THE UNIVERSITY OF CHICAGO

PROTON TRANSPORT IN MOLECULAR SYSTEMS: SAMPLING, MODELS, AND APPLICATIONS

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

CHENGHAN LI

CHICAGO, ILLINOIS

DECEMBER 2021

To my family

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would want to express my profound gratitude to Professor Gregory A. Voth, my Ph.D. adviser, for his constant support and excellent academic training during my graduate school years. Professor Voth has given me the latitude to pursue different new avenues in scientific topics that I was interested in, and he has been extremely encouraging and supportive of my research ideas. I have thoroughly enjoyed my time working with him.

I would also like to thank Professor Jessica M. J. Swanson for her guidance and mentorship throughout my early years of graduate school, and Professor Simon Newstead for our research collaboration.

It is my pleasure to have interactions with the past and current members of the Voth group, Dr. Paul B. Calio, Dr. Yining Han, Dr. Yuxing Peng, Dr. Zhi Wang, Dr. Zhi (Shane) Yue, Dr. Zheful Li, Siyoung Kim, and Yu (Grace) Liu. I have been having a great time with them, engaging in helpful and insightful discussions on science as well as delightful conversations on life and other topics with them.

Aside from that, I want to express my gratitude to my parents, Xusheng Li and Qiong Zhou, for their unconditional love and support.

Abstract

Proton transport (PT) is ubiquitous in aqueous and biomolecular systems. One possible and popular simulation method to model PT reactions is *ab initio* molecular dynamics (AIMD) which computes the electronic structure explicitly on the fly. The generalized gradient approximated (GGA) density functional theory (DFT) represents a widely used and computationally tractable electronic structure method in AIMD, but compromises the accuracy of modeling charge delocalization and hydrogen bond strength. A correction built upon the experimental directed simulation (EDS) technique to hydrogen bonds of water and hydrated proton was developed and used to interpret the long-lived proton anisotropy observed in two-dimension infra-red (2D-IR) experiments. Due to the light mass of protons, the nuclear quantum effects (NQEs) are commonly thought to be important for describing proton and water solvation structures and dynamics. The use of a machine learning (ML) potential in conjunction with the ring polymer contraction (RPC) scheme was presented, as a method for achieving accurate and efficient quantum simulation at the same computing cost as classical simulations. The method was used to demonstrate that a typical approach of increasing simulation temperature in classical AIMD is ineffective in simulating the missing NQEs.

As an alternative approach, the multiscale reactive molecular dynamics (MS-RMD) method benefiting greatly from its computational efficiency serves as a powerful simulation tool for proton solvation and transport. Constrained DFT (CDFT), a diabatic electronic structure method, was used to establish a systematic procedure for parameterizing MS-RMD models. The pK_a's of residues in both water and protein environments are reliably predicted by the amino acid models that were parametrized to match the CDFT charge transfer behavior.

Proton transport reactions usually involve high free energy barriers, and for sufficient sampling of such rare events, importance sampling techniques are often required to accelerate key collective motions relevant to PT processes. Commonly used enhanced sampling methods achieve the acceleration by applying bias potentials on one or more degrees of freedom (DOFs) of the system which are referred to as collective variables (CVs). In the context of PT processes, the most relevant CVs are indeed the position of the excess proton, as well as the associated hydration required for a proton to travel through hydrophobic confined systems. A definition of center of excess charge (CEC), the effective position of the excess proton, was proposed based on charge transfer calculations from CDFT and the encoded proton collective motions were revealed by its IR spectrum. In addition, a differentiable CV was derived to represent the water connectivity in confined space allowing for direct quantification of the connecting between PT and water networks.

PT and how it is connected to related collective motions in various complicated systems were explored thanks to advancements in simulation models and sampling techniques. The case studies include the proton intake mechanism via Sacro/endoplasmic reticulum Ca^{2+}-ATPase (SERCA), the proton-hydration coupling in a Cl^-/H^+ antiporter, ClC-ec1, and the proton-ligand-conformation coupling in a member of the proton-coupled oligo-peptide transporters (POTs), $PepT_{Sh}$.

1 Introduction

1.1 Proton solvation and transport

Protons solvate as hydronium ions (H_3O^+) by bonding to a water molecule in a simplified diagram commonly found in general chemistry textbooks, although they are better described by dominant species such as Eigen ($H_9O_4^+$)[1] and Zundel ($H_5O_2^+$)[2]. The Eigen cation has a core hydronium that is surrounded by three tightly hydrogen-bonded water partners, whereas the Zundel cation has a proton shared by two waters. Although their chemical compositions appear to be separate, Eigen and Zundel are not structurally distinct in solution. The reason for this is that a Zundel can be thought of as a motif of an Eigen (the hydronium plus the strongest hydrogen-bonded water), and an Eigen can be thought of as the complex created by a Zundel with the other two hydrogen-bonded solvation water molecules. Experimental[3] and computational studies[4] have revealed that a geometrically perfect Zundel (with the proton perfectly between the two waters) is rarely found in solutions, with the shared proton favoring one of the waters being more common. When only two water molecules are considered, the proton preference suggests a distorted Zundel. Eigen has a similar asymmetry, in which the three solvation waters are typically not equivalent, meaning that one of the waters forms a stronger hydrogen bond with the central hydronium, while the other two waters are further away, resulting in distorted Eigen.[5] When these distortions are considered, the distinction between Eigen and Zundel gets further blurred, and the distorted Eigen and distorted Zundel appear to be just two synonyms for the same hydrated proton solvation structure. This static, geometric picture, on the other hand, ignores the dynamics of the system. The strongest hydrogen-bonded hydronium-water pair (the special pair) in an Eigen cation is dynamically rotating among the three waters of the Eigen species, according to computational studies[5]. While there is always only one special pair in a distorted Eigen at a given moment, its identity changes

over time, and different waters can be identified as the special partner at various points in time. While two additional solvation waters can always be detected for a Zundel (which structurally resembles an Eigen), these two waters do not become the hydronium's special partner, and a proton is only rattled between the two core waters. As a result, such Eigen/Zundel dynamics can be used to define Eigen and Zundel, resolving the geometrical ambiguity. In the special-pair dance dynamics, the special partner is found to rotate among all three first solvation waters of H_3O^+ in a tens-of-fs timescale[5], making them undistinguishable in a long time limit. Hence, the three waters are better all included in the description of the hydrated proton complex, resulting in the conventional Eigen cation formula ($H_3O^+ \cdot 3H_2O$). Because only one special partner is engaged in proton rattling, the associated proton species is designated by the Zundel formula ($H_3O^+ \cdot H_2O$).

One long-debated topic in proton solvation is which form of a hydrated proton, the Eigen or the Zundel, is the most populated. Early simulations suggested a distorted Zundel picture, in which the hydronium structure formed a special pair with a strongly hydrogen-bonded "special partner"[6, 7]. Later computational studies discovered a distorted Eigen structure to be dominant in which the special partner identity dynamically switches among the three hydrogen-bonded water molecules of H_3O^+[5]. Recent nonlinear infra-red (IR) spectroscopy experiments, however, revealed the possibility of Zundel to be the dominant species. Thamaer and co-authors[8] estimated the population of Zundel to be 40% of the total solvated protons in 4 M HCl solution by (1) identifying the water/Zundel bending frequency to be 1650 cm^{-1} and 1760 cm^{-1} respectively from two-dimensional IR spectroscopy; (2) computing the IR intensity ratio at Zundel bending frequency to the intensity at water frequency in the linear spectrum; (3) converting the intensity ratio to the concentration ratio by assuming a 10-fold larger transition dipole of Zundel bending compared to water bending. Carpenter and coauthors[9, 10] determined the anisotropic decay of the special pair

water bending mode using 2D-IR spectroscopy and discovered that the timescales of the decay are significantly longer than those of the special pair dance, the characteristic dynamics of Eigen cations. The long timescales were used as one of the key pieces of evidence to support the dominance of Zundel cations. A follow-up 2D-IR spectroscopy work by Fournier and coauthors[3] examined the lineshape of the stimulated emission of the proton stretching mode and proposed the "potential energy surface" for the proton transfer is asymmetric, revealing the distortion in the hydrated proton complex.

There are fewer disagreements over the proton transport (PT) mechanism and the Grotthuss mechanism[11], which states the excess proton (H^+) is transported by breaking and forming O-H bonds and rearranging covalent and hydrogen bonds, is widely recognized. However, the precise mechanism is dependent on whether Eigen or Zundel is the dominating hydrated proton species. Simulations[5] suggested an Eigen-Zundel-Eigen (EZE) transport mechanism in which a Zundel-like configuration serves as the transition structure for a distorted Eigen to transfer the excess proton to its neighboring water and produce a new Eigen. The recent 2D-IR experiments[3], in contrary, supposed a Zundel-Zundel transport mechanism to be consistent with the proposed distorted Zundel picture.

1.2 Proton transport in biomolecular systems

Protons are transported via water, titratable ligands, and ionizable residues in proteins, and the Grotthuss hopping mechanism in water is altered by replacing one or more water molecules with titratable moieties that accept and/or donate protons. Due to its ability to alter electrostatic and hydrogen bonding interactions, PT is a critical component of the functional cycle of various channels, enzymes, and transporters.[12-19]

The Ca^{2+} pump SERCA transports proton in the opposite direction of Ca^{2+} to maintain the charge imbalance in its Ca^{2+} binding site.[20, 21] This proton flow via SERCA to the cytosol must be compensated by a proton flow back to the reticulum lumen for the pH homeostasis of reticulum. Previous simulations suggested that a hydrophobic pore connecting the Ca^{2+} binding site to the lumen identified in the regulated state of SERCA provides a potential proton pathway.[22] Transient hydration with lifetime of ~100 ps is observed in the pore between a binding site residue E908 and luminal bulk bypassing the H944 in the middle. These findings shed new light on the proton transport from cytosol to lumen, and revealed the possibility that SERCA intakes protons to reticulum lumen in additional to the well-known PT to the cytosol.

As an another example, POT family proteins are oligo-peptide transporters utilizing the cross-membrane proton gradient as the driving force.[23, 24] The transporter features two gates located at the extracellular and intracellular side respectively, and their opening and closing are considered to be controlled by proton movements. The transport of ligands is achieved by the alternating access mechanism[25, 26] where the extracellular gate and intracellular gate are opened alternatively and allow the incoming of ligand from extracellular side and the release of ligand into the intracellular environment. Structural and mutagenesis studies[27-29] identified several residues to be the proton binding site but how the proton drives the whole functional cycle is not clear yet.

1.3 Modeling explicit proton solvation and transport

The many-body interaction between atoms are typically decomposed into bonded and non-bonded potentials in conventional molecular mechanics force fields (FF). The bonded interactions include bonds, angles, and torsions are modeled by simple energy functions such as springs, sines and cosines. The bonding topology is pre-defined and cannot be changed in the course of simulation,

and thus excess protons are modeled as classical "hydroniums" but charge delocalization or explicit proton transfer cannot be represented appropriately.

Due to the reactive nature of PT, the ab initio molecular dynamics (AIMD)[30] becomes a suitable choice since the electronic structure is calculated explicitly and the nuclei are evolved on the potential energy surface (PES) formed by the electronic ground state. The adiabatic separation between nuclei and electrons is assumed and such treatment is termed as the Born-Oppenheimer (BO) approximation. In principle, any electronic structure method that computes the energy gradient can be used to drive the MD, while in practice, DFT at the level of generalized gradient approximation (GGA) has been the common choice as a reasonable balance between accuracy and computational efficiency. The GGA, as an approximation to the exact exchange-correlation functional suffers an overestimation of intermolecular charge transfer, resulting into too strong hydrogen bonds and underestimated PT barriers.[31] One promising solution is to employ a force-field like correction to the functional that corrects the hydrogen bond strength while introducing negligible additional computational cost. Pitera and Chodera[32] used the maximum entropy principl to demonstrate that a linear form of potentials forms a minimal bias that makes the corrected PES reproduces observables of interest to specified reference values. A follow-up work by White and coauthors[33, 34] utilized the experimental oxygen-oxygen radial distribution function (RDF) as reference to determine a correction to a GGA-DFT description of water. The bias scheme and its parametrization procedure via stochastic gradient descent was then termed as the experiment directed simulation (EDS). Despite of this name, the reference data is not limited to experimentally determined quantities. Calio and coauthors[35] developed a new correction to DFT water making use of the information of oxygen-hydrogen RDF calculated from MB-Pol simulations. The MB-Pol is a highly accurate water model parametrized from coupled cluster (CC) theory, and was found to

reproduce various water properties.[36-38] The correction was found to greatly improve the solvation structures such as O-H and O-O RDFs as well as the O-O-O three body correlation, and dynamical properties such as the water self-diffusion constant.

Even at the compromising GGA-DFT level, the high demanding AIMD limits the temporal and spatial scales it can be applied to. One useful technique for applying AIMD to large biomolecular systems is the hybrid quantum mechanics/molecular mechanics (QM/MM) method, which partitions the system into a reactive part treated by quantum mechanics and a non-reactive environment treated by classical mechanics. When calculating the Coulombic interaction between QM electrons and MM partial charges, the MM charges are typically modeled as a Gaussian distributed charge clouds to mimic electron-screened nuclei.[39] Conventional QM/MM requires QM and MM atom identities predefined and unchanged over the course of MD simulation, making it difficult to handle the situations in which the reaction center translocates a long distance beyond the ranges of the QM box, which could be the case of PT between two distant ionizable amino acid residues in a protein channel.

As an alternative approach, the MS-RMD method[40, 41] developed in the Voth group benefits from its computational efficiency, representing a promising way of simulating proton solvation and transport in large scales. The basic idea of MS-RMD is to expand the ground state of the system on the basis of a number of valence bond (VB) resonance states (diabatic states), and to write the Hamiltonian of the system in the VB representation:

$$\mathbf{H} = \sum_{ij} |i\rangle h_{ij} \langle j| \qquad 1\text{-}1$$

Each VB state $|i\rangle$ corresponds to a distinct bonding topology that can be described by molecular mechanics, so the diagonal energies h_{ii}'s are computed from a classical FF. The off-diagonal h_{ij} is the overlap between two VB states and describes the proton transfer mechanism between the

proton donor and acceptor pair in the two states. The h_{ij} is assumed to take a FF-like form, usually a physics inspired ansatz, but can also be represented by a flexible neural network (NN). In each timestep of the MS-RMD simulation, a state search algorithm is first evoked to find the possible resonance states that contribute non-negligibly to the Hamiltonian. This is usually achieved by a search of up to three solvation shells of the central hydronium or protonated amino acid and rearrange the bonding topology via a chain of proposed proton transfers. The ground-state energy of the system is obtained by the eigen-value problem

$$\mathbf{H}\boldsymbol{c} = E\boldsymbol{c} \tag{1-2}$$

where $\boldsymbol{c} = \{c_i\}$ is the ground-state vector in VB representation. The forces acting on atoms ($\boldsymbol{F} = -\nabla E$) are computed by noting that $E = \sum_{ij} c_i c_j h_{ij}$ and applying the Hellman-Feynman theorem,

$$\boldsymbol{F} = \sum_{ij} c_i c_j \boldsymbol{F}_{ij} \tag{1-3}$$

where $\boldsymbol{F}_{ij} = -\nabla h_{ij}$.

1.4 Enhanced sampling techniques and collective variables

The classical dynamics of a system with Hamiltonian $H = \boldsymbol{p}^T \mathbf{M}^{-1} \boldsymbol{p}/2 + V(\boldsymbol{r})$ follows the equations of motion

$$\dot{\boldsymbol{p}} = -\frac{\partial H}{\partial \boldsymbol{r}} \tag{1-4}$$

$$\dot{\boldsymbol{r}} = \frac{\partial H}{\partial \boldsymbol{p}} \tag{1-5}$$

In the case of position-independent, diagonal mass matrix **M**, the equations of motion degenerate to the Newton's second law for any atom *I*

$$m_I \ddot{\boldsymbol{r}}_I = -\frac{\partial V}{\partial \boldsymbol{r}_I} = \boldsymbol{F}_I \tag{1-6}$$

When considering a system embedded in a thermal reservoir, the dynamics of the system is affected by the surrounding environment and thus deviated from the above Hamiltonian dynamics. One possible way to model the thermostatted dynamics under temperature T is the Langevin equation

$$m_I \ddot{\boldsymbol{r}}_I = -\gamma \dot{\boldsymbol{r}}_I + \boldsymbol{F}_I + \sqrt{2\gamma k_B T}\frac{\mathrm{d}\boldsymbol{B}_t}{\mathrm{d}t} \tag{1-7}$$

where γ is the friction coefficient, k_B is the Boltzmann constant and $\boldsymbol{B}_t$ is the standard Brownian motion. The stationary distribution generated by the Langevin dynamics is the Boltzmann distribution

$$\rho(\boldsymbol{r}) \propto \exp\left(-\frac{V(\boldsymbol{r})}{k_B T}\right) \equiv \exp\big(-\beta V(\boldsymbol{r})\big) \tag{1-8}$$

With the ergodicity assumption, any thermodynamical observables are approximated by the time averages along a MD simulation trajectory $\{\boldsymbol{r}(\tau)|0 \le \tau \le t\}$:

$$\langle A \rangle \equiv \frac{\int A(\boldsymbol{r}) \exp\big(-\beta V(\boldsymbol{r})\big)\,\mathrm{d}\boldsymbol{r}}{\int \exp\big(-\beta V(\boldsymbol{r})\big)\,\mathrm{d}\boldsymbol{r}} \approx \frac{1}{t}\int_0^t A\big(\boldsymbol{r}(\tau)\big)\mathrm{d}\tau \tag{1-9}$$

In particular, when the equilibrium distribution of a collective variable (CV) $\xi(\boldsymbol{r})$ is of interest, we have

$$\rho(\xi') = \langle \delta(\xi(\boldsymbol{r}) - \xi') \rangle = \frac{\int \delta(\xi(\boldsymbol{r}) - \xi') \exp\big(-\beta V(\boldsymbol{r})\big)\,\mathrm{d}\boldsymbol{r}}{\int \exp\big(-\beta V(\boldsymbol{r})\big)\,\mathrm{d}\boldsymbol{r}} \tag{1-10}$$

where $\delta(x - x_0)$ is the Dirca delta function. The free energy, or termed as the potential of mean force (PMF) of ξ is defined as

$$F(\xi) = -k_B T \ln\big(\rho(\xi')\big) \tag{1-11}$$

In practice, high free energy region precludes a brute-force sampling to quantify $F(\xi)$ accurately, the sampling time needed in eq 1-9 will far surpass the computation capability of modern computers. In such scenario, enhanced sampling techniques become advantageous for

bringing down the sampling cost associated with rare events. One major category of enhanced sampling is the replica exchange (RE) method.[42] The basic idea of RE is to construct a series of auxiliary Hamiltonians $\{H_i\}$ with smoother free energy landscapes compared to H_0, and to perform exchange between the Hamiltonians to assist H_0 in overcoming the free energy barriers. The most common choice of $\{H_i\}$ is the scaled potential, $H_i = \alpha_i H_0$ where $\alpha_i < 1$, which is equivalent to sample Boltzmann distributions in higher temperatures according to eq 1-8. It is clear that the exchange between replicas cannot be arbitrary, or otherwise the Boltzmann sampling associated with the original H_0 cannot be recovered. Commonly, the exchange is performed by a Metropolis Monte Carlo to ensure that the replica of H_0 samples the correct Boltzmann distribution. In such exchange scheme, the more overlap between H_i's, the higher rate of exchange will be. Because the exchange rate is considered to be a crucial measure for RE's efficiency, the differences between H_i's need to be sufficiently small. As a result, RE is powerful if the free energy landscape is relatively smoothly varying but may not be very efficient confronted with high free energy barriers.

Another category is the CV-based methods. The umbrella sampling (US)[43], metadynamics (MTD)[44] and adaptive biasing force (ABF)[45] are representatives. These methods employ a bias potential or force on some chosen degrees of freedom (DOFs), i.e. CVs, to accelerate the sampling along those directions. For the sake of simplicity, we consider the bias is a potential $u(\xi)$ instead of a force, which is the case of US and MetaD. The basic idea is to construct $u(\xi)$ so as to bring down the higher free energy region so that the rare but important events are more frequently visited. The sampled ensemble with $u(\xi)$ is thus deviated from the Boltzmann distribution but a simple correction is available as long as $u(\xi)$ is additive to the Hamiltonian,

$$\rho(\boldsymbol{r}) \propto \rho^{\text{biased}}(\boldsymbol{r}) e^{\beta u(\xi(\boldsymbol{r}))} \qquad 1\text{-}12$$

If the marginal distribution of ξ is of interest,

$$\rho(\xi') = \frac{\int \delta(\xi(\boldsymbol{r}) - \xi') e^{-\beta V(\boldsymbol{r}) - \beta u(\xi)} e^{\beta u(\xi)} \mathrm{d}\boldsymbol{r}}{\int e^{-\beta V(\boldsymbol{r})} \mathrm{d}\boldsymbol{r}} \propto \rho^{\mathrm{biased}}(\xi') e^{\beta u(\xi')} \qquad 1\text{-}13$$

The sampling efficiency strongly depends on the choice of CVs. In general, CVs that carry as much as information are appreciated, and in this sense, choosing CVs is essentially a dimension reduction problem from the full phase space of $\boldsymbol{r}$ into a low-dimensional CV space of $\boldsymbol{\xi} = \{\xi_1, \xi_2, \xi_3, \ldots\}$. The reduction methods can be linear, such as principal component analysis (PCA) [46, 47], and time-lagged independent component analysis (tICA)[48, 49], as well as nonlinear, such as auto-encoders[50] and their time-lagged variants[51]. In the context of PT, rough guesses on the required motions to be biased are typically available *a priori*, e.g. the position of the excess proton (H^+) is needed to track the PT progress. Since the frequent bond changing for condensed phase PT, the excess proton is not associated with any specific hydrogen, so instead of tracking the position of a given proton, the center of electron defect introduced by the H^+ is a more appropriate choice. The latter one is also called the center of excess charge (CEC)[52]. In the case of PT through hydrophobic region, such as a proton channel or synthesized nano-pores, the hydration is clearly another DOF crucial to proton solvation and transport, since a naked H^+ or H_3O^+ is by no means to travel through vacuum without the assistance of solvation water.

1.5 Outline of thesis

The work presented in this thesis continues to develop and use reactive MD approaches, including AIMD and MS-RMD, combined with enhanced free energy sampling, to study proton solvation and transport in aqueous as well as in biomolecular systems. The remaining chapters are organized as follows:

Chapter 2 summarizes several developments in reaction coordinates that are essential for performing enhanced sampling of explicit proton transport simulations. Chapters 3 and 4 include advances in methodology of simulating and modelling proton transport. To be more specific,

Chapter 3 introduces the EDS correction developed for proton transport in water and presents an approach of accelerating path integral methods via machine learning (ML) to include the nuclear quantum effects (NQEs). Chapter 4 presents accurate and transferrable MS-RMD models for amino acids derived from an *ab initio* diabatic method. In Chapter 5, the widely used approach of elevating simulation temperature in AIMD to model NQEs was examined and benchmarked against real quantum simulations. In Chapter 6, I provide the work that used the MS-RMD approach and the EDS-AIMD method to simulate solvated excess proton in water with both classical and quantized nuclei to interpret the recent 2D-IR experiments. In Chapters 7 and 8, the PT-hydration coupling was studied in SERCA, a carbon nanotube (CNT) and ClC-ec1. In Chapter 9, the proton movement, PT-conformation, and PT-ligand coupling in $PepT_{Sh}$ was presented.

2 Reaction Coordinates for Tracking the Excess Proton and the Associated Water Connectivity

2.1 Introduction

In the context of proton transport, a proper reaction coordinate reflecting the position of the hydrated excess proton is arguably of central importance. Due to frequent bond breaking and forming events in Grotthuss proton shuttling, the identity of the charge carrier species (hydronium-like or protonated weak acids) is dynamically changing and the excess protonic charge defect tends to be distributed among several solvation shells instead of localizing on a central hydronium structure or on weak acid. As such, and as discussed in the Introduction Chapter, a CV cannot be associated with any specific "proton" in the system but is more appropriately assigned in some way to be the charge defect associated with the excess proton, often referred to as the "center of excess charge" (CEC). In MS-RMD simulations, due to the available diabatic information, the CEC is well-defined as a weighted average of the diabatic "CEC" of each state:

$$\mathbf{r}_{\mathrm{CEC}} = \sum_i c_i^2 \mathbf{r}_i^{\mathrm{COC}} \qquad 2\text{-}1$$

where the "CEC" in each diabatic state is simply the center of charge (COC) of the moiety that carries the excess proton.

For AIMD simulations, there are several CEC definitions that have been proposed, namely mCEC[53], the proton indicator[54], and the more recent rCEC[55]. However, a more rigorous definition based on *ab initio* theory is preferred. In this section we present a variant of the CEC definition derived from a diabatic electronic structure method, the constrained DFT (CDFT)[56] and apply it to two case studies: an excess proton in water and glutamic acid in water. We conclude by unraveling

the collective motions encoded in the CEC via computing its IR spectrum, as well as illustrating its ability to accelerate the sampling of proton transfer when combined with metadynamics.

Another dimension that is closely related to PT is the water hydration. Stable or at least transient "water wires" are believed to be required for proton permeation through confined regions in these systems by exploiting the Grotthuss proton hopping mechanism. For many years, PT pathways, e.g., in proteins, have been inferred from crystal structures that may show intercalated water molecules and/or from MD simulations that study "water wires" in the absence of explicit PT. Indeed, MD simulations have been used extensively to provide an atomistic description of water permeation and solvation in nonpolar confined spaces. The presence of water alone has been commonly used to predict possible PT pathways, but predicting PT behavior based on hydration alone in the absence of an excess proton in the water structure can be very misleading as we explain below.

Though the explicit PT process including PT-hydration coupling can in principle be simulated in certain cases[57-60] using a method such as MS-RMD (and perhaps direct QM/MM with time), we are still lacking a proper CV to fully describe the hydration component as directly coupled to the PT, which also often involves the formation and breakage of multiple hydrogen-bonded water networks. Thus, breakthroughs in computational methodology in addition to the subtle complexity of the PT processes as described in the paragraph above presents an opportunity to rigorously and quantitatively define a mathematical description of water wire connectivity – and as an excess proton is being transported through it – which is continuously differentiable and thus the CV can be sampled along with the explicit excess proton translocation via an enhanced free energy sampling approach. As such, a proper and general CV can arguably put an end to a great deal of speculation about the relationship between PT processes and internal hydration of confined spaces,

i.e., various speculations and/or partial or incorrect conclusions that have been omnipresent in various fields for many years largely attributable to missing or incomplete information.

There are several challenges associated with defining and identifying a CV for this purpose stated above. In order to ensure the continuous nature of the CV, a smooth transition in its value is required when waters are entering and leaving the confined space or channel. This issue can be problematic due to the discrete nature of the number of water molecules. Moreover, the CV must be invariant to the water molecule identity and thus be unaffected by frequent exchanges of waters between confined and bulk-like water, and even among the waters within the channel. While systematic CV discovery methods such as PCA , tICA, spectral gap optimization of order parameter (SGOOP)[61, 62], auto-encoder/decoder, Markov state model (MSM)[63], and variational approach for learning Markov processes (VAMP)[64, 65] all represent possible approaches for addressing this problem, each requires an input of possible useful descriptors to be linearly or non-linearly combined into the output CVs. Identifying such descriptors can be straightforward in some instances; for example, in protein dynamics where the protein conformation can be well described by dihedral angles and contact maps. In contrast, the analogous approach for water network connectivity (especially when it contains an excess proton as is our present focus) is challenging given that there are no *a priori* descriptors available; additionally, the smoothness and identity exchange invariance are also requirements for the input descriptors. More importantly, existing methods typically seek to identify either the largest fluctuating or kinetically slowest degree of freedom, which provide no guarantee for capturing water network formation or breakage. For example, CVs representing end-to-end distance or the rotation of the entire water wire can potentially be a “learned” output because they are slow motions due to confinement, but they can be minimally or not at all related to the PT behavior.

Herein, we present two CVs derived from graph descriptions[66, 67] of water networks, namely the shortest path length (denoted as $\log(S)$) and the principal curve connectivity (denoted as ϕ). We show that $\log(S)$, although directly derived from graph theory, is inherently non-differentiable and therefore inappropriate for generating free energy profiles. On the other hand, ϕ, derived from "coarse-graining" the water graph (cf. Figure 2-2), can serve as a differentiable alternative of $\log(S)$ and is able to drive efficient free energy sampling of water connectivity via CV-based enhanced sampling methods. In turn, the principal curve connectivity provides the long-sought-after quantitative measure of the PT "capacity" of a given protonated water wire structure in a confined space such as a protein channel or a narrow nanotube. Furthermore, the free energy of forming such a structure can now be calculated and thus the facility of a PT process quantified (i.e., not just speculated upon), as demonstrated in below and in Chapter 8 through several non-trivial examples.

2.2 Methods

2.2.1 Theory of constrained DFT

The CDFT framework was proposed for solving the electronic structure of a system subject to the following constraint on electron density:

$$\int w(\boldsymbol{r})\rho(\boldsymbol{r})\mathrm{d}\boldsymbol{r} = N \tag{2-2}$$

where $w(\boldsymbol{r})$ is a weighting function that defines the constraint, and N is the constraint target value. The constrained lowest-energy state can be obtained from an optimization problem via the standard method of Lagrange multipliers:

$$E(N) = \min_{\rho} \max_{\lambda} \left(E[\rho(\boldsymbol{r})] + \lambda \left(\int w(\boldsymbol{r})\rho(\boldsymbol{r})\mathrm{d}\boldsymbol{r} - N \right) \right) \tag{2-3}$$

Herein, $E[\rho(\boldsymbol{r})]$ is the density functional, which in this Chapter is the Becke-Lee-Yang-Parr functional (BLYP)[68, 69] and ωB97X[70], and λ is the Lagrange multiplier. The electron density determined from eq 2-3 thus deviates from the constraint-free adiabatic ground-state density, making it a so-called diabatic state. As molecules in the condensed phase sample more compact geometries on average, the promolecule formalism approach[56, 71] was employed and the system was partitioned into two molecular fragments, A and B. The constraint target value in eq 2-3 was then calculated from the total promolecule density by summing the ground-state electronic density of the two fragments as if they were independent:

$$N \equiv \int w(\boldsymbol{r})\big(\rho_{\mathrm{A}}(\boldsymbol{r}) + \rho_{\mathrm{B}}(\boldsymbol{r})\big)\mathrm{d}\boldsymbol{r} \qquad 2\text{-}4$$

Here, the Becke population[72] scheme was used to define the weighting function as

$$w(\boldsymbol{r}) = \sum_{I} w_I^{\mathrm{Becke}}(\boldsymbol{r}) \qquad 2\text{-}5$$

where summation index *I* refers to the atoms of protonated species in each diabatic state, i.e., a hydronium or a neutral glutamate. The expected behavior of the promolecule constraint is that the resulting diabatic electronic density will resemble as much as possible the superposition of two pure fragments, such as, e.g., a pure water and a pure hydronium in the case of the Zundel cation $H_5O_2^+$.

The coupling between two diabatic states is calculated from the integral using the Kohn-Sham determinant $|\Phi_i\rangle$[56, 73]

$$H_{12} = \langle\Phi_1|\widehat{H}|\Phi_2\rangle \qquad 2\text{-}6$$

A 2×2 Hamiltonian can be constructed by using the diabatic energies from eq 2-3 as diagonal terms

$$\mathbf{H} = \begin{pmatrix} E_1 & H_{12} \\ H_{21} & E_2 \end{pmatrix} \quad 2\text{-}7$$

Similarly, the overlap matrix is defined as

$$\mathbf{S} = \begin{pmatrix} 1 & \langle \Phi_1 | \Phi_2 \rangle \\ \langle \Phi_2 | \Phi_1 \rangle & 1 \end{pmatrix} \quad 2\text{-}8$$

Then, the so-called CDFT configurational interaction (CI)[56, 71] can be performed by solving the generalized eigenvalue problem:

$$\mathbf{H}\boldsymbol{c} = E\mathbf{S}\boldsymbol{c} \quad 2\text{-}9$$

The resulting eigenfunctions, $\boldsymbol{c} = \{c_1, c_2\}$, determine the degree to which each of the two diabatic states contributes to the CI ground state of the system; we will use the $\boldsymbol{c}$ vector to define our CEC, as described below.

2.2.2 Theory of center of excess charge

The CEC for AIMD can be defined similar to the one in MS-RMD (eq 2-1), with the help of CDFT. Assuming that each diabatic state defines a bonding topology, e.g. in Figure 2-1, state $|1\rangle$ defines a neutral glutamate and two neutral water molecules, while state $|2\rangle$ defines a hydronium, a charged Glu, and a neutral water molecule as two of the possible topologies. Given the bonding topology of every diabatic state, the "diabatic" CEC within state $|i\rangle$ is simply the center of charge (COC) of the species that carries protonic charges, i.e., the hydronium or protonated weak acid in $|i\rangle$ in Figure 2-1, such that

$$\boldsymbol{r}_i^{\mathrm{COC}} = \sum_I q_I^i \boldsymbol{r}_I \quad 2\text{-}10$$

Here, we assume that the "diabatic" excess charges are associated with atomic positions $\boldsymbol{r}_I$ and are modeled by fixed charges q_I^i, the charge of atom I in state $|i\rangle$. In diabatic states with well-defined bonding topologies, the fixed force field charges are considered to be a reasonable description of the system, while the excess charge delocalization and the polarization due to the excess proton

are characterized by the $\boldsymbol{c}$ vector. Since the coefficient vector $\boldsymbol{c}$ obtained from eq 2-9 represents the population of each diabatic state in the final CI ground state, the "adiabatic" CEC is naturally defined as the weighted average of each "diabatic" CEC (i.e., COC), eq 2-1. In this sense, the c_i's measure the extent of excess charge transfer. Accordingly, we define the charge transfer factor in what follows to represent the excess charge distribution between state $|i\rangle$ and state $|j\rangle$

$$f_{\mathrm{CT}}^{ij} = \frac{c_j^2}{c_i^2} \qquad 2\text{-}11$$

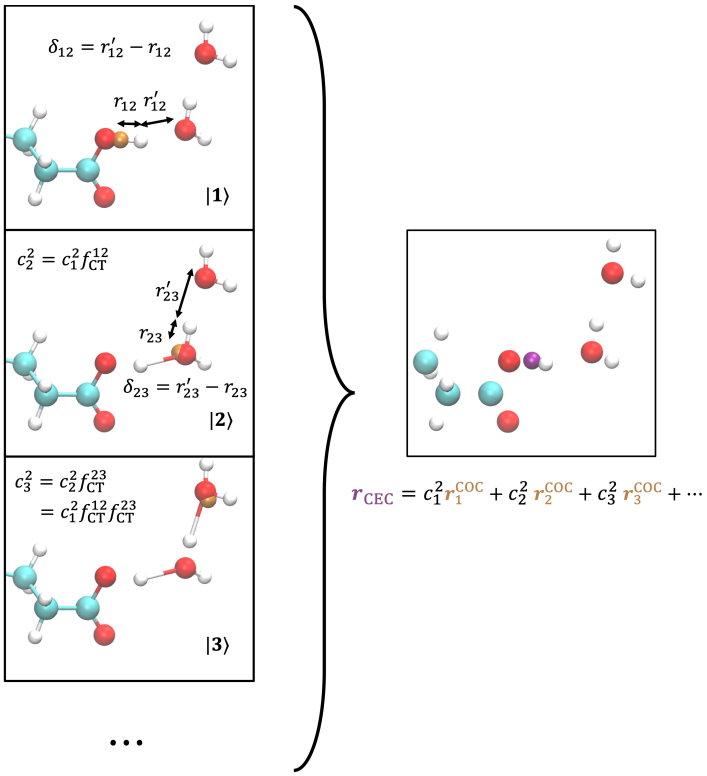

Figure 2-1. Illustration of CEC calculation for Glu in water. For the sake of clarity, only three diabatic states are shown, even though 20-30 states (on average) will be resolved in the condensed phase by searching up to three solvation shells of the excess proton. The COC in each state is rendered by an orange sphere. The resulting CEC as a linear combination of COCs is rendered in purple in the right panel. The f_{CT}^{12} and f_{CT}^{23} are computed using eq 2-12 as a function of δ_{12} and δ_{23}, respectively.

Due to the extended searching space introduced by the multiplier λ, the CDFT calculation (eq 2-3) is typically more expensive in comparison to the adiabatic electronic structure method used

in AIMD. Therefore, we adopted an approximation of the ground state vector $\boldsymbol{c}$ in order to compute the CEC on-the-fly in the AIMD simulations. It was found that an exponential function of proton transfer coordinate δ can provide a good fit for the charge transfer factor:

$$f_{\mathrm{CT}}^{ij} \approx \exp\left(-k\left(\delta_{ij} + \delta_0\right)\right) \tag{2-12}$$

The δ_{ij} is defined here as the difference between two O-H distances

$$\delta_{ij} = r'_{ij} - r_{ij} \tag{2-13}$$

where r_{ij} denotes the distance between the shared proton and the proton donor oxygen in state $|i\rangle$, and r'_{ij} denotes the distance between the proton and the proton acceptor in state $|j\rangle$. The parameters k and δ_0 were calibrated to match the exact f_{CT}^{ij} from CDFT-CI calculations between protonated species and water in the gas phase using BLYP or ωB97X functionals in this case. The list of fitting parameters for the CEC is provided in Table 2-1. Further details pertinent to the parametrization procedure can be found below.

Table 2-1. Fitted parameters of the CEC for water and glutamic acid.

	CDFT Functional	k (Å^{-1})	δ_0 (Å)
$H_3O^+ - H_2O$	BLYP	4.234	0
	ωB97X	4.898	0
Glu$-H_2O$	BLYP	2.946	0.5361

In order to generalize the CEC to the condensed phase environment, we assume that the solvating waters of the hydronium or protonated acid propagates the excess charge to further solvation shells following the same exponential rule (eq 2-12) as illustrated in Figure 2-1. After resolving all the charge transfer factors between each proton donor-acceptor pair, the approximated c_i^2 was then computed from f_{CT}^{ij} by applying the normalization condition ($\sum_i c_i^2 = 1$):

$$c_1^2 = \frac{1}{1/c_1^2} = \frac{1}{\sum_i c_i^2 / c_1^2} = \frac{1}{1 + \sum_{i \neq 1} f_{\mathrm{CT}}^{1i}} \tag{2-14}$$

$$c_i^2 = c_1^2 f_{\mathrm{CT}}^{1i} \tag{2-15}$$

In summary, eqs 2-1 and 2-10 define the CEC, while eqs 2-12, 2-14 and 2-15 yield an approximation to the exact CDFT-CI $\boldsymbol{c}$, thereby facilitating CEC calculations at a reasonable computational cost.

Another way of viewing the CEC is the dipole moment of excess charges. Following eqs. 2-1 and 2-10, we have

$$\begin{aligned} \boldsymbol{r}_{\mathrm{CEC}} &= \sum_i c_i^2 \sum_I q_I^i \boldsymbol{r}_I = \sum_I \left(\sum_i c_i^2 q_I^i \right) \boldsymbol{r}_I \\ &= \sum_I q_I^{\mathrm{ex}} \boldsymbol{r}_I = \boldsymbol{\mu}^{\mathrm{ex}} \end{aligned} \tag{2-16}$$

where we define the excess charge of atom I as its weighted average charge $\sum_i c_i^2 q_I^i$. Hence, the excess charge contribution to the IR spectrum can be calculated directly from the CEC velocity correlation function

$$\begin{aligned} A(\omega) &\propto \int \langle \dot{\boldsymbol{\mu}}^{\mathrm{ex}}(0) \dot{\boldsymbol{\mu}}^{\mathrm{ex}}(t) \rangle \, e^{-i\omega t} \mathrm{d}t \\ &= \int \langle \dot{\boldsymbol{r}}_{\mathrm{CEC}}(0) \dot{\boldsymbol{r}}_{\mathrm{CEC}}(t) \rangle \, e^{-i\omega t} \mathrm{d}t \end{aligned} \tag{2-17}$$

2.2.3 Graph CV to describe water wire connectivity

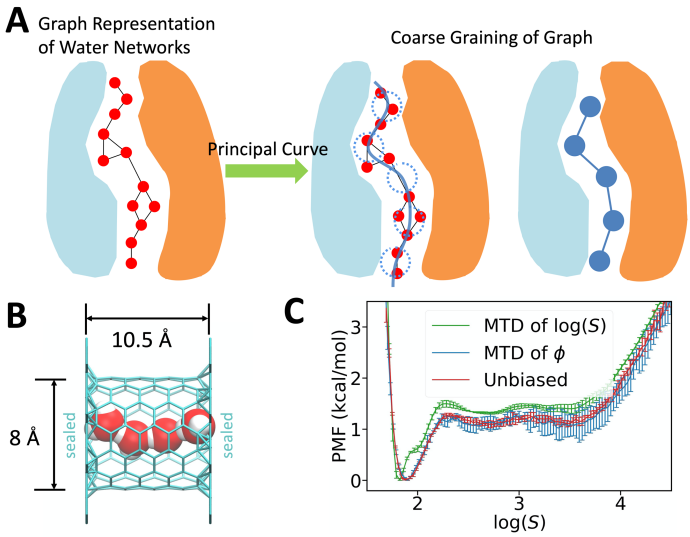

Figure 2-2. Illustration on the definition of water connectivity and benchmarks in a carbon nanotube. (A) An illustration of water connectivity collective variables (CVs). Water molecules are represented as red spheres. Dotted blue circles along the path represent the "coarse-grained" nodes in the simplified graph. (B) System setup of a sealed carbon nanotube containing 4 SPC/Fw waters. (C) Potential of mean forces of log(S) of the sealed CNT system computed from metadynamics of log(S) (green), reweighted from metadynamics of ϕ (blue) and the reference computed from a long unbiased MD run (red).

Consider that each water molecule is a node in a graph (Figure 2-2), and that water networks within a confined system can be fully described by the adjacency matrix of the graph $A_{ij} = A_{ji} = f(r_{ij})$, which represents the connectivity between node i and node j as a function of their distance r_{ij}. The switching function used in this work to approximate infinity when two waters are far apart is

$$f(r) = \frac{1 - \left(\frac{r}{r_0}\right)^{12}}{1 - \left(\frac{r}{r_0}\right)^{6}} \qquad 2\text{-}18$$

where $r_0 = 3$ Å as a typical oxygen-oxygen distance of two h-bonded waters. A path on the graph is defined as a collection of connected nodes, and thus represents a water chain. We refer to each graph path as a microscopic path because of atomistic details it provides. We consider that each graph path contributes to the overall connectivity between two given end points via water chains

measured by the length of the path[74]. The overall connectivity can thus be defined in terms of the shortest path length, indicating the least effort required to travel via waters from one end to the other. Since the range of the shortest path length will vary several orders of magnitude due to the nonlinear functions used in A_{ij}, we define the graph connectivity CV as $\log(S)$, where S denotes the shortest path length.

We now consider the water wires to be fluctuating around a 3D curve within the channel (Figure 2-2). Representing an average of microscopic water pathways, this 3D curve describes the "macroscopic" pathway that water is able to permeate. In this way, we effectively "coarse-grain" the atomistic water graph into a simpler graph consisting of a single macroscopic path; accordingly, defining the graph connectivity can be simplified as described below.

