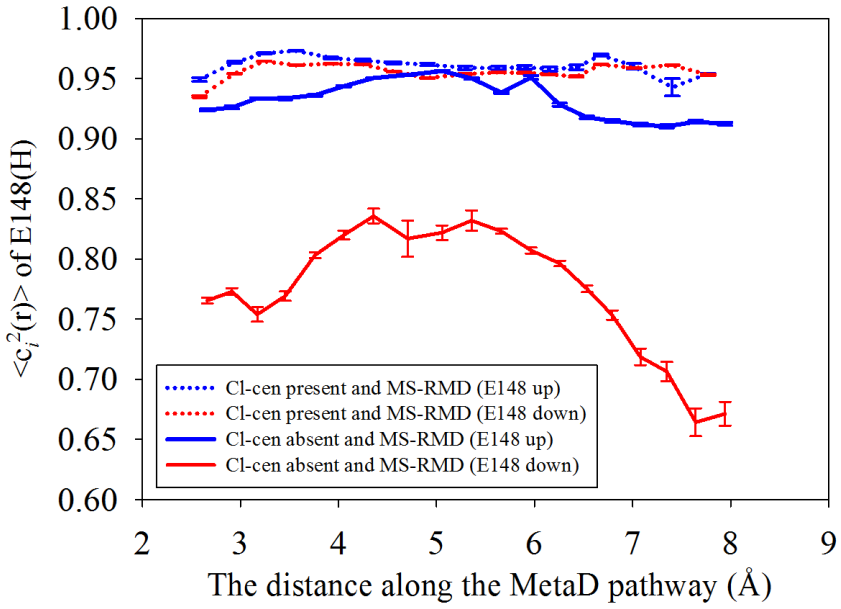


Figure 3.4. The average value of $c_i^2(\mathbf{r})$ of the state for protonated E148 during the rotation of E148 in the PMF in Fig. 2.

As shown in Fig. 3, when E148 goes to the down conformation in the absence of Cl^-_{cen}, its carboxyl group is bound to the vacated S_{cen} site. Previous computational study (51), calculating the electrostatic potential energy profile along the Cl^- pathway, showed that the S_{cen} site favors a negative charge on it. When the carboxyl group is bound to the S_{cen} site, the positively charged excess proton would be more likely delocalized into surrounding waters, decreasing the population of protonated E148. On the other hand, the E148 rotation with Cl^-_{cen} present is sterically more hindered by Cl^-_{cen}. The E148 down/up conformational change is more limited with Cl^-_{cen} present, where some torsional angles of the carboxyl group is changed, and without any large translational motion of the whole residue, the carboxyl hydrogen makes a new hydrogen bond with the water at the extracellular region.

The second step of PT in WT: The deprotonation of E148 to the extracellular region.

The second step in the PMFs with both Cl^-_{cen} present and absent in Fig. 2, starting from the local energy minima for E148 up conformation, corresponds to the deprotonation of E148 and the escape of the excess proton to the extracellular region. The PMFs become a plateau at a long distance at the extracellular region, where the excess proton does not effectively interact with the protein. The height of the free energy barrier for the second step is higher with Cl^-_{cen} present (12.9 kcal/mol), compared to that with Cl^-_{cen} absent (9.3 kcal/mol). The S_{cen} site is located at ~4 Å away from E148 to the central region, which is in the opposite direction of PT of the second step. The presence of Cl^-_{cen} increases the free energy barrier for the deprotonation of E148, as the positively charged excess proton escapes away to the extracellular region. The influence of the presence of Cl^-_{cen} was different in the PMF for PT in the central region in previous work (21),

where Cl^-_{cen} is located at the middle of PT pathway between E203 and E148, decreasing the deprotonation energy barrier of E203 in the PMF. Although the free energy barrier for the second step is higher with Cl^-_{cen} present, the barrier height of the overall process, including both the first and the second steps, is higher with Cl^-_{cen} absent (15.7 kcal/mol) than with Cl^-_{cen} present (13.5 kcal/mol). As discussed earlier, protonated E148 energetically favors the down conformation with Cl^-_{cen} absent, but the rotation of E148 is sterically hindered with Cl^-_{cen} present, causing that the free energy barrier for the overall process is increased, when the system starts from the state for E148 down conformation.