We first discretize the curve into a string of equally separated beads $\{\boldsymbol{x}_i\}$. A smooth water coordination number for each bead, s_i, is calculated to reflect the solvation profile of the curve. Similar to the approach of using an adjacency matrix to describe an atomistic graph, we begin by defining the two-body connectivity in a coarse-grained graph. First, we transform the coordination numbers into the occupancies I_i ranging from 0 to 1 using the Fermi function

$$I_i = \frac{1}{1 + \exp\left(-\frac{s_i - s_w}{\sigma}\right)} \qquad 2\text{-}19$$

where parameters s_w and σ indicate the degree to which a bead along the path is occupied given its water coordination number. In this work, $s_w = 1.5$ and $\sigma = 1.0$ were used for the two CNTs, and $s_w = 1.25$ and $\sigma = 2/3$ were used for ClC-ec1. Two adjacent beads on the principal curve are connected when they are both occupied, and thus two-body connectivity $f_{i,i+1}$ can be defined as $(I_i + I_{i+1})/2$. The curve is considered to be fully connected only when all pairs are connected, meaning that the curve connectivity is a logical conjunction (logical AND) of all the two-body

connectivity. Based on this fact, we take the product of all $f_{i,i+1}$ to represent a smooth version of the conjunction to define the final CV as:

$$\phi = \left(\prod_{i=1}^{N-1} f_{i,i+1} \right)^{1/(N-1)} \quad \text{2-20}$$

In cases when there are n fabricated water pathways in the system, $\{\phi_j \mid j = 1, 2, \ldots, n\}$, each pathway is represented by a principal curve and is combined by a softmax function to represent the connectivity of the best-connected path:

$$\phi = \text{softmax}(\{\phi\}) = \frac{1}{\kappa} \log \left(\frac{1}{n} \sum_j \exp(\kappa \phi_j) \right) \quad \text{2-21}$$

where $\kappa > 0$ to smoothly select the maximum value among $\{\phi_j\}$. The softmax function here works as a smooth version of logical disjunction (logical OR), meaning that the whole system is allowed to be passed through when any of the pathways are connected. We note that determining an analogous softmin combination of all microscopic path lengths would be computationally formidable because an atomistic water graph contains a factorial number of microscopic paths as a function of the number of waters. Accordingly, we utilize a coarse-grained graph that contains a small number (n) of macroscopic paths to overcome this computational impasse.

When the connectivity around the hydrated excess proton is of particular interest, a screening function $a_i = f_{\text{SC}}(|\boldsymbol{r}_{\text{CEC}} - \boldsymbol{x}_i|)$ (eq 2-22) can be applied to each bead as the exponent of the occupancy.

$$f_{\mathrm{SC}}(x) = \begin{cases} 1 & , \quad x < d_0 \\ \left(\frac{(x-d_0)}{r_0} - 0.5\right)\left(2\left(\frac{(x-d_0)}{r_0} - 0.5\right)^2 - 1.5\right) + 0.5, & d_0 \le x \le r_0 + d_0 \\ 0 & , \quad x > r_0 + d_0 \end{cases} \qquad 2\text{-}22$$

where $d_0 = r_0 = 5$ Å. The exponent a_i will remain at 1.0 within d_0 of CEC associated with the excess proton charge defect but decays to zero when distant from the CEC, effectively eliminating the dependency on those less related beads. The resulting connectivity around the hydrated excess proton is denoted as

$$\phi_p = \left(\prod_{i=1}^{N-1} \tilde{f}_{i,i+1}\right)^{1/(N-1)} \qquad 2\text{-}23$$

where $\tilde{f}_{i,i+1} = \frac{\left(I_i^{a_i} + I_{i+1}^{a_{i+1}}\right)}{2}$.

2.3 Simulation Details

2.3.1 CDFT calculations and AIMD setup

The CDFT calculations were conducted for hydronium-water (Figure 2-3A) and glutamate-water (Figure 2-4A); the electronic structure settings were identical to those used for AIMD simulations, shown below. Both the ωB97X and the BLYP functionals were used for hydronium-water CDFT calculations, while the latter was adopted for the glutamate-water. For hydronium-water, the total promolecule density of states $|1\rangle$ and $|2\rangle$ was calculated by adding the ground-state density of the hydronium defined in that state plus the water density, i.e., the hydronium and the water were chosen to be the fragments A and B respectively in eq 2-4. For Glu-water, the promolecule density of state $|1\rangle$ is the sum of neutral Glu density plus the neutral water density, and for state $|2\rangle$, is the sum of deprotonated Glu plus the hydronium. Established literature values[75] for the atomic radii,

including 0.75 Å for carbon, 0.32 Å for hydrogen, 0.63 Å for oxygen, and 0.71 Å for nitrogen were used for calculating the Becke population. The fitting procedure involving the hydronium-water molecular pair was based on a training set comprised of a series of fixed values for the oxygen-oxygen distance, r_{OO}, including 2.2 Å, 2.4 Å, 2.6 Å, 2.8 Å, 3.0 Å, and 3.2 Å. For each value of r_{OO} in this set, 6 shared proton positions were sampled evenly from $r_{\mathrm{OH}} =$ 0.9 Å to $r_{\mathrm{OH}} = r_{\mathrm{OO}}/2$, resulting in $6 \times 6 = 36$ data points. The training set for glutamate-water consists of 7 oxygen-oxygen distances evenly distributed from 2.2 Å to 2.8 Å and 9 oxygen-hydrogen distance values ranging from 1.0 Å to $r_{\mathrm{OO}} - 1.0$ Å, resulting in $7 \times 9 = 63$ data points. The CDFT calculations were performed by the CP2K[76, 77] implementation of CDFT[78] combined with Libxc[79].

The AIMD simulation of the excess proton in water was performed for 128 water molecules and one excess proton in a 15.64 Å ×15.64 Å ×15.64 Å box. The electronic structure was described by the BLYP-D3 density functional[80] with Goedecker-Teter-Hutter pseudopotentials.[81] The Gaussian and plane waves (GPW) method[82] was used, and the Kohn-Sham orbitals were expanded using Gaussian basis set TZV2P, and the electronic density was expanded in plane waves with a cut-off of 400 Ry. An EDS correction[33] was also employed as a minimal add-on bias to correct the overly strong hydrogen bonding in most DFT functional. It was found that excess proton and water diffusion better match experimental values after including the EDS correction. This EDS method for excess protons in water, which follows earlier work for pure water,[35] will be discussed in Chapter 3.

The system was first equilibrated in the constant NVT ensemble at 298 K for 30 ps, and then it was switched to the constant NVE ensemble for 200 ps for collecting non-thermostatted dynamical data. A timestep of 0.5 fs was used to integrate the system MD. All of the AIMD simulations were

carried out with the CP2K program package, coupled with a modified version of PLUMED2[83] for the EDS correction. Both the BLYP-fitted and the ωB97X-fitted CEC parameters were employed in the analysis of AIMD of proton in water.

The Glu-water system consisted of 1 neutral glutamate with 110 water molecules in a 16 Å ×16 Å ×16 Å box. We set up the electronic structure calculation similar to that used for the proton-water system, except that no EDS correction was used. The unbiased and well-tempered metadynamics[84] simulations were conducted with the constant NVT ensemble at 300 K in CP2K with PLUMED2 for computing and biasing the CEC. In order to be consistent with the underlying density functional used to perform AIMD simulations, the BLYP-fitted hydronium-water and glutamate-water CEC parameters were used. The collective variable used in metadynamics was the minimum distance between the CEC and the two Glu carboxyl oxygens:

$$\xi = \mathrm{softmin}(|\boldsymbol{r}_{\mathrm{CEC}} - \boldsymbol{r}_{\mathrm{OE1}}|, |\boldsymbol{r}_{\mathrm{CEC}} - \boldsymbol{r}_{\mathrm{OE2}}|) \tag{2-24}$$

where softmin is a smooth version of minimum function

$$\mathrm{softmin}(r_1, r_2) = -\frac{1}{\kappa}\ln\left(e^{-\frac{\kappa}{2}(r_1 - r_2)} + e^{-\frac{\kappa}{2}(r_2 - r_1)}\right) + \frac{r_1 + r_2}{2} \tag{2-25}$$

where $\kappa = 40$ Å^{-1}. The Gaussians in metadynamics were deposited every 50 fs with an initial height of 0.2 kcal/mol and a width of 0.1 Å. The bias factor of $\gamma = 12$ was used to account for a roughly 9 kcal/mol proton dissociation barrier.

2.3.2 Water connectivity CV benchmarks in carbon nanotubes (CNTs)

The10-Å-long CNT system consisted of 4 SPC/Fw water molecules sealed in a (6,6) armchair CNT by two layers of 16 Å × 16 Å graphene placed in a 25 Å × 25 Å × 25 Å simulation box. The Lennard-Jones interactions between carbons and hydrogens were set to zero. The LJ parameters between the tube carbons and the water oxygens were $\epsilon = 0.1$ kcal/mol and $\sigma = 3$ Å. The LJ

parameters between the graphene carbons and water oxygens were $\epsilon = 0.4$ kcal/mol and $\sigma = 2$ Å. All the carbons were fixed at initial positions and the 4 waters were integrated by a Nose-Hover chain with a chain length of 3, a timestep of 0.5 fs, and a temperature relaxation time of 250 fs at 310 K. The long-range electrostatic was computed by the particle-particle particle-mesh method (PPPM) with an accuracy of 10^{-4}. The simulations were conducted in LAMMPS patched with PLUMED 2.

The 28-Å-long system setup and simulation details were the same as ref [85], except that the EVB 3.2 proton-water model [86] was used instead of the original EVB 3 model [40].

2.4 Results

2.4.1 CEC derived from CDFT

The CDFT calculated charge transfer factors f_{CT}^{12} and the fitted curves using eq 2-12 are shown in

Figure **2-3**. The ωB97X-based value of the charge transfer factor was found to decay more quickly compared to that calculated using the BLYP functional. This finding is not surprising because the range-separated hybrid functional ωB97X produces less charge delocalization in comparison to the GGA BLYP functional. However, for both cases the exponential function (eq 2-12) provides a good fit. Given that $c_1^2 = c_2^2$ when the proton is equally shared between two water molecules, the δ_0 parameter was set to be zero. Interestingly, the fitted parameter k from ωB97X was found to share a similar value with the one used in the rCEC parameters[55] (4.898 Å^{-1} vs. 4.984 Å^{-1}), which is based on the multi-state empirical valence bond method,[40, 86, 87] which better justifies the use of the rCEC variable for AIMD simulations.

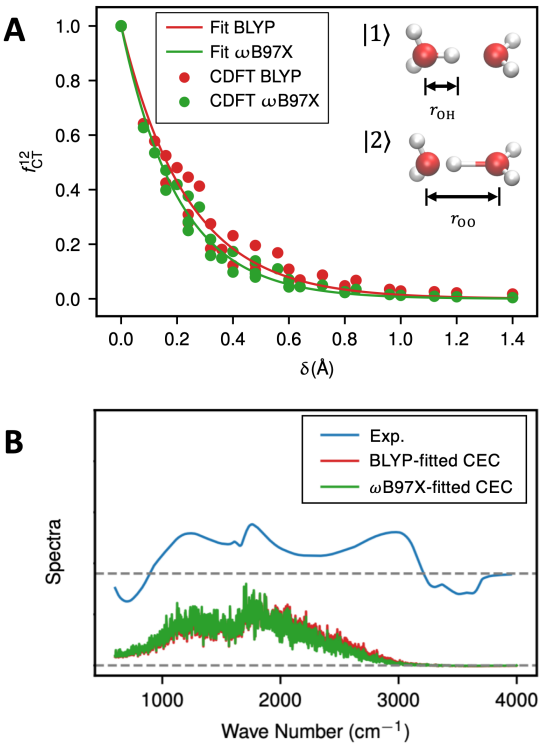

Figure 2-3. Charge transfer between hydronium and water and the IR spectrum of CEC. (A) The charge transfer factor between hydronium and water calculated by CDFT with BLYP and ωB97X functional and fitted curves. (B) The calculated AIMD excess charge spectrum of a hydrated excess proton in water using the two fitted CECs. The experimental IR spectrum is the acid solution spectrum subtracted by the pure water spectrum, taken from ref[88].

As shown in an earlier study,[55] the IR spectrum of the excess proton charge provides a systematic evaluation of the CEC by revealing the encoded collective motions in that CV. Note that although the two functionals exhibited different charge transfer behaviors (

Figure **2-3**A), the BLYP-fitted CEC and the ωB97X-fitted CEC produced very similar spectra, shown in

Figure **2-3**B; this correlation implies that the CEC parametrized by one functional may be applied to AIMD simulations using other functionals. Compared to the experimental IR difference spectrum, the CEC spectra reproduce the acid continuum 600-3200 cm^{-1}, which is the signature feature of the acid solutions that arise from the hydrated excess proton. In particular, the proton transfer mode (PTM) at around 1200 cm^{-1} and the flanking water bending at around 1750 cm^{-1}

are faithfully reproduced in the excess proton CEC spectra, indicating that these modes are well reflected in the encoded collective motions of CEC. It should be noted that the CEC spectrum decays in the range of 2300-3200 cm^{-1}, as opposed to the peak present in the experimental spectrum. This difference is due to the decaying excess proton charge in outer solvation shells, thereby reducing the intensity of this region, which is associated with the red-shifted O-H stretching in the second and third solvation shells of the excess proton, as detailed in a prior paper.[55] Importantly, the CEC spectrum decays to zero at the same position as the experimental difference spectrum at around 3200 cm^{-1}, revealing that the CEC excludes any bulk-like water O-H stretching in its motions. In summary, the CEC mostly represents the inner core motions of the protonated water complex, including the PTM and the flanking water bending, smoothly scales down its weight in the outer solvation shells, and is completely shut off for the bulk-like waters, which is the ideal behavior of a CV to focus only on PT and its related collective motions different from bulk water fluctuations.

We then studied glutamate in water to examine the CEC as a collective variable to be used in enhanced free energy sampling. Accordingly, we first parametrized the CEC from CDFT calculations of charge transfer between glutamate and water. The δ_0 parameter in eq 2-12 is needed for this case to account for the asymmetry in the PT between Glu and water. Figure 2-4A shows that the exponential function indeed outlines the charge transfer behavior for Glu-water, suggesting that this functional form can be applied to PT involving other weak acids. We note that other forms of switching functions could also be employed if is eq 2-12 deficient in providing an accurate fit to the CDFT behavior for a particular system.

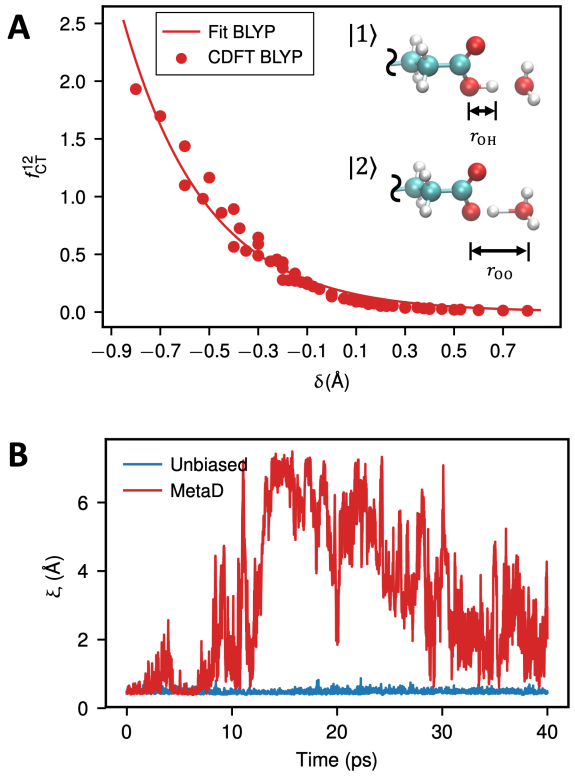

Figure 2-4. Charge transfer between Glu and water and the sampling efficiency gain by driving CEC. (A) The charge transfer factor between glutamate and water calculated by CDFT with the BLYP functional and its fitted curve. (B) The time series of the proton disassociation CV ξ in unbiased AIMD (blue) and a metadynamics run (red) of glutamate solution.

The proton disassociation barrier for Glu is large[89] given the experimental pK_a of around 4.2. Thus, dissociation tends to be a rare event compared to the achievable timescale of AIMD simulations. In Figure 2-4B, we show the sampling efficiency gained from a metadynamics run using the CV ξ, representing the distance between the CEC and the closest carboxyl oxygen (see section 2.3.1 for the definition). During the 40-ps AIMD run, the unbiased simulation was found to sample only the free energy well corresponding a protonated Glu, while the metadynamics drives the CEC to easily sample the dissociation of the proton from Glu and almost completes a "round-trip".

2.4.2 Efficient free energy sampling of water wire connection by driving principal curve connectivity ϕ

We first tested our CVs for a model system (Figure 2-2B), wherein 4 SPC/Fw waters [90] were sealed in a short CNT by two graphene sheets. The system was designed so that both the connected and disconnected water wires could be effectively sampled by long yet affordable unbiased simulations. Hence, these unbiased simulations will provide a reference PMF for any CV, specifically the $\log(S)$ and ϕ CVs of interest, by directly making histograms.

Among the various enhanced free energy sampling methods, metadynamics is known to be a robust tool for studying various chemical and biomolecular processes. The well-tempered metadynamics (WT-MTD) method, which benefits from its asymptotic convergence properties, is one of the most popular variants of the original metadynamics approach [84, 91], and was employed as the enhanced sampling method herein.

Prior to actually running the metadynamics simulations, we noted from its definition that $\log(S)$ is not strictly differentiable when two or more paths are all the shortest but have the same path length. In other words, $\log(S)$ becomes non-differentiable when the identity of the shortest path is about to exchange to another path. Since the probability is zero for the system to visit these singularities, one may expect that non-differentiability has only a minimal effect on the free-energy calculation. However, we found a significant deviation of the metadynamics PMF of $\log(S)$ when compared to the correct one computed from an unbiased molecular dynamics (MD) run (Figure 2-2C). We note that the WT-MTD barrier height for connecting water wires for the $\log(S)$ CV is 1.5 kcal/mol, which is 25% higher than the true (unbiased) value and exceeds statistical error. Additionally, an artificial "shoulder" appears at the position where $\log(S) = 1.8$, while only a smooth single well was observed in the unbiased reference. It is therefore striking that biasing

$\log(S)$ resulted in relatively significant errors in such a simple system featuring only 24 permutations for 4 waters. Based on these findings, we eliminated the possibility of using $\log(S)$ in biased free energy simulations in any real system where the number of permutations is expected to grow factorially as a function of the number of water molecules.

In contrast, the correct PMF for $\log(S)$ was obtained (Figure 2-2C) by reweighting the ϕ metadynamics data using Tiwary and Parrinello's time-independent estimator [92]. This outcome suggests that our new ϕ CV is a differentiable collective variable appropriate for enhanced free energy sampling, and validates its ability to drive efficient sampling of water wire connectivity.

2.5 Conclusions

We developed an appropriate CV to define the charge defect location and transfer properties for proton transfer and transport processes, resulting in a more rigorous *ab initio* definition of the center of excess charge. We also showed that the charge transfer behavior of CDFT can be approximated by an exponential function. We further examined the encoded collective motions in this newly defined CEC via calculating its IR spectrum. The full acid continuum was reproduced, suggesting the ability of this new CEC CV for capturing the excess proton motions without any contamination from other irrelevant degrees of freedom.

We also simulated a glutamate-water solution as an example illustrating the use of the new CEC in enhanced free energy sampling for amino acid ionization in water. An AIMD metadynamics run driving the CEC was found to explore a much larger CV space than an unbiased AIMD run, providing efficient sampling of the proton disassociation of Glu.

We note that the present CDFT formalism is suitable for describing any charge transfer reaction in which bonding topology changes between the diabatic states; thus, it is not limited to the proton

transfer processes studied herein. Therefore, the method can be generalized to identify appropriate CVs for other charge transfer reactions, e.g., ATP or GTP hydrolysis.

Additionally, we proposed a new measure that provides the long sought-after fully quantitative definition of facile water wire connectivity for water-assisted proton transport in confined spaces such as proteins and nanomaterials. From graph theory, we have defined a differentiable CV, ϕ, to represent water wire connectivity along a principal curve, and demonstrated its ability to drive the efficient sampling of water wire formation and breakage.

3 Accurate and Efficient Quantum Simulation of Proton Transport in Water

3.1 Introduction

Because of its ability to seamlessly model reactive nature of proton solvation and transport in aqueous, the AIMD technique remains one of the most commonly used approaches for modeling hydrated excess protons. However, the GGA-level of DFT, which is usually used as the underlying electronic structure technique for AIMD, overestimates charge transport and covalency. The resulting water model exhibits overly structured solvation, sluggish dynamics, and even glass-like behavior.

Higher-level quantum chemical approaches are one way to improve the AIMD description of water and the associated proton solvation and transport, but this comes at the cost of higher computational demands and lower scalability to bigger systems. To attain quantitative accuracy for structural and spectroscopic properties of pure water, it has been reported that the employment of a hybrid functional paired with accurate modeling of NQEs is required.[93] The employment of computationally more expensive quantum approaches, on the other hand, does not guarantee improved physical properties but limits the sufficient sampling of the condensed phase water and hydrated proton. As an example, the second-order Møller-Plesset perturbation theory (MP2) has been employed to simulate pure water, but an apparent discrepancy was seen in the first solvation peak of O-O RDF compared to the experimental one.[94]

An alternative approach is to elevate the simulation temperature of water and interpret the observations in a higher temperature as if they were observed at ambient conditions to "mimic" the missing NQEs for treating nuclei as classical particles in AIMD.[95] As we will discuss in Chapter 5, the elevated temperature does alleviate the over-structuring and slow diffusion issues, but only inadvertently, rather than because of proper modeling of the missing NQEs.

Another approach is to add some force-field-like corrections to the *ab initio* PES. The introduced correction can be evaluated quickly, adding little to the computing cost of the original method. One successful example is the Grimme's D3 dispersion correction for DFT[80]. The inclusion of the correction was found to significantly improve the solvation structures, equilibrium density, and diffusion coefficient of DFT water.[31] Despite the improvements, the corrected DFT functional, such as BLYP-D3, still overestimates the hydrogen bond strength and results in slow diffusion of water.

The recently developed EDS approach, represents another way of parametrizing a force field correction to the density functional used in AIMD. The EDS was built upon the work by Pitera and Chodera that uses a maximum entropy approach to derive an addictive energy bias. Among the infinitely many bias potentials that would correct the original Hamiltonian to reproduce experimental observables, the derived addictive energy bias is the closest one to the original Hamiltonian in the sense that the biased ensemble carries the least excess entropy compared to the unbiased one. The EDS method was applied to pure water to correct the overly strong hydrogen bonds of BLYP and BLYP-D3 water using the highly accurate MB-Pol water model as the "experimental" reference. The corrected BLYP-D3 functional (termed as EDS-BLYP-D3) was found to better reproduce the water RDFs, three-body correlations, as well as dynamical properties such as self-diffusion and hydrogen bond dynamics.[35]

Even with the less expensive GGA level DFT, the AIMD is still inefficient for large-scale simulations over long time scales, and the issue worsens if one wants to model the NQEs appropriately. Feynman's imaginary time path integral approach [96] has been adopted widely to compute the quantum statistics by mapping the sampling of quantum partition function to fictitious beads (replicas) of the corresponding classical nuclear system. The energy and force calculations

must be performed for all such replicas of the system, which typically increases the computational cost by an order of magnitude depending on the number of beads required to converge the path integral, limiting its pervasive use until the recent developments of some computational approximations.[97]. Representative work includes the ring-polymer contraction (RPC)[98], the ring-polymer interpolation[99], and combined path integral and generalized Langevin equation (PI+GLE) approach[100]. The system Hamiltonian is divided into two terms in RPC: a dominant term and a slowly varying term. This partitioning can be accomplished in AIMD by employing a lower level quantum approach as the reference potential, with its deviation from the precise Hamiltonian being regarded slowly varying. RPC speeds up path integral simulations because the exact Hamiltonian is only assessed on a "contracted" ring polymer with fewer beads.

One commonly used low level method in RPC is the self-consistent-charge density-functional tight-binding (SCC-DFTB). However, the SCC-DFTB method was known to overestimate the water coordination number to the hydrated proton while result in less structured water in its second solvation shell.[101] In principle, the accuracy of the RPC approximation is subject to the accuracy of the underlying reference potential, and the efficiency gain is heavily influenced by the efficiency of the low level. Hence, it is essential to seek an accurate enough reference potential which can be evaluated efficiently.

ML is increasingly recognized as a powerful technique in various fields of computational chemistry. One notable application is to use an artificial neural network (ANN) to learn a PES from quantum chemical data allowing the system of interest to be simulated efficiently with near *ab initio* quality. However, it is impossible to learn perfectly the *ab initio* model fed in the network training, due to the many-body nature of molecular systems and the finite expressive ability of an ANN. The situation becomes more challenging for a solvated proton because a data driven

approach is highly affected by the data quantity, but it is harder to sample proton-water interaction than water-water interaction using acid solution simulations at common concentrations. As such, the trained model is possible to describe the water part of the system with high accuracy but becomes less satisfying for the excess proton complex.

In this work, we addressed the difficulty of simulating proton solvation and transport in water with GGA-level of DFT by generalizing the EDS methodology developed for correcting BLYP-D3 water hydrogen bonds. The key challenge was to quantify the extent of hydrogen bond or covalency nature for a O-H pair and accordingly scale the applied EDS bias. We show the resulted EDS-BLYP-D3 functional significantly improve the gas-phase PT energetics and the condensed phase proton diffusion, especially the proton-to-water diffusion ratio. We then utilized the EDS-BLYP-D3 functional to illustrate how to optimize the RPC approach by replacing the low level electronic structure method by a ML potential. Since it is only used as the reference potential in RPC but not to directly propagate the system, the accuracy of the ML potential becomes a less concern. On the other hand, the RPC approach benefits from the efficiency of the ML potential, and exhibits a two-fold acceleration compared to the more conventionally adopted SCC-DFTB method. An excellent overlap of the water and proton properties was found for the ML-RPC approach compared to the full *ab initio* path integral MD.

3.2 Methods

3.2.1 Experiment Directed Simulation Method

The EDS is a correction to the system's Hamiltonian as an additive bias potential which is parametrized to reproduce target observables. The previous work of EDS for pure water was parametrized to reproduce the O-H radial distribution function (RDF) of the MB-Pol water model. The biasing approach for O-H pairs herein is referred as EDS(OH) in contrary to the original

EDS(OO) that targets the O-O RDF. The EDS bias potential is a function of statistical moments $f_k(r_{IJ})$, their target averaged values $\widehat{f_k}$, and coupling constants α_k:

$$V^{\mathrm{EDS}} = \sum_{I}^{n_H} \sum_{k} \frac{\alpha_k}{\widehat{f_k}} f_k(r_{IJ}) = \sum_{I}^{n_H} \sum_{k} \frac{\alpha_k}{\widehat{f_k}} \sum_{J}^{n_O} r_{IJ}^k [1 - u(r_{IJ})] \qquad 3\text{-}1$$

where *I* and *J* labels hydrogens and oxygens respectively, and *k* labels the order of moment of RDF. The coupling constants α_k's were determined from the EDS parametrization procedure using a gradient-based approach, and the $f_k(r_{IJ}) = r_{IJ}^k [1 - u(r_{IJ})]$ is the product of a power of O-H distance and a step-wise function that reads

$$1 - u(r_{IJ}) = \begin{cases} \dfrac{1 - \left(\dfrac{r_{IJ} - r_0}{d_0}\right)^6}{1 - \left(\dfrac{r_{IJ} - r_0}{d_0}\right)^{12}}, & r_{IJ} > r_0 \\ 1, & r_{IJ} \leq r_0 \end{cases} \qquad 3\text{-}2$$

The d_0 was set to 0.7 Å and $r_0 = 2.125$ Å following the previous work. The target moment values were determined by integrating the MB-Pol O-H RDF:

$$\widehat{f_k} = \rho \int_0^{+\infty} \mathrm{d}r \ [1 - u(r)] 4\pi r^{2+k} g_{\mathrm{OH}}(r) \qquad 3\text{-}3$$

When simulating pure water, the EDS potential (eq 3-1) is set to zero for O-H pairs whose distances are within 1.2 Å, which is the zero-point of O-H RDF that separates the intramolecular peak from the first intermolecular peak. The rationale is to turn off any perturbation to bonded O-H pairs but only focuses on the non-bonded O-H pairs and thus corrects the hydrogen bonds in DFT water. This hard cutoff is obviously less justifiable to proton in water since the proton of a hydronium is able to sample closer distances to oxygen atoms within 1.2 Å. More importantly, the excess proton delocalizes its charge to surrounding waters and thus it becomes unclear if a O-H pair is covalently

bonded or hydrogen bonded. We will show that these difficulties can be resolved with the help of CDFT.

3.2.2 Combine CDFT information into EDS (OH)

The CDFT methodology has been described in Section 2.2.1, and an approximation to the CDFT charge transfer behavior was developed in Section 2.2.2 to enable an efficient computation of the ground-state wave function in a diabatic representation.

Since each diabatic state is associated with a well-defined bonding topology, there is no confusion about whether a O-H bond is covalent or not, and thus the EDS potential developed originally for rectifying hydrogen bonds is directly applicable within each diabatic state. That is, the EDS bias is turned off between any bonded O-H pair (i.e. the O-H pairs within every water and hydronium) according to the corresponding bonding topology of every diabatic state. The final EDS bias force is a weighted average of each diabatic EDS bias force $\boldsymbol{F}_i^{\mathrm{EDS}}$ in state $|i\rangle$ by the population coefficient c_i^2,

$$\boldsymbol{F}^{\mathrm{EDS}} = \sum_i c_i^2 \boldsymbol{F}_i^{\mathrm{EDS}} \qquad 3\text{-}4$$

The behavior of this weighting can be better understood by considering an O-H pair in some extreme cases. For a given O-H pair, consider all the diabatic states in which the pair is non-bonded, and denote the sum of c_i^2's of these states as c_{NB}^2. For an intermolecular O-H pair between two waters that are both distant from the solvated H^+ and not directly involved in the excess proton solvation, the sum c_{NB}^2 becomes 1 as the O-H pair is seen non-bonded in all the diabatic states. According to eq 3-1, the EDS potential for hydrogen-bond correction in this case is completely switched on, which turns out to be the case for non-bonded O-H pairs in pure water. On the other hand, the c_{NB}^2 becomes 0 for the intramolecular oxygen and hydrogen of a far water and thus the

EDS correction for this O-H pair is completely turned off. Therefore, for the distant water, this weighted EDS scheme degenerates to the previous pure-water EDS correction. The more common case is an intermediate between the two extremes, and the continuous functions of eq 2-14 and eq 2-15 enable an automatic and smooth adjustment of the EDS bias for solvated proton in water.

3.2.3 Combine ML in RPC

The RPC approach is a computational approximation to the full path integral simulations. The simulated system's Hamiltonian is split into two parts, a dominant term and a slowly varying term. In the context of AIMD, the partitioning is usually achieved by employing a lower level electronic structure method as the reference potential and its deviation from the original higher level Hamiltonian is regarded as a perturbing term.

$$H_{\text{high}} = H_{\text{low}} + H_{\text{diff}} \qquad 3\text{-}5$$

The reference potential H_{low} is evaluated for all the beads while the difference between high level and low level, $H_{\text{diff}} = H_{\text{high}} - H_{\text{low}}$, is only evaluated on a "contracted" or "coarse-grained" ring polymer with reduced number of beads. As such, the high-level potential becomes available for each bead in the full ring-polymer by combining the evaluated reference potential with the H_{diff} extrapolated to all the beads. As such, RPC approach saves computational costs by reducing the number of H_{high}evaluations but introduces additional overhead of evaluating H_{low} on the full ring polymer. The efficiency of the low level H_{low} is thus crucial to the RPC overall efficiency.

Herein, we made use of the DPMD approach[102]. to machine learn a many-body PES to approximate the EDS-BLYP-D3 description for proton in water. The motivation is to make use of the accuracy and computational efficiency of the ML potential to accelerate path integral simulations by employing the learned potential as the reference potential (H_{low}) in the RPC scheme.

3.2.4 Simulation Details.

The EDS parameters were taken from ref[35] for BLYP-D3 water. The constrained DFT calculations using the ωB97X functional for proton-water were used to fit the charge transfer factor as detailed in Section 2.2.2.

The AIMD of proton in water consists of 128 water and one excess proton in a 15.64 Å ×15.64 Å ×15.64 Å box resulting into a density of 1.00 g/cm^3. The electronic structure was described by the BLYP-D3 density functional with the Goedecker-Teter-Hutter pseudopotentials. The Gaussian and plane waves (GPW) scheme was used where Gaussian TZV2P basis set was used to expand the Kohn-Sham orbitals while a plane wave basis set with a cut-off of 400 Ry was used to represent the electronic density. The system was first equilibrated in NVT ensemble at 298 K for 90 ps using EDS-BLYP-D3. Three EDS-BLYP-D3 simulations were initiated from the configurations and velocities sampled in the NVT equilibration every 30 ps and were run in the NVE ensemble for 200 ps to collect non-perturbed dynamical data.

A path integral MD (PIMD) run was conducted without RPC to provide a reference of the accurate quantum statistics. The PIMD was first equilibrated by a Langevin thermostat with a relaxation time of 15 fs started from the classical NVT equilibrated configuration. The production PIMD was run after that for ~32 ps with the Langevin relaxation time softened to 100 fs. The PIMD was thermostatted at 298K and 30 beads were used to represent the ring polymer. A timestep 0.5 fs was used in all the simulations. The *ab initio* energy and force calculation was carried out by CP2K and the EDS correction was implemented in a modified version of PLUMED 2. The PIMD was conducted with i-Pi[103] coupled to CP2K.

Initiated from independent samples from the classical NVT equilibration, six thermostatted ring polymer molecular dynamics (TRPMD) were conducted using the ML-RPC approach. The ML

reference potential was computed by LAMMPS coupled to DeepMD-kit, the *ab initio* calculation was performed by CP2K, the EDS correction was computed by PLUMED 2, and the MD integration was performed by i-Pi.

3.3 Results

3.3.1 EDS (OH) improves BLYP-D3 significantly

We first checked the energetic influence of the EDS (OH) correction via a gas-phase Zundel cation ($H_5O_2^+$). The PT PES as a function of the proton sharing coordinate δ is shown in Figure 3-1. The reference PES was calculated from the CCSD(T)/cc-pVTZ level of theory, as well as a hybrid functional revPBE0-D3, which was found to reproduce experimental water properties[93]. The original BLYP-D3 was also plotted for comparison. As shown in the figure, the BLYP-D3 functional underestimates the PT barrier significantly, while the EDS correction brings the PES much closer to the CCSD(T) and revPBE0-D3 reference. It is worth noting that the EDS correction was parametrized to correct hydrogen bonds in condensed-phase BLYP-D3 water, but promisingly exhibits the transferability to gas-phase protonated water herein.

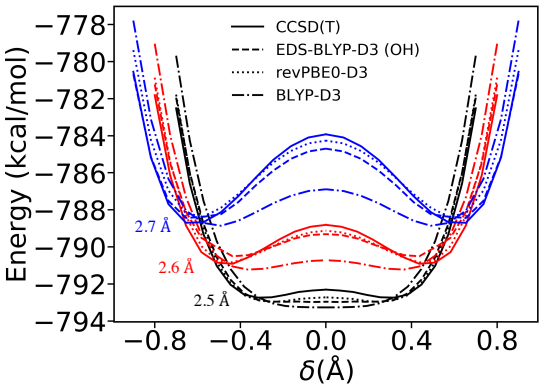

Figure 3-1. Gas-phase PES of proton transfer in Zundel.

The EDS-BLYP-D3 model was then applied to simulate the condensed phase hydrated proton, and the calculated diffusion constants of H^+ and water are summarized in Table 3-1, which shows that water diffusion, proton diffusion, as well as the proton/water diffusion ratio are all greatly

improved by the EDS correction. Notably, the EDS potential represents an intermolecular O-H repulsion as shown in the gas-phase PES (larger PT barrier in EDS-BLYP-D3 compared to BLYP-D3), which could reduce proton hopping and contributes negatively to proton diffusion. On the other hand, the weakened intermolecular hydrogen bonds make the water less glassy and increase its diffusion constant. The more flexibility of water facilitates the environmental preparation for proton hopping and transport, such as the hydrogen bond rearrangements in the proton's second solvation shell. The two competing effects merged into a net positive contribution to the proton diffusion as seen in EDS-BLYP-D3 simulations.

Table 3-1. Diffusion constants of proton and water.

	$D_{\mathrm{H2O}}\ (\mathrm{Å}^2/\mathrm{ps})$	$D_{\mathrm{H^+}}\ (\mathrm{Å}^2/\mathrm{ps})$	$D_{\mathrm{H^+}}/D_{\mathrm{H2O}}$
BLYP-D3[a]	0.049±0.004	0.54±0.25	10.9±5.1
EDS-BLYP-D3[a]	0.15±0.01	0.73±0.13	4.9±0.7
EDS-BLYP-D3[b]	0.11±0.003	1.46±0.31	13±3
Experiment	0.23[c]	0.94[d]	4.1

[a] Classical nuclei
[b] TRPMD
[c] Taken from ref[104]
[d] Taken from ref[105]

Although the EDS potential corrects the PT barrier of BLYP-D3 in the right direction, the resulted PES still underestimates the barrier slightly as shown in Figure 3-1. This explains why the proton and water diffusion ratio in classical EDS-BLYP-D3 is slightly larger than the experimental value. However, the issue is more evident when including NQEs as shown in Table 3-1. The quantized nuclei exacerbate the already overestimated hydrogen bonds (shown by the underestimated PT barrier) and as a result water diffusion is slowed down compared to the classical simulation. On the other hand, proton transport is increased, as evidenced by an exaggerated proton diffusion constant, with the aid of delocalized nuclei in overcoming PT barriers. We note that the EDS correction was parametrized from classical AIMD simulations. The parametrization

attempted to identify a solution that would allow the classical ensemble averages of the observables (O-H RDF moments) to match the reference values, but it was missing information on the quantum ensemble averages. To improve the present EDS correction in the future, PIMD should be conducted to optimize the EDS parameters using the centroid (center of beads) ensemble averages.

3.3.2 ML-RPC models NQEs accurately

As shown in Figure 3-2, the ML potential faithfully reproduces the EDS-BLYP-D3 water two-body correlations. We note that the trained ML potential is not perfect and shows some subtle but visible differences from the reference EDS-BLYP-D3 especially in the H-H RDF. However, when using the ML potential as the reference potential in RPC instead of directly driving MD, we observe excellent matching in the water RDFs compared to the exact PIMD that computes the EDS-BLYP-D3 functional on all the beads. Notably, the deviation in the H-H RDF becomes very little in the case of ML-RPC TRPMD compared against the full *ab initio* PIMD.

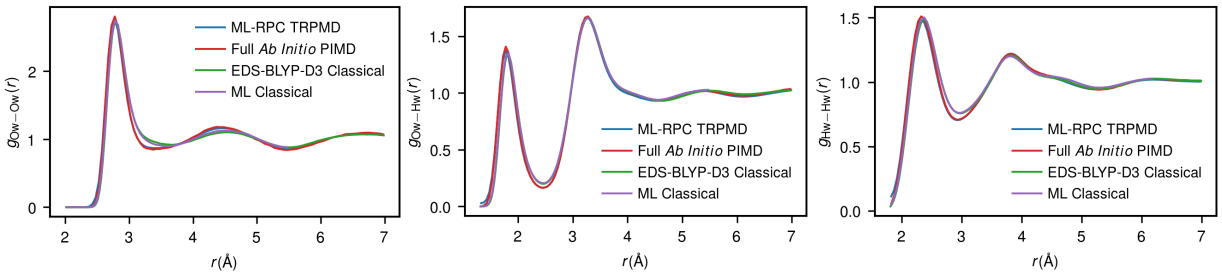

Figure 3-2. Radial distributions of water in the TRPMD using the ML-RPC approach (blue), the PIMD without RPC approximation (red), the EDS simulation with classical nuclei (green), and the classical simulation directly using the ML potential (purple).

We next investigated the proton solvation structure measured by the proton sharing coordinate δ (eq 2-13). As shown in Figure 3-3, the ML MD exhibits deviated PT barrier as well as the most probable δ value (location of the PMF well) from the exact EDS-BLYP-D3 simulation. These deviations are arguably more announced than the ones found in the water structures, possibly due

to the less sampled proton-related interactions than water ones in a dilute solution as the training set. Despite the deficiency in the trained ML potential, use of ML in RPC TRPMD predicts the correct δ PMF within statistical errors from the full PIMD.

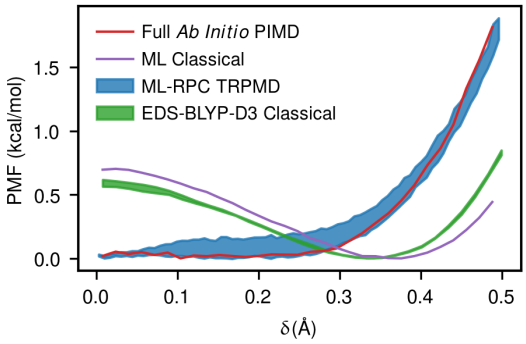

Figure 3-3. PMF of δ. The error intervals were estimated by the standard deviation between independent runs.

3.3.3 ML-RPC models NQEs efficiently

In Figure 3-4, we show the simulation speed of RPC TRPMD as a function of number of cores using ML or SCC-DFTB as the reference potential. In general, the use of ML potential in RPC results in a two- to three-fold computational efficiency compared to the SCC-DFTB one. The full PIMD runs at roughly 30 fs/hour on 40 cores/node×15 nodes, and thus the ML-RPC approach speeds up the quantum simulation by about 130-fold, bringing the computational cost down to that of a classical simulation.

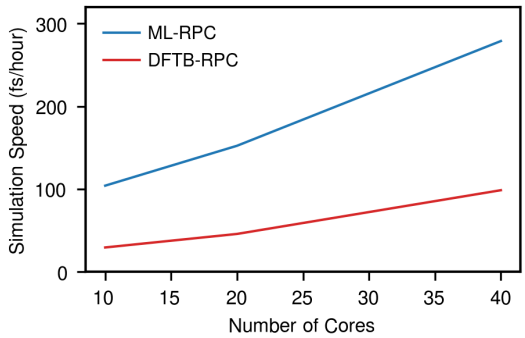

Figure 3-4. Computational speed benchmarks of RPC. The simulation speed is defined as simulation time per wall time.

3.4 Conclusion

We presented a generalization of the EDS approach originally defined for correcting DFT water hydrogen bonds to simulate hydrated excess proton in water. We found the resulting EDS correction significantly improve the original BLYP-D3 functional both for gas-phase PT energetics and condensed-phase water and proton diffusion. We note that the corrected BLYP-D3 functional still underestimates the PT barrier, and overestimates hydrogen bond strength. As a result, the quantum EDS-BLYP-D3 simulation shows too fast proton diffusion and too slow water diffusion.

We then employed the EDS-BLYP-D3 functional to benchmark the approach of accelerating path integral simulations utilizing ML potentials. We show that even when the ML potential is not optimal, the quantum simulations using the ML-RPC approach reproduce the results of full *ab initio* PIMD, but are only at the same computational cost as a classical simulation. RPC approximation assumes the deviation of the ML potential from the *ab initio* PES is slowly varying with respect to atomic motions. However, the current ML training methodology aims to train an ANN that minimizes energy and force differences between the model and the training reference, and thus focuses on minimizing the deviation itself instead of its fluctuations. As a result, we anticipate the ML-RPC technique to be more accurate if the loss function to be minimized in the ML training procedure includes not only the energy and force deviations but also their variances.

4 Accurate and Transferrable Reactive Molecular Dynamics Models Derived from the Constrained Density Functional Theory

4.1 Introduction

MD is a powerful computational tool for studying the kinetics and thermodynamics of chemical, materials science, and biomolecular systems and processes. The conventional MD (also referred to classical MD in this Chapter) requires a pre-defined bonding topology of the systems being simulated. Since this topology remains fixed during simulation, conventional MD is only suited for simulating strictly speaking physical processes without chemical reactions. In contrast, the AIMD approach represents a natural choice for simulating reactive processes because it solves the electronic Schrodinger equation on-the-fly and does not rely on the bonding structure of a system. Note that chemical reactions can be intrinsically multiscaled, i.e., they can be coupled with multiple collective and corporative motions that span spatial and temporal scales of several magnitudes. Given the multiscale nature of reactions, an efficient simulation model is needed to sample the required space and time scales in order to deliver statistically meaningful results. However, AIMD calculations can be prohibitively expensive and therefore often limited to sub-nanosecond timescales for a system consisting of hundreds of electrons. In addition, the accuracy of AIMD is highly correlated with the adopted electronic structure method. A popular approach is the GGA within the framework of DFT, which often overestimates charge transfer and partial covalency. Using significantly more computational expensive hybrid functionals was found to be necessary for achieving quantitative accuracy, in the description of somewhat simple and non-reactive systems such as pure water.[93]

An alternative approach is the MS-RMD method, which is three-orders of magnitude more computationally efficient compared to AIMD. MS-RMD enables microsecond sampling of complex systems, making it effective for accurately modeling condensed phase chemical reactions with sufficient sampling of the coupled, but relatively slow, processes. MS-RMD was previously formulated as a force-matching (FM) algorithm,[41, 89, 106] in which the models were parametrized to best reproduce the atomic forces of *ab initio* calculations in a least-square sense. The FM-based MS-RMD model was shown to successfully reproduce the benchmark results of the reference *ab initio* model in both aqueous[106] and biomolecular environments[41], as well as replicate experimentally derived thermodynamic and kinetic data.[57-60, 107] The models were parameterized in a case-by-case manner, whereby training *ab initio* calculations were conducted each time the MS-RMD model was applied to a new chemical environment.

In the condensed phase, the MS-RMD approach describes the system as a linear combination of as many as 20-30 diabatic states, each of which corresponds to a resonance form of the system with a different bonding topology. Conventional DFT was adopted as the reference *ab initio* method; however, due to the lack of diabatic state information, only the ground state of MS-RMD was calibrated in the FM scheme. To address this deficit, *ab initio* diabatic methods were employed in this work to facilitate a more systematic parametrization of our RMD models at the diabatic state level. Furthermore, the newer diabatic matching (DM) approach can be regarded as a generalization of our original FM method that focuses only on the ground state. The quantum diabatic method we chose was CDFT that is flexible in dealing with various atomic charge prescriptions such as Mulliken,[108] Hirshfeld,[109] and Becke,[72] but we note that other diabatic methods such as the multistate DFT (MS-DFT)[110, 111] can also be employed.

As an important category of chemical reactions, PT along confined water and titration of protonable residues in biomolecular systems, represent key steps in many protein machine functional cycles.[12-17, 19, 26, 112] Employing an efficient and accurate computational approach that explicitly models this PT reaction is crucial for understanding the full working mechanisms associated with these enzymes, channels, and transporters. As detailed in this chapter, the proton disassociation of ionizable amino acids in water and in staphylococcal nuclease (SNase) was examined using the new DM method. The computational efficiency of MS-RMD allowed the computation of the free energy profile for the coupled PT process and local conformational changes in SNase. As a result, our model can predict the pK_a of glutamate and lysine in both water and SNase with good agreement with experimental results without the need to reparametrize. Due to the aforementioned transferability, we anticipate that the MS-RMD Glu/Lys models will be widely applied to more complicated biomolecular systems for studying PT and its coupled phenomenon, notably hydration change, global protein conformational change, as well as ligand transport. Moreover, since neither MS-RMD nor CDFT is limited to PT reactions, our approach could significantly advance the systematic development of accurate and transferrable MD models to simulate other chemical reactions with greater efficiency.

4.2 Methods

4.2.1 MS-RMD force field

The general MS-RMD framework was reviewed in section 1.3, while more detailed descriptions on force field functional forms are provided herein.

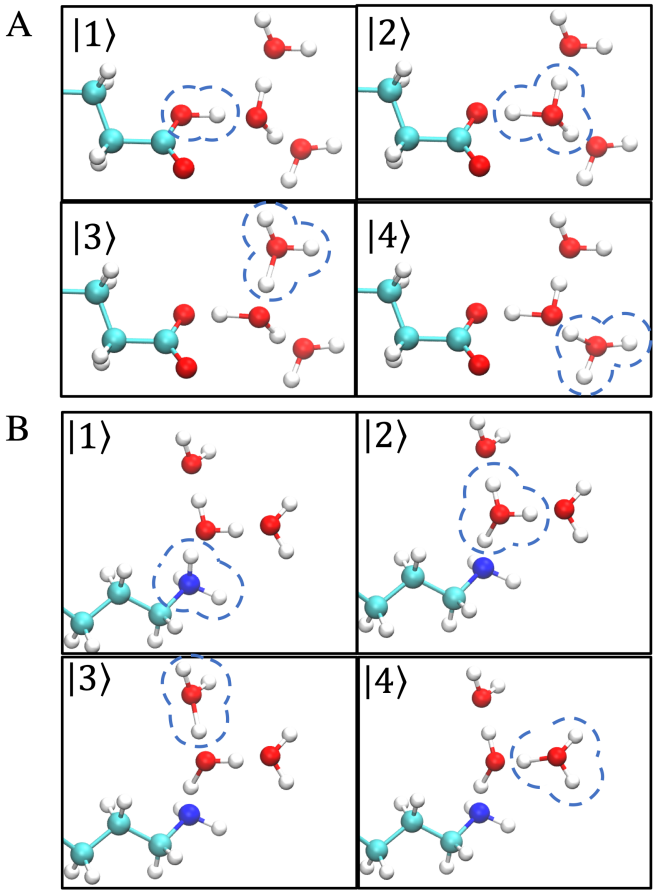

Figure 4-1. Illustration of MS-RMD states for (A) Glu in water and (B) Lys in water. The protonated moiety in each diabatic state is circled in blue. Note that typically 20-30 states are included in the condensed-phase calculations, but only four states are shown here for clarity.

The diagonal h_{ii} terms in MS-RMD Hamiltonian are described by the standard CHARMM36 FF[113] with several adaptations clarified below. The original harmonic O-H bond in a protonated Glu carboxylic group as well as the N-H bond in the Lys ammine group are replaced by a Morse bond to better describe the proton disassociation curve:

$$U_{\text{Morse}}^{\text{bond}} = D\left(1 - e^{-\alpha(r-r_0)}\right)^2 \qquad 4\text{-}1$$

where r represents the bond length, D, α, r_0 are bond parameters, taken from ref[89] for Glu. The bond parameters for Lys were optimized to match an energy scan of the bond disassociation, in which the proton was scanned with respect to nitrogen within a distance range of 0.9 Å to 1.9 Å

with 0.1 Å intervals. The reference energies were computed with ωB97X[70]/TZV2P in CP2K coupled with Libxc. In any diabatic state wherein a protonated Glu or Lys was found, such as state $|1\rangle$ shown in Figure 4-1, a trainable energy offset V_{ii}^{const} was added to the total energy of the system to bridge zero points of the classical MM PESs, that were originally parametrized for nonreactive protonated and deprotonated forms of the amino acids.

The hydronium FF parameters including bonds, angles, and Lennard-Jones (LJ) were obtained from ref[86]. Additionally, additive repulsions between hydronium and Glu carboxyl/Lys ammine were employed to correct the over-attraction between point charges described in the MM FFs, as indicated in the following equations:

$$U_{\text{OX}}^{\text{rep}} = B \exp\left(-b(r_{\text{OX}} - d_{\text{OX}}^{0})\right) \cdot \sum_{j=1}^{3} \exp\left(-b' \boldsymbol{q}_{\text{H}_j\text{X}}^{2}\right) \tag{4-2}$$

and

$$U_{\text{HX}}^{\text{rep}} = C \exp\left(-c(r_{\text{HX}} - d_{\text{HX}}^{0})\right) \tag{4-3}$$

The r_{OX} represents the distance between hydronium oxygen and Glu carboxylic oxygen/Lys ammine nitrogen, and r_{HX} represents the associated distance with hydronium hydrogen. The vector $\boldsymbol{q}_{\text{H}_j\text{X}}$ is the proton-sharing vector defined as follows:

$$\boldsymbol{q}_{\text{H}_j\text{X}} = \boldsymbol{r}_{\text{OH}_j} - \frac{1}{2}\boldsymbol{r}_{\text{OX}} \tag{4-4}$$

where the atoms O and X follow the same meaning above, while H_j represents one of the three hydrogens of the hydronium. The B, b, b', C, c are tunable parameters while $d_{\text{OX}}^{0} = 2.4$ Å and $d_{\text{HX}}^{0} = 1.0$ Å are fixed and do not need fitting since they can be absorbed into the prefactors B and C.

As part of the MM FF, some of the nonbonded LJ interactions were tuned as well to better describe proton-transfer reactions. While the standard 12-6 LJ functional form was used as shown in eq 4-5

$$U_{\mathrm{LJ}} = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \quad 4\text{-}5$$

the LJ parameters σ and ϵ for interactions between the carboxyl/ammine and water/hydronium were adjusted. To be specific, the modified interactions were between eight atom pairs: (1) protonated Glu carboxyl oxygen (OEP) and water oxygen (Ow), (2) Glu carboxyl proton (HEP) and Ow, (3) deprotonated Glu carboxyl oxygen (OE) and hydronium oxygen (OH), (4) OE and hydronium proton (HH), (5) protonated Lys ammine nitrogen (NKP) and Ow, (6) protonated Lys ammine proton (HKP) and Ow, (7) deprotonated Lys ammine nitrogen (NK) and OH, and (8) NK and HH.

The off-diagonal energy term of h_{ij} between amino acid and water proposed in ref[89] was used herein:

$$h_{ij} = c_1 \exp(-c_2(r_{\mathrm{HX}} - c_3)^2) \quad 4\text{-}6$$

The atom X represents the carboxyl oxygen or the ammine nitrogen, and the atom H represents the shared proton between Glu/Lys with water. The c_1, c_2, and c_3 are trainable parameters.

4.2.2 Parametrization of MS-RMD using CDFT

The original FM-based MS-RMD parametrization was designed to minimize the force residual between ground-state MS-RMD forces and the ground-state ab initio forces:

$$\chi^2 = \left\langle \left| \boldsymbol{F}^{\mathrm{RMD}} - \boldsymbol{F}^{\mathrm{QM}} \right|^2 \right\rangle \quad 4\text{-}7$$

where the bracket indicates an ensemble average. The MS-RMD atomic forces $\boldsymbol{F}^{\mathrm{RMD}}$ are computed from the Hellmann-Feynman theorem (eq 1-3) and depend on the FF parameters noted

in the section 4.2.1. We note from eq 1-3 that the ground-state forces depend on the contributions from both ground-state wavefunction $\boldsymbol{c}$ and the diabatic forces $\boldsymbol{F}_{ij}$. The parametrization based solely on the ground-state force residual (eq 4-7) is likely to overfit due to the error cancellation in $\boldsymbol{c}$ and $\boldsymbol{F}_{ij}$. The rationale of our new approach is to separate the ground state wavefunction $\boldsymbol{c}$ and diabatic forces in the training procedure. Accordingly, two residuals, defined utilizing the diabatic information computed by CDFT, should be minimized with respect to MS-RMD parameters:

$$\chi_c^2 = \left\langle \left| \boldsymbol{c}^{\mathrm{RMD}} - \boldsymbol{c}^{\mathrm{CDFT}} \right|^2 \right\rangle \tag{4-8}$$

and

$$\chi_F^2 = \left\langle \sum_{ij} \left| \boldsymbol{F}_{ij}^{\mathrm{RMD}} - \boldsymbol{F}_{ij}^{\mathrm{CDFT}} \right|^2 \right\rangle \tag{4-9}$$

We note that as an eigenvector, the ground state wavefunction $\boldsymbol{c}$ is determined by the diabatic energies in $\mathbf{H}^{\mathrm{RMD}}$ and $\mathbf{H}^{\mathrm{CDFT}}$. Thus, minimizing χ_c^2 can be regarded as an implicit energy matching process. In this sense, the DM approach systematically matches the energy and force through minimization of χ_c^2 and χ_F^2, simultaneously. The aforementioned technique could be beneficial since it was reported that including both energy and forces in the loss function results in a more reliable machine-learned FF.[102]

The EVB 3.2 model[86] has shown success in describing proton transfer between water and water. Therefore, our parametrization of MS-RMD kept the hydronium-water PT model unchanged, focusing instead on diagonal and off-diagonal terms that correspond to the proton transfer between Glu/Lys and its first solvation shell proton-acceptor water. The training set was generated by placing one water at r_{OO} (distance between water oxygen and carboxyl oxygen of Glu) or r_{ON} (distance between water oxygen and ammine nitrogen of Lys) ranging from 2.2 Å to 2.8 Å with a

0.1 Å spacing, and for each r_{OO} or r_{ON}, the shared proton was placed at nine equally spaced positions with r_{OH} ranging from 1.0 Å to $r_{\mathrm{OO}} - 1.0$ Å for Glu and r_{NH} ranging from 1.0 Å to $r_{\mathrm{ON}} - 1.0$ Å for Lys. The CDFT calculations were based on these configurations at a level of BLYP/TZV2P and ωB97X/TZV2P followed by a CDFT-CI performed using the DFT Kohn-Sham surrogates for the diabatic state wavefunctions. To reduce computational costs and achieve better convergence, the ωB97X-level CDFT was conducted with Glu/Lys sidechains only. Calculations that included backbones at the BLYP level showed no difference in $\boldsymbol{c}$ from sidechain-only ones. All CDFT calculations were performed using the CP2K package. Optimizing MS-RMD parameters was conducted using an in-house Python script to minimize the residual (eq 4-8); all obtained parameters are summarized in Table 4-1. Note that the models obtained from minimizing eq 4-8 exhibited good accuracy and transferability even prior to applying further improvements using force residual eq 4-9, which is detailed in the following sections.

4.3 Simulation Details

The Glu/Lys in water simulations were conducted with one protonated solute solvated in 241 water and in 237 water in a cubic box with a side length of 20 Å. Temperature was controlled by a Nose-Hoover chain[114, 115] at 300 K, and a timestep of 1 fs was used to integrate the system. To enhance the sampling of proton disassociation from Glu/Lys, WT-MTD was performed. The RC for Glu driven in WT-MTD was defined as the distance between the CEC and the closer Glu carboxyl oxygen, implemented as

$$\xi_{\mathrm{CEC}}^{\mathrm{Glu}} = -\frac{1}{\kappa}\log\left[\exp\left(-\kappa(r_1 - \bar{r})\right) + \exp(-\kappa(r_2 - \bar{r}))\right] + \bar{r} \qquad 4\text{-}10$$

where $\kappa = 40$ Å^{-1}, r_1 and r_2 denote the CEC separation from the two carboxyl oxygen atoms, and $\bar{r} = (r_1 + r_2)/2$. The RC, $\xi_{\mathrm{CEC}}^{\mathrm{Lys}}$ for Lys was the distance between CEC and ammine nitrogen. The

initial gaussian height for WT-MTD was 0.8 kcal/mol, and the bias factor was set to 12. The gaussians were deposited every 1 ps with a fixed width of 0.1 Å. To restrain the sampling in regions of interest, a harmonic potential wall with a 25 kcal/mol/Å^{-2} force constant was added on the RC ξ_{CEC} if its value exceeded 8 Å. The WT-MTD was run for ~20 ns for both Glu and Lys, and the PMF was obtained by summing the deposited gaussians. The error bar of the PMF was estimated by partitioning the full trajectory into 8 blocks and calculating the standard deviation of PMFs computed from the last 5 blocks. The pKa of Glu/Lys was computed from their PMFs via[116]

$$\mathrm{p}K_{\mathrm{a}} = \log\left(c_0 \int_0^{\dagger} 4\pi\xi_{\mathrm{CEC}}^2 \, e^{-\beta\left(F(\xi_{\mathrm{CEC}})-F(+\infty)\right)} \mathrm{d}\xi_{\mathrm{CEC}} \right) \qquad 4\text{-}11$$

where $c_0 = 1/1660$ Å^{-3} is the standard state concentration (1 M) expressed in number density, $F(\xi_{\mathrm{CEC}})$ is the PMF, and $F(\infty)$ is the value when ξ_{CEC} is sufficiently large and $F(\xi_{\mathrm{CEC}})$ reaches a plateau. The integral was evaluated up to the dividing surface (denoted by †) between protonated and deprotonated Glu/Lys. However, the integral value did not change if the integral went further because the free energy passing the dividing surface was sufficiently high for weak acids and the exponential integrand at further distances vanishes. Note that the $F(\xi_{\mathrm{CEC}})$ used in pK_{a} calculation (shown in Figure 4-2) differs from the PMF directly summed from gaussians (denoted as $F'(\xi_{\mathrm{CEC}})$) via the relation

$$4\pi\xi_{\mathrm{CEC}}^2 e^{-\beta F(\xi_{\mathrm{CEC}})} = e^{-\beta F'(\xi_{\mathrm{CEC}})} \qquad 4\text{-}12$$

In essence, eq 4-12 converts the probability density for the system to visit $[\xi_{\mathrm{CEC}}, \xi_{\mathrm{CEC}} + \mathrm{d}\xi_{\mathrm{CEC}}]$ into the probability density of visiting the infinitesimal spherical shell $4\pi\xi_{\mathrm{CEC}}^2\mathrm{d}\xi_{\mathrm{CEC}}$. The resulting $F(\xi_{\mathrm{CEC}})$ thus becomes a plateau with a sufficiently large ξ_{CEC}, in contrast to an ever-decaying $F'(\xi_{\mathrm{CEC}})$ that reflects the increasing probability density in $[\xi_{\mathrm{CEC}}, \xi_{\mathrm{CEC}} + \mathrm{d}\xi_{\mathrm{CEC}}]$ due to the larger accessible volume of $4\pi\xi_{\mathrm{CEC}}^2\mathrm{d}\xi_{\mathrm{CEC}}$ spherical shell.

Our simulations of SNase consisted of one SNase (PBD id: 1U9R[117]) solvated in a cubic water box of 70 Å on a side with 0.15 M of sodium chloride added. A buried residue V66 was mutated into Glu/Lys to form the V66E and V66K mutants. For both the protonated and deprotonated E66/V66 mutants, classical MD equilibrations were conducted in the NPT ensemble at 298 K under 1 atm for 200 ns. The temperature and pressure were controlled by a Nose-Hoover chain and by the Parrinello-Rahman barostat[118], respectively. All the bonds involving hydrogens were constrained using the LINCS algorithm[119] and a timestep of 2 fs was used to propagate the system. All classical equilibrations were carried out in the GROMACS package.[120, 121]

The SNase MS-RMD simulations, initiated from classical equilibrations, were conducted in NVT ensemble at 298 K using a timestep of 1 fs. The RC for proton disassociation from E66/K66 was defined as the same as the ξ_{CEC} for Glu/Lys in water. Another RC d_{SC} was defined as follows

$$d_{\mathrm{SC}} = \boldsymbol{v}_{\mathrm{SC}} \cdot \boldsymbol{n}_{\mathrm{prot}} \tag{4-13}$$

to characterize the sidechain rotation of E66/K66, given that protonated/deprotonated forms of E66/K66 were found to display distinct sidechain orientations via classical equilibrations (Figure 4-3). In eq 4-13, $\boldsymbol{v}_{\mathrm{SC}}$ represents the vector pointing from E66/K66 α-carbon to the geometric center of carboxylic group (COO) of E66 or the ammine nitrogen of K66, and $\boldsymbol{n}_{\mathrm{prot}}$ is a vector defined to reflect the protein overall orientation:

$$\boldsymbol{n}_{\mathrm{prot}} = \frac{\boldsymbol{r}_{\mathrm{C2}} - \boldsymbol{r}_{\mathrm{C1}}}{|\boldsymbol{r}_{\mathrm{C2}} - \boldsymbol{r}_{\mathrm{C1}}|} \tag{4-14}$$

where $\boldsymbol{r}_{\mathrm{C1}}$ is the center of backbone atoms of residues 15-19 and 61-65, and $\boldsymbol{r}_{\mathrm{C2}}$ is the center of backbone of residues 90-94 (illustrated in Figure 4-4). To restrain the lateral diffusion of solvated proton when completely disassociated from E66/K66, a harmonic potential wall u_{res} was added on a collective variable $r_{\perp}$ defined as the length of the orthogonal part of $\boldsymbol{v}_{\mathrm{CEC}}$ with respect to $\boldsymbol{n}_{\mathrm{prot}}$:

$$r_{\perp} = \left|\boldsymbol{v}_{\mathrm{CEC}} - (\boldsymbol{v}_{\mathrm{CEC}} \cdot \boldsymbol{n}_{\mathrm{prot}})\boldsymbol{n}_{\mathrm{prot}}\right| \quad 4\text{-}15$$

where $\boldsymbol{v}_{\mathrm{CEC}}$ is the vector pointing from E66 COO center or K66 nitrogen to the CEC. The force constant of the wall was 10 kcal/mol/Å^2, and the restraining potential was switched on when $r_{\perp} \geq$ 7 Å. To enhance the sampling of both ξ_{CEC} and d_{SC}, 2D-US on the two RCs was performed. Each US window was run for ~1 ns and the total simulation time was ~1 μs for both V66E and V66K. The weighted histogram analysis method (WHAM) [122] was used to combine the 2D-US data and compute the PMF $F(\xi_{\mathrm{CEC}}, d_{\mathrm{SC}})$, from which the one-dimensional (1D) PMF (Figure 4-3E) for proton disassociation was obtained by integrating out the d_{SC} degree of freedom (DOF). Specifically,

$$F(\xi_{\mathrm{CEC}}) = -\beta^{-1} \ln\left(\int e^{-\beta F(\xi_{\mathrm{CEC}}, d_{\mathrm{SC}})}\, \mathrm{d}d_{\mathrm{SC}}\right) \quad 4\text{-}16$$

Then, following the derivation in ref[123], the p*K*a of E66/K66 was given by

$$\mathrm{p}K_{\mathrm{a}} = \log\left(c_0 S_u \int_0^{\dagger} e^{-\beta\left(F(\xi_{\mathrm{CEC}}) - F(+\infty)\right)} \mathrm{d}\xi_{\mathrm{CEC}}\right) \quad 4\text{-}17$$

where the meaning of c_0 and the integral range were the same as eq 4-11, and the $S_u = \int_0^{\infty} 2\pi r_{\perp} e^{-\beta u_{\mathrm{res}}(r_{\perp})} \mathrm{d}r_{\perp}$ corrects for the introduced radial restraint on CEC. The errors reported for PMFs and pK_{a} were obtained from partitioning the trajectories of all US windows into 6 equally sized blocks and calculating the standard deviation using the final 4 blocks. All MS-RMD simulations were performed using the LAMMPS package[124] coupled with RAPTOR[125] for reactions and PLUMED 2[83] for the free energy sampling.

4.4 Results

4.4.1 Benchmark of reactive models in water

Table 4-1. Optimized MS-RMD Model Parameters Using ωB97X CDFT Data.

	Glu	Lys		Glu	Lys
B	3.94793	1.01096	V_{ii}	-153.284	-67.0979
b	1.41638	1.41969	$\epsilon^{\mathrm{LJ}}_{\mathrm{OE/NK-HH}}$	0.23117	0.0175896
b'	1.08444	1.07948	$\sigma^{\mathrm{LJ}}_{\mathrm{OE/NK-HH}}$	1.39561	1.68765
C	3.8605	0.989666	$\epsilon^{\mathrm{LJ}}_{\mathrm{Ow-HEP/HKP}}$	0.72595	0.0107436
c	1.14669	1.14233	$\sigma^{\mathrm{LJ}}_{\mathrm{Ow-HEP/HKP}}$	1.29196	1.77364
c_1	-25.0422	-25.015	$\epsilon^{\mathrm{LJ}}_{\mathrm{OE/NK-OH}}$	0.125249	0.115627
c_2	2.99968	3.02538	$\sigma^{\mathrm{LJ}}_{\mathrm{OE/NK-OH}}$	3.00742	3.20591
c_3	1.40533	1.37087	$\epsilon^{\mathrm{LJ}}_{\mathrm{OEP/NKP-Ow}}$	0.162054	0.162548
D	143.003	157.014	$\sigma^{\mathrm{LJ}}_{\mathrm{OEP/NKP-Ow}}$	3.07772	3.23758
α	1.8	1.70825			
r_0	0.975	1.02389			

The units of the listed parameters use kcal/mol as the energy unit and Å as the length unit.

The MS-RMD parameters obtained using the DM approach with ωB97X CDFT data are provided in Table 4-1, and model parameters determined from the BLYP data are given in Table 4-3. The resulting MS-RMD Glu/Lys models were benchmarked in water by computing the PMF of their ionization (Figure 4-2). Potential of mean force is arguably the most valuable information about the system as it provides the full free energy profile as a function of the reaction progress (monitored by the RC value). At $\xi_{\mathrm{CEC}} \approx 0.5$ Å, a narrow free energy well was observed in both the Glu and Lys PMFs. The molecular configurations corresponds to the well are protonated Glu/Lys (Figure 4-2 insets), and the well in PMF reveals the weak acid nature of both Glu and Lys. The steep wall of $\xi_{\mathrm{CEC}} < 0.5$ Å is caused by the large energy penalty of compressing O-H/N-H bond of Glu/Lys. The smooth uphill increase in the area of $0.5 \text{ Å} < \xi_{\mathrm{CEC}} < 1.4$ Å arises from the energy barrier of proton disassociation from Glu/Lys to its first solvation shell water. The free energy was observed to reach a peak at around 1.4 Å, the point at which the proton is shared by Glu/Lys and the water, forming a Glu/Lys-replaced Zundel ($H_5O_2^+$) configuration (Figure 4-2 insets). The second free energy well was noted at $\xi_{\mathrm{CEC}} \approx 2.3$ Å when the excess proton is found on the first

solvation shell water. The resulting H_3O^+ and negatively charged Glu closely interact with each other to form a CIP (Figure 4-2A inset). The favorable Coulombic interaction yields a deeper potential well for Glu compared to that of Lys. In the latter case, a CIP cannot be formed between a neutral deprotonated Lys and H_3O^+, but the system is stabilized by forming an Eigen cation ($H_3O(H_2O)_3^+$) with one water substituted by the Lys (Figure 4-2B inset). We observed that a second free energy potential well at around 2.3 Å was shallow within the ωB97X model in comparison with that of the BLYP model, which is consistent with the expectation that the GGA functional BLYP over-stabilizes the CIP due to its overestimation of charge transfer than the hybrid functional ωB97X. After passing the well, the value of the free energy potential increases within the range of $2.3\ \text{Å} < \xi_{\text{CEC}} < 4\ \text{Å}$, which corresponds to proton disassociation from the first shell water. Although the excess proton mostly resides on a second shell water when $\xi_{\text{CEC}} \approx 4\ \text{Å}$ (Figure 4-2 insets), this layer of water is sufficiently diffusive to become indistinguishable from the bulk water, given that the PMFs reach a plateau after this point.

For a direct comparison with the experimental results, we calculated the pK_a of Glu and Lys in water from the PMF using eq 4-11. The resulting data are summarized in Table 4-2. We found that the ωB97X models accurately predicted the pK_a values for both Glu and Lys. However, our findings based on the BLYP models were slightly less accurate, but still within ~1 pH unit of error. The success of the DM-based MS-RMD models encouraged us to apply them to a more complicated system, SNase. Given that the ωB97X models showed better agreement with the experimental pK_a than the BLYP models, the latter models were not tested in subsequent SNase simulations.

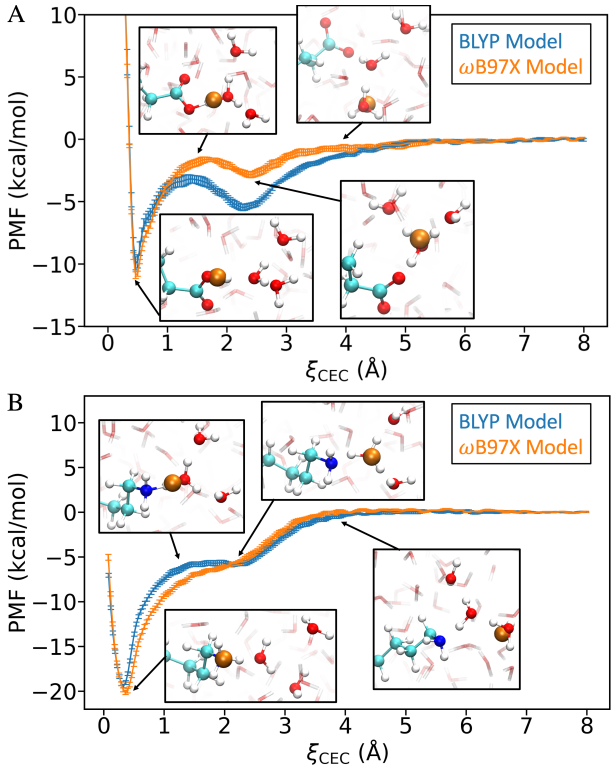

Figure 4-2. The proton disassociation potential of mean force of (1) Glu in water, and (2) Lys in water. computed from BLYP or ωB97X parametrized MS-RMD models. The insets show representative molecular configurations at the corresponding positions on the PMFs. The center of excess charge (effective position of H^+) is rendered as an orange sphere.

Table 4-2. MS-RMD Predicted and Experimentally Determined pK_a of Glu and Lys.

		In Water	In SNase V66 Mutants
Glu	Simulation (ωB97X)	4.1±0.2	9.8±0.3[a]
	Simulation (BLYP)	3.4±0.4	
	Experiment	4.15[b]	9.00-9.10[c] (8.73-9.28[d])
Lys	Simulation (ωB97X)	10.7±0.2	5.8±0.3[a]
	Simulation (BLYP)	10.0±0.1	
	Experiment	10.67[b]	5.61-6.05[e] (6.25-6.45[f])

[a] The PDB structure (1U9R) used in the simulation was the PHS form of SNase (engineered with three substitutions: P117G, H124A, and S128L). Note that the missing 45-50 loop in the solved structure was not modeled in our simulations, and thus the simulated system more resembles the Δ+PHS form (additional G50F, V51N and 44-49 deletion from PHS).
[b] Taken from ref[126].
[c] pK_a in Δ+PHS by potentiometry taken from ref[127].
[d] pK_a in PHS from chemical denaturation taken from ref[127].

Table 4-2. MS-RMD Predicted and Experimentally Determined pK_a of Glu and Lys continued.
[e] pK_a in Δ+PHS from chemical denaturation taken from ref[128].
[f] pK_a in PHS by potentiometry take from ref[129].

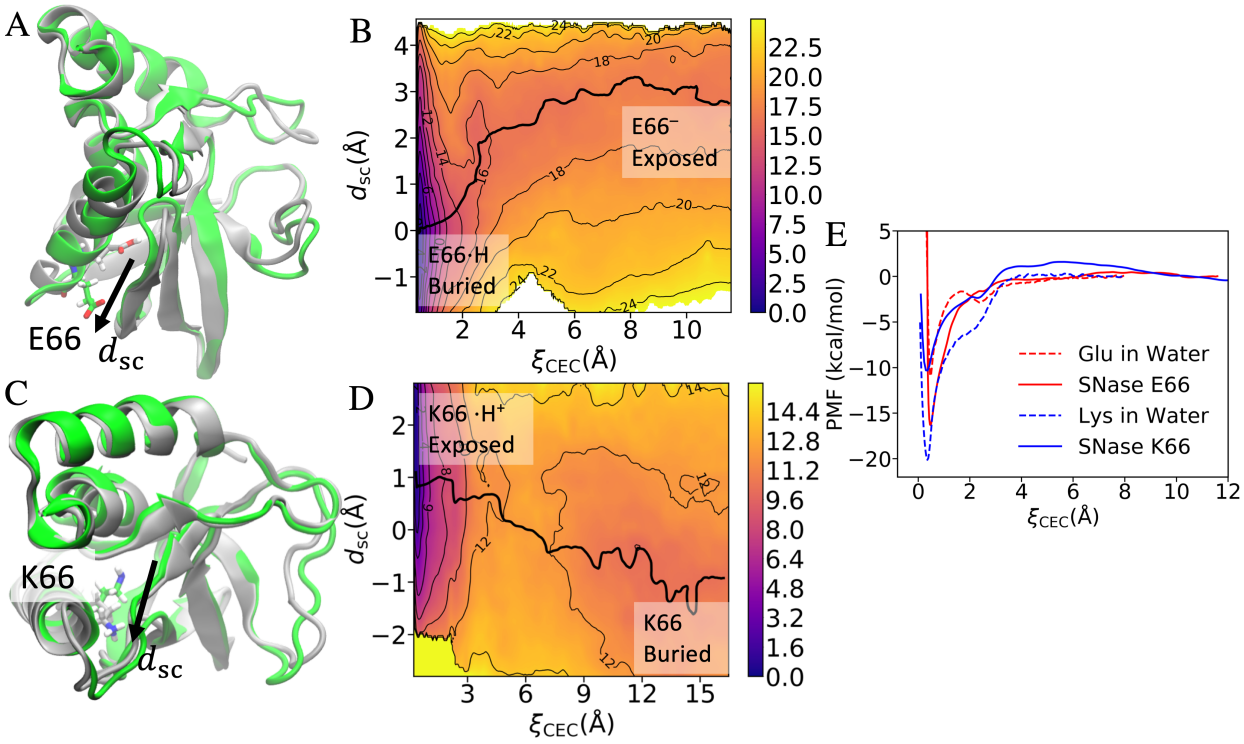

Figure 4-3. The coupling between proton disassociation and sidechain orientation in SNase. (A) Classical equilibrated configurations of V66E mutants when E66 is deprotonated (green) and protonated (gray). (B) The potential of mean force in kcal/mol of E66 ionization and its sidechain rotation. The minimum free energy path is shown as a black curve. The PMF statistical error can be found in Figure S2A. (C) Classical equilibrated configurations of V66K mutants when K66 is deprotonated (green) and protonated (gray). (D) The PMF of K66 ionization and its sidechain rotation. The minimum free energy path is shown in black. The PMF error can be found in Figure S2B. (E) Comparison of ionization PMFs of Glu (red) and Lys (blue) in water (dashed) and in SNase (solid; computed via eq 4-16).

4.4.2 Benchmark of reactive models in SNase

SNase is a well-known protein model system for proton titration of internal ionizable groups. Several SNase mutants have been generated by focusing on one of the buried hydrophobic residues, V66, and substituting it with Asp, Glu and Lys. The pK_a of the mutated residue has been measured experimentally and large pK_a shifts were observed in favor of the neutral forms of the

residues.[127-129] Therefore, the SNase mutants provide an ideal and well-regulated protein environment for benchmarking reactive models against experimental characterizations.

Classical equilibrations of V66E and V66K mutants were performed for both protonated and deprotonated forms of E66 and K66. For the neutral form of E66 (deprotonated) and K66 (protonated), the residue adopted a buried conformation, while a charged E66/K66 was found to be exposed to water through the rotation of its sidechain to stabilize its charges in the higher polarizable medium (Figure 4-3A & Figure 4-3C). This observation is indeed consistent with the experimental pK_a shifts for Glu and Lys toward stabilizing a neutral form.

Considering the distinct conformations that are dependent on the protonation state, a RC d_{SC} reflecting sidechain orientation was defined and explicitly sampled along with the RC ξ_{CEC}, which describes proton disassociation in the reactive 2D-US (see section 4.3 for details). The next section provides a detailed discussion of the informative 2D PMF findings (Figure 4-3B & Figure 4-3D) calculated by the 2D-US. We consider the pK_a of E66 and K66 calculated from the PMFs (Figure 4-3E) as a byproduct of the free energy calculations but enables a direct comparison to experimental measurements. According to Table 4-2, the calculated pK_a values were found to be in good agreement with those of their experimentally determined counterparts, with better agreement for the Lys model, which achieved a nearly perfect match. The Lys and Glu models were parametrized following the same procedure, except that the Morse bond of Lys was fit to an energy scan, while Glu was taken from our previous work[89]. The better performance of our Lys model suggests that the current Glu model may be readily improved by a bond energy scan followed by a reparameterization. Notably, we used the exactly same MS-RMD parameters of Glu/Lys in water when simulating SNase, but our models demonstrated high accuracy and transferability in predicting the acidic and basic residue pK_a in the two distinct environments.

4.4.3 Proton transport and conformation coupling in SNase

As discussed in the prior section, we compared the model predicted pK_a with experimental values. It is worth noting that a reactive MD model is not only useful for calculating reaction equilibrium constants, but can also predict reaction rates when combined with appropriate kinetic theory and modeling, such as the transition state theory[130] and the Markov state model.[131] More importantly, atomistic details obtained from reactive MD simulations enable a detailed exploration of the reaction mechanisms, as well as the identification of crucial affecting factors and interactions. An important question arises regarding SNase is that how proton transport is coupled with the protein conformation. As discussed, the classical equilibrium data indicate that the sidechain orientation of E66/K66 depends on its protonation state. However, in the absence of knowledge of the reaction pathways, classical simulations can only determine the metastable states. In contrast, reactive MD can determine whether the sidechain rotation takes precedence in facilitating the protonation/deprotonation process or is actually a result responding to a protonation state change. Interestingly, the answer is likely neither according to our 2D PMF results (Figure 4-3B and Figure 4-3D). The most probable reaction pathway (or, equivalently, the minimum free energy path; MFEP) shows that the protonation disassociation and the sidechain rotation are not two independent processes that happen in a step-by-step manner. Instead, these two processes are highly and reciprocally coupled, which can be verified by the ramped slope of the MFEP connecting the two ending points of the charged, exposed E66/K66 state and the neutral buried state. If the PT DOF ξ_{CEC} is driven by some external force such as a proton gradient, and if the pulling is a quasistatic process, then the movement of ξ_{CEC} will result in a responsive movement in d_{SC} along the curvy MFEP (and vice-versa). In addition to the sidechain orientation observed in this study, we found in our previous work [57, 59, 85, 107] and will discuss in Chapters 7 and 8 that

this type of coupling typically occurs between PT and hydration in confined hydrophobic channels. This outcome could also be the case for PT and ligand transport in some proton-driven transporters; indeed, if this should be the case then it is understandable why the proton and ligand can both be the driving force to the transport of each other.

4.5 Conclusions

As a generalization to the previous FM scheme, we proposed a DM framework to systematically parameterize reactive MD models from CDFT. We used the proton disassociation reactions of Glu and Lys as our case studies and found the DM-based models can reproduce the experimental absolute pK_a of the acid and the basic amino acid in water. We also confirmed that the same reactive models were able to accurately capture the large pK_a shifts introduced by the apolar SNase protein environment. In addition, the efficiency of the reactive MD enabled us to compute the joint free energy surface of the proton coordinate, along with the sidechain orientation of E66/K66, from which we found that PT is coupled to this local conformational change. In other words, the two processes can be considered as mutually interactive, in contrast to being a sequential process whereby the protonation state change occurs either before or after a conformational change. Since our results indicated that our models are transferrable, we believe they can readily be applied to other biomolecular and biomaterial systems. Moreover, the noted efficiency of the models can facilitate the discovery of other DOFs coupled to PT in those systems, such as hydration, global conformational change, and ligand transport. Finally, we anticipate that the DM approach provides a potentially powerful tool for developing accurate and transferrable reactive modes for other chemical reactions in the future.

4.6 Appendix

Table 4-3. Optimized MS-RMD Model Parameters Using BLYP CDFT Data.

	Glu	Lys		Glu	Lys
B	3.3439	1.10946	V_{ii}	-152.795	-69.4226
b	1.44315	1.35386	$\epsilon^{\mathrm{LJ}}_{\mathrm{OE/NK-HH}}$	0.192089	0.00769145
b'	1.10416	1.06392	$\sigma^{\mathrm{LJ}}_{\mathrm{OE/NK-HH}}$	1.31464	1.68974
C	3.3731	1.04393	$\epsilon^{\mathrm{LJ}}_{\mathrm{Ow-HEP/HKP}}$	0.724763	0.0106015
c	1.18011	1.18869	$\sigma^{\mathrm{LJ}}_{\mathrm{Ow-HEP/HKP}}$	1.22985	1.69318
c_1	-29.1037	-24.953	$\epsilon^{\mathrm{LJ}}_{\mathrm{OE/NK-OH}}$	0.108658	0.141469
c_2	3.02482	2.93992	$\sigma^{\mathrm{LJ}}_{\mathrm{OE/NK-OH}}$	3.08549	3.19191
c_3	1.53429	1.51078	$\epsilon^{\mathrm{LJ}}_{\mathrm{OEP/NKP-Ow}}$	0.168629	0.19681
D	143.003	157.014	$\sigma^{\mathrm{LJ}}_{\mathrm{OEP/NKP-Ow}}$	3.06885	3.19355
α	1.8	1.70825			
r_0	0.975	1.02389			

The units of the listed parameters use kcal/mol as the energy unit and Å as the length unit.

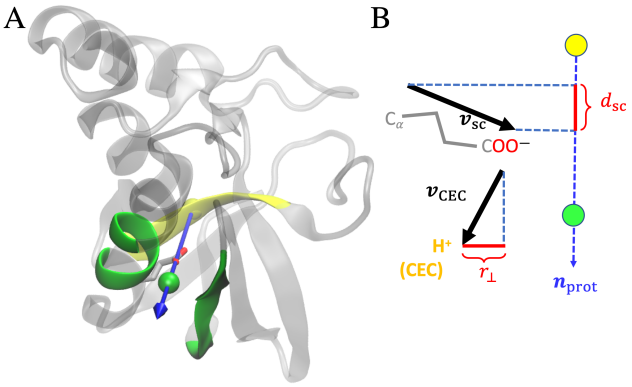

Figure 4-4. Illustration of reaction coordinate definitions in SNase. (A) A molecular figure of V66E showing the definition of $\boldsymbol{r}_{\mathrm{C1}}$ and $\boldsymbol{r}_{\mathrm{C2}}$. Residues 15-19 and 61-65 that define $\boldsymbol{r}_{\mathrm{C1}}$ are in green and residues 90-94 that define $\boldsymbol{r}_{\mathrm{C2}}$ are in yellow. The positions of $\boldsymbol{r}_{\mathrm{C1}}$ and $\boldsymbol{r}_{\mathrm{C2}}$ are shown as a green and yellow sphere. The blue arrow indicates the direction of $\boldsymbol{n}_{\mathrm{prot}}$. (B) A cartoon of V66E explaining the definition of d_{SC} and $r_\perp$. The sidechain of E66 is represented in grey sticks, the $\boldsymbol{r}_{\mathrm{C1}}$ and $\boldsymbol{r}_{\mathrm{C2}}$ are represented as green and yellow plates, and the excess proton CEC is shown in orange.