The rate constants for the overall process were calculated as the effective rate constant, k_{eff}, of the two-step process, where $k_{eff} = k_2 \cdot k_1/k_{-1}$, assuming that the first step reaches the fast equilibrium, when $k_2 \ll k_{-1}$. Table 2 shows that k_{eff} with Cl^-_{cen} present is 1.1 ms^{-1}, and with Cl^-_{cen} absent is $7.7 \times 10^{-3}\ ms^{-1}$. k_{eff} with Cl^-_{cen} present is comparable to the experimental value of the turnover rate for PT, 1.0 ms^{-1} (10, 11). (The PT turnover rate was calculated, using the Cl^- turnover rate, 2.3 ms^{-1}, and the Cl^-:H^+ exchange ratio, 2.2:1 in experiment.) This result tells that PT from E148 to the extracellular region with Cl^-_{cen} present could be a rate-limiting step for the overall PT process. However, k_{eff} with Cl^-_{cen} absent is in the order of sec^{-1}, which may be too slow to be measured in conventional experiments, because it will be difficult to distinguish the actual H^+ flux through the protein from the background leak current through the membrane (8, 9, 52).

	Cl^-_{cen}	PMF figure	Reaction step	k (ms^{-1})	$\Delta F^{\ddagger}$ (kcal/mol)
WT (outward H^+ flux)	present	2 A	k_1	3.5×10^6	5.0
			k_{-1}	8.5×10^6	4.1
			k_2	2.8	12.9
			k_{eff}	1.1	
	absent	2 B	k_1	1.8×10^1	11.4
			k_{-1}	2.5×10^6	4.8
			k_2	1.1×10^2	9.4
			k_{eff}	7.7×10^{-3}	
WT (inward H^+ flux)	present		E148 to E203	3.4×10^{-1}	13.8
	absent		E148 to E203	2.9×10^{-4}	18.0
E148A (outward H^+ flux)	present		E203 to the extra. region	1.2×10^{-7}	22.7
	absent		E203 to the extra. region	3.2×10^{-11}	27.5
E148A (inward H^+ flux)	present		the extra. region to E203	8.2×10^{-4}	17.3
	absent		the extra. region to E203	2.3×10^{-6}	20.9

Table 3.2. The rate constants for PT in the WT ClC-ec1 and E148A mutant. The rate constants were calculated with transition state theory.

Our calculation of the PT rate constants give us some more detailed information of the PT mechanism. Although it has been known that the Cl^-/H^+ exchange mechanism is reversible and can operate in both directions (12), a later experimental study (53) suggested that there is a preferred directionality: the inward Cl^- flux (coupled with the outward H^+ flux) would be more favored than that in the opposite direction. Therefore, when we describe the overall PT process in a more preferred direction: from the intracellular to the extracellular region, our previous study (21) showed that, PT from E203 to E148 is not the rate-determining step, regardless of the presence of Cl^-_{cen}. This study showed that, the next intermediate step, PT from E148 to the extracellular region occurs in the millisecond time scale with Cl^-_{cen} present, and when Cl^-_{cen} is

absent, PT is almost blocked, where the PT rate constant goes below the experimentally measurable range. That deprotonation of E148 is coupled to the presence of Cl^-_{cen} is the key mechanism of the Cl^-/H^+ coupling in the overall process. In addition, even if the overall process is described in the opposite direction, e.g. PT from E148 to E203 in the central region, Table 2 shows that E148 can be deprotonated only with Cl^-_{cen} present, which can be also the rate-limiting step for the overall process in the opposite direction.

The pKa of E148 was calculated from the PMF, using the eq. 5. The pKa of E148 in the PMF with Cl^-_{cen} present, is 6.9 at the local energy minima for E148 down conformation, and 6.4 for E148 up conformation, which are comparable to the apparent pKa measured by the Cl^- uptake assay at different pH in experiment, 6.2 (54). The pKa of E148 in the PMF with Cl^-_{cen} absent, is 6.8 for E148 down conformation, and 2.6 for E148 up conformation.

Structural comparison between the simulation and the crystal structures

The X-ray crystal structures often represent different snapshots captured at different intermediate states, which here correspond to different local minima in the PMFs. Fig. 5 compares the structures obtained at the local energy minima in our PMFs with three crystal structures from WT of ClC-ec1 (PDB ID: 1OTS) (13), E148Q mutant of ClC-ec1 (1OTU) (13), and WT of cmClC (3ORG) (17), which is a eukaryotic ClC homologue. Q148 in E148Q mutant is considered to mimic the protonated state of E148 in WT. Those three crystal structures show different conformations of E148. 1OTS and 1OTU are well overlapped with the simulation structures taken from the windows at the local energy minima in the PMF with Cl^-_{cen} present, for E148 down and up conformations, respectively. 3ORG agrees with the simulation structure of

E148 down conformation in the PMF with Cl^-_{cen} absent. The E148 up conformation with Cl^-_{cen} absent is in a high energy state, and the corresponding crystal structure has not been found yet.

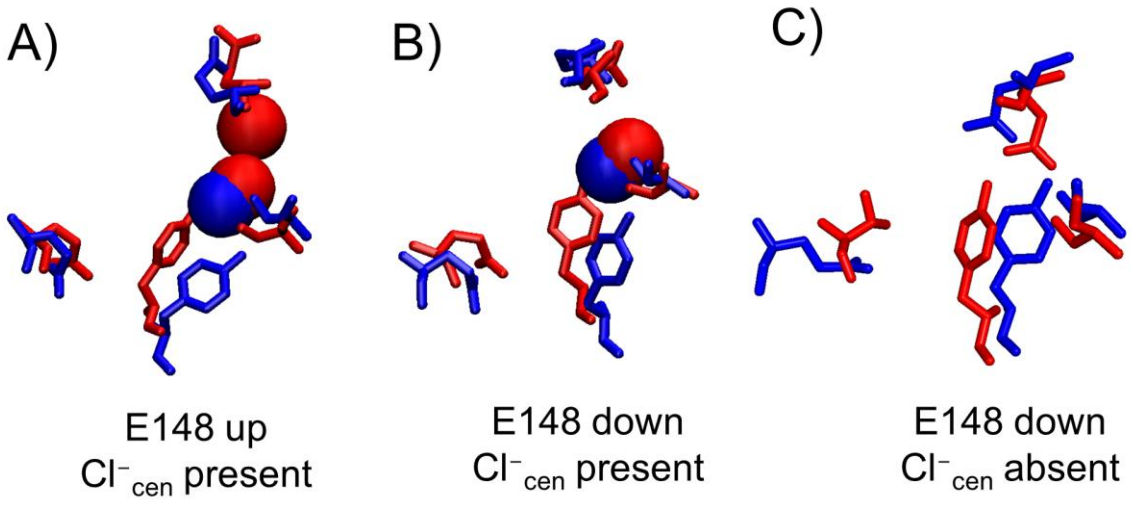


Figure 3.5 The structures from the simulation (blue) and the X-ray crystal structures (red) are overlapped to each other. (A) The simulation structure was obtained from the local energy minima in the PMF with E148 in the up conformation and Cl^-_{cen} present. The corresponding crystal structure is E148Q mutant of ClC-ec1 (1OTU). (B) The simulation structure: with E148 in the down conformation and Cl^-_{cen} present. The crystal structure: WT of ClC-ec1 (1OTS). (C) The simulation structure: with E148 in the down conformation and Cl^-_{cen} absent. The crystal structure: WT of cmClC (3ORG).

PT in E148A mutant

Another two sets of PMFs of PT were calculated in E148A mutant, for both cases with Cl^-_{cen} present and absent, where it starts from E203, through the central region, and escapes to the extracellular region. (See Fig. 6) It has been known that E148A mutant becomes like a Cl^- channel, where PT is blocked, but Cl^- is passively transported following its concentration gradient (7, 12). The pore of the protein channel at E148 is so narrow that, in WT, the excess

proton passes through this region, while E148 is protonated, and rotates from the central region to the extracellular region. However, in E148A mutant, when E148 is mutated to a non-protonatable amino acid, the free energy cost required for the excess proton passing through this narrow region is much increased. The top of the free energy barrier in the PMFs in Fig. 6 corresponds to when the excess proton is located near A148. The two PMFs show that the free energy barrier for PT is decreased with Cl^-_{cen} present. When PT is described in the direction from E203 to the extracellular region, the barrier height in each PMF was 22.5 kcal/mol with Cl^-_{cen} present, and 27.6 kcal/mol with Cl^-_{cen} absent, respectively. The corresponding PT rate constants are $1.2 \times 10^{-7}\ ms^{-1}$ with Cl^-_{cen} present, and $3.2 \times 10^{-11}\ ms^{-1}$ with Cl^-_{cen} absent, telling that PT is blocked, regardless of the presence of Cl^-_{cen}. The conclusion does not change when PT is described in the opposite direction. Our result agrees with the experimental data, where Cl^- is kept bound and unbound at the S_{cen}, while it is transported through the protein, but PT through the protein is always blocked.