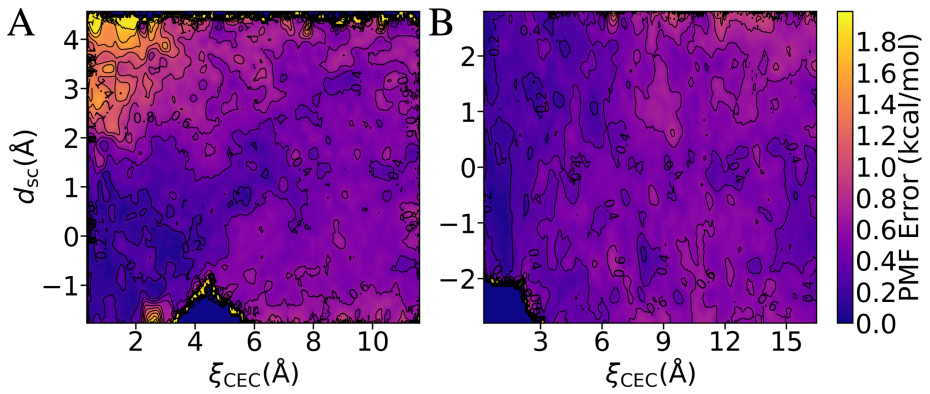

Figure 4-5. (A**)** PMF error of V66E (Figure 4-3B). (B) PMF error of V66K (Figure 4-3D).

5 Classical Ab Initio Molecular Dynamics at an Elevated Temperature Does Not Model Well the Nuclear Quantum Effects at Ambient Temperature

5.1 Introduction

Due to the light mass of the hydrogen atoms, NQEs are usually assumed to be crucial for a quantitative modeling of structural, thermodynamic, and dynamical properties of liquid water[95]. Though NQEs appear to be minimal for intermolecular properties in some water simulations[132, 133], they were found to be non-negligible in some cases[93, 134] as demonstrated by the differences in thermodynamic properties between light and heavy water[95].

In principle, Feynman's imaginary-time path-integral formalism[96] enables the modeling of NQEs in liquid water to numerical accuracy, but the associated high computational cost has hindered widespread application of PIMD simulations until recent developments of more efficient approximations[97], such as RPC[98], the ring-polymer interpolation[99], and the PI+GLE approach[100]. The computational cost of a PIMD simulation of liquid water significantly increases when the underlying Born-Oppenheimer potential energy surface is calculated "on the fly" as in AIMD simulations where Kohn-Sham DFT (KS-DFT)[135] is generally used to solve the (electronic) Schrödinger equation at each step of the MD trajectory. As a consequence, most of the AIMD simulations reported in the literature were conducted ignoring NQEs and treating the nuclei as classical particles.

Among existing exchange-correlation functionals, GGA functionals have been extensively used in AIMD simulations of liquid water due to their relatively lower computational cost. GGA functionals typically overestimate the strength of the hydrogen bonds in water. This results in over-structuring of the liquid phase which is accompanied by slow molecular diffusion and, in some cases, glass-like behavior.[31] The inclusion of dispersion corrections was found to partially alleviate

these problems. An alternative approach adopted in the literature to overcome over-structuring and slow diffusion in AIMD simulations of liquid water consists in performing the simulations at a higher temperature. While simulations at a higher temperature sample a different thermodynamic ensemble, it is implicitly assumed that the extra thermal energy effectively mimics NQEs that are missed in classical simulations at room temperature. Within this assumption, the results obtained from classical AIMD simulations carried out at higher temperature are thus considered to be equivalent to the actual quantum results at room temperature.

Herein, we employ the RPC method to explicitly model NQEs at room temperature and systematically benchmark NQEs on structural correlations as well as dynamical properties of liquid water at room temperature against classical simulations carried out at higher temperature. Our analysis includes three exchange-correlation functionals: (1) the strongly constrained and appropriately normed (SCAN) functional, a meta-GGA functional[136], (2) BLYP-D3, one of the most common GGA functionals, and (3) BLYP-D3 with EDS correction[35]. The EDS correction employs a minimal bias to improve the BLYP-D3 description of hydrogen bonding in liquid water, and has been shown to provide a more accurate water properties[35]. Additionally, we perform the same analyses using MB-pol,[36] arguably the most accurate water model to date. MB-pol is a data-driven model rigorously derived from the many-body expansion of the interaction energies calculated at the coupled cluster level of theory which has been shown to accurately predict the properties of water, from small gas-phase clusters to liquid water and ice. [37, 38] We note that the goal of this study is not to determine which water model and simulation protocol best reproduces the experimental data, but to assess the validity of using an elevated temperature in classical AIMD simulations of liquid water to effectively model NQEs at ambient temperature.

5.2 Methods

5.2.1 DFT simulation details

The BLYP-D3 and EDS-BLYP-D3 simulations were performed with 128 water molecules in a cubic simulation box of side L = 15.64 Å. The GTH pseudopotentials were used to model the core electrons, while a TZV2P basis set was used to expand the Kohn-Sham orbitals and a plane wave basis set with a cutoff of 400 Ry was used to expand the electron density. The orbital transformation (OT) method[137] was used to optimize the wave function at each step, using a self-consistent field (SCF) convergence criterion of 1×10^{-7} a.u. The SCAN simulations were performed with 64 water molecules in a cubic box of side L = 12.66 Å to be consistent with the setup in ref[138]. Similar to the BLYP-D3 simulations, the TZV2P basis set was used but a larger plane wave basis set with a cutoff of 600 Ry was used for better SCF convergence. The GTH pseudopotentials optimized for SCAN were used to model the core electrons (https://github.com/juerghutter/GTH). The OT method was used in SCF and the convergence criteria was the same as BLYP-D3.

For the classical MD simulations with BLYP-D3 and EDS-BLYP-D3, the system was equilibrated in the canonical (NVT) ensemble for 90 ps at 298 K and for 85 ps at 328 K. In the case of SCAN, the NVT equilibration was carried out for 70 ps at 298 K and for 72.5 ps at 328 K. In all simulations, a Nose-Hoover chain thermostat with a characteristic frequency of 3000 cm^{-1} was coupled to all degrees of freedom, and a timestep of 0.5 fs was used to integrate the Newton's equations of motion. All simulations were carried out with the CP2K software package.[76, 77] The EDS correction was added via a modified version of PLUMED 2 coupled to CP2K, with the EDS parameters taken from ref[35], while the simulations with the SCAN exchange-correlation functional were carried out by linking the Libxc library to CP2K.

The dynamical properties were calculated by performing MD simulations in the microcanonical (NVE) ensemble which were started from configurations previously equilibrated in the NVT ensemble. At each temperature, five independent replicas were launched from the NVT trajectories at intervals of 10 ps, and were carried out for 30-40 ps. The MTS scheme was used to accelerate the simulations, with an inner timestep of 0.25 fs and an outer timestep of 2 fs. For the reference potential used in the MTS scheme, we used a deep learning potential (DP) trained on BLYP-D3, EDS-BLYP-D3, and SCAN data. The training set for BLYP-D3 contained 2500 configurations sampled at 298 K and 2500 configurations sampled at 328 K. The training set for EDS-BLYP-D3 contained 132000 configurations sampled at 298 K. The training set for SCAN contained 10000 configurations sampled at 298 K and 10000 configurations sampled at 328 K. Both energies and forces were used in the training. The training was conducted using the DeepMD-kit tool with the smooth edition of DPMD. The BLYP-D3 energies and forces were calculated with CP2K, the forces for the DP potential were computed using LAMMPS coupled to DeepMD-kit, the EDS correction was computed using PLUMED 2, and the MD simulations were performed using the i-Pi force engine. The structural properties obtained from NVE simulations were found to be identical with those obtained in the NVT ensemble. An independent NVE simulation was carried out with a timestep of 0.5 fs, without applying the MTS approximation, to calculate the diffusion coefficient which was found to be identical to the value obtained from an analogous simulation carried out using the MTS scheme.

Quantum dynamics simulations were carried out using TRPMD. Following ref[93], the RPC approach was employed with $P' = 1$, i.e. centroid contraction. The path integral of each atom was discretized with $P = 30$ beads, and all the analyses were performed on the centroid coordinates.

The same reference potential used in MTS was used in RPC. Five independent TRPMD trajectories of 30-40 ps each were performed to calculate the dynamical properties.

5.2.2 MB-pol simulations

Classical and quantum MB-pol trajectories were taken from refs[37, 133].

5.3 Results and discussion

We analyzed simulations with four water models (BLYP-D3, EDS-BLYP-D3, SCAN, and MB-pol) at ambient temperature (298 K) and an elevated temperature (328 K) commonly used to effectively mimic NQEs. The simulations at 298 K were performed with both classical and quantized nuclei, while classical nuclei were used at 328 K.

Figure 5-1 to 5-3show the radial distribution functions of the four water models. As seen in Figure 5-1(a) to Figure 5-3(a), NQEs only have a minimal effect on the two-body correlations in MB-pol water, with the first peak in the O-O RDF being slightly less structured with quantized nuclei, which, in turn, slightly improves the agreement with the experimental data. NQEs show an opposite effect in the three DFT models where all the O-O, O-H, and H-H RDFs become more structured with quantized nuclei. Notably, when NQEs are explicitly accounted for, the hydrogen bonds in DFT water shrink as shown by the first peak in the O-H RDFs (Figure 5-2(b)-(d)) moving towards shorter distances (dotted vs. solid curves). This agrees with the experimental observation of shorter hydrogen bond length in light water than heavy water.[139] On the other hand, the effect of elevated temperature is consistent among the four water models – the extra thermal energy reduces the solvation structure in all the RDFs. It follows that performing classical MD simulations at an elevated temperature apparently mimics NQEs in simulations with MB-pol, while has opposite effects in simulations with the three DFT models. As a result, NQEs make the RDFs

calculated with BLYP-D3 and SCAN even more structured. It should be noted that including NQEs improves the agreement between the EDS-BLYP-D3 and experimental RDFs.

We next examined the three-body correlation by computing the tetrahedral order parameter q defined as[140]

$$q = 1 - \frac{3}{8}\sum_{i\neq j}\left(\cos\theta_{ij} + \frac{1}{3}\right)^2, \qquad 5\text{-}1$$

where θ_{ij} is the $\mathrm{O}_i - \mathrm{O} - \mathrm{O}_j$ angle centered on a given oxygen O, and the sums are over the four closest oxygen atoms around O. The value of q provides a measure of tetrahedral order, with a value of 1 corresponding to a perfect tetrahedral arrangement and a value of 0 representing the ideal gas limit. Figure 5-4 summarizes the distribution of q for the four water models. NQEs still play a small role in determining the solvation structure in the MB-pol simulations at ambient temperature. More pronounced differences are found in the simulations with the three DFT models, with the peak at $q \approx 0.85$ increasing and the peak at $q \approx 0.5$ decreasing, which is consistent with the more structured RDFs observed in Figure 5-1 to Figure 5-3. Increasing the temperature reduces the tetrahedral structure in all four water models, which makes the distributions obtained with BLYP-D3 and SCAN qualitatively more similar to the distribution calculated with MB-pol at ambient temperature. It should be noted that this apparent better agreement with the MB-pol distribution is the result of fortuitous error cancellation associated with intrinsic deficiencies in the BLYP-D3 and SCAN ability to represent water and not a consequence of NQEs since, when NQEs are explicitly taken into account in the simulations, both BLYP-D3 and SCAN predicts a significantly more tetrahedral structure.

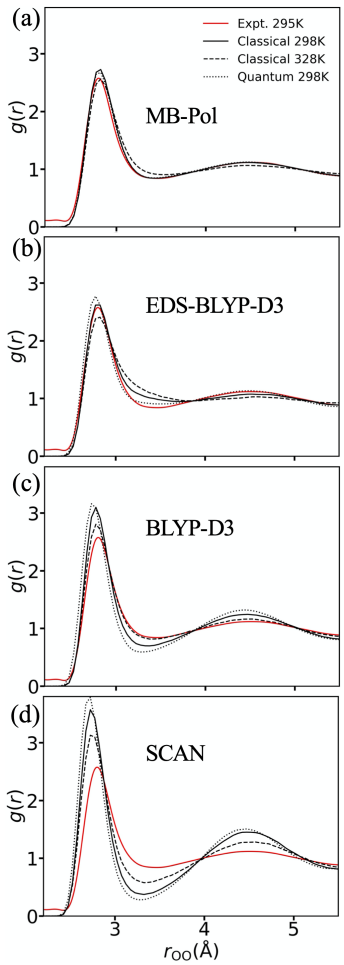

Figure 5-1. O-O RDFs of (a) MB-Pol, (b) EDS-BLYP-D3, (c) BLYP-D3 and (d) SCAN water model at 298K with classical nuclei (solid) and with quantized nuclei (dotted), and at 328K (dashed). The experimental value[141] at 295K was plotted in red.

To determine how the additional thermal energy available at 328 K perturbs the system's dynamics depending on the underlying water model, we compute the hydrogen-bond dynamics and water diffusion constant at both temperature, with and without including NQEs. In this analysis, we consider that molecule j is hydrogen-bonded to molecule i if the O-O distance is shorter than 3.5 Å and the $\mathrm{H}_i - \mathrm{O}_i - \mathrm{O}_j$ angle is smaller than 30°, where H_i is one of the two bonded hydrogen atoms to O_i. At a given time t, the hydrogen bond matrix is computed as

$$h_{ij} = \begin{cases} 1, & j \text{ hydrogen} - \text{bonded to } i \\ 0, & \text{otherwise} \end{cases}, \qquad 5\text{-}2$$

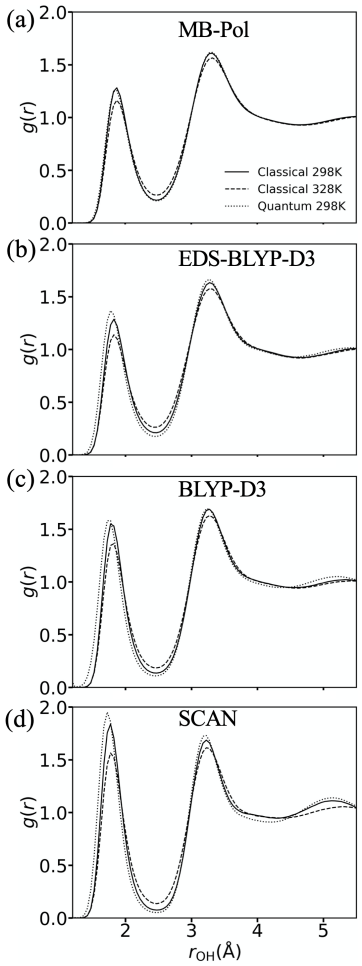

Figure 5-2. O-H RDFs of (a) MB-Pol, (b) EDS-BLYP-D3, (c) BLYP-D3 and (d) SCAN water model at 298K with classical nuclei (solid) and with quantized nuclei (dotted), and at 328K (dashed).

The corresponding hydrogen-bond correlation function is defined as

$$\langle h(0)h(\tau)\rangle = \frac{1}{N_{\mathrm{wat}}(N_{\mathrm{wat}}-1)}\sum_{i\neq j}\langle h_{ij}(0)h_{ij}(\tau)\rangle. \qquad 5\text{-}3$$

In the analysis of the quantum simulations with the DFT models, we observe water auto-disassociation in approximately 5% of the time. Due to the ambiguity of assigning bonded hydrogens to oxygens with auto-ionized water, the hydrogen bond matrix h_{ij} is set to be the value at the closest time when the bonding topology is well defined, i.e. when water auto-ionization does not happen. This transient auto-ionization was also reported in ref[142] and is likely the consequence

of NQEs further reducing the proton transfer barrier between two water molecule which is already underestimated in the DFT models of water.[143] As such, NQEs strengthen the hydrogen bonds and makes DFT water more structured as shown in Figure 5-1 to Figure 5-4, which results in a slower hydrogen-bond dynamics as shown in Figure 5-5. As expected, simulations carried out at 328 K display an accelerated hydrogen-bond dynamics as seen in more rapid decays of the corresponding hydrogen-bond correlation functions.

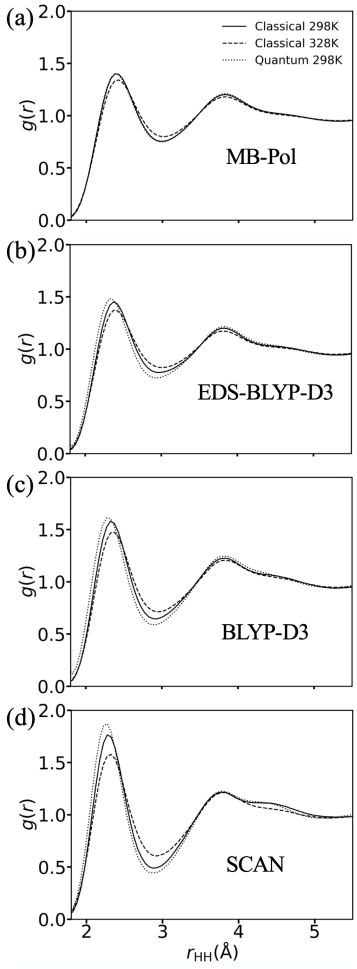

Figure 5-3. H-H RDFs of (a) MB-Pol, (b) EDS-BLYP-D3, (c) BLYP-D3 and (d) SCAN water model at 298K with classical nuclei (solid) and with quantized nuclei (dotted), and at 328K (dashed).

We further investigate water dynamics by computing the self-diffusion constant from the linear fit to the 5 ps -15 ps segment of MSD,

$$\mathrm{MSD}(t) = \langle (\boldsymbol{r}_{\mathrm{O}}(t) - \boldsymbol{r}_{\mathrm{O}}(0))^2 \rangle, \qquad 5\text{-}4$$

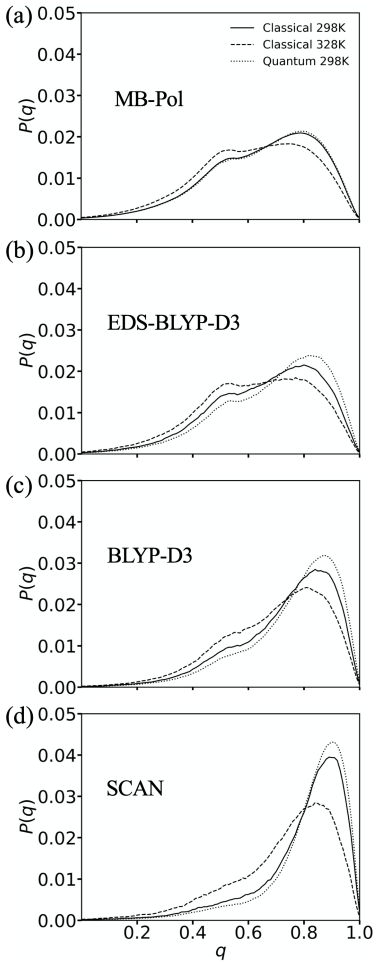

Figure 5-4. Tetrahedral order parameter q distribution of (a) MB-Pol, (b) EDS-BLYP-D3, (c) BLYP-D3 and (d) SCAN water model at 298K with classical nuclei (solid) and with quantized nuclei (dotted), and at 328K (dashed).

where $\boldsymbol{r}_{\mathrm{O}}$ represents the oxygen position of a water molecule. The computed values are summarized in Table 5-1. We also report the diffusion constants after a correction for the finite simulation box used in simulations via

$$D(\infty) = D(L) + \frac{\xi k_B T}{6\pi\eta L}, \qquad 5\text{-}5$$

where $\xi = 2.837297$ is a constant for cubic boxes, k_B is the Boltzmann constant, T is the simulation time, and L is the simulation box side length. In solving eq 5-5, the experimentally

determined viscosity of water η is used,[144] which results in an overestimation of $D(\infty)$ for the BLYP-D3 and SCAN models since the over-structuring predicted by these two models would actually be associated with a viscosity higher than the experimental value. Clearly, for the three DFT models, the NQE slows down the water self-diffusion while the elevated temperature accelerates all the dynamics. For MB-pol, in spite of a slower hydrogen-bond dynamics, NQEs do not introduce significant difference in the diffusion constant at ambient temperature, while, as expected, a much faster diffusion is observed at 328 K.

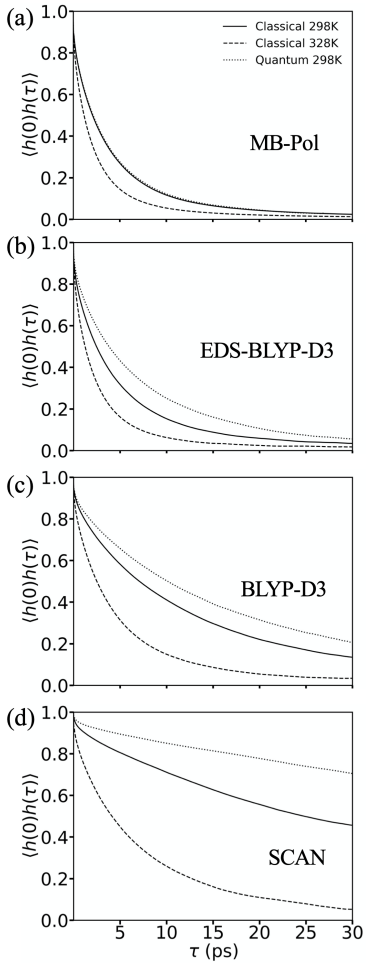

Figure 5-5. Hydrogen bond correlation function of (a) MB-Pol, (b) EDS-BLYP-D3, (c) BLYP-D3 and (d) SCAN water model at 298K with classical nuclei (solid) and with quantized nuclei (dotted), and at 328K (dashed).

Table 5-1. Self-diffusion constants of water in $Å^2/ps$.

Model		MB-Pol	EDS-BLYP-D3	BLYP-D3	SCAN
Diffusion Constant	Classical 298K	0.23	0.19	0.08	0.03
	Classical 328K	0.38	0.33	0.20	0.14
	Quantum 298K	0.23	0.1	0.06	0.01
Diffusion Constant after Size Correction	Classical 298K	0.27	0.23	0.12	0.09
	Classical 328K	0.42	0.38	0.25	0.20
	Quantum 298K	0.27	0.17	0.10	0.07
Experimental[145]	298K	0.2299			
	329K	0.4444			

Although it is not the main focus of this study to compare water models relative to the experimental data, it is worth noting that, among the four models considered in our analyses, MB-pol provides the most accurate description of water at both temperatures, followed by the EDS corrected BLYP-D3 model. Both structural and dynamical properties of BLYP-D3 at 328 K are accidentally close to the experimental values determined at 298 K but not as the result of effectively mimic NQEs.

5.4 Conclusions

In this Chapter, we systematically investigated the effect of an elevated temperature in simulating water, with a particular focus on the empirical relation between temperature and nuclear quantum effects. We considered three DFT water models, the widely used BLYP-D3 functional, with and without the EDS correction, and the meta-GGA SCAN functional. For all three DFT water models, the analysis of several structural and dynamical properties indicates that performing classical MD simulations at a higher temperature (328 K) introduces distinct and often opposite effects compared to performing quantum simulations at 298 K. For MB-pol, the elevated temperature seems to have a similar softening effect as NQEs at the two-body level, but the three-body correlation, hydrogen-bond dynamics, and diffusivity are clearly disrupted by the high temperature. These findings suggest that "mimicking" NQEs in water by performing classical MD

simulations at an elevated temperature may be problematic and, in some cases, misleading. Importantly, we consistently found stronger hydrogen bonds when NQEs are explicitly taken into account in simulations with all four water models considered in this study. This implies that the over-structuring predicted by the DFT models is further emphasized by NQEs. We, however, noted that performing classical MD simulations with DFT models at an elevated temperature accidentally fixes the over-structuring issue in an *ad hoc* way. Based on our analyses, we conclude that the elevated temperature approach does not represent a correct way of effectively mimicking NQEs and possibly needs further careful characterizations, including higher-order correlations and both molecular and collective dynamics, in a case-by-case manner before being applied to other systems or new computational models.

6 Understanding the Essential Nature of the Hydrated Excess Proton Through Simulation and Interpretation of Recent Spectroscopic Experiments

6.1 Introduction

The hydrated excess proton (aka “hydronium cation” plus nearby solvating water molecules) is pervasive in complex systems, e.g., proteins[13, 18, 146, 147] and renewable energy materials.[148-150] For more than two centuries, researchers have been studying its fundamental solvation and transport characteristics. The hydrated excess proton has an abnormally high diffusion coefficient in bulk water when compared to other +1 cations[146, 151], which is usually explained by the Grotthuss mechanism. The solvation structure of hydrated excess proton is sometimes described by the limiting cases of either an Eigen cation or a Zundel cation, but the precise nature of the excess proton hopping mechanism and most stable solvation structures is a topic of ongoing research (see, e.g., refs[152, 153]).

The solvation and transport properties of hydrated excess protons have been the topic of numerous simulation studies, see, e.g., refs[6, 87, 154-157] The most prevalent PT mechanism is thought to be an Eigen-Zundel-Eigen (EZE) mechanism, in which the most stable structure is a "distorted" Eigen cation and the Zundel cation is primarily an intermediate complex.[5, 158] On the other hand, some simulation literature suggests Zundel-Zundel conversions,[6, 155, 159, 160] although the underlying model appears to introduce a bias toward such a picture in some cases.

In reality, the hydrated proton solvation structure is more complex than either Eigen or Zundel cations. In fact, the hydrated excess proton can be found in a wide variety of configurations, making Eigen and Zundel cations the limiting structures that can hardly be deconvoluted. For example, Tuckerman et al.[6, 7] utilized AIMD simulations of an excess proton in 32 waters to demonstrate the presence of a “special pair” between the hydronium and a nearby water molecule,

which seems to be very characteristic of a Zundel cation. Later MD investigations identified the structure of the solvated proton as a distorted Eigen cation, in which the three-fold symmetry is disrupted due to the proton preference to one of the three coordinated water.[5, 87, 161] According to multistate empirical valence bond (MS-EVB) and AIMD simulations, the identity of the special partner is not static, but rather swaps with the other two hydrating water molecules in the Eigen cation on timescales of tens of femtoseconds[5], termed as the "special pair dance". Hence, these studies suggested that the primary structure of the hydrated proton in water is, on average, a "distorted" Eigen cation, but also that this structure is quite dynamic among the three possible special pairs.

The difficulty in matching observed infrared frequencies with structural information derived from MD simulations further complicates elucidating the hydrated proton solvation structure. Although gas-phase studies have been useful in connecting frequencies with structures,[162-165] the condensed phase adds to the complexity owing to thermal fluctuations.[87, 142] The acidic IR spectra has four distinct characteristics when compared to the absorption spectrum of pure water[88, 166, 167]: (1) a red shift in the O-H peak of bulk water due to stronger hydrogen bonding environments around the excess proton; (2) an acid continuum ranging from 2000-3200 cm^{-1}, which is most recently ascribed to more distorted Eigen-like configurations;[167] (3) a peak at 1200 cm^{-1} corresponding the PTM between two flanking waters; and (4) a peak at 1750 cm^{-1} corresponding to flanking water bend.

In the past few years, non-linear spectroscopy experiments have been pioneering efforts to understand the hydrated excess proton.[3, 8-10, 168, 169] For example, the PTM $|1\rangle \rightarrow |2\rangle$ transition was shown to be larger than the $|0\rangle \rightarrow |1\rangle$ transition in experimental studies of acid clusters in acetonitrile mixtures[168, 169]. This finding was utilized to suggest a 1D PES picture for a PTM with

a symmetric double-well structure, which could be a Zundel cation feature. In water-acid solutions,[8] Tokmakoff and co-workers used two-dimensional infrared spectroscopy that excited the O-H stretching vibrations around 3150 cm^{-1} and detected spectral responses within a spectrum ranging from 1500 – 4000 cm^{-1}. They hypothesized that the population of the Zundel cation is larger than previously predicted in theoretical studies by attributing 1750 cm^{-1} to the bending vibration of the flanking waters of the Zundel complex, leading them to infer that it may serve as more than a simple PT intermediate. Their further 2D IR experiments [3] revealed that the effective PES is not a symmetric double-energy well like that of a symmetric Zundel cation, but rather contains a distortion into the underlying PES of the hydrated excess proton (and these authors called it a "distorted Zundel" structure). Additionally, two anisotropy timescales of 2 ps and 200 fs were detected for 2M HCl solutions using data collected from parallel and perpendicular 2D IR spectra at 1750 cm^{-1},[9], which are also beyond the timescale of special pair dance. In these studies, researchers defined the excess proton as being "Zundel-like," emphasizing that the excess proton was shared between two flanking waters, synonymously known as a special pair.

In this Chapter, we more clearly elucidate the dynamics of the hydrated excess proton by utilizing acid solution trajectories of hydrated excess protons with multiple modeling approaches, and at same time provide a more illuminating interpretation of recent spectroscopic data. In particular, based on these trajectories, we are able to capture specific processes and mechanisms that give rise to the anisotropy decay. As detailed herein, we show that the long-lived timescale corresponds to proton transport observed by recent nonlinear spectroscopy, while confirming the the hydrated excess proton is best described as a distorted Eigen cation.

6.2 Methods

6.2.1 Simulation Methods

Two different and complementary simulation methods, namely MS-EVB and EDS-AIMD were used to carry out this study. The MS-EVB approach has been described in section 1.3, and the EDS-AIMD approach has been illustrated in Chapter 3.

In this Chapter, we used the MS-EVB 3.2 and (anharmonic MS-EVB) aMS-EVB 3.2 models.[86] In comparison to earlier MS-EVB models for the excess proton in water,[40, 87, 170] an additional Lennard-Jones potential energy term is incorporated in the 3.2 version to better account for the fourth water pre-solvation around the hydronium core.[171] The anharmonic aMS-EVB 3.2 model also incorporates non-harmonic vibrations in the solvation water molecules. In previous work,.[86, 88, 172] the MS-EVB 3.2 model was used to help interpret infrared spectroscopy data of hydrated excess protons in bulk HCl acid and isotopically substituted water solutions. The model also provided cluster configurations extracted from those systems to calculate instantaneous normal modes using DFT at the B3LYP functional level.

6.2.2 Simulation Details

In total, ten independent MS-EVB 3.2 and aMS-EVB 3.2 simulations of 1 HCl in 256 H_2O with a box side length of 19.73 Å were first equilibrated in the constant NVT ensemble using a Nose-Hoover chain thermostat with chain length of 3 with a temperature set to 298 K and a time-constant of 50 fs using LAMMPS + RAPTOR. Water molecules were modelled using SPC/Fw[90] and aSPC/Fw[173] for MS-EVB 3.2 and aMS-EVB 3.2, respectively. A state searching algorithm that selects up-to three solvation shells of H_3O^+ was employed to construct the MS-EVB Hamiltonian matrix. After a 1 ns non-reactive equilibration period, each simulation was equilibrated using our reactive molecular dynamics simulation code for 500 ps in the constant NVT ensemble, and all

production runs were carried out in the constant NVE ensemble for 1 ns. Each MS-EVB simulations used a timestep of 0.5 fs with a long-range cutoff of 9.0 Å and an Ewald summation with an error of 10^{-5}. In order to include NQEs, additional TRPMD of MS-EVB 3.2 was conducted by the i-PI force engine coupled to LAMMPS. The RPC and MTS approximations were employed. The MS-EVB methodology described earlier with three solvation shell searching was used to compute the forces on the centroid while the same MS-EVB settings but with two solvation shell state searching were used as the reference force. The inner timestep for the MTS integrator was 0.15 fs while the outer timestep was 1.2 fs. Ten independent runs were performed for 504 ps each initiated from independent samples in classical MS-EVB simulations.

In total, three EDS-AIMD(OO) simulations of 1 HCl in 128 water molecules with a box side length of 15.72 Å and two EDS-AIMD(OH) simulation of an H^+ in 128 water molecules with a box side length of 15.64 Å were equilibrated in the constant NVT ensemble using a Nose-Hoover chain thermostat with a time-constant of 11.12 fs, followed by 80 ps and 200 ps in the constant NVE ensemble for EDS-AIMD(OO) and EDS-AIMD(OH), respectively. EDS parameters were identified from our previous work[34, 35] for the BLYP exchange-correlation functional with the D3 Grimme dispersion interaction. All EDS-AIMD simulations were carried out with Quickstep module in CP2K and PLUMED 2 packages using GTH pseudopotentials with a TZV2P basis set and an auxiliary plane-wave basis set with a cutoff of 400 Ry. The three EDS-AIMD(OO) runs had an average temperature of (1) 301 K $\pm$ 10.3 K, (2) 290 K $\pm$ 9.9 K and (3) 292 K $\pm$ 9.7 K, and the three EDS-AIMD(OH) runs had an average temperature of (1) 293 K $\pm$ 10.1 K, (2) 298 K $\pm$ 10.2 K, and (3) 294 K $\pm$ 10.1 K. In order to examine the effect of NQE on proton solvation thermodynamics, an additional PIMD was performed with EDS-AIMD(OH) in i-Pi coupled to CP2K. The PIMD was first equilibrated by a Langevin thermostat with a relaxation time of 15 fs

started from the NVT equilibrated configuration of classical nuclei. The production PIMD was run after that for ~32 ps with the Langevin relaxation time softened to 100 fs. A timestep 0.5 fs was used to integrate the dynamics and the full *ab initio* calculations were conducted on the total 30 beads. For quantum dynamics of EDS(OH), six independent TRPMD simulations were performed using the RPC and MTS approach for 200 ps per run. The reference potential was a machine learned (ML) potential trained with energies and forces sampled in the EDS-BLYP-D3(OH) NVT equilibration using the DeepMD-kit tool. An inner timestep of 0.5 fs and an outer timestep of 2 fs was used to integrate the dynamics.

6.3 Results and discussion

6.3.1 Solvation structure of hydrated excess proton

Figure 6-1a illustrates the solvation structure of hydrated proton in EDS(OH) simulations. The O-O RDFs include all oxygen atoms in the system, those centered on the most probable hydronium-like (O^*, black), the special partner oxygen ($O_{1x,}$ red), and the remaining water molecules in the hydrated proton complex (O_{1yz}, blue). In the analysis, O_{1x}, O_{1y} and O_{1z} are defined as the neighboring water molecules with increasing value of the proton sharing parameter, δ (eq 2-13) around the most probable hydronium.

The O^*-O RDFs show a unimodal peak centered at ~2.5 Å, due to the three water molecules around the hydronium core, as found by integrating the first peak (the coordination number of ~3). However, the O_{1x} RDFs display a prominent peak at a shorter distance, which corresponds to the interaction of the special partner O_{1x} with the hydronium-like oxygen atom, while longer distance peaks relate to the O_{1x} interaction with second solvation shell of the O^*.

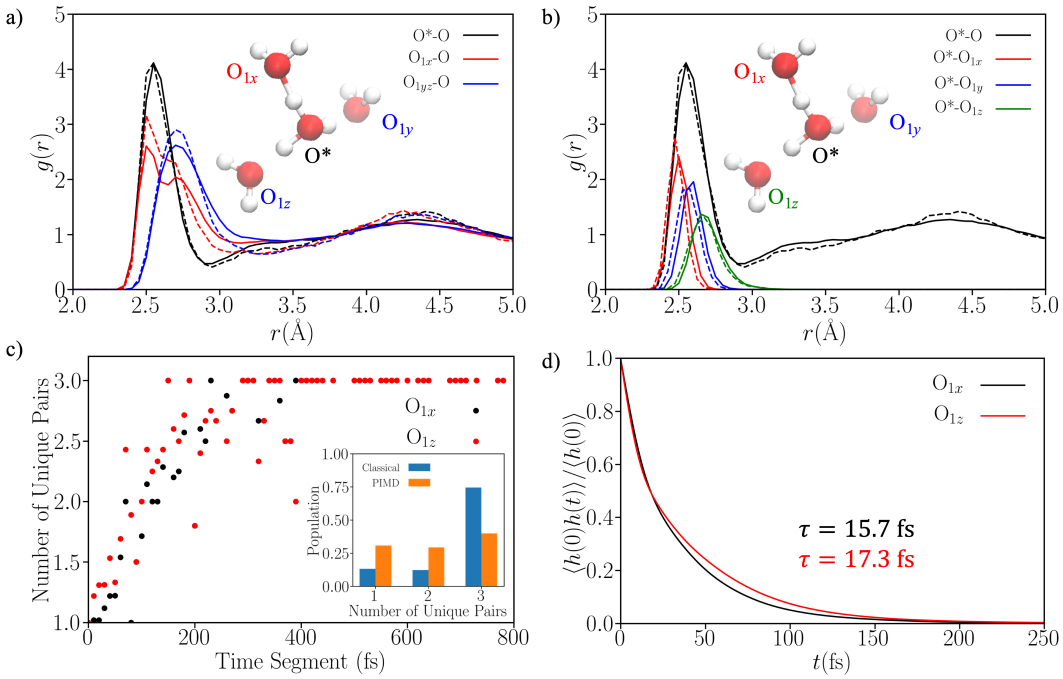

Figure 6-1. Solvation structure of the excess proton and the special pair dance dynamics. (a) O-O radial distribution functions of the O^*-O (black), O_{1x}-O (red), and O_{1yz}-O (blue) in EDS-AIMD(OH) classical simulations (solid) and PIMD simulations (dashed). O^* is the oxygen with the most hydronium like character (most probable), while O_{1x}, O_{1y} and O_{1z} are the 1st solvation shell oxygens of H_3O^+ in the order of strongest to weakest hydrogen bond. (b) Classical (solid) and quantum (dashed) EDS-AIMD(OH) O^*-O RDF (black) decomposed into three water molecules, O_{1x} (red), O_{1y} (blue), and O_{1z}(green). (c) The number of unique O_{1x} (black) and O_{1z} (red) identities in time segments where the proton does not hop, as a function of segment length. The inset shows the population of time segments where the number of unique special partner visited equals to 1, 2, and 3 corresponding to Zundel, $H_7O_3^+$, and Eigen, respectively. (d) Continuous correlation function of the O_{1x} (black) and O_{1z} (red) of eq 4-1. The time constant was obtained from $\int \mathrm{d}t\ C(t)$.

As evidenced by differing O^* and O_{1x} RDFs, the solvation structures of the hydronium-like structure and the special partner oxygen are not identical. The excess proton in a symmetric Zundel cation would remain in the middle of the two flanking water molecules, making the flanking waters almost similar. Even if the excess proton "rattling" is taken into account in this Zundel picture, the ensemble average of this rattling would still lead to identical solvation environments for the two flanking water molecules, and hence almost identical solvation structures for the O^* and O_{1x}. We observed the O_{1x} becomes more similar to O^* in Figure 6-1a after including the NQE, but still O^* and O_{1x} exhibit clearly distinct solvation structures. We also point out that recent studies that have

proposed a symmetric Zundel cation as the dominate species in solution have distinguished Zundel and Eigen cation differently from the present work, with a criterion that is based on the O^*-O_{1x} distance being less than or greater than 2.7 Å, respectively, for Zundel and Eigen. [169] This classification method is questionable, however, since we note that for the simulation methods reported here, all O_{1x} distances were found to be less than 2.7 Å, so the criterion used on this other cited work would never identify a distorted Eigen cation as has been done here. A purely Zundel cation picture also is at odds with the difference in the in O^* and O_{1x} RDFs seen here.

The calculated RDFs in this work indeed indicate a distinction between the hydronium-like structure having the excess proton (O^*) and the special partner (O_{1x}), which is a clear characteristic of a distorted Eigen cation. (We note that the distorted Eigen cation has also been suggested as the most thermodynamically stable structure in recent AIMD studies which additionally used activated rate theory to characterize proton transfers in water.[174]) In a Zundel-like proton configuration picture,[175] we see clear evidence of a single water molecule positioned closer to the hydronium-like cation, which is in agreement with the RDFs found in Figure 6-1a. However, these various static snapshots ignore the important dynamics of the protonated complex. The O^*-O RDF would match the O_{1x} RDF if the Zundel-like picture also represented the real dynamics of the system, corresponding to a dominant first peak for the special partner and integrating to a single water molecule coordination. The O^*-O RDF in the ensemble and time-averaged structure, on the other hand, reveals a single peak and an average coordination of about three water molecules. Given these observations, as well as the excess proton's preference for associating with one water molecule at a time, the simulation strongly suggests that the hydrated excess proton is best described by a distorted Eigen. By virtue of the distorted Eigen picture, we can incorporate the special-pair (i.e., instantaneous two-water shared proton picture), while additionally accounting

for the dynamics of the hydrated excess proton involving other water molecules in the larger Eigen cation complex. In fact, the picture of a distorted Eigen cation as best representing the hydrated excess proton structure first emerged from simulation more than twenty years ago.[87]

6.3.2 Special-pair dance revisited

It has previously been shown that the hydrated excess proton is not a static complex, but is a very dynamic one. In prior work,[5] it was revealed that the identity of the hydrated proton's special partner switches with other waters in the hydrated complex on a short timescale of tens of femtoseconds. Here we revisit this result using the EDS-AIMD(OH) hydrated excess proton model, which is the most recent model and one we consider to be the most accurate (including the use of the highly accurate MB-pol model[36, 37, 133, 176] as the EDS reference).

In Figure 6-1b we show more detail of the O^*-O RDF of the EDS-AIMD(OH) model. We further decompose the first peak into its special partner (O_{1x}) and two, less hydrogen-bonded water molecules (O_{1y} and O_{1z}). We again defined these water molecules with increasing δ value. When examining Figure 6-1b, three distinct peaks are found with varying distances from the most hydronium-like oxygen, which correspond with the closest water being the special partner. The overall statistically combined RDF peak (black line) is clearly unimodal. This result is another indication of the distorted Eigen cation as being the predominant hydrated proton species, where the three water molecules in the 1st shell are statistically unique when decomposed into three component peaks, and one water molecule on average is found to be closer to the hydronium core. However, this is an average over all configurations and the specific water molecules contributing to these three peaks vary with time.

We next investigate the dynamics of the closest and furthest water molecule in Figure 6-1c and Figure 6-1d. We begin by examining the number of unique O_{1x} and O_{1z} identities for periods of

time where the most hydronium-like oxygen does not change, i.e. no proton hopping (Figure 6-1c). For short times, there is only one unique water molecule for the special-partner and the third furthest water molecule. As the length of time in which the proton resides on a single water molecule increases, the number of water molecules that could be identifies as the special-partner (O_{1x}) and furthest water molecule (O_{1z}) increases to 3. These findings indicate that when the proton resides on a single water molecule for short times – as is the case for single configuration snapshots and proton rattling events – only one water molecule is found to be the special-partner and another water molecule can be identified as O_{1z}. On the other hand, during long periods of no proton transfer events, the special-partner and corresponding O_{1z} water molecules are dynamically switching in the first solvation shell of the H_3O^+ motif, and thus point to the significance of accounting for all 3 water molecules in the proton complex instead of using static configurations, such as is often the case for a Zundel-like configurations.