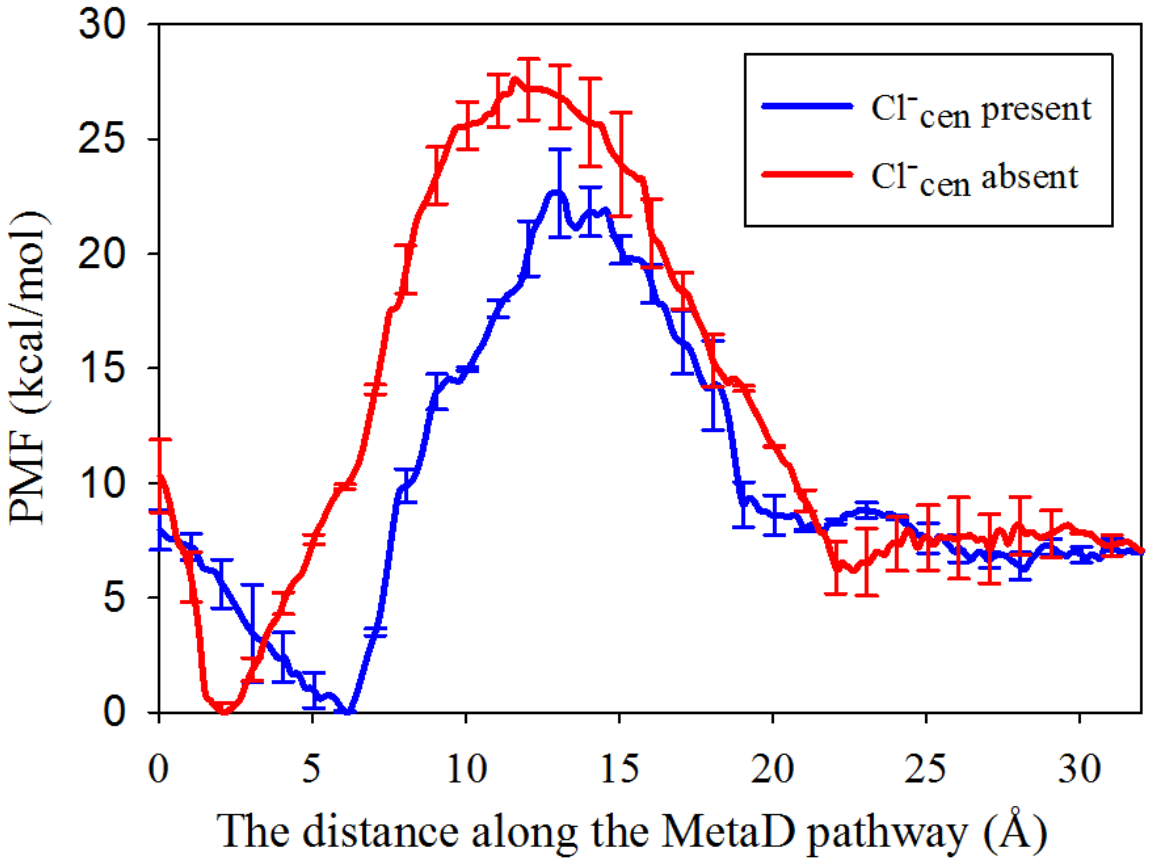


Figure 3.6. The PMF for PT in E148A mutant, from E203 to the extracellular region, with Cl^-_{cen} present (blue) and absent (red).

3.4. Conclusions

In order to answer some of unsolved issues from our previous work (21), about the PT mechanism in the central region of ClC-ec1, we extended our efforts to perform the MS-RMD simulations for the next PT step from E148 to the extracellular region. This step consists of two intermediate steps, which include the rotation of E148 from down to up conformation, followed by deprotonation of E148 to the extracellular region. Because the MS-RMD model for E148 depend on the conformational changes, the final PMF was calculated by parameterizing two

different MS-RMD models for E148 down and up conformations, calculating the PMF for each region, then merged into a single PMF.

Our calculations of the PMF and the PT rate constant showed that, the key mechanism of the Cl^-/H^+ coupling in the overall process is that Cl^-_{cen} facilitates the deprotonation of E148. The calculated PT rate constant was comparable to the experimental result corresponding to the overall PT rate, indicating that this PT step can be the rate-limiting step for the overall process. However, when Cl^-_{cen} is absent, E148 favors the down conformation and is likely to be bound to the S_{cen} site, which increases the free energy barrier for the overall PT to the extracellular region and decreases the PT rate constant below the experimentally measurable range. The Cl^-/H^+ exchange mechanism is reversible in experiment, and we showed that our mechanism holds when the PT steps are described in the opposite direction. The structures from our simulations at the local energy minima in the PMF are well overlapped with several X-ray crystal structures, showing the conformational change of E148.

The PMF for PT was also calculated in E148A mutant, where E203 transfer a proton to the extracellular region. The PMF and the PT rate constants showed that, the free energy barrier for PT is high enough to block PT in both directions, regardless of the presence of Cl^-_{cen}. The simulation results also agree with the behavior of E148A mutant found in experiment, where PT is blocked, while Cl^- is still passively transported through the protein.

Following previous work, the PMF calculation with the MS-RMD model successfully revealed some of more detailed information of the PT mechanism in ClC-ec1, of not only WT, but also E148A mutant of this study, and E203K mutant of previous work. The PT rate constant calculated from the PMF provided us the time scale of each intermediate step, and eventually enabled to obtain the overall PT mechanism. However, it is worth noting that the overall PT rate

was measured in experiment with the concentration gradient applied in the system. However, in the simulations, we have the assumptions that the intermediate PT steps occur in the protein interior, and the rate constants of those steps are free from the concentration in the solution. In order to obtain a more realistic model, where the concentration gradient is also taken account in, our future efforts will aim to calculate more intermediate steps, which would be more sensitive to the concentration in the solution, such as PT from the solution to E148 or E203, and eventually obtain a full kinetic map, showing the overall Cl^-/H^+ exchange process, in the network of all intermediate steps.

3.5. References

1. Pusch, M. 2004. Structural insights into chloride and proton-mediated gating of CLC chloride channels. *Biochemistry* 43:1135-1144.

2. Chen, T. Y. 2005. Structure and function of ClC channels. *Annu. Rev. Physiol.* 67:809-839.

3. Maduke, M., C. M. and, and J. A. Mindell. 2000. A Decade of CLC Chloride Channels: Structure, Mechanism, and Many Unsettled Questions. *Annu. Rev. Biophys. Biomol. Struct.* 29:411-438.

4. Jentsch, T. J., I. Neagoe, and O. Scheel. 2005. CLC chloride channels and transporters. *Curr. Opin. Neurobiol.* 15:319-325.

5. Miller, C. 2006. ClC chloride channels viewed through a transporter lens. *Nature* 440:484-489.

6. Jentsch, T. J. 2015. Discovery of CLC transport proteins: cloning, structure, function and pathophysiology. *The Journal of Physiology* 593:4091-4109.

7. Accardi, A., and C. Miller. 2004. Secondary active transport mediated by a prokaryotic homologue of ClC Cl- channels. In Nature. 803-807.

8. Nguitragool, W., and C. Miller. 2006. Uncoupling of a CLC Cl−/H+ Exchange Transporter by Polyatomic Anions. *J. Mol. Biol.* 362:682-690.

9. Walden, M., A. Accardi, F. Wu, C. Xu, C. Williams, and C. Miller. 2007. Uncoupling and Turnover in a Cl−/H+ Exchange Transporter. *J. Gen. Physiol.* 129:317-329.

10. Lim, H. H., T. Shane, and C. Miller. 2012. Intracellular proton access in a Cl(-)/H(+) antiporter. *PLoS Biol.* 10:e1001441.

11. Basilio, D., K. Noack, A. Picollo, and A. Accardi. 2014. Conformational changes required for H(+)/Cl(-) exchange mediated by a CLC transporter. *Nat. Struct. Mol. Biol.* 21:456-463.