It is then worth checking which of proton rattling or the special pair dance is the dominate dynamics for the solvated excess proton. This is done by splitting the trajectory in segments of time in which hopping events do not occur, and the number of unique special partners O_{1x} is determined for each segment. The no-hopping trajectory segments were then grouped together according to the number of unique special partners visited in the segment, and the total time length of each group was determined by summing the length of each segments in the group. The population of each group is defined as the group total length over the length of the full trajectory. As shown in the inset of Figure 6-1c, the Eigen dynamics (3 unique O_{1x} visited, i.e. the special pair dance) is the most populated process in EDS-AIMD(OH) simulations, while the NQE brings up the population of Zundel dynamics (1 unique O_{1x} visited) and $H_7O_3^+$ dynamics (two O_{1x} visited) but does not change the dominance of Eigen. We note that although EDS-AIMD(OH) improves

BLYP-D3 significantly to a comparable accuracy with CCSD(T) for gas-phase PT energies (Figure 3-1), it slightly underestimates the PT barrier, meaning that the Zundel population is likely overestimated in EDS-BLYP-D3(OH) simulations. The dominance of Eigen, as concluded from dynamics analysis, is consistent with the result of solvation structure analysis shown in Figure 6-1a.

We additionally calculated the continuous correlation function in Figure 6-1d for the special partner and the O_{1z} water molecule according to the following equation

$$C(t) = \frac{\langle h(0)h(t)\rangle}{\langle h(0)\rangle} \qquad 6\text{-}1$$

where $h(t)$ = 1 during segments of the trajectory where the O_{1x} or O_{1z} do not change and 0 for all other times. This is similar to the continuous correlation function used for the excess proton structure.[177] We find that the continuous correlation function for both the O_{1x} and O_{1z} are very similar. By integrating these correlation functions, we can gain an estimate of the lifetime of these species in the distorted Eigen cation, and we find a lifetime of 15.7 fs and 17.3 fs for O_{1x} and O_{1z}, respectively. These again point to the dynamical nature of the surrounding water molecules in the distorted Eigen cation, with the special partner continuously evolving between the hydronium core and several (three) water molecules dynamically solvating it, with excess proton rattling events occurring between all of these three waters but at different times.

6.3.3 Special pair anisotropy decay

Recent nonlinear infrared spectroscopy data from Carpenter et al. indicates that the reorientation of the hydrated excess proton complex is bi-phenomenological featuring two timescales of ~2 ps and ~200 fs,[9] which was confirmed by measuring the intensity via parallel and perpendicular pulses at 1740-1790 cm^{-1}. The vibrational bending of the flanking water molecules in the special pair was ascribed at 1750 cm^{-1} in this important new experimental work. These authors also discounted

certain phenomena as possible structural reorientations that could explain the 2 ps timescale, such as the complete reorientation of the hydrated excess proton complex without proton transfer, rapid structural fluctuations, and energy and thermal transfer from the hydrated complex to the surrounding aqueous environment. Instead, they suggested that the long reorientation timescale corresponds to irreversible proton transfer. Additionally, they claimed that special pair dance could not play a role in their results, as one might have expected a change in the identity of the special partner to reorient the transition dipole more rapidly than either of the ~2 ps or ~200 fs timescales. Since the special pair dance phenomenon is uniquely associated with Eigen species, the long timescales were used as a crucial evidence to derive a distorted Zundel picture.[3]

Here, we seek to understand the structural reorientations of the hydrated excess proton complex that correlate to the observed experimental anisotropy timescales through simulation, with a particular focus on the nature of special pair dynamics and irreversible proton transport. We opted to define the unit vector along the O^*-O axis of the special pair, given recent IR experimental discoveries using a two-water or special pair approach. The second Legendre polynomial ($P_2(x) = (3x^2 - 1)/2$) of the unit vector describing the special pair was then used to calculate anisotropy, such that

$$C_2(t) = \frac{\langle P_2(\hat{\boldsymbol{u}}(0)\hat{\boldsymbol{u}}(t))\rangle}{\langle P_2(\hat{\boldsymbol{u}}(0)\hat{\boldsymbol{u}}(0))\rangle} \qquad 6\text{-}2$$

Here, the $\langle \dots \rangle$ denotes an ensemble average of the unit-vector.

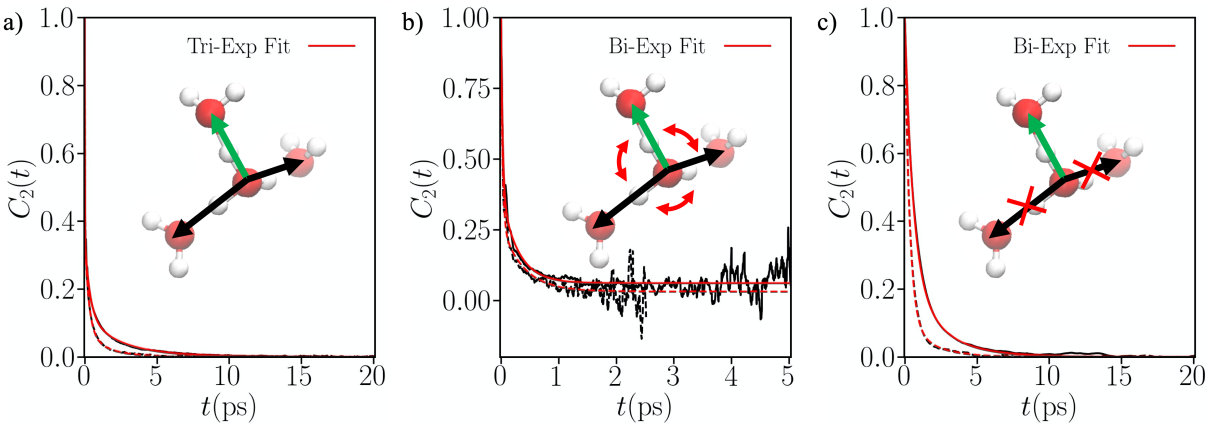

Figure 6-2. Anisotropy plots using eq 6-2 for the O*-O unit vector. The anisotropy plots are broken down based on (a) total anisotropy, (b) special pair dance (no proton transfer), and (c) long-lived special pair (no special pair dance). The solid curves represent the classical MS-EVB results while the dashed represent the quantum MS-EVB results.

Figure 2-1a illustrates the anisotropy decay of the O*-O special pair from classical and quantum MS-EVB 3.2 simulations with the special pair defined as the O*-O that has the lowest δ parameter. We obtained time constants of 12 fs, 0.36 ps, and 2.5 ps, with corresponding exponential term amplitudes of 0.65, 0.23, and 0.12, by fitting a triple exponential function (eq 6-3) to total anisotropy decay from MS-EVB classical simulations.

$$P_2(t) = a_1 \exp\left(-\frac{t}{\tau_1}\right) + a_2 \exp\left(-\frac{t}{\tau_2}\right) + a_3 \exp\left(-\frac{t}{\tau_3}\right) + C \qquad 6\text{-}3$$

where in the fitting procedure, $a_3 = 1 - a_1 - a_2$

Comparable values were obtained from the other simulation methods, as indicated in Table 6-1. Of particular importance is that these time constants and amplitude data can be replicated across various simulation methods, confirming that our findings are not unique to the MS-EVB simulations. It is worth noting that, with NQE, the time constants τ_2 and τ_3 of MS-EVB TRPMD are in good agreement with the time constants determined from 2D-IR experiments[9, 10], while an additional fast timescale τ_1 with a non-negligible amplitude is discovered herein from our simulations. To identify the precise structural phenomena corresponding to these time constants,

we therefore removed specific structural dynamics from the trajectory input for the anisotropy decay calculations and refit that data to a bi-exponential function. Since we observed good agreement between the MS-EVB and EDS-AIMD simulations, we conducted subsequent calculations using only the more computationally efficient MS-EVB approach, as it can be run much longer and thus provide better statistics in comparison to the EDS-AIMD approach.

Table 6-1. Total anisotropy decay timescales and amplitudes.

System	a_1	τ_1 (fs)	a_2	τ_2 (ps)	a_3	τ_3 (ps)	C
MS-EVB[a]	0.65	12	0.23	0.36	0.12	2.5	0.00
aMS-EVB[a]	0.66	10	0.23	0.32	0.11	2.3	0.00
EDS-AIMD(OO) [a]	0.74	17	0.17	0.49	0.09	3.2	-0.01
EDS-AIMD(OH) [a]	0.74	17	0.19	0.38	0.08	2.6	0.00
MS-EVB TRPMD	0.63	15	0.29	0.27	0.08	1.4	0.00
Experiment[b]				~0.2[c]		~1.5[d]	

[a] Nuclei are treated classically.
[b] Experimental data from ref[10].
[c] The timescale (referred as "fast component" in ref[10]) shows weak dependency on concentration. The 1 M experimental data is used here.
[d] The data for 1M, 2M and 4M (referred as "slow component" in ref[10]) are used to extrapolate to the MS-EVB simulation concentration 0.22 M.

First, any contribution in the trajectory associated with proton transfer was removed from the anisotropy calculation, but other phenomena such as the special pair dance were maintained. As previously stated, this is a process in which the identification of the special partner changes in a distorted Eigen cation, while the identity of the central hydronium-like core remains unchanged. The unit vector for the special pair was computed after parsing the trajectory into segments where the proton remained on a single water molecule. After that, the anisotropy was computed for each of these no-proton-transfer segments and averaged across them all. The special pair dance anisotropy calculation is shown in Figure 6-2b for MS-EVB 3.2; as indicated therein, we obtained time constants of 28 fs and 0.29 ps from classical simulations, and 27 fs and 0.38 ps from quantum simulations. Comparing to time constants listed in Table 6-1, we found that the long timescale τ_3

vanishes while the τ_1 and τ_2 are retained. Since the proton transfer is excluded from the anisotropy calculation here, we recognized the fading timescale τ_3 as corresponding to the special pair vector reorientation associated with proton transfers. Previous MS-EVB simulations have revealed that the identity of the special pair on tens of fs timescales[163], which is consistent with the 28 fs found in this anisotropy calculation. This rapid timescale is too fast for current experimental techniques to resolve; in fact, it is faster than the shortest pulse utilized in experiments.

Table 6-2. Bi-exponential fits to the Special pair Dance Anisotropy Calculations

System	a_1	τ_1 (fs)	a_2	τ_2 (ps)	C
MS-EVB	0.68	28	0.32	0.29	0.06
aMS-EVB	0.70	29	0.30	0.31	0.05
MS-EVB TRPMD	0.76	27	0.24	0.38	0.03
MS-EVB TRPMD[a]	0.68	15	0.32	0.27	0.00

[a] Taken from Table 1 of total anisotropy. The amplitudes a_1 and a_2 are renormalized to unitary for better comparison.

In a similar manner, we examined the slower time constants by removing the special pair dance from the total anisotropy calculations. To do so, we define the special pair as the O-O vector between the hydronium oxygen and the water oxygen to which the excess proton hops, rather than the O^*-O pair with the lowest δ value. This mimics a laser pulse with a finite time window longer than the special pair dance timescale resolves the special pair in a time-averaged manner. The anisotropy calculations for MS-EVB 3.2 are found in Figure 6-2c, and amplitude and time constant data are provided in Table 6-3. Note that for MS-EVB 3.2 we identified time constants of 0.56 ps and 2.17 ps from classical simulations, and 0.32 ps and 1.58 ps from quantum simulations, revealing that the τ_1 listed in Table 6-1 vanishes in this analysis. It should also be pointed out that in all three tables there is seen an intermediate timescale τ_2 of sub-picosecond. We did not specifically analyze this motion but since it involves the decay of an angular correlation we presume it reflects the diffusive rotation of the overall Eigen complex.

Table 6-3. Bi-exponential fits to the Long-Lived Anisotropy Calculations

System	a_2	τ_2 (ps)	a_3	τ_3 (ps)	C
MS-EVB	0.74	0.56	0.26	2.2	0.00
aMS-EVB	0.67	0.53	0.33	1.8	0.00
MS-EVB TRPMD	0.88	0.32	0.12	1.6	0.00
MS-EVB TRPMD[a]	0.78	0.27	0.22	1.4	0.00

[a] Taken from Table 1 of total anisotropy. The amplitudes a_2 and a_3 are renormalized to unitary for better comparison.

It should be appreciated that we were able to recover these anisotropy timescales while fully keeping the distorted-Eigen cation picture of the hydrated excess proton complex. We were able to regain the long-lived time constant by eliminating the special pair dance from the distorted-Eigen cation; similarly, we were able to retain the special pair dance by eliminating the proton transport component. In addition to confirming the agreement between the time constants and amplitudes between Table 6-2, Table 6-3 and Table 6-1, these findings indicate that (a) the fast time constant correlates to the special pair dance, and (b) the slow time constant corresponds to irreversible proton transfer. These hypotheses are verified by the strong agreement of timescales and physical processes observed from recent 2D-IR studies. All of this agreement, however, can be attained within the distorted Eigen cation framework.

6.4 Conclusions

We used MS-EVB and EDS-AIMD simulations in this chapter to further investigate the structure of the hydrated excess proton in light of recent spectroscopic experiments. As a result, we were able to obtain anisotropy decay data that are quite comparable to those found using nonlinear spectroscopy. In addition, we were able to discover anisotropy timescales that give rise to the special pair dance and the long-term decay of irreversible proton transfer by decomposing anisotropy based on structural phenomena. While reproducing the long-lived anisotropy decay found experimentally.[9], these processes associated with these timescales coincide with earlier

theoretical investigations of the hydrated excess proton[5, 158]. Most notably, all of these results were achieved using different simulation methods that demonstrate the distorted Eigen cation, continuously undergoing a special pair dance, as the dominant "core" hydrated proton structure in dilute acid solution, which is always conducting a peculiar pair dance.

7 Multiscale Simulation Reveals Passive Proton Transport Through SERCA on the Microsecond Timescale

7.1 Introduction

SERCA pump is a critical component of Ca^{2+} transport in cells and is also an extensively studied member of the large family of P-Type ATPases. SERCA plays a central role in muscle contraction and intracellular Ca^{2+} homeostasis by clearing cytosolic Ca^{2+}. At the cost of one ATP hydrolyzed, SERCA pumps two Ca^{2+} from the cytoplasm into sarcoplasmic reticulum lumen and, at the same time, transports two or three protons in the opposite direction due to the need for charge balance of the binding site in the absence of Ca^{2+} [20, 21, 178-182]. During its functional cycle, SERCA prominently populates two types of states, a cytoplasmic facing E1 state and a luminal facing E2 state [183-185]. The electroneutrality across the ER membrane during the Ca^{2+} intake is compensated by the proton counter-transport of SERCA as well as other fluxes of ions abundant in the cell, such as K^+, Na^+, and Cl^- [186]. Among them, the Cl^- influx was reported to play an essential role in balancing luminal positive charges [187, 188]. Several ClC family proteins, which were identified as Cl^-/H^+ exchangers, were found in the ER membrane colocalized with SERCA. The ClC as well as SERCA mediated proton efflux must be compensated in some way in order to maintain the luminal pH neutralization. Early studies provided evidence for the existing proton influx towards the ER/SR lumen [189, 190] and it was estimated that around 10% of the countercurrent needed for compensating charge imbalance during Ca^{2+} release comes from proton movements [191, 192]. However, this ER proton intake mechanism remains unclear [193]. One possible contributor could be the K^+/H^+ exchangers [186] while Na^+/H^+ exchangers or Ca^{2+}-leak channels are also potential candidates [193].

Recent studies on SERCA's regulation have shed new light on the proton intake into the ER/SR membrane [22, 194]. SERCA is prominently regulated by a transmembrane 52-residue-long protein, phospholamban (PLB), which binds and inhibits SERCA's Ca^{2+} pumping activity [195]. Structural and computational studies [22, 196, 197] have suggested the calcium regulation mechanism originates from the stabilization of a Ca^{2+}-free E1 intermediate state of SERCA. Two protons were predicted to be in the binding-site and bind to E771 and E908 according to empirical pK_a estimations. A transient-water-occupied pore was also identified by classical MD simulations[22, 194]. This observation suggested the presence of a PT pathway for the release of a proton from residue E908 in the binding site to an intermediate luminal residue H944 and down to the luminal environment, However, classical MD simulations treat excess proton, water molecules, and protein in a non-reactive, fixed bonding topology manner with a fixed charge distribution, thus ignoring the well-known Grotthuss shuttling mechanism of PT, the delocalization of the net positive excess protonic charge defect [161, 198, 199], and the altered hydration introduced by an explicit excess proton in the water structures [57, 59, 85, 200]. Alternative approaches, such as the QM/MM method, can provide a reactive description of excess proton solvation and transport in protein channels due to the explicit treatment of the electronic structure. However, the QM/MM method is computationally expensive and allows for sampling on the tens to hundreds of picosecond timescales, thus limiting its ability to carry out the extensive free energy sampling required to fully understand PT processes in proteins [41]. To overcome these sampling limitations, the MS-RMD method have been developed, which are three orders of magnitude more computationally efficient compared to QM/MM. The MS-RMD models are developed from and calibrated against QM/MM data by utilizing a force matching algorithm in a "machine learning" type methodology. The MS-RMD approach can

efficiently and accurately simulate explicit PT in proteins, as mediated by water molecules and amino acids while including Grotthuss proton shuttling, see, e.g., refs [57-60, 107, 200-202].

In this Chapter, MS-RMD simulations are employed along with free energy sampling (umbrella sampling) to test the classical MD-based hypothesis that PT occurs in SERCA from residue E908 into the lumen of the sarcoplasmic reticulum [22, 194]. The free energy profile (potential of mean force, or PMF) was computed for the excess proton migration as well as its coupling to the fluctuations in water hydration of the transport pathway. Recent work in our group has revealed that a 2D PMF is required to fully understand the PT mechanism and pathway in many proteins [57-60, 85, 107, 200]. The two CVs defining the 2D PMF coordinates are the excess proton net positive charge defect location along the PT pathway and the water hydration (number of water molecules) occupying that pathway. It has been universally found so far that the excess proton motion and the water hydration structures are intrinsically coupled, i.e., the water hydration alone in the absence of an excess proton in the "water wire" is not sufficient for understanding the PT pathway and its mechanism, nor the rate of the proton translocation along that pathway. From either the 1D or 2D PMF, the rate of PT can be estimated from transition state theory (TST) [130, 203]. By explicitly calculating the 2D PMF for the excess proton translocation and its hydration, the results presented in this work provide a quantitative description and molecular detail of one PT mechanism through SERCA, and our results support the hypothesis that this passive PT occurs on the microsecond timescale.

7.2 Materials and methods

7.2.1 Classical equilibration

Classical MD was continued from the equilibrated configuration of SERCA from previous studies [22, 194]. The system was embedded in a lipid bilayer of 368 POPC lipids and solvated with 49412

water molecules and 115 K^+ and 93 Cl^- ions in an 11.68 Å × 11.68 Å × 15.4577 Å simulation box. Molecular interactions were described using the CHARMM36 force field and the TIP3P water model. The temperature was controlled by a modified velocity rescale thermostat [204] at 310 K and the pressure was controlled by the Berendsen barostat [205] at 1 atm. The system was integrated with a 2-fs timestep with all bonds involving hydrogen atoms constrained by the LINCS algorithm. The short-range interactions were smoothly switched off between 10 Å and 12 Å using the force-switching scheme and the long-range interactions were computed by the particle mesh Ewald method. The classical simulation was carried out with the GROMACS MD package.

7.2.2 MS-RMD method and model development

The MS-RMD methodology is described in more detail elsewhere [40, 41, 86, 106] as well as in section 1.3. Here we outline the essential aspects of MS-RMD. A reactive molecular system is described by a quantum-like Hamiltonian:

$$\mathbf{H} = \sum_{ij} |i\rangle h_{ij} \langle j| \qquad 7\text{-}1$$

Each basis state $|i\rangle$ corresponds to a given bonding topology with a distinct localized protonated species. The reactive process (change in bonding topology) and excess proton charge defect delocalization is described using a linear combination of basis states so that the ground state of the system $|\psi\rangle$ is expanded in the basis set $\{|i\rangle\}$:

$$|\psi\rangle = \sum_{i} c_i |i\rangle \qquad 7\text{-}2$$

The coefficients c_i are solved from the eigen-value problem at each MD integration timestep:

$$\mathbf{H}\boldsymbol{c} = E_0 \boldsymbol{c} \qquad 7\text{-}3$$

The diagonal terms h_{ii} in **H** can be described by a fixed bonding topology MM potential using, in the present case, a modified version of the CHARMM36 force field. The classical force field is

not capable of reactive simulations, so h_{ii} also includes an additional correction for shifting the diabatic surfaces referred to different zero points [89, 106]. The off-diagonal matrix terms h_{ij}, which provide the proton transfer mechanism between waters and waters with amino acid was chosen as the same functional form as previous studies, see ref [40, 116] for the amino acids and also the Appendix for details. The excess proton in the water molecules was described by the MS-EVB 3.2 model [86]. The parameters in the Glu/His-water off-diagonal and the diabatic corrections for protonated Glu and His states were fit to QM/MM forces by minimizing the force residual $\chi^2 = \left\langle \sum_i \left| \boldsymbol{F}_i^{\mathrm{MS-RMD}} - \boldsymbol{F}_i^{\mathrm{QM/MM}} \right|^2 \right\rangle$ [41]. The fitted parameters were summarized in Table 7-1. The configurations used in the training set were sampled by MS-RMD umbrella sampling biasing the distance between the excess proton CEC (eq 2-1) and the proton-acceptor atoms in Glu and His, namely the carboxylic oxygens of Glu and the imidazole nitrogen atoms of His. For E908, the umbrella windows range from 1.75 Å to 4.00 Å with a 0.25 Å separation. For H944, the umbrella windows range from 1.25 Å to 4.25 Å separated by a 0.25 Å spacing. The configurations were collected every 2 ps in each window, resulting in ~1000 frames in total in both cases. Single-point QM/MM calculations were performed on the sampled configurations as reference forces. The MM part in QM/MM calculation was described by the standard CHARMM36 force field. The QM part was described by DFT at the BLYP/TZV2P level of theory. Besides the two titratable residues E908 and H944, sidechains of T799, V798, W794, V905, M909, S940, T763, V795, S767, S902, and N911 were all included as QM atoms. The electronic structure of three solvation shells of water around E908, H944, T799, T763, S767, and W794 were also treated explicitly. The QM/MM electrostatic coupling was computed by the Gaussian Expansion of the Electrostatic Potential (GEEP) method with periodic boundary conditions [39, 206]. The broken alpha carbon and beta carbon

bonds that crossed the QM/MM boundaries were capped with hydrogen atoms and the QM and MM forces were incorporated by the IMOMM scheme [207] with a scaling factor of 1.50.

7.2.3 Umbrella sampling molecular dynamics

The MS-RMD umbrella sampling was conducted using LAMMPS + RAPTOR with PLUMED 2. The non-bonded interactions were truncated at 10 Å and the long-ranged interactions were computed by the particle-particle particle-mesh (PPPM) algorithm [208]. The system was integrated using a 1-fs timestep by the Nose-Hoover chain thermostat at 310 K under constant volume.

In order to define the proton pathway, a short WT-MTD [209] run (~3 ns) was conducted using the following collective variable:

$$R = \frac{r_{\mathrm{H944}}}{r_{\mathrm{H944}} + r_{\mathrm{E908}}} \qquad 7\text{-}4$$

where r_{H944} indicates the distance between CEC and H944 and r_{E908} is the distance between CEC and E908. The Gaussian height was 0.25 kcal/mol with $\sigma = 0.01$ and placed every 1 ps. During the MTD, 4-5 transitions between E908 and H944 were observed and Hastie's algorithm [210] was used to extract the principal curve through the point cloud formed by the CEC positions sampled. The projection of the CEC onto the curve [211] was then defined as the CV used in the production runs of US for the proton transport between E908 and H944. The proton release pathway from H944 to the lumen is relatively straightforward and thus the CV was defined as the distance between the excess proton CEC and H944, as projected on the averaged vector pointing from H944 nitrogen moiety towards the water in the pore between H944 and the lumen. The hydration CV in the 2D PMFs was taken as the water occupancy number [85] between E908 and H944 and between H944 and the lumen, respectively. Each umbrella sampling window was run for 150 ps to 4 ns, depending on convergence, resulting in a cumulative simulation time of 0.6 μs.

7.2.4 PT rate calculation

The PT rate calculation was based on transition state theory [130, 203]:

$$k_{PT} = \sqrt{\frac{\langle \dot{r}^2_{\mathrm{CEC}} \rangle}{2\pi}} \frac{\int_S e^{-\beta F(R_1,R_2)}\,\mathrm{d}S}{\int_V e^{-\beta F(R_1,R_2)}\mathrm{d}R_1\mathrm{d}R_2} \qquad 7\text{-}5$$

where the velocity of the excess proton CEC was sampled from the PMF minimum of the starting (reactant) state, the denominator is a two-dimensional integral over the reactant basin with respect to the two CVs (CEC position and water hydration occupancy, denoted here as R_1 and R_2), and the integral in the numerator was performed over the one-dimensional dividing surface at the transition state.

7.3 Results and discussion

Previous microsecond-long classical MD simulations have shown that in the presence of bound PLB, SERCA populates a metal-ion-free E1 state (Figure 7-1A) where transport site residues E771 and E908 are protonated [22]. Based on these studies, a transient water pathway connecting residues E908 and H944 was seen in SERCA (Figure 7-1B). This water pathway runs through transmembrane helices TM6, TM8, and TM9, and was stable for about 100 ps [22]. Here, we used the microsecond-equilibrated structure reported in these studies as a starting structure to investigate the hydration environment of the pore from H944 to the lumen formed by transmembrane helices TM8, TM9, and TM10 (Figure 7-1B). During the 200 ns of simulation, the pore between E908 and H944 showed only transient solvation, while the hydration of the pore between H944 and bulk was much more stable yet not always hydrated. Owing to these distinct hydration profiles of the two pores found in this study, we split the whole PT process into two steps (i) PT from E908 to H944, and (ii) from H944 down to the lumen.

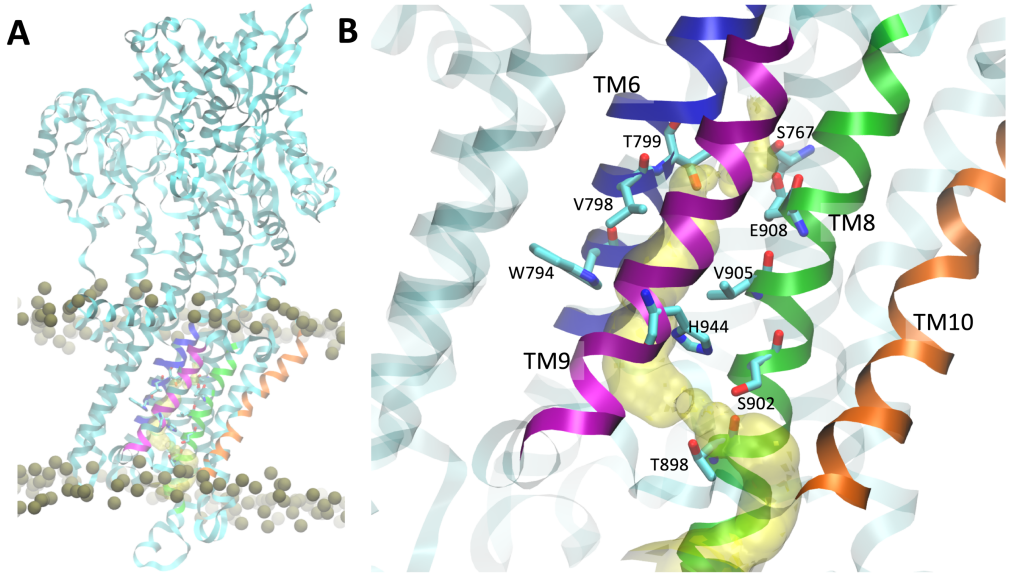

Figure 7-1. Location of the luminal water pore of SERCA. (A) Structure of SERCA bound to PLB and embedded in a lipid bilayer. The proteins are shown as ribbons, and the phosphate groups of the lipids are shown as gold spheres. For clarity, the TM helices are colored as follows: TM6, blue; TM8, green; TM9, purple; TM10, orange. (B) Luminal water pore of SERCA located between transmembrane helices TM6, TM8, TM9, and TM10; the pore is shown as a yellow surface.

7.3.1 Proton transport from E908 to H944

The 2D PMF (Figure 7-2A) for PT from E908 to H944 reveals a very clear coupled mechanism of the excess proton CEC movement with the pathway hydration for the overall PT mechanism. By following the MFEP (black curve) in Figure 7-2A, it is seen that the channel first becomes hydrated (movement in the vertical direction) to provide a hydrated electrostatic environment and adequate hydrogen bonds to help the breakage of the hydrogen bond between residues E908 and S767 (i.e., transition A→B, Figure 7-3). The coordinated waters also allow charge defect delocalization of protonated E908, facilitating the proton dissociation from that residue. At the same time, the hydrated excess proton draws more water into the channel, inducing an increase in water occupancy between step B and the formation of the transition state (T) of the PT process (Figure 7-3). The free energy barrier up to the transition state arises from two contributions: the deprotonation of E908 and unfavorable protonic charge localization induced by poor hydration of

hydrophobic residues V798 and V905. This finding is clearly illustrated in Figure 7-3T, showing that water molecules above or below the two hydrophobic gating residues are either hydrogen-bonded to other water molecules or to residues T799 and H944, contrary to the less solvated waters close to residues V798 and V905. This solvation imbalance results in a tendency for the proton to go back towards E908 or head to H944 and thus makes the transition state overlapping with the positions of V798 and V905. After overcoming the barrier, the system releases 5.9 kcal/mol of free energy from protonating H944 (state C). Then the channel becomes dehydrated (transition C→D), causing the free energy to decrease by 1.9 kcal/mol. The total reaction barrier for PT from E908 to H944 (A→T) is 8.7±0.3 kcal/mol, corresponding to a rate constant of 0.70±0.08 μs^{-1}, indicating that PT occurs through this pathway on the microsecond time scale, in agreement with the suggestion that passive proton transport can occur via this pore [194].

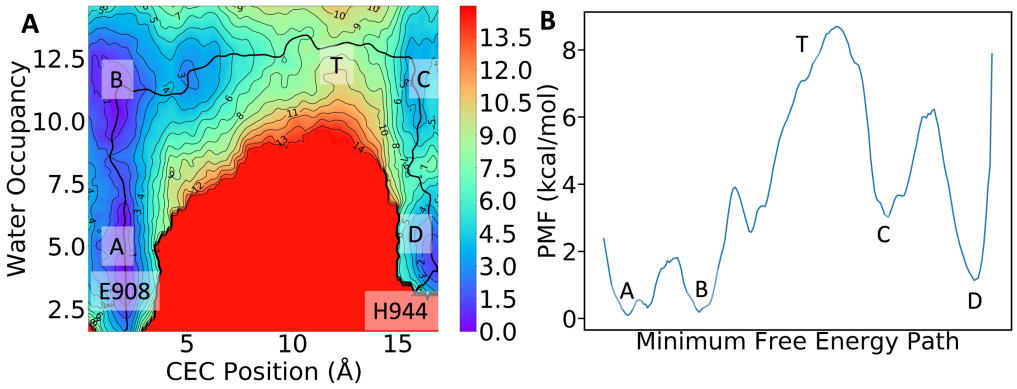

Figure 7-2. (A) Two-dimensional potential of mean force for proton transport from E908 to H944. (B) Free energy along the minimum free energy path (MFEP).

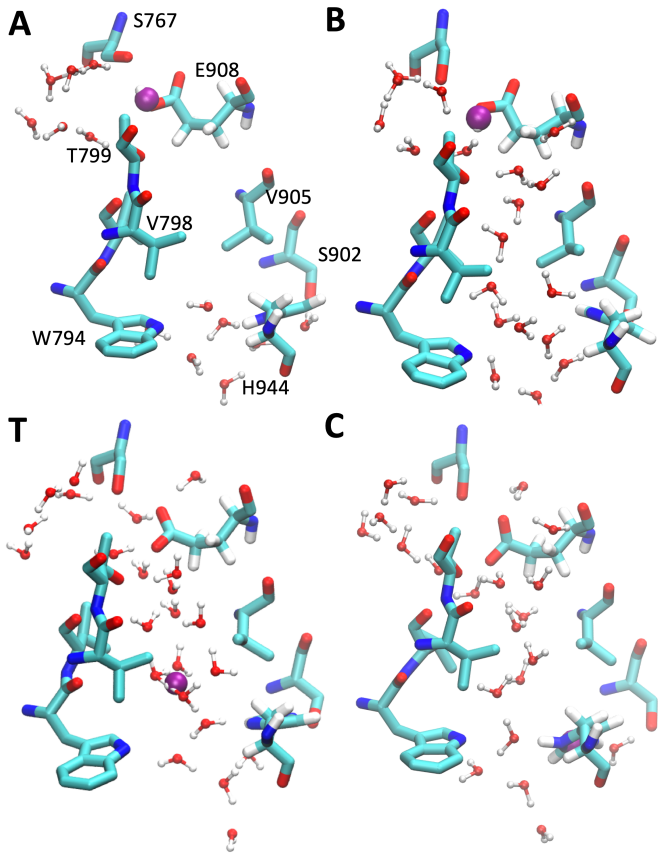

Figure 7-3. Representative configurations along the minimum free energy path, with labels showing the positions on the 2D PMF (Figure 7-2A). The position of the excess proton defect CEC is rendered as a purple sphere. The hydrogen atoms of S767, T799, V798, V905, S902, and W794 are not shown for clarity. (A) The protonated E908 forms a hydrogen bond with S767. (B) The channel becomes hydrated and the E908-S767 hydrogen bond breaks. (T) The transition state of the PT reaction where the excess proton is solvated in the water close to the hydrophobic V798 and V905 residues. (C) The excess proton shuttles to H944 via water wires.

7.3.2 Proton transport from H944 to the lumen

We next focused on the second PT step corresponding to proton permeation from H944 to the luminal bulk via the pore formed by transmembrane helices TM8, TM9, and TM10. We found in our classical MD equilibration that the pore is better hydrated due to the hydrophilic environment created by residues S902 and T898. Due to this hydrophilic pore nature, the 2D PMF of excess proton CEC motion and water occupancy for this PT step (Figure 7-4A) follows a simpler mechanism compared to the PT from E908 to H944 reported in the prior section. Specifically, the

PMF features a free-energy well corresponding to a protonated H944 and solvation water number ranging from 5 to 7. The transition state is located 3.6 Å away from H944, which coincides with the position of the bottleneck of the pore formed by residues S902 and T898. As shown in Figure 7-4C, around the transition state the two gating residues, S902 and T898, replace two waters in the first and second solvation shell of the proton-water motif. As Ser and Thr are less basic than water, they destabilize the excess proton solvation and increase the free energy in this region. Compared to the gating residues V798 and V905, the S902 and T898 gate results in a much lower barrier of 3.9±0.4 kcal/mol. From the transition state theory, we obtained a rate constant of 3.1 ±1.5 ns-1. Compared to the microsecond timescale of the previous PT step, this nanosecond timescale indicates the proton can easily exit the protein after it reaches H944 from E908. Based on these findings, we propose that the E908→H944 PT step in the previous section serves as the rate-limiting step of the entire PT process.

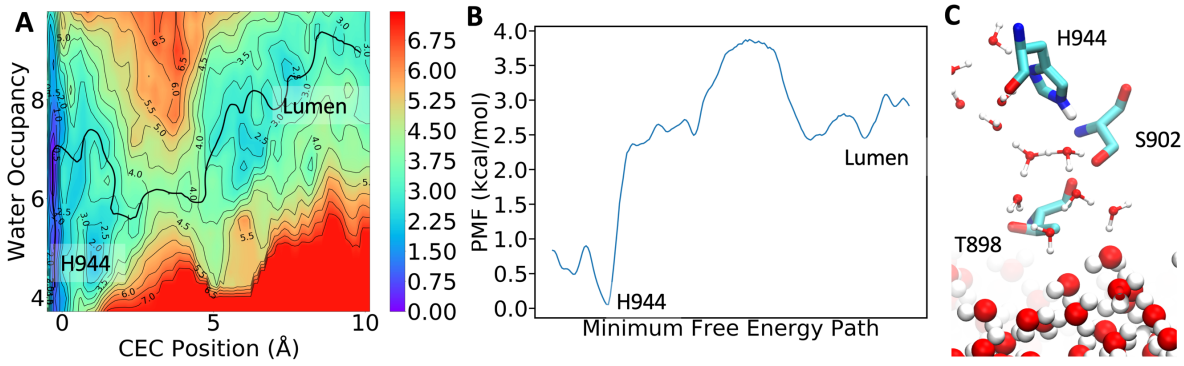

Figure 7-4 (A) 2D PMF for PT from H944 to the lumen. (B) Free energy along the MFEP. (C) Representative configuration at the transition state.

7.4 Conclusions

Extensive multiscale simulations have been reported in this paper to study the PT process through SERCA. To the best of our knowledge, this is the first computational study of this large biomolecular system that has included the explicit physical process of proton transport via

Grotthuss shuttling. The efficiency of the MS-RMD simulation approach has enabled the simulation of the proton translocation along a 25 Å-long pathway at atomistic-level detail for nearly a microsecond of total simulation time, which is well beyond the achievable timescale of QM/MM simulation. The simulations included the calculation of 2D PMFs for the complex PT process, revealing the coupled role of hydration with the excess proton translocation in the pathway. The calculated rate constant along the minimum free energy path of the 2D PMF reveals a microsecond timescale for proton transport from the Ca^{2+}-binding site to the lumen. This result thus highlights the crucial role of a pore that was discovered in a Ca^{2+}-free E1 state of SERCA for deprotonating the binding site and reactivating SERCA into the Ca^{2+}-affinitive E2 state. More importantly, the pore was shown to be a feasible passive proton transport pathway mediated by SERCA that may explain the proton flux towards the SR/ER lumen observed in experiments. The PT process involves the breakage of the hydrogen bond between S767 and E908 and proton permeation through the V798-V905 gate. It is thus proposed that the residues S767, V798, and V905 – which are conserved in the SERCA family – are possible targets for future experimental mutagenesis studies.

7.5 Appendix

7.5.1 Functional form of the MS-RMD models

The off-diagonal interaction used in this work between Glu and water was

$$h_{ij} = V_{ij}^{\text{const}} \cdot \exp(-\gamma \boldsymbol{q}^2) \cdot \exp\left(-\alpha(|\boldsymbol{r}_O - \boldsymbol{r}_{O'}| - R_{OO}^0)\right)$$

where $\boldsymbol{q} = (\boldsymbol{r}_O + \boldsymbol{r}_{O'})/2 - \boldsymbol{r}_{H^*}$ is the asymmetric stretch coordinate ($\boldsymbol{r}_O$ and $\boldsymbol{r}_{O'}$ are the positions of the carboxylic oxygen of Glu and the water oxygen, and $\boldsymbol{r}_{H^*}$ is the shared proton of the two oxygens). The diabatic correction added on the protonated Glu state was a constant, V_{ii}^{const}, while the correction for deprotonated Glu state was eqs 7-9 in ref [40].

The off-diagonal between His and water was eqs. 4-8 in ref [116]. The diabatic correction for protonated His state and deprotonated state are both constants following the ref [116], denoted as V_{ii}^{HSP}, V_{ii}^{HSD} and V_{ii}^{HSE} respectively in.

Table 7-1. The MS-RMD model parameters.

A)

E908 parameters	
V_{ij}^{const}	-27.754
γ	1.3056
α	0.0392
R_{OO}^{0}	3.5999
V_{ii}^{const}	-145.84

B)

H944 off-diagonal parameters		
	N-ϵ	N-δ
V_{ij}^{const}	-31.177	-12.854
r_{sc}^{0}	1.3056	1.2057
λ	0.59978	0.31334
R_{DA}^{0}	2.6544	2.4549
C	0.57498	0.68372
α	1.2677	0.32141
a_{DA}	2.7994	2.3469
β	0.25061	0.42373
b_{DA}	1.9964	2.0944
ϵ	1.2459	12.972
c_{DA}	3.1494	1.0983
γ	1.8713	4.0248

C)

H944 diabatic corrections	
V_{ii}^{HSP}	-102.41
V_{ii}^{HSD}	-12.700
V_{ii}^{HSE}	0

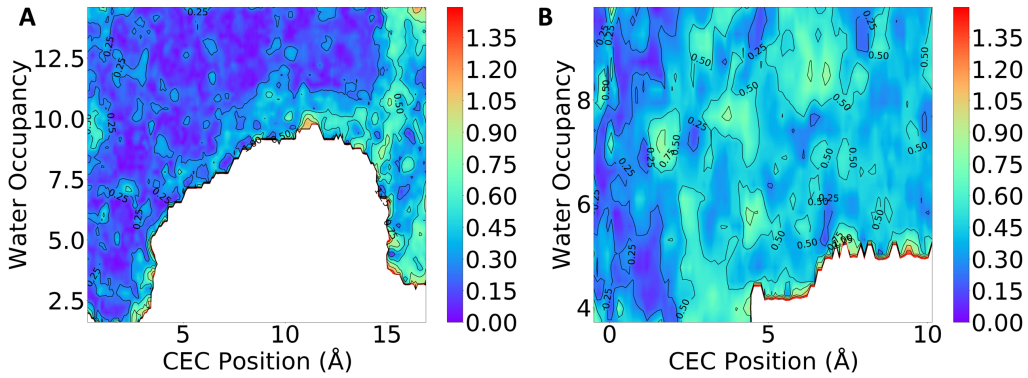

Figure 7-5. (A) The uncertainty of the potential of mean force (PMF) in Figure 7-2A. (B) The uncertainty of PMF in Figure 7-4A

8 Proton Transport and Water Wire Coupling in CNT and ClC-ec1

8.1 Introduction

PT plays a pivotal role in, e.g., the functioning of various biomolecules such as proton exchangers, transporters, and pumps [13, 18, 212], as well as certain nanomaterials [150, 213]. Stable or at least transient "water wires" are believed to be required for proton permeation through confined regions in these systems by exploiting the Grotthuss proton hopping mechanism. In Grotthuss hopping the positive charge defect associated with the excess proton in the water structure is transported by dynamically rearranging the chemical-bonding and hydrogen-bonding topology. This process is thus a "chemical" one, involving the breaking and making of chemical bonds.

It should first be noted that simulating the explicit process of PT through water networks exceeds the capability of traditional MD approaches as widely used in molecular simulations, due to the challenges of modeling the charge delocalization of the excess protonic charge defect and the chemically reactive nature of the excess proton migration [146, 214]. *Ab initio* MD, which treats the electronic structure and nuclear motion both explicitly, benefits from its chemically reactive nature, but it can suffer from the shortcomings of insufficiently accurate underlying electronic density functionals and inadequate statistical sampling due to its high computational cost [31]. The MS-RMD method [4, 41, 87, 106, 125] (an evolution of the earlier MS-EVB method) has proven to be capable and successful in simulating PT in a number of biomolecular systems (see, e.g., Refs. [57-60, 107, 201, 202]). MS-RMD (and MS-EVB before it) has a computational efficiency comparable to regular classical MD so that extensive free energy sampling can be carried out for a PT process in large, realistic biomolecular systems.

Indeed, to date relatively few simulation studies have actually included explicit PT behavior, and even fewer experimental measurements have directly probed the PT phenomenon at a detailed

molecular level. As such, much speculation has occurred on the nature of proton transport in confined systems such as proteins. Moreover, computer simulations that can treat explicit proton transport have more recently shown a non-trivial coupling of the excess proton translocation to the water hydration of the translocation pathway via a transient and dynamic coupling mechanism [57-60, 85, 107, 215-217]. In essence, it has been found that a hydrated excess proton (hydronium-like) structure can in effect "grow" its own water wire for a subsequent PT process via Grotthuss hopping through that water wire. This occurs because the hydrogen bonding emanating from the excess proton hydronium-like structure is stronger and longer range than normal hydrogen bonding, which can more than compensate for the loss of entropy associated with the formation of the ordered hydration structure (including in a hydrophobic pore or channel). In the majority of these simulation studies, it had appeared to be largely sufficient to simply "count" or "bin" the number of hydrating water molecules in the PT pathway to account for the dynamical coupling of those water structures to the explicit excess proton being present in the water wire (i.e., the water hydration structures are distinctly different – even non-existent – when a hydrated excess proton is not actually in the water wire). A two-dimensional free energy sampling of the excess proton charge migration path in space as one of the coordinates and the degree of hydration (number of water molecules) as the second coordinate has proven especially revealing as to the coupled nature of the charge translocation and water structures in the PT process. Moreover, rate calculations of certain PT processes on this 2D free energy surface has proven to be largely in good agreement with experimental measurements of the proton conduction.

Yet, as can often be the case, a "smoking gun" example emerged [107] that strongly suggested the situation is not so straightforward and that the coupling of PT to hydration requires a more precise and quantitatively powerful description. This particular example involved the ClC-ec1 Cl^-/H^+

antiporter protein, but when NO_3^- and SCN^- anions are passed through it rather than Cl^-. The H^+ transporting activity in these alternative anion cases is significantly disrupted versus the wildtype Cl^- limit, and the novel 2.2/1 stoichiometry between chloride and proton transport is also changed. However, the calculation of a 2D free energy surface as described earlier – in this case as a function of excess proton location and the simple hydration of the cavity between the critical E203 and E148 amino acids –provided little explanation for the coupling of the proton antiport to the hydration. Instead, it became clear that the hydration of the PT pathway was simply not enough to account for the overall behavior of the antiporter with an alternative anion flux. And, upon further inspection, it became more obvious that the *connectivity of the water wire hydrogen bonding* – as tied *also* to the location of the excess proton charge defect in the water structure – was a critical feature needed to explain the data. As such, logic suggests that this more complex behavior should be a universal feature of PT processes in confined spaces of proteins and in materials in general, as it certainly includes simple hydration (number of hydrating water molecules) but also goes well beyond just that measure to include structural features (connectivity) of the water wire, and all in the presence of an explicit excess proton.

In this Chapter, we demonstrate the application of the graph CV ϕ derived in Chapter 2 to a CNT) system (Figure 8-1) and a Cl^-/H^+ antiporter, ClC-ec1 (Fig. 3A), from the ClC family [218-222]. We combined the MS-RMD method to simulate the explicit proton transport with the ϕ CV applied to enhance the sampling of the transient water wire connectivity. We show that hydration itself is generally not sufficient for facile PT but, rather, PT primarily occurs when water wires are also fully connected. We further find that excess protons substantially change the water wire conformations and thermodynamics and, consistent with earlier results, that a hydrated excess proton can sometimes create its own water wire where one did not exist before in the absence of

the hydrated proton. In doing so the excess proton can reduce the free energy barrier for forming the water wire by ~10 kcal/mol.

8.2 Results

8.2.1 Proton transport and hydration coupling in CNT

Our investigation involved the use of a more complicated system that was previously utilized to study PT-hydration coupling [85]. Specifically, the system consists of two water slabs separated by two layers of inert graphene-like material with a 28-Å-long armchair-type (6,6) single-walled CNT aligned with its z-axis going through the empty space between the two layers. The ϵ_{LJ} parameter for carbon was reduced to produce a CNT wherein continuous water networks have rarely been sampled. To study the proton permeation mechanism through the hydrophobic tube with quantitative accuracy, we conducted 2D US to compute the PMF of (1) the z coordinate of the excess proton CEC and (2) the water connectivity ϕ. Figure 8-1A shows the resulting 2D-PMF and the MFEP, revealing a 3-step PT mechanism. First, the hydrated excess proton is transiently trapped at the surface near CNT mouth, as shown by Point A in Figure 8-1A and the bottom panel of Figure 8-1B. Then, continuous water wires are spontaneously formed when the proton is at the mouth of the tube, as shown by the almost vertical transition A→B→C on the 2D-PMF along the ϕ direction, and from middle to top panel in Figure 8-1B. Finally, the proton permeates through the fully connected water wire (Figure 8-1B, top panel) following the horizontal valley on the PMF near the top. Interestingly, this free energy trough can only be seen when the water wire is fully connected corresponding to a ϕ of ~ 0.9. This contrasts with any horizontal slice along lower ϕ values, which correspond to a partially connected wire, and suggests that our newly defined water connectivity plays a critical role for PT through even a simple nano-confined water channel. Moreover, it must be emphasized that the parameters of the CNT were chosen to make it

hydrophobic, so water does not occupy it at all in the absence of the excess proton, i.e., the excess proton "grows" its own water wire for facilitating its transport through the CNT

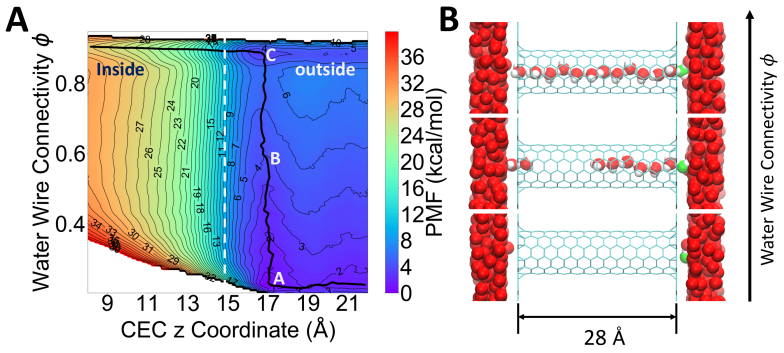

Figure 8-1. Proton transport and water wire connection coupling in CNT. (A) Two-dimensional potential of mean force as a function of ϕ and CEC z coordinate. The zero point of z-axis is set at the middle of the carbon nanotube and the mouth of the tube is around 14.8 Å in the z position. The minimum free energy path is shown as a black curve. (B) Representative configurations at the positions A, B and C denoted on the 2D-PMF. The most probable hydronium oxygen is shown in green. Hydrogens of water outside the tube are not shown for clarity.

Similar proton-induced hydration was first discovered in previous work [85] by biasing a different hydration CV, namely the simple water occupancy number of the CNT (also referred to as water density). Though the water density CV, which simply counts the number of waters in CNT, does not directly reflect connectivity as does the present CV, it is not surprising for it to correlate well with connectivity in this simple case because the CNT is modeled as a rigid body that permit only a single file water wire. However, this is not the case for real protein channels and transporters and more complex materials that have other coupled molecular motions and can accommodate the formation of multiple water wires. This is where the connectivity CVs defined in this work will come into play, as shown in the next section.

8.2.2 The connectivity CV reveals proton transport coupled hydration in ClC-ec1

ClC-ec1 is a proton antiporter (Figure 8-2A) that also transports anions, such as Br^-, I^-, NO_3^-, SCN^- and Cl^-, with distinct H^+ coupling for each anion [223-228]. Electrophysiological experiments have

determined in descending order the proton coupling for these anions to be: Cl^-, NO_3^-, and SCN^-. This behavior is indicated by anion-proton stoichiometry ratios of 2.2:1 for Cl^- to 7-10:1 for NO_3^-, while showing no measurable H^+ transport for SCN^- [228]. Prior simulation studies [107, 229] have shed some light on the molecular-level activity of the anion modulation mechanism of ClC-ec1 with respect to proton transport. It must be noted, however, that Jiang et al. [229] did not treat proton transport explicitly in their work, and hence did not consider the influence of hydrated excess protons on the transient hydration by water wires. By contrast, Wang et al. [107] did include explicit modeling of excess protons via MS-RMD also with umbrella sampling, but as noted earlier their simple water density CV only provided ancillary evidence for the correlation between PT and water wire connectivity. The results also required an additional qualitative hydrogen bond analysis of the umbrella sampling trajectories in order to better understand the underlying physical behavior. This system is thus the "smoking gun" example alluded to earlier that helped to motivate the search for a quantitative, paradigm-defining analysis of proton transporting water wires presented in this work.

When also sampling the water wire connectivity CV, ϕ, along with the progress of the hydrated excess proton CEC, the theoretical developments contained herein reveal a significantly complex hydration mechanism arising from coupling with an explicit excess proton. The PT progress from E203 to E148 is revealed from a 2D PMF calculated in these two CVs, especially as a function of the MFEP on that 2D PMF. Figure 8-2B shows this 2D PMF for ClC-ec1 in its NO_3^--antiporting state. We first note that the MFEP (black curve in Figure 8-2B) features a bell curve-like shape, displaying a simultaneous increase and then decrease in the water wire connectivity ϕ as the proton transport progresses, as described by the CEC position. This observation clearly shows that the water wire connectivity within the channel is strongly coupled with the excess proton

transporting in the water wire (the CEC position in Figure 8-2B), which is different from a "PT-occurs-after-water-wire-formation" mechanism that has often been assumed in a great deal of prior literature. The behavior shown here, which can now be clearly demonstrated and quantified by the combination of a reactive MD simulation method involving the actual (explicit) proton transport plus the new water wire connectivity CV, ϕ, thus calls into question the validity of speculation on PT behavior using only water wire existence alone from crystal structures and/or from standard classical MD trajectories, the latter not including chemical reactivity (i.e., with a fixed bonding topology and hence no possibility of Grotthuss proton shuttling). Likewise, the absence of a pre-existing water wire in a crystal structure or a classical MD simulation *does not necessarily mean that such a wire cannot transiently form*, as a result of a hydrated proton coming into a certain region of a molecular structure and forming its own water wire along the way.

Figure 8-2C depicts the value of $\log(S)$ with respect to the MFEP, showing a clear anti-correlation with the 1D PMF extracted along the MFEP. The other two descriptors, namely the probability of forming continuous water wires and the water wire gap length employed in prior work [107], show similar correlations (Figure 8-5A and Figure 8-5B of Appendix). It must be stressed that water wires are fully connected around (and only around) the transition state (smallest value of the path CV $\log(S)$), as seen in Figure 8-2B, Figure 8-2C, Figure 8-5A and Figure 8-5B, while the simpler water number density already hits high values ($\sim 4-5$) roughly at the halfway point of the up-hill part of the MFEP, as shown in Figure 8-2D, confirming again that PT not only requires enough hydration of the channel but also requires suitably continuous water wires.

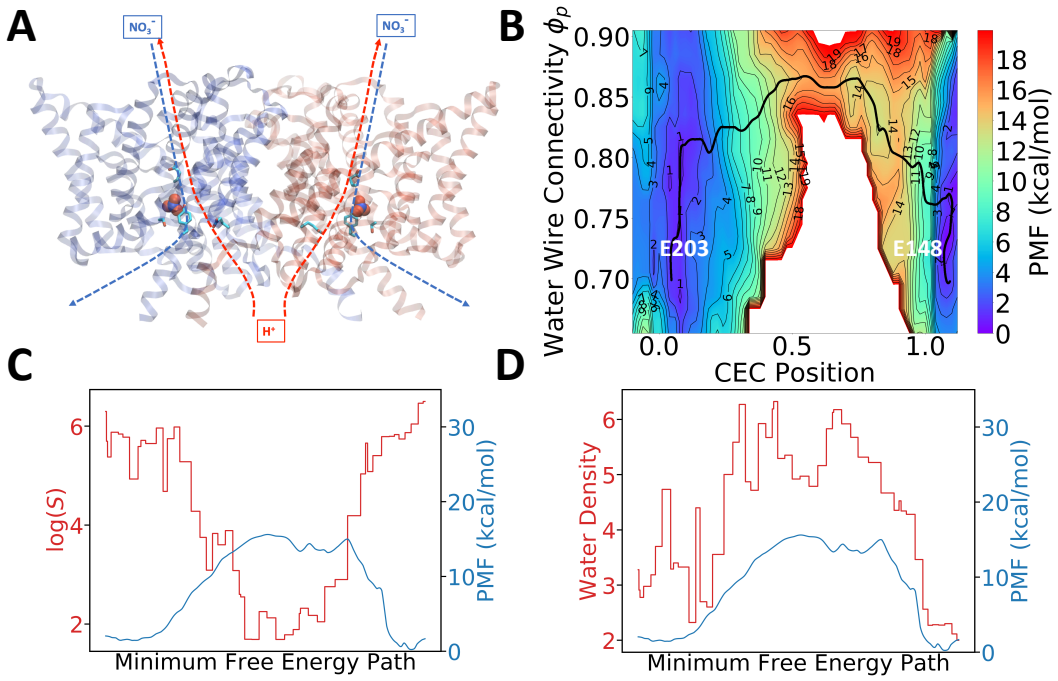

Figure 8-2. Proton-hydration coupling in. ClC-ec1. (A) Image of ClC-ec1 in the NO_3^--antiporting state. Proton and anion pathways are indicated by dotted lines. (B) Two-dimensional PMF as a function of water CV, ϕ, and the hydrated excess proton CEC position. The black curve shows the MFEP on the 2D PMF (C) The shortest path CV log(S) plotted against PMF along the MFEP (same as black curve in panel B). (D) The water wire gap (see ref [107] for a detailed definition) along the MFEP.

Table 8-1. Calculated proton transport rate.

	Simulation with ϕ	Experiment	Simulation with water density
Reaction Rate Constant (s^{-1})	62.6±20.7	~70	14 ± 5

The theoretical value was calculated from the transition state theory (eq 8-9). The experimental value was calculated from the anion transport rate combined with anion/proton stoichiometry [107, 228]. The previous simulation used water density as the second CV [107].

The reaction-rate constant computed from transition state theory applied to the 2D-PMF is summarized in Table 1. Note the good agreement with the experimental value [228] relative to the previous simulation in which only the water density CV was used, thus confirming that ϕ is a better CV for quantifying hydration changes relevant to PT. The prior study in which water density was employed in a 2D-PMF calculation also did not directly reveal a discernible coupling between PT and hydration [107]. This result again highlights that water density is often not an optimal choice

for capturing the essential slower motions of water networks as they pertain to proton translocation, which also explains the underestimated reaction rate in that work (Table 8-1).

8.2.3 Distinct water wires in ClC-ec1 respond to excess protons

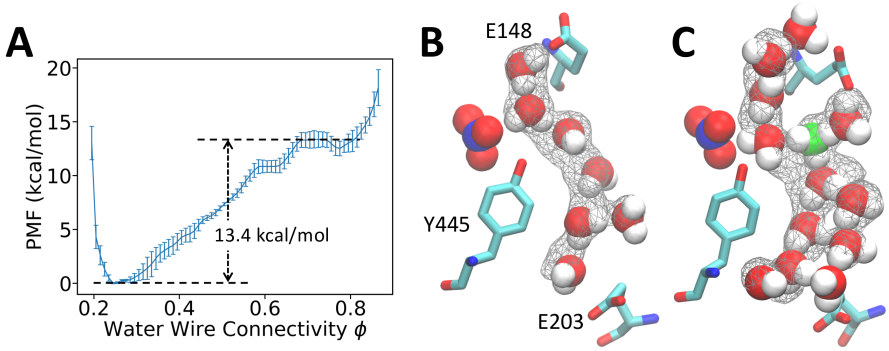

Figure 8-3. Water connecitivity in ClC-ec1. (A) Potential of mean force of ϕ calculated from standard non-reactive classical MD without an excess proton. (B) A representative configuration of connected water wire sampled in the classical MD. The nitrate anion is rendered in VdW representation. The gray wireframe indicates over 40% water occupancy. (C) A representative configuration of connected water wire sampled with explicitly treating an excess proton by reactive MS-RMD. The most probable hydronium-like structure is shown in green.

In order to determine how the hydrated excess proton affects water network connectivity, we performed 1D umbrella sampling of ϕ with non-reactive classical MD in the absence of an explicit excess proton (Figure 8-3A). Figure 8-5B of Appendix also shows a representative configuration of a single connected water wire obtained from umbrella sampling, consistent with the configurations resulting from unbiased MD (Figure 8-3B), thus validating that ϕ does not introduce artifacts in biased simulations. In contrast to purely classical MD studies, water conformations in the presence of an explicit hydrated proton in reactive MS-RMD simulations show much "broader" water wire structures to facilitate the charge defect delocalization of the excess proton (Figure 8-3C). Additionally, reactive simulations with the explicit proton revealed a new water pathway that classical simulations have failed to capture directly, but which was previously suggested by crystal structure analysis [230]. These significantly distinct hydration

network conformations further highlight the correlation between PT and hydration—namely, that the excess proton not only delivers more solvated water into a hydrophobic channel but can also create new water wires. These findings are also consistent with the observations associated with the CNT model of this work and a prior related study [85].

The PMF for water connectivity as defined by ϕ using classical MD is shown in Figure 8-3A, which has a water wire connection barrier at 13.4±0.7 kcal/mol. This large barrier makes the formation of fully connected water wires without explicit protons a rare event that requires enhanced sampling to draw accurate quantitative conclusions within affordable computational cost. In order to compare with the cases when an excess proton is free to move (hop or otherwise translocate), we decomposed the total reaction barrier of the 2D-PMF (Figure 8-2B) into separate contributions from hydration and from the proton transport according to the gradient theorem,

$$\Delta F = \int_{\mathrm{R}\to\mathrm{TS}} \nabla F \cdot \mathbf{d}l = \int_{\xi^{\mathrm{R}}}^{\xi^{\mathrm{TS}}} \frac{\partial F}{\partial \xi}\mathrm{d}\xi + \int_{\phi_p^{\mathrm{R}}}^{\phi_p^{\mathrm{TS}}} \frac{\partial F}{\partial \phi_p}\mathrm{d}\phi_{\mathrm{p}} \qquad 8\text{-}1$$

where the integration was evaluated from the reactant minimum to the transition state. The hydration barrier is defined as the integration of mean force in the direction of ϕ_p:

$$\Delta F_{\phi_p} = \int_{\phi_p^{\mathrm{R}}}^{\phi_p^{\mathrm{TS}}} \frac{\partial F}{\partial \phi_p}\mathrm{d}\phi_{\mathrm{p}} \qquad 8\text{-}2$$

Surprisingly, when an explicit excess proton is present in the water structures, the hydration barrier drops greatly to 1.7 ± 0.8 kcal/mol, which is a few k_BT at ambient conditions. The origin of the overall free energy barrier for the PT therefore comes more from the electrostatics plus chemical bonding rearrangements of the proton translocation process in the confined region and much less from the cost of forming a water wire alone. When comparing observations with and without an excess proton, the water wire between the E203 and E148 residues was actually mostly

disconnected or absent when the excess proton had yet to reach this region; nonetheless, water wires naturally formed as a result of thermal fluctuations during proton transport between the two glutamates. These results again highlight that studies based on standard non-reactive classical MD simulations can be misleading because both the conformations and thermodynamics of the water wires can be substantially impacted by an excess proton being in them.

8.3 Conclusions

In this Chapter applied the new measure ϕ CV defined in section 2.2.3 to a hydrophobic CNT and a Cl^-/H^+ antiporter, and as combined with reactive molecular dynamics and free energy sampling, we are able to identify a novel coupling mechanism between proton transport and water wire connectivity. We find that, even when enough hydration is provided, the hydrated excess proton may often "wait" for fully connected water wires to form before transporting through hydrophobic channels in an activated process fashion. When comparing the reactive MD with classical MD, we also discovered a hydrated excess proton can reduce the free energy barrier of forming continuous water wires by ~10 kcal/mol and can even "create" its own water network. This mathematical description of proton transporting water wire connectivity now provides a powerful and quantitative tool for identifying and characterizing excess proton permeation and hydration coupling behavior in confined spaces such as biomolecular and nanomaterial channels.

8.4 Appendix

8.4.1 Supporting figures

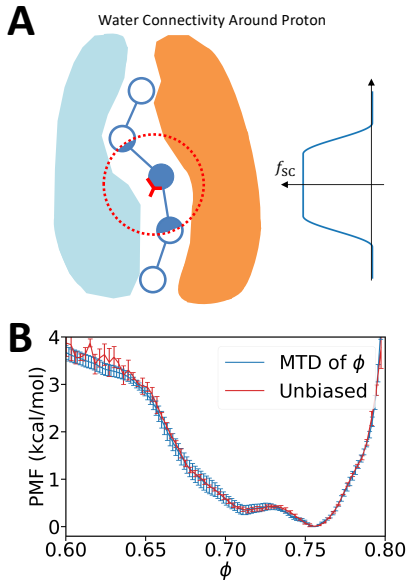

Figure 8-4. Illustration of water connectivity around proton and PMF of ϕ in short CNT. (A) Illustrates how the water connectivity around the excess proton can be calculated. Note that the fictitious beads are spaced more compactly in practice, and the sparse beads here are for clarity of illustration. (B) Potential of mean force of ϕ in the short CNT computed from well-tempered metadynamics and unbiased molecular dynamics. Note that the small $\log(S)$ value indicates connected water wires, while the large ϕ represents connected water wires so that the PMFs of $\log(S)$ and ϕ have contrary positions of wells.

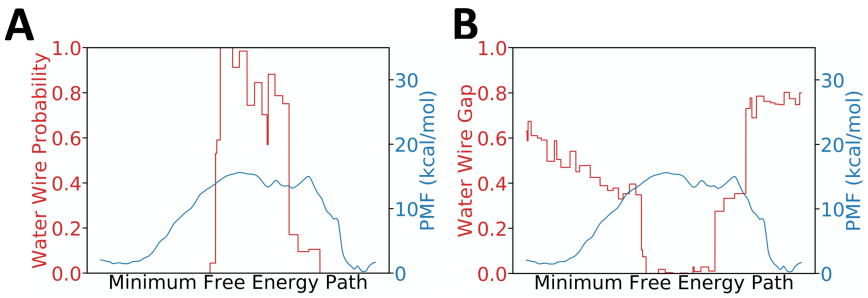

Figure 8-5. Water wire connectivity along MFEP. (A) Probability of forming continuous water wire along the minimum free energy path (MFEP). (B) Water density along the MFEP.

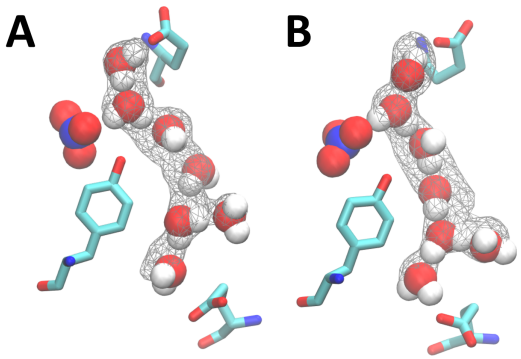

Figure 8-6. Water wire conformations in ClC-ec1 in unbiased and biased simulations. (A) A representative configuration of connected water wire sampled from unbiased MD of ClC-ec1. Note that this figure is identical to Figure 8-3B of the main text. (B) A representative configuration of a connected water wire sampled from umbrella sampling of ϕ. In both figures, the gray wireframe indicates over 40% water occupancy.

8.4.2 Collective variable definitions

Here we provide detailed discussion of the collective variables used in this work, in addition to the more general descriptions in the main text.

Definition of shortest path length $\log(S)$

The adjacency matrix A_{ij} describes the connectivity between any two waters and was given by a switching function in eq 2-18. In addition to the specific choice here, we anticipate that any positive-valued, monotonically non-decreasing switching function $f(r)$ that satisfies $\lim_{r\to+\infty} f(r)/r = +\infty$ would work.

Given two nodes a, b on a graph, S is defined as the shortest path length connecting a and b:

$$S = \min_{P\in\wp} \sum_{i=1}^{|P|} A_{P_i,P_{i+1}}\,, \qquad 8\text{-}3$$

where $\wp$ denotes the collection of all paths satisfying $P_1 = a$ and $P_{|P|} = b$. The shortest path is resolved on-the-fly by an implementation (https://github.com/iszczesniak/yen) of Yen's algorithm [231], in the boost graph library [232], and incorporated in PLUMED 2. The end points a and b are chosen to be the carboxyl oxygens of E203 and E148 in ClC-ec1. Due to the absence of

protonatable residues in the nanotubes, two virtual atoms located at the center of the circular mouths of the tube are used as end points.

Definition of principal curve connectivity ϕ

In general, the principal curve was computed from the point cloud formed by solvation water using Hastie's algorithm. Such was the case in the study of ClC-ec1, but central lines were used for the two CNT systems as they represent the obvious principal curves in rigid straight tubes. The path was discretized at a resolution of 3.2 Å and 2.8 Å in the two CNTs, roughly equivalent to the typical oxygen-oxygen distance of two h-bonded waters. A higher resolution of 1.5 Å was used in ClC-ec1 in order to more accurately capture the curved paths. The water coordination number for each bead $\boldsymbol{x}_i$ was calculated from a summation of all water oxygens $\{\boldsymbol{x}_I\}$:

$$s_i = \sum_{I}^{N_w} f_{\mathrm{CN}}(|\boldsymbol{x}_I - \boldsymbol{x}_i|) \qquad 8\text{-}4$$

where the switching function f_{CN} has the same form of eq 2-22. In this work, $d_0 = 1.5$ Å and $r_0 = 3.0$ Å for ClC-ec1 were chosen in order to allow the water wires to fluctuate around the principal curve but not enough to introduce a change into the CV value. A d_0 of 3.0 Å was used for the CNTs to address its broader effective width due to the modified Lennard-Jones interaction between carbons and water. The coordination number ranging from 0 to infinity was transformed into the interval $(0,1)$ using the Fermi function (eq 2-19). In order to filter out the contribution from less related beads to proton transport, a screening function (eq 2-22) was employed as the exponent of occupancy. The parameters $d_0 = 5$ Å and $r_0 = 5$ Å for eq 2-22 were used to account for multiple solvation shells of the hydrated excess proton. Generally, it is acceptable to use any parameter for switching functions, as long as the resulting ϕ is able to distinguish connected and disconnected water wires. For example, we still obtained the correct $\log(S)$ PMF from reweighting ϕ for the

short CNT when using the exact same settings as in ClC-ec1, even if the choice of parameters for ClC-ec1 may not be optimal for CNTs. Since a bifurcated path was observed in ClC-ec1, the overall connectivity was combined from the independent connectivity of each of the two paths by a softmax function (eq 2-21) and, to be more specific, was the following for ClC-ec1:

$$\phi = \mathrm{softmax}(\phi_1, \phi_2) = \frac{1}{\kappa}\log\left(\frac{1}{2}(\exp(\kappa\phi_1) + \exp(\kappa\phi_2))\right) \quad 8\text{-}5$$

, where $\kappa = 5$ was used in this work to balance smoothness and accuracy of maximum.

Reaction coordinates for proton transport

The CEC is a virtual "atom" or "site" that tracks the effective position of the net positive charge defect arising from the hydrated excess proton. Its definition and detailed discussion can be found elsewhere[52]. The z coordinate of CEC was identified as the obvious coordinate to reflect PT progress along the z-axis-aligned carbon nanotube.

For ClC-ec1, we adopted the same reaction coordinate defined in prior work[107]:

$$\xi = \frac{\min\limits_{i\in\{1,2\}}\left(\boldsymbol{r}_{\mathrm{CEC}} - \boldsymbol{r}_{\mathrm{E203},i}\right)\cdot\hat{\boldsymbol{n}}_{\mathrm{PT}}}{\min\limits_{i,j\in\{1,2\}}\left(\boldsymbol{r}_{\mathrm{E148},j} - \boldsymbol{r}_{\mathrm{E203},i}\right)\cdot\hat{\boldsymbol{n}}_{\mathrm{PT}}} \quad 8\text{-}6$$

, where $\hat{\boldsymbol{n}}_{\mathrm{PT}}$ is a predefined unit vector pointing from E203 to E148; $\boldsymbol{r}_{\mathrm{CEC}}$, $\boldsymbol{r}_{\mathrm{E148},i}$ and $\boldsymbol{r}_{\mathrm{E203},j}$ are the positions of CEC, one of the carboxyl oxygens of E148 and one carboxyl oxygen of E203. Using this definition, $\xi = 0$ when E203 is protonated and $\xi = 1$ when E148 is protonated, thus indicating PT progress between the two glutamates. More detailed discussions about this CV can be found in supplemental ref[107].

8.4.3 Free energy and related calculations

This section provides details regarding enhanced free energy sampling simulations and other related calculations.

Metadynamics

The well-tempered metadynamics of $\log(S)$ and ϕ both used a Gaussian height 0.025 kcal/mol with a pace of 500 fs. The σ of Gaussian was 0.1 for $\log(S)$ and 0.005 for ϕ for counting the different scales of the CVs. A bias factor $\gamma = 8$ was used because of a roughly 4 kcal/mol barrier estimated from our unbiased run. The error bars were calculated from block-average by the last 3 blocks of an even partitioning of the full trajectory into 4 blocks.

The PMF of $\log(S)$ under Boltzmann ensemble was recovered from the metadynamics data of ϕ by a weighted histogram [92]:

$$\mathrm{Prob}(\log S = s^*) = \left\langle \delta(\log S - s^*) \mathrm{e}^{\beta\left(V(\phi,t)-c(t)\right)} \right\rangle \qquad 8\text{-}7$$

, where the δ function was implemented as a Gaussian kernel with a bandwidth of 0.02; the bracket indicates the average under the ensemble generated by metadynamics; $V(\phi, t)$ is the instantaneous bias energy at time t and $c(t)$ was defined as

$$c(t) = \frac{1}{\beta} \log \frac{\int \mathrm{d}\phi \ \mathrm{e}^{-\beta F(\phi)}}{\int \mathrm{d}\phi \ \mathrm{e}^{-\beta\left(F(\phi)+V(\phi,t)\right)}} \qquad 8\text{-}8$$

, where the $F(\phi)$ is the PMF of ϕ estimated on-the-fly as $F(\phi, t) = -\left[\frac{\gamma}{\gamma-1}\right] V(\phi, t)$. Both the metadynamics and reweighting were performed using PLUMED 2.

Umbrella sampling

The US functionality was provided by PLUMED 2 software. The initial configuration of each umbrella window was either the last frame of the umbrella windows in ref[85] and [107], or from equilibration from an adjacent window. In the long CNT, window spacing for CEC z coordinate was 0.5 Å, and as 0.035 for ϕ, resulting in a total of 420 windows. The force constant for z was 8-10 kcal/mol/Å^2 and was 1250-2500 kcal/mol for ϕ depending on the curvature of the free energy surface. In ClC-ec1, window spacing for MS-RMD US was 0.025-0.05 for both ξ and ϕ_p. The

force constant was 1400-4600 kcal/mol and 2500-5000 kcal/mol, respectively. The window spacing of ϕ for classical US was 0.025 and the force constant ranged from 2500 to 5000 kcal/mol. All the PMFs were computed from the WHAM using Grossfield's implementation[233]. The statistical errors were estimated by block analysis of repeating WHAM on 5 blocks.

Rate constant calculation

The rate constant was calculated from the 2D-PMF from TST [203], which has demonstrated adequate accuracy in prior studies of ClC-ec1[57, 58]. The rate constant is given by

$$k_{\mathrm{E203}\rightarrow\mathrm{E148}} = \sqrt{\frac{\langle\dot{\xi}^2\rangle}{2\pi}} \frac{\int_L \mathrm{e}^{-\beta F(\xi,\phi_p)}\ \mathrm{d}l}{\int_A \mathrm{e}^{-\beta F(\xi,\phi_p)}\, d\xi d\phi_p} \qquad 8\text{-}9$$

, where $\dot{\xi}$ is the velocity of PT reaction coordinate; $F(\xi, \phi_p)$ denotes the 2D-PMF; the integral in the numerator was evaluated along the dividing curve *L* at the transition state; and the integral in the denominator was evaluated over the reactant basin *A* corresponded to E203.

8.4.4 Simulation details and system setup

10-Å-long CNT

The system consisted of 4 SPC/Fw water molecules sealed in a 10-Å-long (6,6) armchair CNT by two layers of 16 Å × 16 Å graphene placed in a 25 Å × 25 Å × 25 Å simulation box. The Lennard-Jones interactions between carbons and hydrogens were set to zero. The LJ parameters between the tube carbons and the water oxygens were $\epsilon = 0.1$ kcal/mol and $\sigma = 3$ Å. The LJ parameters between the graphene carbons and water oxygens were $\epsilon = 0.4$ kcal/mol and $\sigma = 2$ Å. All the carbons were fixed at initial positions and the 4 waters were integrated by a Nose-Hoover chain with a chain length of 3, a timestep of 0.5 fs, and a temperature relaxation time of 250 fs at 310 K. The long-range electrostatic was computed by the PPPM method with an accuracy of 10^{-4}. The simulations were carried out with the LAMMPS MD package patched with PLUMED 2.

28-Å-long CNT

The system setup and simulation details were the same as ref[85], except that the MS-EVB 3.2 proton-water model[86] was used instead of the original MS-EVB 3 model[40].

ClC-ec1

The system setup and MS-RMD simulation details were the same as ref[107]. The classical umbrella sampling of ϕ was carried out employing the same simulation settings as the MS-RMD simulations, except with the reactive functionality turned off.

8.4.5 Graphics details

All the molecular images were rendered by VMD [234]. The hydration profile in molecular figures were computed by VolMap plugin of VMD. All the data plots were made by matplotlib [235].

9 Unraveling the Proton Coupling Mechanism of a Peptide Transporter

9.1 Introduction

POTs utilize a membrane pH gradient to drive cellular uptake of di-/tri-peptides and their analogs with homologs found in all bacterial and eukaryotic genomes[23, 24]. Mammalian cells contain two POT family transporters, PepT1 and PepT2, which are responsible for the bulk uptake and retention of dietary peptides in the small intestine and kidneys respectively[23]. PepT1 and PepT2 can also transport prodrug molecules and are increasingly recognized as important targets for rational drug design to improve drug pharmacokinetics[236-238]. The POT proteins use an alternating access mechanism, where peptide transport is realized through conformational switching between two major conformations, termed inward-facing (IF) and outward-facing (OF) states[25]. An occluded (OC) state has also been observed[239, 240], although this is likely to be transitory. In OF and IF conformations the ligand-binding site is accessible either from the extracellular or the intracellular environment, while access is prohibited from both sides in the OC state[241]. The alternating access cycle in POTs has been determined following structural and biochemical studies on bacterial homologs [27] and is rationalized on the basis of alternating formation and breaking of conserved salt bridge interactions, which drive the structural changes following peptide and proton binding[26, 27].

Several conserved side chains have been identified that play key roles in the transport mechanism, in particular it was found that the salt bridge between a highly-conserved glutamate/aspartate on transmembrane helix (TM) 7 and the arginine/lysine on TM1, stabilizes the closed state of the extracellular gate[25, 242, 243]. Protonation of the glutamate/aspartate was predicted to break the salt bridge and trigger the IF-to-OF conformational change, allowing ligand access from the extracellular side[25, 26], but direct evidence for this crucial part of the transport mechanism

is still lacking. MD simulations have been used to investigate the protein and ligand dynamics and coupling in POTs[26, 244-248]. However, many of these MD based studies contain contradictory findings, ascribing different functions for these side chains in different POT family homologs, and highlighting the challenges of sufficient sampling the extended phase space spanned by protein, ligand, protons and their associated hydrations. It was reported in an MD study on GkPOT from *Geobacillus kaustophilus* for example, that the charge state of the TM7 Glu has no effect on the conformational state of the transporter[245]. A later study that employed enhanced conformational sampling found the OF conformation of $PepT_{So}$ from *Shewanella oneidensis* to be more stable than the IF one when the TM7 Glu is deprotonated[246], suggesting the TM1-TM7 salt bridge cannot determine the extracellular gate closure. A more recent computational study systematically evaluated the impact of the TM7 Glu and ligand on the conformational free energy landscape of $PepT_{St}$ from *Streptococcus thermophilus*, and found the IF and OC states were the most stable, independent of the TM7 Glu and ligand[247], inconsistent with the expectation for a stable apo OF state to allow efficient ligand periplasmic access. All of these studies were conducted in a fixed charge and non-reactive manner that limited their ability to fully sample the complete transport cycle and especially the proton coupling mechanism.

Previously, we discovered a homolog of PepT1 from the bacterium *Staphylococcus hominis*, termed as $PepT_{Sh}$, transports a peptide-like thioalchohol precursor, cysteinylglycine-3-methyl-3-sulfanylhexan-1-ol (S-Cys-Gly-3M3SH)[28]. The compound is secreted by the apocrine gland under human axilla skin, uptake by the transporter into bacteria, and bio-transformed into odorous volatiles responsible for human body odor. $PepT_{Sh}$ shares high sequence similarity with PepT1 and PepT2, and can also transport prodrugs such as valacyclovir and 5-aminolevulinic acid. Hence, $PepT_{Sh}$ represents an ideal model for understanding the proton coupling mechanism in its human

counterparts. Our previous work captured $PepT_{Sh}$ IF state bound with S-Cys-Gly-3M3SH as well as several prodrugs[28, 29], but the OF and OC states were not captured, and it remains challenging to complete the functional circle of proton-coupled ligand transport starting from static IF structures.

Here, we performed extensive all-atom MD and MS-RMD[40, 41] combined with enhanced free energy sampling to elucidate how PT through key residues drives conformational changes and ligand translocation. We have demonstrated that it is indeed the titrations on the TM7 and TM10 glutamates that trigger the global conformational change and ligand release in $PepT_{Sh}$. Extending previous studies, we have also evaluated the energetics associated with proton movements and found the PT free energies are subtly modulated by the transporter conformational state and the bound ligand. Our findings, which are consistent with a previous study looking at extracellular gate dynamics in $PepT_{Xc}$[26], reveal a coupled and cooperative mechanism between the transporter conformation, proton, and ligand motions, which represents a major step toward a complete and quantitative description of the full transport cycle.

9.2 Methods

9.2.1 Classical molecular dynamics simulations

The MD computational model was constructed using CHARMM-GUI[249, 250] from the holo IF crystal structure of $PepT_{Sh}$ (PDB: 6EXS) by embedding the protein into a 90-Å × 90-Å bilayer composed of 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) lipids and then solvating the system with a 20-Å layer of TIP3P water[251] on both sides of the membrane with 0.15 M sodium chloride. The protonation states of ionizable residues were assigned according to pK_a predictions from constant-pH MD (cpHMD) simulations (detailed in below) and are listed in Table 9-2. The convergence of cpHMD pK_a calculations for crucial residues are shown in Figure 9-12. The

CHARMM 22/CMAP[252-254] force field was employed to describe the protein and the CHARMM36 was used for lipid interactions. The ligand was modeled as a dipeptide consisting of S-Cys-3M3SH (CSM) and a glycine. CSM was added as a non-standard amino acid to the force field, and was parameterized by CHARMM General Force Field (CGenFF, version 4.0)[255] using ParamChem (version 2.2.0). The CGenFF parameters for the backbone and sidechain atoms up to S^{γ} were replaced by those for cysteine from CHARMM22 force field. The atomic charge of C^{β} was adjusted for charge neutrality. PME [256, 257] with a cutoff of 12.0 Å and a precision of 10^{-5} was used to compute the electrostatic interactions. The LJ non-bonded interactions was force-switched from 10 Å to 12 Å. The system was equilibrated following the CHARMM-GUI protocol[258], followed by a 200-ns additional equilibration with 1000 kJ/nm^2 harmonic restraints on protein heavy atoms. In the additional equilibration and production runs, the system was integrated by the leap-frog algorithm with a 2-fs long time step, and all the bonds involving hydrogen atoms were restrained using the LINCS algorithm. The temperature was controlled at 303.15 K by the Nosé-Hoover thermostat and the pressure was controlled at 1 atm by the Parrinello-Rahman semiisotropic barostat. The simulation time and the initial configurations of production runs were summarized in Table 9-2. All of the classical MD simulations were conducted in the GROMACS package.

9.2.2 MS-RMD simulations

The MS-RMD approach was well documented in refs[40, 41] and in section 1.3. In brief, the MS-RMD approach provides an efficient way to simulate molecular systems with explicit modeling of chemical reactions. This was achieved by considering the system as a linear combination of diabatic states $\{|i\rangle\}$, each of which corresponds to a different bonding topology. The Hamiltonian of the system is then expressed in the following diabatic state representation:

$$\mathbf{H}^{\mathrm{RMD}} = \sum_{ij} |i\rangle\, h_{ij} \langle j| \qquad 9\text{-}1$$

The diagonal term h_{ii} is taken to be the energy function described by the classical force field, namely the CHARMM22/CMAP for proteins and CHARMM36 for lipids. The off-diagonal element, h_{ij}, is modeled by a physically inspired ansatz in a MM form. The detailed definitions of these terms are provided in section 4.2.1. The ground state of the reactive system can be obtained through solving the following secular equation "on the fly" as a function of nuclear configuration, such that

$$\mathbf{H}^{\mathrm{RMD}}\boldsymbol{c} = E\boldsymbol{c} \qquad 9\text{-}2$$

The eigenvector $\boldsymbol{c} = \{c_i\}$ with the lowest eigen energy is the adiabatic wave-function of the ground state. The atomic forces, as the energy gradient, are c_i weighted diabatic forces according to the HF theorem (eq 1-3). The diabatic matching approach (Chapter 4) was used to parametrize the ionizable MS-RMD model for glutamates described by the CHARMM22 force field and the parameters are summarized in Table 9-3.

The MS-RMD simulations were initiated from classical MD equilibrations. The electrostatics was computed by the particle-particle particle-mesh method[208] with a cut-off of 10 Å and an accuracy criterion of 10^{-4}. The non-bonded LJ potential was energy-switched from 8 Å to 10 Å. The system was integrated by the Nosé-Hoover chain thermostat to maintain a 303.15 K temperature in the NVT ensemble using a timestep of 1 fs. The MS-RMD simulations were performed by the LAMMPS MD package coupled to RAPTOR to enable chemical reactions.