12. Accardi, A., M. Walden, W. Nguitragool, H. Jayaram, C. Williams, and C. Miller. 2005. Separate ion pathways in a Cl-/H+ exchanger. *J. Gen. Physiol.* 126:563-570.

13. Dutzler, R., E. B. Campbell, and R. MacKinnon. 2003. Gating the selectivity filter in ClC chloride channels. *Science* 300:108-112.

14. Miller, C., and W. Nguitragool. 2009. A provisional transport mechanism for a chloride channel-type Cl−/H+ exchanger. *Phil. Trans. R. Soc. B* 364:175-180.

15. Lim, H.-H., R. B. Stockbridge, and C. Miller. 2013. Fluoride-dependent interruption of the transport cycle of a CLC Cl−/H+ antiporter. *Nat. Chem. Biol.* 9:721-725.

16. Accardi, A., S. Lobet, C. Williams, C. Miller, and R. Dutzler. 2006. Synergism between halide binding and proton transport in a CLC-type exchanger. *J Mol Biol* 362:691-699.

17. Feng, L., E. B. Campbell, Y. Hsiung, and R. MacKinnon. 2010. Structure of a eukaryotic CLC transporter defines an intermediate state in the transport cycle. *Science* 330:635-641.

18. Feng, L., E. B. Campbell, and R. MacKinnon. 2012. Molecular mechanism of proton transport in CLC Cl-/H+ exchange transporters. *Proc Natl Acad Sci U S A* 109:11699-11704.

19. Ko, Y. J., and W. H. Jo. 2010. Secondary water pore formation for proton transport in a ClC exchanger revealed by an atomistic molecular-dynamics simulation. *Biophys. J.* 98:2163-2169.

20. Han, W., R. C. Cheng, M. C. Maduke, and E. Tajkhorshid. 2014. Water access points and hydration pathways in CLC H+/Cl- transporters. *Proc Natl Acad Sci U S A* 111:1819-1824.

21. Lee, S., J. M. Swanson, and G. A. Voth. 2016. Multiscale Simulations Reveal Key Aspects of the Proton Transport Mechanism in the ClC-ec1 Antiporter. *Biophys. J.* 110:1334-1345.

22. Lee, S., R. Liang, G. A. Voth, and J. M. Swanson. 2016. Computationally Efficient Multiscale Reactive Molecular Dynamics to Describe Amino Acid Deprotonation in Proteins. *J Chem Theory Comput* 12:879-891.

23. Elvington, S. M., C. W. Liu, and M. C. Maduke. 2009. Substrate-driven conformational changes in ClC-ec1 observed by fluorine NMR. *The EMBO Journal* 28:3090-3102.

24. Bell, S. P., P. K. Curran, S. Choi, and J. A. Mindell. 2006. Site-Directed Fluorescence Studies of a Prokaryotic ClC Antiporter†. *Biochemistry* 45:6773-6782.

25. Khantwal, C. M., S. J. Abraham, W. Han, T. Jiang, T. S. Chavan, R. C. Cheng, S. M. Elvington, C. W. Liu, I. I. Mathews, R. A. Stein, H. S. McHaourab, E. Tajkhorshid, and M. Maduke. 2016. Revealing an outward-facing open conformational state in a CLC Cl–/H+ exchange transporter. *eLife* 5:e11189.

26. Miloshevsky, G. V., A. Hassanein, and P. C. Jordan. 2010. Antiport mechanism for Cl(-)/H(+) in ClC-ec1 from normal-mode analysis. *Biophys. J.* 98:999-1008.

27. MacKerell, A. D., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* 102:3586-3616.

28. Feller, S. E., and A. D. MacKerell. 2000. An Improved Empirical Potential Energy Function for Molecular Simulations of Phospholipids. *J. Phys. Chem. B* 104:7510-7515.

29. Wu, Y., H. Chen, F. Wang, F. Paesani, and G. A. Voth. 2008. An improved multistate empirical valence bond model for aqueous proton solvation and transport. *J. Phys. Chem. B* 112:467-482.

30. Nelson, J. G., Y. Peng, D. W. Silverstein, and J. M. Swanson. 2014. Multiscale Reactive Molecular Dynamics for Absolute pKa Predictions and Amino Acid Deprotonation. *J. Chem. Theory Comput.* 10:2729-2737.

31. Faraldo-Gomez, J. D., and B. Roux. 2004. Electrostatics of ion stabilization in a ClC chloride channel homologue from Escherichia coli. *J. Mol. Biol.* 339:981-1000.

32. Laio, A., and M. Parrinello. 2002. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* 99:12562-12566.

33. Yamashita, T., Y. Peng, C. Knight, and G. A. Voth. 2012. Computationally Efficient Multiconfigurational Reactive Molecular Dynamics. *J. Chem. Theory Comput.* 8:4863-4875.

34. Plimpton, S. 1995. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* 117:1-19.

35. Tribello, G. A., M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi. 2014. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* 185:604-613.

36. Zhang, Y., and G. A. Voth. 2011. Combined Metadynamics and Umbrella Sampling Method for the Calculation of Ion Permeation Free Energy Profiles. *J. Chem. Theory Comput.* 7:2277-2283.

37. Sugita, Y., A. Kitao, and Y. Okamoto. 2000. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* 113:6042-6051.

38. Kumar, S., J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. 1992. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13:1011-1021.

39. VandeVondele, J., M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter. 2005. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* 167:103-128.

40. Becke, A. D. 1988. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* 38:3098-3100.

41. Lee, C., W. Yang, and R. G. Parr. 1988. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* 37:785-789.

42. Schäfer, A., C. Huber, and R. Ahlrichs. 1994. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.* 100:5829-5835.

43. Grimme, S., J. Antony, S. Ehrlich, and H. Krieg. 2010. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* 132:154104.

44. Laino, T., F. Mohamed, A. Laio, and M. Parrinello. 2006. An efficient linear-scaling electrostatic coupling for treating periodic boundary conditions in QM/MM simulations. *J. Chem. Theory Comput.* 2:1370-1378.

45. Laino, T., F. Mohamed, A. Laio, and M. Parrinello. 2005. An efficient real space multigrid OM/MM electrostatic coupling. *J. Chem. Theory Comput.* 1:1176-1184.

46. Blochl, P. E. 1995. Electrostatic Decoupling of Periodic Images of Plane-Wave-Expanded Densities and Derived Atomic Point Charges. *J. Chem. Phys.* 103:7422-7428.

47. Maseras, F., and K. Morokuma. 1995. Imomm - a New Integrated Ab-Initio Plus Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition-States. *J. Comput. Chem.* 16:1170-1179.

48. Liang, R. B., H. Li, J. M. J. Swanson, and G. A. Voth. 2014. Multiscale simulation reveals a multifaceted mechanism of proton permeation through the influenza A M2 proton channel. *Proc. Natl. Acad. Sci. U. S. A.* 111:9396-9401.

49. Chandler, D. 1987. Introduction to modern statistical mechanics. Oxford University Press, New York. 234-270.

50. Roux, B., O. S. Andersen, and T. W. Allen. 2008. Comment on "Free energy simulations of single and double ion occupancy in gramicidin A" [J. Chem. Phys.126, 105103 (2007)]. *J. Chem. Phys.* 128:227101.

51. Yin, J., Z. Kuang, U. Mahankali, and T. L. Beck. 2004. Ion transit pathways and gating in ClC chloride channels. *Proteins: Struct. Funct. Bioinf.* 57:414-421.

52. Lisal, J., and M. Maduke. 2008. The ClC-0 chloride channel is a 'broken' Cl-/H+ antiporter. *Nat. Struct. Mol. Biol.* 15:805-810.

53. Picollo, A., M. Malvezzi, J. C. Houtman, and A. Accardi. 2009. Basis of substrate binding and conservation of selectivity in the CLC family of channels and transporters. *Nat. Struct. Mol. Biol.* 16:1294-1301.

54. Picollo, A., Y. Xu, N. Johner, S. Berneche, and A. Accardi. 2012. Synergistic substrate binding determines the stoichiometry of transport of a prokaryotic H(+)/Cl(-) exchanger. *Nat. Struct. Mol. Biol.* 19:525-531, S521.