9.2.3 Enhanced sampling and rate calculations

The PT between E311 and E418 in the apo form was enhanced by the WT-MTD approach. The CV used in the WT-MTD was defined as a distance ratio:

$$\xi_{\mathrm{CEC}} = \frac{r_{\mathrm{E311}}}{r_{\mathrm{E311}} + r_{\mathrm{E418}}} \tag{9-3}$$

where r_{E311} and r_{E418} are the minimum distance between CEC and carboxyl oxygen atoms of E311 or E418. Due to the charge delocalization nature of a solvated proton, its effective position is tracked by the center of excess charges introduced by the excess proton, or equivalently, the "electron hole" created by the proton nuclear. In MS-RMD, the CEC is defined as the c_i weighted COC of the proton-carrier species in each diabatic state[52],

$$\boldsymbol{r}_{\mathrm{CEC}} = \sum_i c_i^2 \boldsymbol{r}_i^{\mathrm{COC}} \tag{9-4}$$

The minimum distance between CEC and carboxyl oxygens was implemented by a softmin function

$$\mathrm{softmin}(r_1, r_2) = -\frac{1}{\kappa}\log\left[\exp\left(-\kappa(r_1 - \bar{r})\right) + \exp(-\kappa(r_2 - \bar{r}))\right] + \bar{r} \tag{9-5}$$

where $\kappa = 40$ Å^{-1}, r_1 and r_2 denote the CEC separation from the two carboxyl oxygen atoms, and $\bar{r} = (r_1 + r_2)/2$. The initial Gaussian height in WT-MTD was 0.8 kcal/mol and scaled according to a bias factor of 12. The Gaussians were deposited on the ξ_{CEC} dimension every 1 ps with a fixed width of 0.01. In both apo IF and apo OF states, two replicates of metadynamics were run for at least 10 ns.

The PT between E311 and E418 in holo IF state was enhanced sampled by umbrella sampling [259] for its convenience of distributing sampling tasks on multiple computers. The umbrella window centers were placed from $\xi_{\mathrm{CEC}} = 0$ to $\xi_{\mathrm{CEC}} = 1$ every 0.025 with harmonic force constants ranging from 2000 kcal/mol to 3500 kcal/mol. The simulation time for each window ranged from 370 ps to 7 ns depending on its convergence, resulting in a 57-ns simulation time in total. All the enhanced sampling was performed by PLUMED 2 coupled to LAMMPS and RAPTOR.

The free energy surfaces (PMFs) were computed from metadynamics or umbrella sampling data using the dynamic histogram analysis method (DHAM)[260], and so was the Markov transition matrix in the CV space. The CV-position-dependent diffusion constants $D(\xi)$ were computed following the same procedure in ref[261] using the transition matrix, and the reaction rate constant was computed as the inverse of the mean first passage time (MFPT)[262],

$$k_{\mathrm{E311}\to\mathrm{E418}} = \tau_{\mathrm{MFPT}}^{-1} = 1/\int_{\xi_{\mathrm{E311}}}^{\xi_{\mathrm{E418}}} \mathrm{d}\xi' \int_{\xi_l}^{\xi'} \mathrm{d}\xi'' \, D(\xi')^{-1} e^{-\beta\left(F(\xi')-F(\xi'')\right)} \qquad 9\text{-}6$$

assuming a Smoluchowski dynamics of ξ_{CEC}. In eq 9-6, the ξ_{E311} and ξ_{E418} are the CV values that correspond to free energy minima of protonated E311 and E418 respectively, ξ_l is the CV value that corresponds to the lower boundary of E311 free energy well, $\beta = 1/k_B T$ is the inverse temperature, and $F(\xi)$ is the PMF. The reverse PT rate was computed via the detailed balance relation

$$k_{\mathrm{E418}\to\mathrm{E311}} x_{\mathrm{E418}} = k_{\mathrm{E311}\to\mathrm{E418}} x_{\mathrm{E311}} \qquad 9\text{-}7$$

where x is equilibrium concentration computed as the integral of Boltzmann factor $e^{-\beta F(\xi)}$ in the E311 basin or the E418 basin. The errors reported for WT-MTD PMFs and PT rates were computed as the standard deviation between two replicates, while the errors in umbrella sampling counterparts were computed from the standard deviation of the last 5 blocks of equally partitioning the trajectories into 6 blocks.

9.2.4 Characterization of gate sizes and gate hydration

The extracellular and intracellular gate sizes were calculated as the tip distance between TM1,2 and TM7,8 and between TM4,5 and TM10,11. We first defined four virtual atoms as backbone geometric centers of (1) residues 46-52 and 64-70, (2) residues 316-324 and 342-347, (3) residues 140-146 and 154-160, and (4) residues 426-432 and 438-444, to represent the tip positions of

TM1,2, TM7,8, TM4,5 and TM10,11 respectively. The extracellular gate size was then defined as the distance between virtual atoms (1) and (2), and the intracellular gate size was similarly defined as the distance between virtual atoms (3) and (4).

The intracellular gate water density was defined as the number of water molecules in a quadrangular prism divided by its volume. The prism base was defined as the *xy*-plane (membrane plane) projection of the tip residues of TM4,5,10 and 11, represented by four virtual atoms defined as the backbone geometric center of residues 140-146, 154-160, 426-432 and 438-444. The height of the prism was defined as the *z* range of the backbone atoms used to define the virtual centers.

9.2.5 CpHMD simulation

To enable titration, doubly protonated HIS residues were used, dummy hydrogen atoms were added to ASP, GLU, and C-terminus (CT) carboxylates in *syn* positions. The membrane-enabled[263] hybrid-solvent CpHMD[264] simulations were conducted using CHARMM[265] (version c42b2) with the pH-based replica exchange (pH-REX) enhanced-sampling protocol[264]. The simulations represented protein, lipids, and waters using the CHARMM22/CMA force field, the CHARMM36 model, the CHARMM-modified TIP3P model, respectively. The CpHMD parameters for the N-terminus (NT) (Figure 9-13) and CT (Figure 9-14) were derived using the protocols described by Lee *et al*.[266] and Khandogin *et al.*[267], respectively. The LJ interactions were force-switched[268] from 8 to 12 Å. Electrostatic interactions were computed using the PME method with a real-space cutoff of 12 Å and a sixth-order interpolation with 1-Å grid spacing. The non-bonded neighbor list was updated heuristically. To allow for a 2-fs step length, SHAKE[269] was used to restrain the bonds connecting hydrogen atoms. All simulations were conducted with periodic boundary conditions at 303.15 K by the Nosé-Hoover thermostat, and 1 atm by the Langevin piston pressure-coupling algorithm[270].

In CpHMD, a fictitious titration coordination λ that updates simultaneously with the spatial coordinates is coupled to each ionizable site to describe its charge state. The λ particles with a mass of 10 atomic mass unit are propagated using the Langevin algorithm[271] with a collision frequency of 5 ps^{-1} and updated every 10 MD steps to allow for water relaxation. The hybrid-solvent CpHMD uses the leapfrog Verlet integrator[272] to propagate the conformational dynamics in explicit solvent and lipid molecules. The implicit solvent and membrane modelled by the generalized-Born (GB) model GBSW[273, 274] with optimized GB input radii[275] is used to calculate the electrostatic hydration forces on λ particles. An infinite low-dielectric-constant (ε = 2) slab with a high-dielectric-constant (ε = 80) exclusion cylinder aligned with the membrane normal is used to represent protein-embedded membrane in GBSW. The implicit-membrane had a thickness of 35 Å and the radius of the exclusion cylinder was 25 Å. The dielectric constant (ε) was switched from 2 to 80 within 2.5 Å from both membrane surfaces. An ionic strength of 0.150 M was used for the Debye–Hückel term[276] in GB. All Asp, Glu, His, Lys, Arg, Cys, Tyr residues, as well as the NT and CT of the ligand were allowed to ionize. A cylindrical restraint with a force constant of 1 kcal/(mol•$Å^2$) was applied to the center of mass of the protein heavy atoms via the MMFP utility in CHARMM[265] to prevent the protein from lateral drift. Ions were excluded from the hydrophobic membrane region (–16.5 Å < Z < 16.5 Å) by a planar restraint of 5 kcal/(mol•$Å^2$) via MMFP. In pH-REX, exchanges between neighboring pH replicas were attempted every 500 MD steps. Decision of accepting or refusing an exchange attempt was determined using the Metropolis criterion[277]. Two independent runs were conducted for apo-$PepT_{Sh}$: one with and the other without titrations on Lys/Arg/Cys/Tyr residues (Table 9-4). The 1st apo run placed 40 replicas from pH 2.0 to 11.75 with an interval of 0.25 and lasted 20 ns per replica. The 2nd apo run placed 40 replicas from pH 2.0 to 9.0 with an interval of 0.125 or 0.25 and lasted 20 ns per replica. The apo run of

dipeptide CSM-GLY (i.e., ligand) placed 20 replicas from pH 7.0 to 16.5 with an interval of 0.5 and lasted for 5 ns. The holo run (i.e., $PepT_{Sh}$ with CSM-GLY) placed 40 replicas from pH 2.0 to 11.75 with an interval of 0.25 and lasted 10 ns per replica.

9.2.6 Calculation of pK_a

The pK_a was computed by fitting the unprotonated fraction S^{unprot} *vs.* pH to the Hill equation[278],

$$S^{\text{unprot}} = \frac{1}{1 + 10^{n(\text{p}K_a - \text{pH})}} \quad 9\text{-}8$$

where the Hill coefficient n describes the steepness of the transition region in a titration curve. S^{unprot} was determined by counting the population of protonated (defined as those with $\lambda \leq 0.1$) and deprotonated ($\lambda \geq 0.9$) states for every pH.

9.3 Results

9.3.1 TM10 Glu is crucial for ligand binding and its titration controls ligand release

Our earlier $PepT_{Sh}$ crystal structure shows the transporter coupled to S-Cys-Gly-3M3SH in the IF state (Figure 9-1A). The hydrophobic tail of the ligand was discovered to be inserted into a pocket created by TM7 and TM10 (Figure 9-1B). Although the E418 residue is not directly coordinated to the ligand in our crystal structure, mutagenesis of the TM10 Glu in $PepT_{Sh}$, as well as in $PepT_{St}$[25] and human PepT1[279] has suggested the Glu is important for proton coupling and ligand recognition. This prompted us to run two simulations with protonated and deprotonated E418 (TM10 Glu in $PepT_{Sh}$) from the crystal structure. We defined a CV as the minimal distance between TM9 backbone atoms and the ligand heavy atoms in order to assess the binding of the ligand tail in the pocket (see Figure 9-1B). We found that if E418 is protonated, simulations could reproduce the crystal ligand binding pose, as demonstrated by the time evolution and distribution of the distance CV (Figure 9-1C). When E418 is deprotonated, on the other hand, the ligand departs the TM7-

TM10 binding pocket, yielding a lower binding posture that differs somewhat from the crystal (Figure 9-2A).

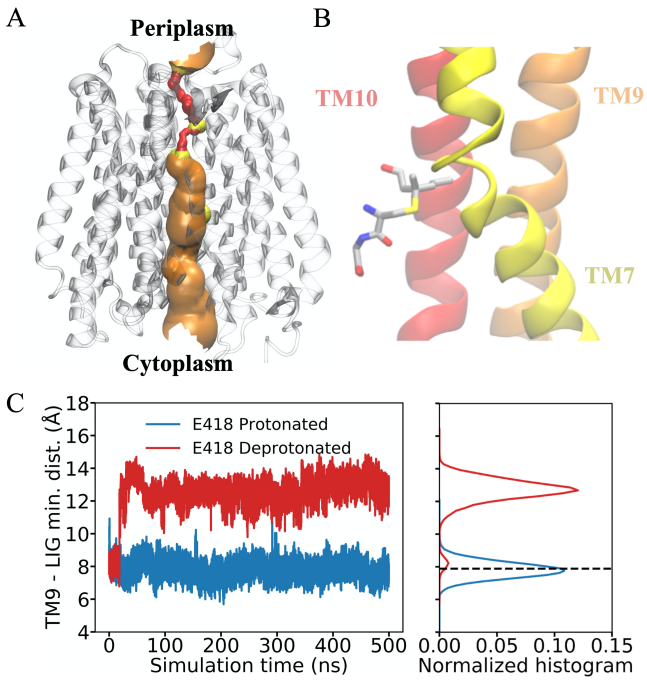

Figure 9-1. Role of E418 in ligand binding. (A) The inward-facing holo crystal structure captured in our previous work [28]. The pore radius profile was computed by the HOLE program [280]. The region that forbids water (pore radius < 1.15 Å) is colored red, the region that allows single water permeation (1.15 Å < pore radius < 2.30 Å) is colored yellow, and orange indicates pore radius > 2.30 Å. This color scheme will be used for the following molecular figures. (B) The position of the ligand and the TM7-TM10 pocket as well as TM9 in the crystal structure. (C) The minimum distance between the ligand and TM9 backbone atoms in simulations with protonated E418 and deprotonated E418.

To further understand the role of E418 in ligand binding, we computed the minimal distance between the carboxyl of E418 and the N-terminus of the ligand. The protonated E418 simulation, which serves as a control, faithfully reproduces the loose interaction pattern between E418 and ligand seen in the crystal structure, as evidenced by the crystal distance value falling within the MD distribution measured. A deprotonated E418, however, grasps the ligand fast (~50 ns) by forming a salt bridge with it (Figure 9-2B), causing its hydrophobic tail to leave the TM7-TM10 pocket as previously stated. Interestingly, despite these discrepancies in its binding pose, the bound

ligand remains stable in the binding site over the course of the 3.3-μs simulation, possibly due to the stability gained from forming the salt bridge with E418 compensating for the loss from leaving the TM7-TM10 pocket. The residue E418 is well-solvated in the IF conformational state, and the water networks fully connect the Glu to the cytosolic bulk, so a pH gradient should be able to drive the proton of E418 down to the cytosol. As such, the crystal structure could be a prelude to the deprotonated E418 state, in which the ligand transfers into the lower binding location after the proton dissociates from E418.

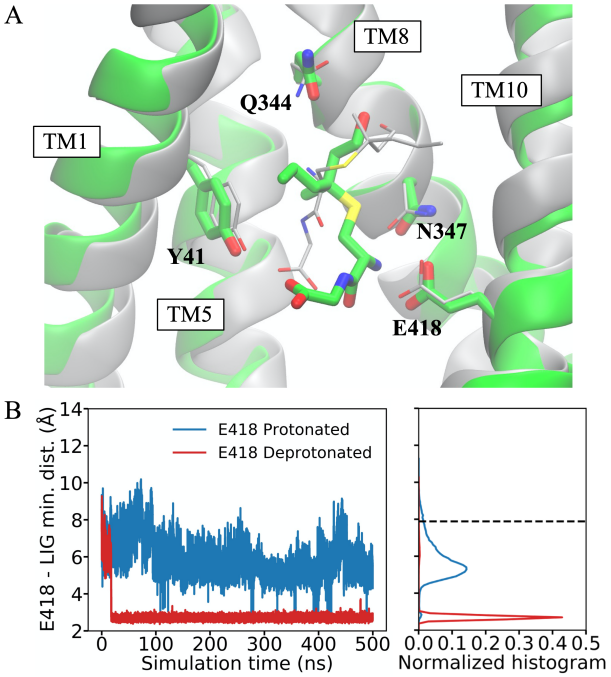

Figure 9-2. The binding mode of the ligand when E418 is deprotonated. (A) A super-position of the crystal structure (grey) and the equilibrated structure with deprotonated E418. (B) The minimum distance between E418 carboxyl and the ligand N-terminus in simulations with protonated and deprotonated E418.

We next protonated the TM10 Glu to model a PT driven by the pH gradient from the periplasm side to the Glu, and initiated MD trajectories from equilibrated configurations sampled when the Glu is deprotonated. We will quantify and illustrate the feasibility of this hypothetical PT in the following sections, but we now report the consequences of this protonation state change. As shown in Figure 9-3A, protonation of E418 makes the ligand unstable and eventually triggers the ligand

release into the cytosol. Notably, the salt bridge between E418 and the ligand breaks almost instantly upon the protonation, but the ligand remains near the binding site for several hundreds of nanoseconds. In this metastable period, we looked at the interacting residues with the ligand (Figure 9-3C) and compared them to the interaction pattern in the stable ligand-bound state when E418 is deprotonated (Figure 9-3B). We found that the protonation indeed weakens the interaction between the ligand and E418, and as a result, the Q344 becomes the dominant residue interacting with the ligand via a hydrogen bond to its sidechain hydroxyl group. The missing hydroxyl in di-/tri-peptides could be one of the reasons why S-Cys-Gly-3M3SH is transported at a slower pace than these other ligands. The lack of contacts with E418 and N347 (Figure 9-3D) resulted in quicker ligand release in a replicate than the other two (Figure 9-3A red vs. green and blue), emphasizing the importance of the residues for ligand binding, which is consistent with the mutagenesis results that E418A and N347A decreased or eliminated transport efficiency [28].

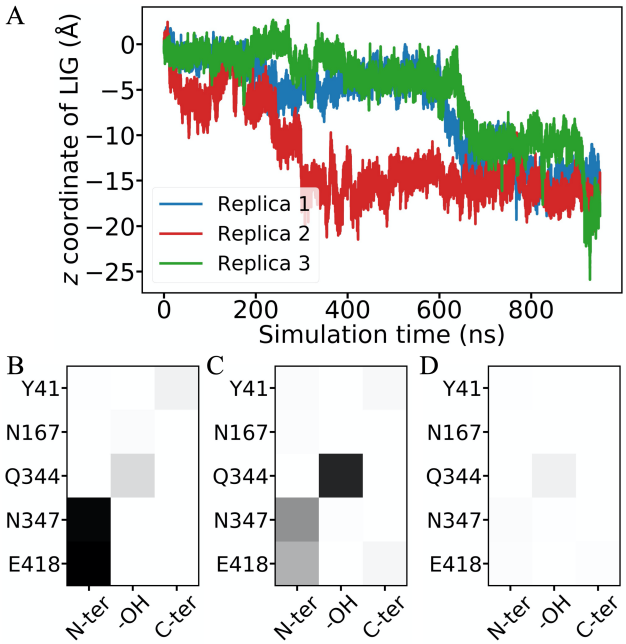

Figure 9-3. Proton-induced ligand release. (A) The *z* coordinate of ligand geometric center with respect to the middle of the membrane. (B) The contact map between ligand functional groups and binding-site residues (B) in the bound state with deprotonated E418, (C) in the first 200-ns metastable state in replicas 1 & 3, and (D) in the first 20-ns of replica 2 with E418 protonated.

9.3.2 Titration of TM7 Glu triggers conformational change to OF state

The highly conserved TM7 glutamate/aspartate–TM1 arginine/lysine pair and the TM7 serine–TM2 histidine pair are two of the critical salt bridge and hydrogen bond interactions that maintain the closure of POT extracellular gate. We have previously shown[26] that the extracellular gate in $PepT_{Xc}$ is regulated by the interaction between histidine and serine, and protonation of histidine causes the gate to open by disrupting the histidine-serine hydrogen bond. The histidine and serine residues are conserved in bacteria $PepT_{Xc}$ and $PepT_{So}$, as well as in mammalian POTs, however, the TM2 histidine is absent from certain bacterial POTs, such as $PepT_{St}$, and $PepT_{Sh}$ studied in this work. Instead, the salt bridge between TM7 Glu/Asp and TM1 Arg/Lys, which can be seen in virtually every IF structure of POTs, might serve as an alternate mechanism for controlling the extracellular gate conformation. We performed MD simulations using deprotonated E311 (TM7 Glu in $PepT_{Sh}$) to stabilize the salt bridge, and as predicted, the salt bridge remained in the simulations and the transporter dominated the IF state (Figure 9-4A blue histogram). Notably, the system also samples an inward-facing occluded (IF-OC) state featuring a partially closed intracellular gate in our 3.3-μs simulation (Figure 9-4B, 4C, and 4D), and we discovered that this state is metastable since the transporter returns to the IF state after ~400 ns in the IF-OC state (Figure 9-4C). The biological relevance of the IF-OC state will be discussed in the following paragraphs.

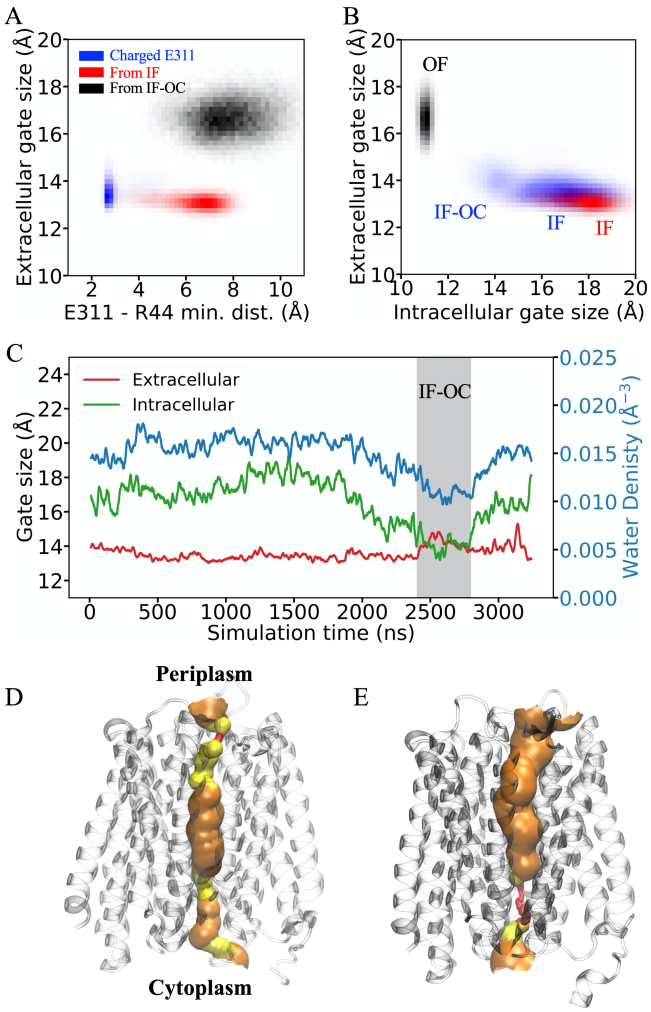

Figure 9-4. Proton-induced conformational change. (A) Two-dimensional histogram of the minimum distance between E311 and R44 heavy atoms, and the extracellular gate size of the simulation with deprotonated E311 (blue), with protonated E311 initiated from an inward-facing conformation (red), and with protonated E311 initiated from an inward-facing occluded conformation (black). (B) Two-dimensional histogram of the intracellular and extracellular gate sizes. (C) The gate sizes and the water density around the intracellular gate in the deprotonated E311 simulation. The region corresponding to the inward-facing occluded state is highlighted by grey. A running average with a 20-ns window was performed on the time series. (D) The pore radius profile of the MD-sampled inward-facing occluded state. (E) The pore radius profile of the MD-sampled outward-facing state.

To model PT from the periplasm to E311, we changed the Glu to be protonated starting from the equilibrated IF structures, and as a result, the salt bridge between E311 and R44 becomes unstable and the two residues separate. Interestingly, in our 500-ns simulations of two independent runs, the transporter remains in the IF state (Figure 9-4A and 4B red histogram; Figure 9-8B), which is

indeed consistent with many prior MD investigations. This contradicts the idea that protonation of the TM7 Glu opens the extracellular gate, since we found only a weak relationship between the TM7 Glu–TM1 Arg distance and the extracellular gate size, shown by a wide range of Glu-Arg distance correlating to a narrow gate size distribution.

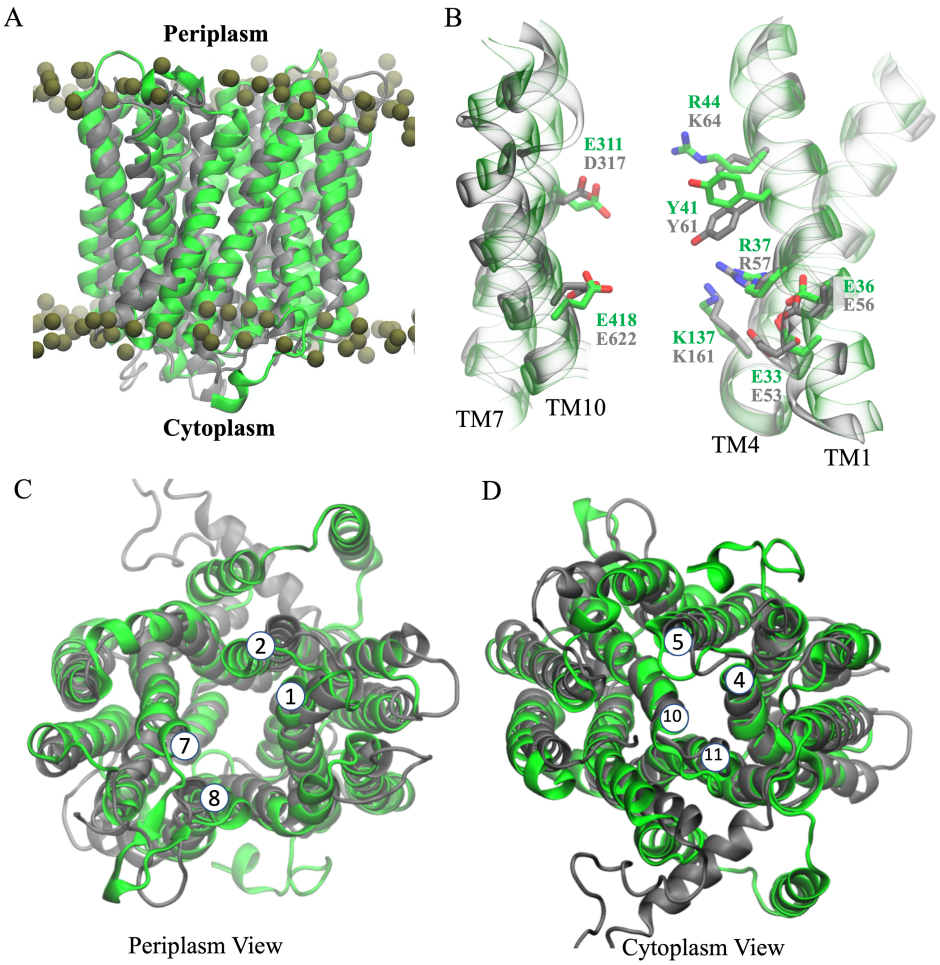

Figure 9-5. Structural comparison between $PepT_{Sh}$ and PepT2. (A) Super-position of MD-sampled outward-facing $PepT_{Sh}$ (green) with the cryo-EM outward-facing PepT2 (grey). The membrane phosphorus atoms are represented by dark yellow spheres. Comparison for (B) key residue, (C) extracellular gate and (D) intracellular gate between $PepT_{Sh}$ and PepT2. The helices that form the gates are labeled with numbers in (C) and (D).

However, in simulations initiated from the IF-OC state, the transporter undergoes a fast conformational change into the OF state (~100 ns) following the protonation of E311, as seen in two separate runs (Figure 9-8A). The slow shutting of the intracellular gate is thought to be the cause for the restriction of a straight transition from IF to OF. Notably, there is a clear correlation

between the intracellular gate size and the water density between helix pair tips TM4,5 and TM10,11, the helices that form the intracellular gate (Figure 9-4C), indicating that hydration fluctuation may be important for intracellular gate closure. The water collective motion can represent a slow degree of freedom, and an enhanced conformational sampling that omits it may suffer from strong hysteresis and produce inaccurate free energetics (see Figure 9-9 as an example of slow convergence of conformational free energies). Interestingly, when starting from the equilibrated OF structures, deprotonation of E311 does not result in the closing of the extracellular gate within our 550 ns × 2 replicas (Figure 9-10), implying that the closing of the gate is consistently slow as the intracellular gate, which may be the rate-limiting step of the entire functional cycle.

The equilibrated OF conformation was superpositioned with a recently resolved OF structure of rat PepT2 (PDB: 7NQK) in order to inspect if this OF structure was an artifact of MD simulations. The two structures aligned extremely well, as illustrated in Figure 9-5, and a deeper look at the critical residues TM 7 and 10 glutamates/aspartates and TM1 arginine/lysine, as well as the well-conserved ExxER/K motif, reveals a significant overlap between the two structures. These observations strongly support that the OF state sampled from MD simulations to be a physical state of the transporter. Merged from the findings presented above, we directly prove that the titration of TM7 E311 initiates the opening of the extracellular gate via the breaking of the E311-R44 salt bridge, and the underlying mechanism is more sophisticated than previously anticipated. Although a protonated E311 loses the salt bridge with R44, potentially allowing the extracellular gate structural flexibility, this is insufficient for the transporter to directly switch from an IF to an OF state. The rate-limiting step for the transition is to find an IF-OC state and potentially the

coupled hydration fluctuation that could take microseconds. Once the IF-OC state has been reached, the transition to the OF state occurs quickly, in roughly 100 nanoseconds.

9.3.3 Facile PT between TM7 and TM10 via water

Given that TM7 E311 requires losing a proton to seal the extracellular gate while TM10 E418 requires a proton to deliver the ligand into the cytosol, it is a plausible assumption that the proton needs to be transported from E311 to E418. According to the results we showed previously, the proton is able to drive the transporter conformational change and ligand transport, and now we want to look at how the conformation and ligand, in turn, effect this PT. We used the MS-RMD methodology to simulate explicit proton transfers on an all-atom reactive potential energy surface in both the IF and OF states, as well as the apo and holo forms, to measure the free energetics of this hypothetical PT step with these coupled motions. Interestingly, as seen in Figure 9-6A, the two potentials of mean force (PMF; free energy profile) of PT in apo OF and apo IF differ only slightly. The difference in the free energy well corresponding to a protonated E311 reveals that the proton on E311 in the IF state is about 1 kcal/mol less stable than that in OF, which is a result of a closer positively charged R44 in the IF state. Another notable variation is the location of the free energy barrier maximum, which is closer to E311 in apo IF than in apo OF. In the IF state, the closure of the extracellular gate restricts hydration around E311, resulting in less solvation for the excess proton (H^+) in that region and as a result, an earlier barrier is seen in the PMF. In spite of these differences, both PMFs reveal that the proton is more thermodynamically stable when bonded to E418 than E311, showing an E311→ E418 proton movement, and this favorability will be even more significant when the inward proton gradient across the membrane is taken into account. In addition to the thermodynamic favorability, the calculated rate constants (Table 9-1) confirm that the PT is facile in both apo conformational states.

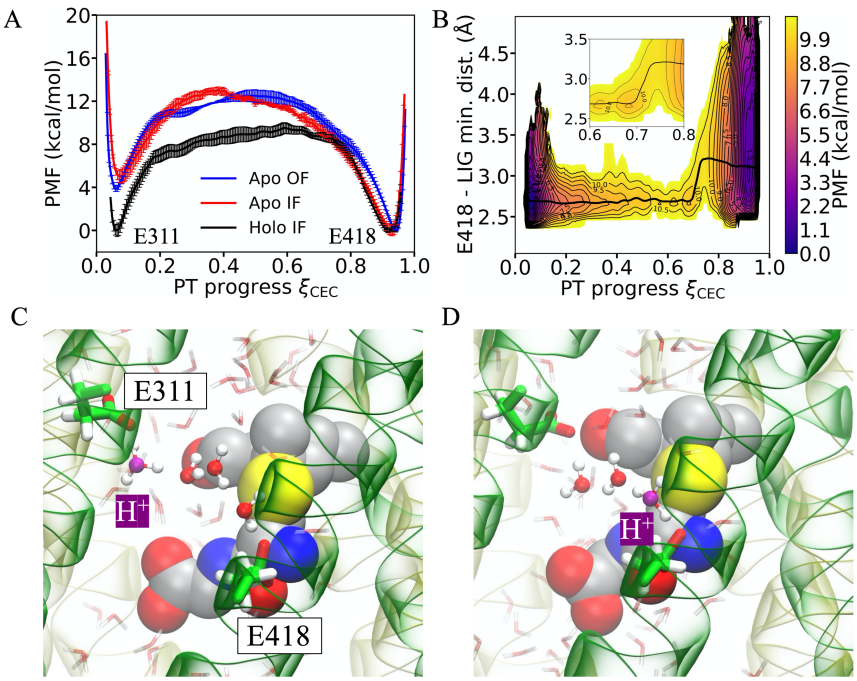

Figure 9-6. Characterization of proton transport between TM7 and TM10. (A) Potential of mean force for proton transport between E311 and E418. (B) Two-dimensional potential of mean force of proton transport between E311 and E418 with the minimum distance between E418 carboxyl and ligand N-terminal nitrogen. The inset is a zoom-in showing the strongly coupled region between proton and ligand. (C) & (D) Molecular figures showing Grotthuss proton shuttling mechanism when the ligand is present. The most probable hydronium oxygen is highlighted in purple. The ligand is shown in the Van-der-Waals representation. The N-terminal bundle of the protein is shown in transparent yellow and the C-terminal bundle is shown in transparent green.

As mentioned in the preceding section, the TM10 E418 is correlated to ligand transport, and its protonation results in cytosolic ligand release. Here, we used the MS-RMD method in conjunction with enhanced sampling to investigate how the proton could be transferred from E311 to E418 in the presence of the ligand. The presence of ligand does not significantly alter the overall shape of the PT free energy, as seen in Figure 9-6A, but it does affect the relative stability between protonated E311 and E418. Because of the energy gained through forming a salt bridge between deprotonated E418 and the ligand N-terminus, the proton is now equally stable on both glutamates, and the PT direction will be fully determined by the direction of the proton gradient. Interestingly, the ligand bulky sidechain does not fully remove water between the glutamates, allowing an excess

proton to travel through connected water wires via Grotthuss shuttling (Figure 9-6C & 6D). As a result, the PT rate is slower than in the apo case (Table 9-1), but it is still a feasible PT and would not constitute the rate-limiting step of the entire cycle. This facile PT mechanism entails a "decoupling" between proton and ligand transport in the sense that the ligand is not needed to participate and has only a little impact on proton movement, adding to our understanding of why the transporter may use the proton gradient to transport a wide range of ligands. The joint free energy profile of PT progress with the E418–ligand distance (Figure 9-6B), on the other hand, depicts the coupling between proton and ligand. When a proton is bonded to E418, the PMF shows more flexibility in the direction of E418–ligand distance, consistent with the observations in classical MD that protonation of E418 eventually triggers ligand release. At equilibrium, the system motion tends to follow the MFEP, and the ramped slope of MFEP, especially in the range of $0.6 < \xi_{\mathrm{CEC}} < 0.8$ (Figure 9-6B inset), reveals that the motion of the proton can drive the motion of the ligand and that the ligand can also be the driving force of the proton.

Table 9-1. Calculated proton transport rates between TM7 E311 and TM10 E418.

System	Apo IF	Apo OF	Holo IF
$k_{\mathrm{E311\rightarrow E418}}$ (μs^{-1})	19.2±0.8	1.6±0.4	0.54±0.26
$k_{\mathrm{E418\rightarrow E311}}$ (μs^{-1})	$1.5 \pm 1.4 \times 10^{-2}$	$3.8 \pm 1.6 \times 10^{-3}$	0.41±0.39

9.4 Discussion

POT family proteins represent secondary active transporters that make use of the cross-membrane proton gradient to transport various peptide analogs into cells. Understanding their transport mechanism and how it is coupled to the proton gradient is critical for improving medication pharmacokinetics and oral bioavailability. However, it remains challenging to experimentally track the molecular motions in real-time at an atomistic resolution, limiting a direct examination of the detailed transport process and discovery of vitally influencing interactions. MD simulations,

on the other hand, are a valuable tool for studying complicated molecular systems and processes to enable direct observation of atom motions as well as the measurement of associated free energetics and kinetics.

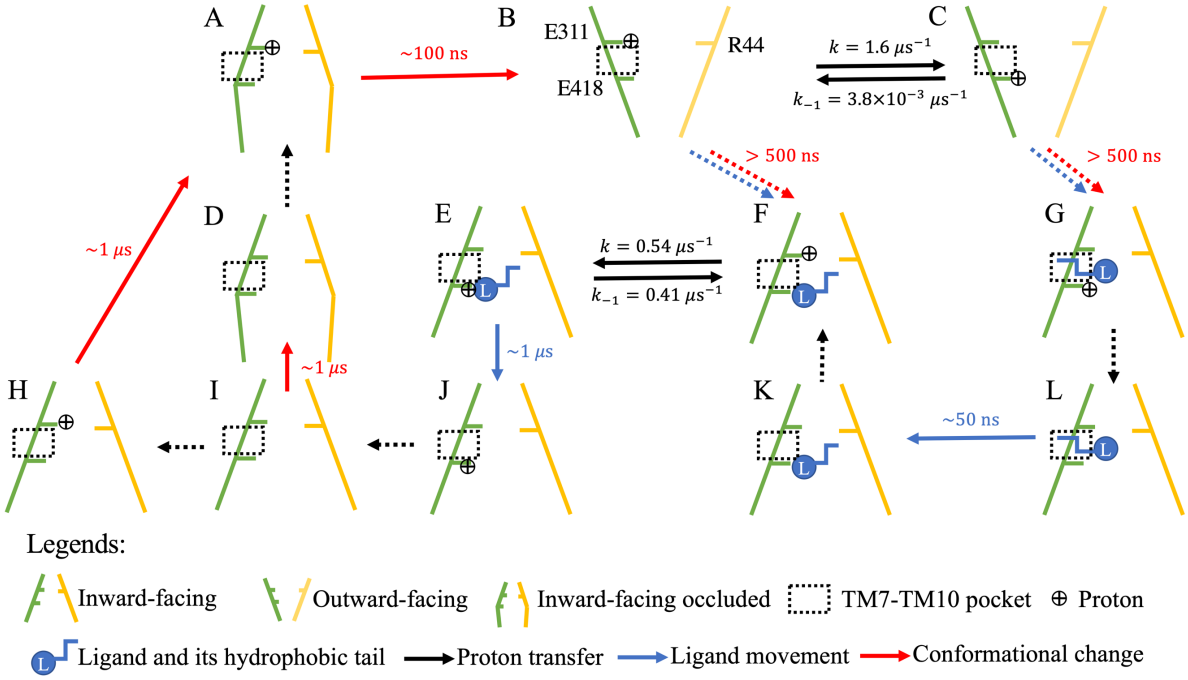

Figure 9-7. Schematic diagram of the transporter functional cycle. The dashed arrows represent the transitions whose reaction rates are not known yet from the performed simulations. The N-terminal bundle is represented by yellow sticks and the C-terminal bundle is colored in green. Note the PT rates were computed without pH gradient.

In this study, building upon our prior structural and biochemical characterizations, we used comprehensive classical and reactive MD along with enhanced sampling approaches to elucidate the proton coupling mechanism of $PepT_{Sh}$, a member of the POT family. As summarized in Figure 9-7, we depicted a schematic functional cycle where the cross-membrane proton flow mediated by the transporter drives the conformational switching and ligand movement via altering the TM7 and TM10 glutamate charge states. However, we need to emphasize that proteins are stochastic molecular machines and the hopping between the microstates shown in Figure 9-7 can be probabilistic. Moreover, there could exist multiple possible transition pathways connecting the states as seen in other transporters, such as ClC-ec1[215, 281] and PiPT[282]. It should also be the case

of $PepT_{Sh}$, and for example, in one pathway (Figure 9-7B→F→E→J), that one single proton reaches TM10 E418 after ligand binding and activates its release. Alternatively, the proton may possibly reach E418 before the ligand binds, and the ligand tail inserts into the TM7-TM10 pocket, resulting in the crystal structure state (Figure 9-7B→C→G). The proton then dissociates from E418 and enters the cytosol (Figure 9-7G→L), forming an E418–ligand salt bridge (Figure 9-7L→K), which is followed by another proton binding to E418 and releasing the ligand (Figure 9-7K→F→E→J). In Figure 9-7, we only outlined the primary functioning pathways that best represents the data we have, but a more complete quantitative description of the whole functional cycle may be required to fully understand the coupling between protein, ligand, and proton motions, as well as their pH dependency and stoichiometry[215, 281]. To achieve that, extensive and converged conformational enhanced sampling is necessary to more accurately estimate the transition energy and rates. Because hydration may play a critical and entangled role in conformational changes, a reaction coordinate describing it should be properly defined and employed in the sampling. The inclusion of additional degrees of freedom in enhanced sampling may result in an excessive computing overhead when sampling the high-dimensional CV space. In this scenario, machine learning and statistical approaches[48, 49, 64, 283] could be extremely effective in reducing the sampling dimensionality and intelligently selecting the most significant phase space region to focus on.

We should point out that the information supplied here provides a limited understanding of the critical importance of the ExxER/K motif. The preliminary data in Figure 9-11 shows that the intracellular gate becomes more flexible with a deprotonated E33 since the IF-OC state was observed earlier than the protonated E33 simulation (Figure 9-4C), but a more rigorous conformational free energy landscape is needed to draw a firm conclusion. In current simulations, the charged E33 serves as an adhesive to stabilize the unusual interaction between R37 and K137

observed in $PepT_{Sh}$ as well as in the cryo-EM PepT2 structure. We recognize that the force field-based description may be inaccurate for a possible proton sharing between R37 and K137, but a quantum mechanical treatment of the motif would be too computationally intensive to achieve adequate conformational sampling. We propose in our future work to train an effective potential from highly accurate electronic structure calculations that corrects the conventional force fields for better modeling this novel Arg-Lys interaction, in light of the recent development of deep learning potential using *ab initio* calculations[102, 284].

9.5 Appendix

Table 9-2. Simulation details of classical MD.

Simulation ID	Protonation States (E33/E311/E418)[a]	Initial Configuration[b]	Total Length (ns)	Data usage (time/Figure)[c]
1	+/-/+	Crystal (6EXS)	500	0-500 ns/Figure 9-1C
				0-500 ns/Figure 9-2B
2	+/-/-	Crystal (6EXS)	3300	0-500 ns/Figure 9-1C
				500 ns/Figure 9-2A
				0-500 ns/Figure 9-2B
				0-500 ns/Figure 9-3B
				0-3300 ns/Figure 9-4A
				0-3300 ns/Figure 9-4B
				0-3300 ns/Figure 9-4C
				2690.2 ns/Figure 9-4D
3-1	+/-/+	550 ns (simulation 1)	1000	0-1000 ns/Figure 9-3A
				0-200 ns/Figure 9-3C
3-2	+/-/+	800 ns (simulation 1)	1000	0-1000 ns/Figure 9-3A
				0-20 ns/Figure 9-3D
3-3	+/-/+	1050 ns (simulation 1)	1000	0-1000 ns/Figure 9-3A
				0-20 ns/Figure 9-3C
4-1	+/+/-	2500 ns (simulation 1)	500	100-500 ns/Figure 9-4A
				100-500 ns/Figure 9-4B
				500 ns/Figure 9-4E
				500 ns/Figure 9-5
				0-500 ns/Figure 9-8A

Table 9-2. Simulation details of classical MD continued.

4-2	+/+/-	2450 ns (simulation 1)	500	0-500 ns/Figure 9-8A
5-1	+/+/-	3250 ns (simulation 1)	500	0-500 ns/Figure 9-4A
				0-500 ns/Figure 9-4B
				0-500 ns/Figure 9-8B
5-2	+/+/-	2100 ns (simulation 1)	500	0-500 ns/Figure 9-8B
6-1	+/-/-	150 ns (simulation 4-1)	550	0-500 ns/Figure 9-10
6-2	+/-/-	250 ns (simulation 4-1)	550	0-500 ns/Figure 9-10
7	-/-/-	Crystal (6EXS)	1000	0-1000 ns/Figure 9-11

[a] Protonated is indicated by "+", and deprotonated is indicated by "-". All the other ionizable residues were assigned their default protonation states except H22, H179, H187, H253, H399, and ligand terminuses were protonated according to CpHMD pK_a calculations. [b] The number indicates the simulation was initiated from that time of the simulation specified in the parenthesis. "Crystal" means the simulation was started from the 200-ns equilibration with protein restraints as described in the Method section. [c] The time segments or time points to be used for making the specified figure.