Chapter 4

Concluding Remark

A series of work have been done to investigate the PT mechanism of ClC-ec1 antiporter in a molecular level. First, the MS-RMD model was developed to explicitly simulate the proton transport (PT) in large biological systems, which involves in the protonation/deprotonation reactions of Glu or Lys residues in the protein interior, along with the Grotthuss shuttling of the excess proton through surrounding water molecules. In order to benchmark the performance of the model, once the MS-RMD model was parameterized by using the QM/MM reference force data, the same free energy profile (the potential of mean forces; PMF) of PT between E203 and E148 was calculated separately, using both the MS-RMD model and the reference QM/MM method, and the two PMFs by different models generally agreed with each other over all range. The advantage of using the MS-RMD model is the efficiency, where its MD simulation speed was faster than typical QM/MM MD methods by several orders of magnitudes. By the virtue of the reduced computational cost, it became feasible to sample trajectories of the reactive MD in the nanosecond time scale in a large system, and calculate the PMF with the multi-dimensional umbrella sampling, which requires to sample over a large number of sampling windows.

Second, the PT mechanism of the central region, where PT occurs from E203 to E148, was studied by calculating several two-dimensional PMFs with different sets of the collective variables (CV): the one was the effective position of the excess proton (defined as the center of the excess charge (CEC) in the simulation), and the other was chosen to be either the water density (the effective number of water molecules) in the central region, or the rotation of E148 and E203. The PMF results revealed several important issues in the PT mechanism in the central

region, which were not fully understood in previous experimental studies. When PT occurs from E203 to E148, the excess proton on E203 is delocalized into the surrounding water molecules, which induces more waters entering the central region, and stabilizes a water wire with continuously linked with hydrogen bond network between E203 and E148, which is used as the proton pathway to E148.

Third, the PMF for the next intermediate step beyond the central region, where PT occurs from E148 to the extracellular solution, was calculated in order to find the rate-determining step for the overall PT process and the molecular mechanism of the Cl^-/H^+ coupling. The PMF was obtained by hybridizing the two PMFs calculated with different MS-RMD models for E148, depending on the up/down conformational changes, since the model depends on the change of the surrounding protein environment. The two PMFs are well overlapped around the boundary region, where E148 changes its conformation, showing that the merged PMF is independent from the choice of the boundary between different models. Based on the PMF results, we proposed the Cl^-/H^+ coupling mechanism as follows: deprotonation of E148 is the rate-limiting step of the overall PT process, which is facilitated by Cl^-_{cen}. When Cl^-_{cen} is absent, protonated E148 goes to the down conformation, and stabilizes at the Cl^-_{cen} binding site, causing that PT is practically blocked, where the PT rate becomes below the experimentally measurable range.

Finally, in addition to the calculation of the PMFs for all intermediate steps for PT, the calculation of the PMFs for Cl^- transport between several intermediate binding sites in the protein and the Brownian dynamics simulation to measure the rate constant of the diffusion of Cl^- from the solution to the protein, are currently on going. Once the rate constants for all intermediate steps for PT and Cl^- transport are obtained, the Markov state model will be built to investigate the overall mechanism of the Cl^-/H^+ exchange process. We expect the model will be

able to provide more detailed explanations on the overall process, for example, how the Cl^-/H^+ exchange ratio is independent from the change of the ionic concentration gradient, or what is the most dominant mechanism of the transports of Cl^- and H^+, while they go through in the network of all intermediate steps.

A series of our approaches shown above can be applied to other systems in the future studies to study the complex mechanism in biological systems, which have many coupled degrees of freedom. For the short-term goal, we can investigate other problems in ClC systems. For example, ClC-0 Cl^- channel has very similar structures with ClC-ec1. Although E148 is located at the same position with ClC-ec1, but experiment showed that the Cl^-/H^+ exchange is blocked. It has been suggested by experiment, that E148 is still protonated in ClC-0, but the PT rate is too slow to be measured by experiment. It would be interesting if we investigate how much the free energy barrier is changed in the PMFs for PT in ClC-0, and the Markov state model can be applied as well in order to see how the overall mechanism is changed in ClC-0, compared to ClC-ec1. For the long-term view, when the Markov state model is developed for each reaction in different protein system, the models from different protein systems can be connected into the system with even a larger scale. For example, ClC-ec1 is one of the transporter in the membrane, which couples the Cl^-/H^+ exchange, and there are many other kinds of other channels or transporters which maintain the concentration of Cl^- or H^+ in a real cell. It may be possible that we simulate the mechanism of the homeostasis of the ionic concentration in a cell, which will be the next level of the system scale beyond the single protein system.