Table 9-3. MS-RMD parameters for Glu.

B	1.94530	V_{ii}	-151.2996
b	1.40003	$\epsilon^{\mathrm{LJ}}_{\mathrm{OE-HH}}$	0.173646
b'	1.08892	$\sigma^{\mathrm{LJ}}_{\mathrm{OE-HH}}$	1.35219
C	1.90167	$\epsilon^{\mathrm{LJ}}_{\mathrm{Ow-HEP}}$	0.544690
c	1.29037	$\sigma^{\mathrm{LJ}}_{\mathrm{Ow-HEP}}$	1.35577
c_1	-25.0477	$\epsilon^{\mathrm{LJ}}_{\mathrm{OE-OH}}$	0.160421
c_2	2.95380	$\sigma^{\mathrm{LJ}}_{\mathrm{OE-OH}}$	3.09726
c_3	1.36184	$\epsilon^{\mathrm{LJ}}_{\mathrm{OEP-Ow}}$	0.0775357
D	143.003	$\sigma^{\mathrm{LJ}}_{\mathrm{OEP-Ow}}$	3.06560
α	1.8		
r_0	0.975		

The units of the listed parameters use kcal/mol as the energy unit and Å as the length unit.

Table 9-4. pK_a's from pH-REX hybrid-solvent CpHMD.

Residue	pK_a values[a]		
	Apo, 1st[b]	Apo, 2nd[c]	Holo[b]
D63	2.8 (0.9)	2.5 (0.7)	2.8 (0.9)
D90	3.9 (0.8)	4.0 (0.9)	4.0 (0.9)
D152	2.1 (0.9)	2.1 (0.8)	2.3 (1.0)
D156	3.6 (0.9)	3.4 (1.0)	3.5 (0.9)
D182	2.5 (0.8)	2.2 (0.8)	2.5 (0.9)
D258	2.5 (0.7)	2.5 (0.9)	2.4 (0.8)
D284	3.6 (0.9)	3.7 (0.8)	3.8 (0.9)
D287	3.5 (0.8)	3.3 (0.8)	2.7 (0.9)
D326	4.3 (0.8)	4.4 (0.9)	4.3 (0.8)
E33	**6.3 (0.6)**	**6.3 (0.7)**	**6.7 (0.6)**
E36	**4.2 (0.7)**	**4.0 (0.8)**	**3.9 (0.8)**
E150	3.7 (1.0)	3.7 (1.0)	3.8 (0.9)
E226	3.4 (0.9)	3.3 (0.9)	3.1 (0.9)
E227	2.3 (0.8)	2.1 (0.9)	2.4 (1.3)
E289	4.1 (0.8)	4.2 (0.8)	4.7 (0.9)
E311	**4.2 (1.0)**	**4.3 (1.0)**	**4.4 (0.8)**
E323	3.1 (0.8)	3.1 (0.9)	3.2 (0.9)
E340	4.7 (0.7)	4.6 (0.8)	4.7 (0.7)
E418	**5.8 (0.8)**	**5.9 (0.8)**	**3.8 (0.6)**
H22	7.7 (0.8)	7.7 (0.8)	7.4 (0.7)
H56	5.8 (0.8)	5.7 (0.8)	5.4 (0.8)
H110	5.2 (0.7)	5.4 (0.7)	5.3 (0.6)
H179	7.6 (0.9)	7.6 (0.9)	7.6 (0.9)
H187	7.8 (0.8)	7.9 (0.8)	7.8 (0.8)
H253	7.6 (0.8)	7.7 (0.9)	7.5 (0.8)
H399	7.1 (0.9)	7.1 (0.9)	7.1 (0.9)
K17	10.2 (0.9)	-	10.3 (0.9)
K64	Stay protonated	-	Stay protonated
K137	Stay protonated	-	Stay protonated
K210	10.4 (0.8)	-	10.3 (0.8)
K225	10.3 (0.8)	-	10.4 (0.8)
K228	9.9 (0.8)	-	9.8 (0.7)
K230	11.4 (0.6)	-	11.5 (0.5)
K283	10.7 (0.9)	-	10.8 (0.9)
K294	9.4 (0.7)	-	9.5 (0.8)
K364	10.7 (0.6)	-	9.6 (0.5)
K367	10.2 (0.8)	-	10.4 (0.8)
K368	10.3 (0.8)	-	10.4 (0.8)
K375	Stay protonated	-	Stay protonated
K431	10.1 (0.7)	-	10.4 (0.8)
K435	10.6 (0.9)	-	10.7 (0.9)

Table 9-4. pK_a's from pH-REX hybrid-solvent CpHMD continued.

K461	11.3 (0.7)	-	11.4 (0.6)
K464	10.5 (0.8)	-	10.6 (0.8)
K492	10.0 (0.7)	-	9.9 (0.7)
K495	10.4 (0.8)	-	10.3 (0.7)
R24	Stay protonated	-	Stay protonated
R37	Stay protonated	-	Stay protonated
R44	Stay protonated	-	Stay protonated
R91	Stay protonated	-	Stay protonated
R96	Stay protonated	-	Stay protonated
R146	Stay protonated	-	Stay protonated
R154	Stay protonated	-	Stay protonated
R184	Stay protonated	-	Stay protonated
R209	Stay protonated	-	Stay protonated
R229	Stay protonated	-	Stay protonated
R281	Stay protonated	-	Stay protonated
R290	Stay protonated	-	Stay protonated
R292	Stay protonated	-	Stay protonated
R324	Stay protonated	-	Stay protonated
R337	Stay protonated	-	Stay protonated
C112	11.1 (0.6)	-	11.1 (0.6)
C414	10.9 (0.8)	-	11.6 (1.3)
C420	9.5 (0.5)	-	Stay protonated
Y40	Stay protonated	-	Stay protonated
Y41	11.3 (0.6)	-	Stay protonated
Y50	Stay protonated	-	Stay protonated
Y52	11.0 (0.7)	-	11.1 (0.8)
Y74	Stay protonated	-	Stay protonated
Y79	10.5 (0.7)	-	11.3 (0.9)
Y148	Stay protonated	-	Stay protonated
Y163	10.7 (0.8)	-	10.4 (1.0)
Y204	Stay protonated	-	Stay protonated
Y231	11.4 (0.4)	-	11.1 (0.3)
Y250	Stay protonated	-	Stay protonated
Y251	Stay protonated	-	Stay protonated
Y275	11.5 (0.6)	-	11.5 (0.9)
Y320	Stay protonated	-	Stay protonated
Y387	Stay protonated	-	Stay protonated
Y397	11.2 (0.6)	-	10.9 (0.7)
Y411	Stay protonated	-	Stay protonated
Y471	Stay protonated	-	Stay protonated
NT-CSM[d]	16.0 (0.9)		Stay protonated
CT-GLY[d]	8.6 (0.9)		9.8 (0.6)

Table 9-4. pK_a's from pH-REX hybrid-solvent CpHMD continued.

[a] Parenthesized are Hill coefficients. "Stay protonated" indicates a residue, though ionizable in CpHMD, did not titrate within the pH range studied. [b] pH range 2.0–11.75. [c] pH range 2.0–9.0. Lys/Arg/Cys/Tyr permanently protonated. [d] pH range 7.0–16.5.

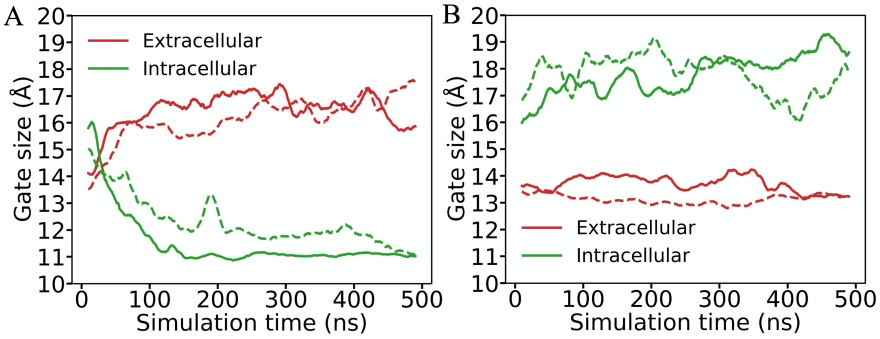

Figure 9-8. Time evolution of gate sizes in protonated E311 simulations. (A) Simulations initiated from the inward-facing occluded state. The solid and dashed curves represent two independent replicas respectively. (B) Simulations initiated from the inward-facing state with two independent replicas (solid and dashed). A running average with a 20-ns window was performed on the time series.

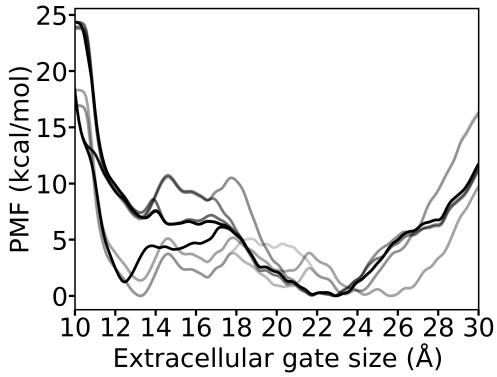

Figure 9-9. Time evolution of potential of mean force of extracellular gate size in a well-tempered metadynamics simulation. The transparency of the curve color indicates the accumulative simulation time from 1 μs to 3 μs with a 200-ns spacing. The metadynamics biased both the extracellular and intracellular gate sizes, and used an initial Gaussian height = 0.6 kcal/mol. The bias factor was 25, Gaussian widths = (0.25 Å, 0.25 Å), and Gaussians were deposited every 100 ps.

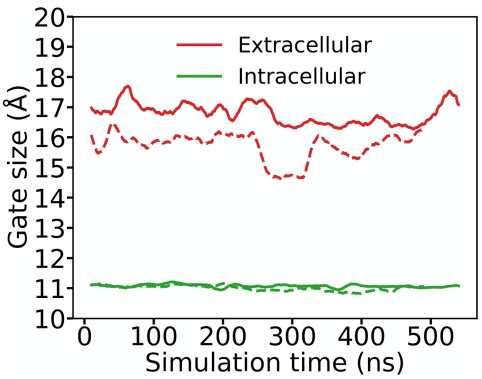

Figure 9-10. Time evolution of gate sizes in deprotonated E311 simulations initiated from the outward-facing state. The solid and dashed curves represent two independent runs. A running average with a 20-ns window was performed on the time series.

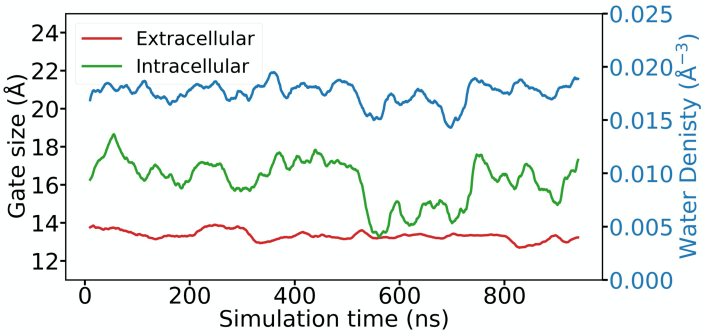

Figure 9-11. Time evolution of gate sizes and the water density around the intracellular gate in a simulation with deprotonated E33. A running average using a 20-ns window was performed on the time series.

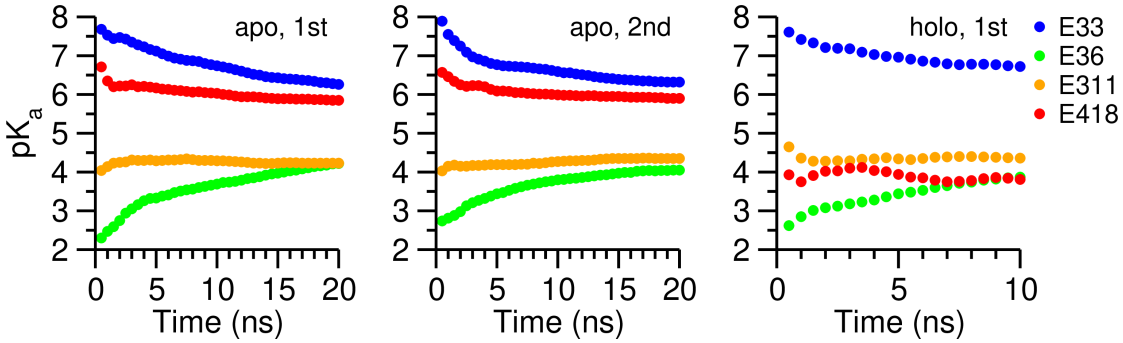

Figure 9-12. Convergence of the pK_a values for E33, E36, E311, and E418 in $PepT_{Sh}$ calculated by pH-REX CpHMD. pK_a's were calculated cumulatively versus simulation time.

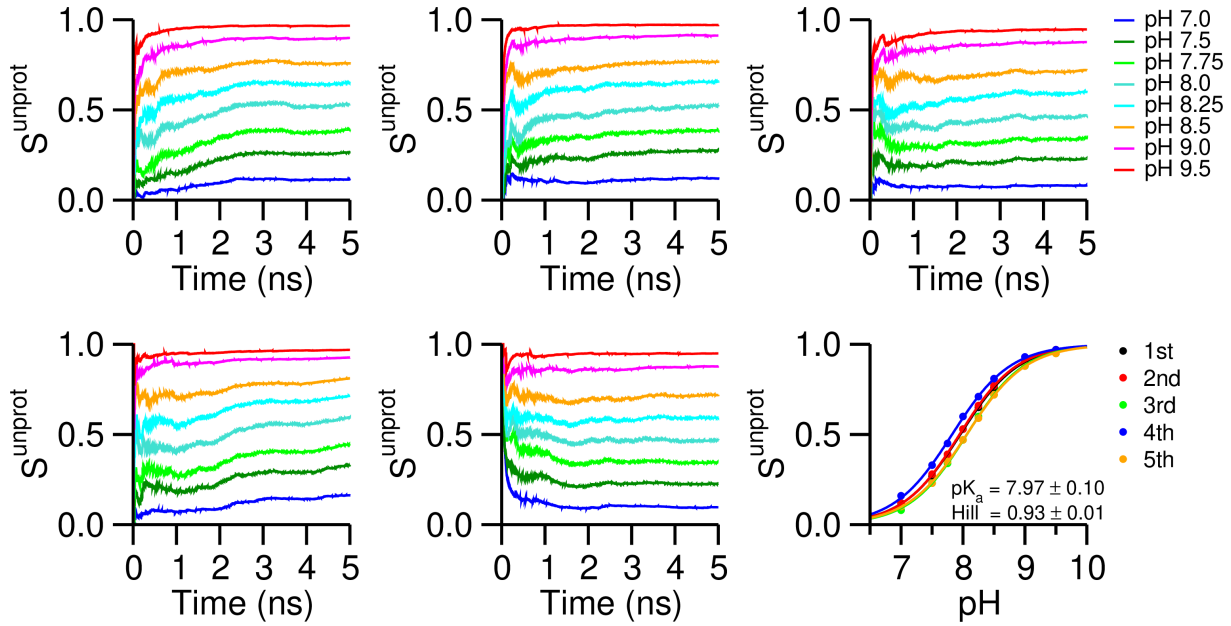

Figure 9-13. Validation of the hybrid-solvent CpHMD parameters for N-terminus (NT). To calculated the pK_a of NT in aqueous, 5 independent pH-REX CpHMD simulations were performed. The titration curves are plotted in the bottom right panel with pK_a and Hill coefficient n reported (mean ± standard deviation). Other panels plot the convergence of the unprotonated fraction S^{unprot} cumulatively calculated *vs.* time.

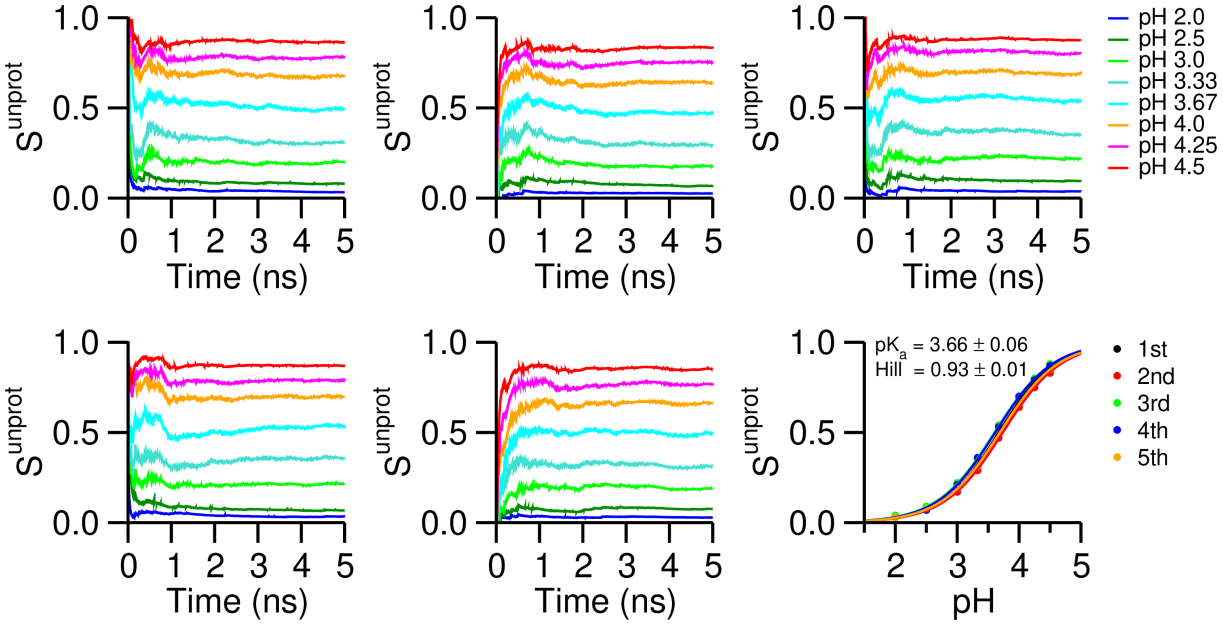

Figure 9-14. Validation of the hybrid-solvent CpHMD parameters for C-terminus (CT). To calculated the pK_a of CT in aqueous, 5 independent pH-REX CpHMD simulations were performed. The titration curves are plotted in the bottom right panel with pK_a and Hill coefficient n reported (mean ± standard deviation). Other panels plot the convergence of the unprotonated fraction S^{unprot} cumulatively calculated *vs.* time.

References

1. Eigen, M., Proton transfer, acid - base catalysis, and enzymatic hydrolysis. Part I: elementary processes. *Angew. Chem., Int. Ed. Engl.* **1964,** *3* (1), 1-19.

2. Zundel, G.; Metzger, H., Energiebänder der tunnelnden überschuß-protonen in flüssigen säuren. Eine IR-spektroskopische untersuchung der natur der gruppierungen H5O2+. *Z. Phys. Chem.* **1968,** *58* (5_6), 225-245.

3. Fournier, J. A.; Carpenter, W. B.; Lewis, N. H. C.; Tokmakoff, A., Broadband 2D IR spectroscopy reveals dominant asymmetric H5O2+ proton hydration structures in acid solutions. *Nat. Chem.* **2018,** *10* (9), 932-937.

4. Voth, G. A., Computer simulation of proton solvation and transport in aqueous and biomolecular systems. *Acc. Chem. Res.* **2006,** *39* (2), 143-150.

5. Markovitch, O.; Chen, H.; Izvekov, S.; Paesani, F.; Voth, G. A.; Agmon, N., Special pair dance and partner selection: Elementary steps in proton transport in liquid water. *J. Phys. Chem. B* **2008,** *112* (31), 9456-9466.

6. Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M., Ab initio molecular dynamics simulation of the solvation and transport of hydronium and hydroxyl ions in water. *J. Chem. Phys.* **1995,** *103* (1), 150-161.

7. Tuckerman, M.; Laasonen, K.; Sprik, M.; Parrinello, M., Ab initio molecular dynamics simulation of the solvation and transport of H3O+ and OH-ions in water. *J. Phys. Chem.* **1995,** *99* (16), 5749-5752.

8. Thämer, M.; De Marco, L.; Ramasesha, K.; Mandal, A.; Tokmakoff, A., Ultrafast 2D IR spectroscopy of the excess proton in liquid water. *Science* **2015,** *350* (6256), 78-82.

9. Carpenter, W. B.; Fournier, J. A.; Lewis, N. H. C.; Tokmakoff, A., Picosecond Proton Transfer Kinetics in Water Revealed with Ultrafast IR Spectroscopy. *J. Phys. Chem. B* **2018,** *122* (10), 2792-2802.

10. Carpenter, W. B.; Lewis, N. H. C.; Fournier, J. A.; Tokmakoff, A., Entropic barriers in the kinetics of aqueous proton transfer. *J. Chem. Phys.* **2019,** *151* (3), 034501.

11. Grotthuss, C. J. D. v., Sur la décomposition de l'eau et des corps qu'elle tient en dissolution à l'aide de l'électricité galvanique. *Ann. Chim.* **1806,** (58), 54-73.

12. Bush, D. R., Proton-Coupled Sugar and Amino-Acid Transporters in Plants. *Annu Rev Plant Phys* **1993,** *44*, 513-542.

13. Decoursey, T. E., Voltage-gated proton channels and other proton transfer pathways. *Physiol. Rev.* **2003,** *83* (2), 475-579.

14. Turina, P.; Samoray, D.; Graber, P., H+/ATP ratio of proton transport-coupled ATP synthesis and hydrolysis catalysed by CF0F1-liposomes. *EMBO J*. **2003,** *22* (3), 418-426.

15. Weber, J.; Senior, A. E., ATP synthesis driven by proton transport in F1F0-ATP synthase. *Febs Letters* **2003,** *545* (1), 61-70.

16. Pinto, L. H.; Lamb, R. A., The M2 proton channels of influenza A and B viruses. *J. Biol. Chem*. **2006,** *281* (14), 8997-9000.

17. Nevo, Y.; Nelson, N., The NRAMP family of metal-ion transporters. *Bba-Mol Cell Res* **2006,** *1763* (7), 609-620.

18. Wraight, C. A., Chance and design - Proton transfer in water, channels and bioenergetic proteins. *Biochim. Biophys. Acta, Bioenerg*. **2006,** *1757* (8), 886-912.

19. Buch-Pedersen, M. J.; Pedersen, B. P.; Veierskov, B.; Nissen, P.; Palmgren, M. G., Protons and how they are transported by proton pumps. *Pflug Arch Eur J Phy* **2009,** *457* (3), 573-579.

20. Levy, D.; Seigneuret, M.; Bluzat, A.; Rigaud, J., Evidence for proton countertransport by the sarcoplasmic reticulum Ca2 (+)-ATPase during calcium transport in reconstituted proteoliposomes with low ionic permeability. *J. Biol. Chem*. **1990,** *265* (32), 19524-19534.

21. Yu, X.; Carroll, S.; Rigaud, J.; Inesi, G., H+ countertransport and electrogenicity of the sarcoplasmic reticulum Ca2+ pump in reconstituted proteoliposomes. *Biophys. J*. **1993,** *64* (4), 1232-1242.

22. Espinoza-Fonseca, L. M.; Ramírez-Salinas, G. L., Microsecond molecular simulations reveal a transient proton pathway in the calcium pump. *J. Am. Chem. Soc*. **2015,** *137* (22), 7055-7058.

23. Daniel, H.; Kottra, G., The proton oligopeptide cotransporter family SLC15 in physiology and pharmacology. *Pflügers Archiv* **2004,** *447* (5), 610-618.

24. Smith, D. E.; Clémençon, B.; Hediger, M. A., Proton-coupled oligopeptide transporter family SLC15: physiological, pharmacological and pathological implications. *Mol Aspects Med* **2013,** *34* (2-3), 323-336.

25. Solcan, N.; Kwok, J.; Fowler, P. W.; Cameron, A. D.; Drew, D.; Iwata, S.; Newstead, S., Alternating access mechanism in the POT family of oligopeptide transporters. *EMBO J*. **2012,** *31* (16), 3411-3421.

26. Parker, J. L.; Li, C.; Brinth, A.; Wang, Z.; Vogeley, L.; Solcan, N.; Ledderboge-Vucinic, G.; Swanson, J. M. J.; Caffrey, M.; Voth, G. A.; Newstead, S., Proton movement and coupling in the POT family of peptide transporters. *Proc. Natl. Acad. Sci. U. S. A*. **2017,** *114* (50), 13182-13187.

27. Newstead, S., Recent advances in understanding proton coupled peptide transport via the POT family. *Curr. Opin. Struct. Biol.* **2017,** *45*, 17-24.

28. Minhas, G. S.; Bawdon, D.; Herman, R.; Rudden, M.; Stone, A. P.; James, A. G.; Thomas, G. H.; Newstead, S., Structural basis of malodour precursor transport in the human axilla. *Elife* **2018,** *7*, e34995.

29. Minhas, G. S.; Newstead, S., Structural basis for prodrug recognition by the SLC15 family of proton-coupled peptide transporters. *Proc. Natl. Acad. Sci. U. S. A.* **2019,** *116* (3), 804-809.

30. Tuckerman, M. E., Ab initio molecular dynamics: basic concepts, current trends and novel applications. *J. Phys.: Condens. Matter* **2002,** *14* (50), R1297.

31. Gillan, M. J.; Alfe, D.; Michaelides, A., Perspective: How good is DFT for water? *J. Chem. Phys.* **2016,** *144* (13), 130901.

32. Pitera, J. W.; Chodera, J. D., On the use of experimental observations to bias simulated ensembles. *J. Chem. Theory Comput.* **2012,** *8* (10), 3445-3451.

33. White, A. D.; Voth, G. A., Efficient and Minimal Method to Bias Molecular Simulations with Experimental Data. *J. Chem. Theory Comput.* **2014,** *10* (8), 3023-3030.

34. White, A. D.; Knight, C.; Hocky, G. M.; Voth, G. A., Communication: Improved ab initio molecular dynamics by minimally biasing with experimental data. *J. Chem. Phys.* **2017,** *146* (4), 041102.

35. Calio, P. B.; Hocky, G. M.; Voth, G. A., Minimal Experimental Bias on the Hydrogen Bond Greatly Improves Ab Initio Molecular Dynamics Simulations of Water. *J. Chem. Theory Comput.* **2020,** *16* (9), 5675-5684.

36. Medders, G. R.; Babin, V.; Paesani, F., A critical assessment of two-body and three-body interactions in water. *J. Chem. Theory Comput.* **2013,** *9* (2), 1103-1114.

37. Reddy, S. K.; Straight, S. C.; Bajaj, P.; Huy Pham, C.; Riera, M.; Moberg, D. R.; Morales, M. A.; Knight, C.; Götz, A. W.; Paesani, F., On the accuracy of the MB-pol many-body potential for water: Interaction energies, vibrational frequencies, and classical thermodynamic and dynamical properties from clusters to liquid water and ice. *J. Chem. Phys.* **2016,** *145* (19), 194504.

38. Reddy, S. K.; Moberg, D. R.; Straight, S. C.; Paesani, F., Temperature-dependent vibrational spectra and structure of liquid water from classical and quantum simulations with the MB-pol potential energy function. *J. Chem. Phys.* **2017,** *147* (24), 244504.

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

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

41. Lee, S.; Liang, R.; Voth, G. A.; Swanson, J. M., Computationally efficient multiscale reactive molecular dynamics to describe amino acid deprotonation in proteins. *J. Chem. Theory Comput.* **2016,** *12* (2), 879-891.

42. Sugita, Y.; Kitao, A.; Okamoto, Y., Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000,** *113* (15), 6042–6051.

43. Torrie, G. M.; Valleau, J. P., Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977,** *23* (2), 187–199.

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

45. Comer, J.; Gumbart, J. C.; Henin, J.; Lelievre, T.; Pohorille, A.; Chipot, C., The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask. *J. Phys. Chem. B* **2015,** *119* (3), 1129-1151.

46. Hayward, S.; Kitao, A.; Gō, N., Harmonicity and anharmonicity in protein dynamics: a normal mode analysis and principal component analysis. *Proteins* **1995,** *23* (2), 177-186.

47. Spiwok, V.; Lipovová, P.; Králová, B., Metadynamics in essential coordinates: free energy simulation of conformational changes. *J. Phys. Chem. B* **2007,** *111* (12), 3073-3076.

48. M. Sultan, M.; Pande, V. S., tICA-metadynamics: accelerating metadynamics by using kinetically selected collective variables. *J. Chem. Theory Comput.* **2017,** *13* (6), 2440-2447.

49. McCarty, J.; Parrinello, M., A variational conformational dynamics approach to the selection of collective variables in metadynamics. *J. Chem. Phys.* **2017,** *147* (20), 204109.

50. Chen, W.; Ferguson, A. L., Molecular enhanced sampling with autoencoders: On‐the‐fly collective variable discovery and accelerated free energy landscape exploration. *J. Comput. Chem.* **2018,** *39* (25), 2079-2102.

51. Wehmeyer, C.; Noé, F., Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *J. Chem. Phys.* **2018,** *148* (24), 241703.

52. Cuma, M.; Schmitt, U. W.; Voth, G. A., A multi-state empirical valence bond model for weak acid dissociation in aqueous solution. *J. Phys. Chem. A* **2001,** *105* (12), 2814-2823.

53. Konig, P. H.; Ghosh, N.; Hoffmann, M.; Elstner, M.; Tajkhorshid, E.; Frauenheim, T.; Cui, Q., Toward theoretical analyis of long-range proton transfer kinetics in biomolecular pumps. *J. Phys. Chem. A* **2006,** *110* (2), 548-563.

54. Pezeshki, S.; Lin, H., Adaptive-Partitioning QM/MM for Molecular Dynamics Simulations: 4. Proton Hopping in Bulk Water. *J. Chem. Theory Comput.* **2015,** *11* (6), 2398-2411.

55. Li, C. H.; Swanson, J. M. J., Understanding and Tracking the Excess Proton in Ab Initio Simulations; Insights from IR Spectra. *J. Phys. Chem. B* **2020,** *124* (27), 5696-5708.

56. Kaduk, B.; Kowalczyk, T.; Van Voorhis, T., Constrained Density Functional Theory. *Chem. Rev.* **2012,** *112* (1), 321-370.

57. Lee, S.; Swanson, J. M. J.; Voth, G. A., Multiscale Simulations Reveal Key Aspects of the Proton Transport Mechanism in the ClC-ec1 Antiporter. *Biophys. J.* **2016,** *110* (6), 1334-1345.

58. Lee, S.; Mayes, H. B.; Swanson, J. M. J.; Voth, G. A., The Origin of Coupled Chloride and Proton Transport in a Cl-/H+ Antiporter. *J. Am. Chem. Soc.* **2016,** *138* (45), 14923-14930.

59. Liang, R.; Swanson, J. M. J.; Peng, Y.; Wikström, M.; Voth, G. A., Multiscale simulations reveal key features of the proton-pumping mechanism in cytochrome c oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113* (27), 7420-7425.

60. Liang, R.; Swanson, J. M.; Wikström, M.; Voth, G. A., Understanding the essential proton-pumping kinetic gates and decoupling mutations in cytochrome c oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **2017,** *114* (23), 5924-5929.

61. Tiwary, P.; Berne, B., Spectral gap optimization of order parameters for sampling complex molecular systems. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113* (11), 2839-2844.

62. Tiwary, P.; Berne, B., Predicting reaction coordinates in energy landscapes with diffusion anisotropy. *J. Chem. Phys.* **2017,** *147* (15), 152701.

63. Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F., Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **2013,** *139* (1), 07B604_1.

64. Mardt, A.; Pasquali, L.; Wu, H.; Noé, F., VAMPnets for deep learning of molecular kinetics. *Nat. Commun.* **2018,** *9* (1), 1-11.

65. Wu, H.; Noé, F., Variational approach for learning Markov processes from time series data. *Journal of Nonlinear Science* **2020,** *30* (1), 23-66.

66. Radhakrishnan, T.; Herndon, W. C., Graph theoretical analysis of water clusters. *J. Phys. Chem.* **1991,** *95* (26), 10609-10617.

67. Tribello, G. A.; Giberti, F.; Sosso, G. C.; Salvalaglio, M.; Parrinello, M., Analyzing and driving cluster formation in atomistic simulations. *J. Chem. Theory Comput.* **2017,** *13* (3), 1317-1327.

68. Becke, A. D., Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys. Rev. A* **1988,** *38* (6), 3098-3100.

69. Lee, C. T.; Yang, W. T.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys. Rev. B* **1988,** *37* (2), 785-789.

70. Chai, J. D.; Head-Gordon, M., Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008,** *128* (8), 084106.

71. Wu, Q.; Kaduk, B.; Van Voorhis, T., Constrained density functional theory based configuration interaction improves the prediction of reaction barrier heights. *J. Chem. Phys.* **2009,** *130* (3), 034109.

72. Becke, A. D., A Multicenter Numerical-Integration Scheme for Polyatomic-Molecules. *J. Chem. Phys.* **1988,** *88* (4), 2547-2553.

73. Wu, Q.; Van Voorhis, T., Extracting electron transfer coupling elements from constrained density functional theory. *J. Chem. Phys.* **2006,** *125* (16), 164105.

74. Latora, V.; Marchiori, M., Efficient behavior of small-world networks. *Phys. Rev. Lett.* **2001,** *87* (19), 198701.

75. Pyykko, P.; Atsumi, M., Molecular Single-Bond Covalent Radii for Elements 1-118. *Chem. Eur. J.* **2009,** *15* (1), 186-197.

76. VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J., QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005,** *167* (2), 103-128.

77. Kuhne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schutt, O.; Schiffmann, F.; Golze, D.; Wilhelm, J.; Chulkov, S.; Bani-Hashemian, M. H.; Weber, V.; Borstnik, U.; Taillefumier, M.; Jakobovits, A. S.; Lazzaro, A.; Pabst, H.; Muller, T.; Schade, R.; Guidon, M.; Andermatt, S.; Holmberg, N.; Schenter, G. K.; Hehn, A.; Bussy, A.; Belleflamme, F.; Tabacchi, G.; Gloss, A.; Lass, M.; Bethune, I.; Mundy, C. J.; Plessl, C.; Watkins, M.; VandeVondele, J.; Krack, M.; Hutter, J., CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020,** *152* (19), 194103.

78. Holmberg, N.; Laasonen, K., Efficient constrained density functional theory implementation for simulation of condensed phase electron transfer reactions. *J. Chem. Theory Comput.* **2017,** *13* (2), 587-601.

79. Lehtola, S.; Steigemann, C.; Oliveira, M. J. T.; Marques, M. A. L., Recent developments in LIBXC - A comprehensive library of functionals for density functional theory. *Softwarex* **2018,** *7*, 1-5.

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

81. Hartwigsen, C.; Gœdecker, S.; Hutter, J., Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **1998,** *58* (7), 3641.

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

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

84. Barducci, A.; Bussi, G.; Parrinello, M., Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008,** *100* (2), 020603.

85. Peng, Y.; Swanson, J. M. J.; Kang, S.-g.; Zhou, R.; Voth, G. A., Hydrated Excess Protons Can Create Their Own Water Wires. *J. Phys. Chem. B* **2015,** *119* (29), 9212-9218.

86. Biswas, R.; Tse, Y. L. S.; Tokmakoff, A.; Voth, G. A., Role of Presolvation and Anharmonicity in Aqueous Phase Hydrated Proton Solvation and Transport. *J. Phys. Chem. B* **2016,** *120* (8), 1793-1804.

87. Schmitt, U. W.; Voth, G. A., The computer simulation of proton transport in water. *J. Chem. Phys.* **1999,** *111* (20), 9361-9381.

88. Biswas, R.; Carpenter, W.; Fournier, J. A.; Voth, G. A.; Tokmakoff, A., IR spectral assignments for the hydrated excess proton in liquid water. *J. Chem. Phys.* **2017,** *146* (15), 154507.

89. Nelson, J. G.; Peng, Y.; Silverstein, D. W.; Swanson, J. M. J., Multiscale Reactive Molecular Dynamics for Absolute pK(a) Predictions and Amino Acid Deprotonation. *J. Chem. Theory Comput.* **2014,** *10* (7), 2729-2737.

90. Wu, Y.; Tepper, H. L.; Voth, G. A., Flexible simple point-charge water model with improved liquid-state properties. *J. Chem. Phys.* **2006,** *124* (2), 024503.

91. Dama, J. F.; Parrinello, M.; Voth, G. A., Well-tempered metadynamics converges asymptotically. *Phys. Rev. Lett.* **2014,** *112* (24), 240602.

92. Tiwary, P.; Parrinello, M., A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B* **2015,** *119* (3), 736-742.

93. Marsalek, O.; Markland, T. E., Quantum dynamics and spectroscopy of ab initio liquid water: The interplay of nuclear and electronic quantum effects. *J. Phys. Chem. Lett.* **2017,** *8* (7), 1545-1551.

94. Del Ben, M.; Schönherr, M.; Hutter, J. r.; VandeVondele, J., Bulk liquid water at ambient temperature and pressure from MP2 theory. *J. Phys. Chem. Lett.* **2013,** *4* (21), 3753-3759.

95. Ceriotti, M.; Fang, W.; Kusalik, P. G.; McKenzie, R. H.; Michaelides, A.; Morales, M. A.; Markland, T. E., Nuclear quantum effects in water and aqueous systems: Experiment, theory, and current challenges. *Chem. Rev.* **2016,** *116* (13), 7529-7550.

96. Feynman, R. P.; Hibbs, A. R., *Quantum mechanics and path integrals*. McGraw-Hill: New York,, 1965; p xiv, 365 p.

97. Markland, T. E.; Ceriotti, M., Nuclear quantum effects enter the mainstream. *Nat. Rev. Chem.* **2018,** *2* (3), 1-14.

98. Markland, T. E.; Manolopoulos, D. E., An efficient ring polymer contraction scheme for imaginary time path integral simulations. *J. Chem. Phys.* **2008,** *129* (2), 024105.

99. Buxton, S. J.; Habershon, S., Accelerated path-integral simulations using ring-polymer interpolation. *J. Chem. Phys.* **2017,** *147* (22), 224107.

100. Ceriotti, M.; Manolopoulos, D. E., Efficient first-principles calculation of the quantum kinetic energy and momentum distribution of nuclei. *Phys. Rev. Lett.* **2012,** *109* (10), 100604.

101. Liang, R.; Swanson, J. M.; Voth, G. A., Benchmark study of the SCC-DFTB approach for a biomolecular proton channel. *J. Chem. Theory Comput.* **2014,** *10* (1), 451-462.

102. Zhang, L. F.; Han, J. Q.; Wang, H.; Car, R.; Weinan, E., Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018,** *120* (14).

103. Kapil, V.; Rossi, M.; Marsalek, O.; Petraglia, R.; Litman, Y.; Spura, T.; Cheng, B.; Cuzzocrea, A.; Meißner, R. H.; Wilkins, D. M., i-PI 2.0: A universal force engine for advanced molecular simulations. *Comput. Phys. Commun.* **2019,** *236*, 214-223.

104. Krynicki, K.; Green, C. D.; Sawyer, D. W., Pressure and temperature dependence of self-diffusion in water. *Faraday Discuss. Chem. Soc.* **1978,** *66*, 199-208.

105. Roberts, N. K.; Northey, H. L., Proton and deuteron mobility in normal and heavy water solutions of electrolytes. *J. Chem. Soc., Faraday Trans. 1* **1974,** *70*, 253-262.

106. Knight, C.; Lindberg, G. E.; Voth, G. A., Multiscale reactive molecular dynamics. *J. Chem. Phys.* **2012,** *137* (22), 22A525.

107. Wang, Z.; Swanson, J. M. J.; Voth, G. A., Modulating the Chemical Transport Properties of a Transmembrane Antiporter via Alternative Anion Flux. *J. Am. Chem. Soc.* **2018,** *140* (48), 16535-16543.

108. Mulliken, R. S., Electronic Population Analysis on Lcao-Mo Molecular Wave Functions .2. Overlap Populations, Bond Orders, and Covalent Bond Energies. *J. Chem. Phys.* **1955,** *23* (10), 1841-1846.

109. Hirshfeld, F. L., Bonded-Atom Fragments for Describing Molecular Charge-Densities. *Theor Chim Acta* **1977,** *44* (2), 129-138.

110. Cembran, A.; Song, L. C.; Mo, Y. R.; Gao, J. L., Block-Localized Density Functional Theory (BLDFT), Diabatic Coupling, and Their Use in Valence Bond Theory for Representing Reactive Potential Energy Surfaces. *J. Chem. Theory Comput.* **2009,** *5* (10), 2702-2716.

111. Ren, H. S.; Provorse, M. R.; Bao, P.; Qu, Z. X.; Gao, J. L., Multistate Density Functional Theory for Effective Diabatic Electronic Coupling. *J. Phys. Chem. Lett.* **2016,** *7* (12), 2286-2293.

112. Zhang, Y.; Voth, G. A., The Coupled Proton Transport in the ClC-ec1 Cl-/H+ Antiporter. *Biophys. J.* **2011,** *101* (10), L47-L49.

113. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. *J. Chem. Theory Comput.* **2012,** *8* (9), 3257-3273.

114. Nosé, S., A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984,** *52* (2), 255–268.

115. Hoover, W. G., Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985,** *31* (3), 1695–1697.

116. Maupin, C. M.; Wong, K. F.; Soudackov, A. V.; Kim, S.; Voth, G. A., A multistate empirical valence bond description of protonatable amino acids. *J. Phys. Chem. A* **2006,** *110* (2), 631-639.

117. Denisov, V. P.; Schlessman, J. L.; Garcia-Moreno, B.; Halle, B., Stabilization of internal charges in a protein: Water penetration or conformational change? *Biophys. J.* **2004,** *87* (6), 3982-3994.

118. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52* (12), 7182–7190.

119. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997,** *18* (12), 1463–1472.

120. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *Softwarex* **2015,** *1–2*, 19–25.

121. Kutzner, C.; Páll, S.; Fechner, M.; Esztermann, A.; de Groot, B. L.; Grubmüller, H., Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. Wiley Online Library: 2015.

122. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M., The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J. Comput. Chem.* **1992,** *13* (8), 1011-1021.

123. Roux, B.; Andersen, O. S.; Allen, T. W., Comment on "Free energy simulations of single and double ion occupancy in gramicidin A" [J. Chem. Phys. 126, 105103 (2007)]. *J. Chem. Phys.* **2008,** *128* (22).

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

125. Yamashita, T.; Peng, Y.; Knight, C.; Voth, G. A., Computationally Efficient Multiconfigurational Reactive Molecular Dynamics. *J. Chem. Theory Comput.* **2012,** *8* (12), 4863-4875.

126. Haynes, W. M., *CRC handbook of chemistry and physics*. CRC press: 2014.

127. Karp, D. A.; Stahley, M. R.; Garcia-Moreno, E. B., Conformational Consequences of Ionization of Lys, Asp, and Glu Buried at Position 66 in Staphylococcal Nuclease. *Biochemistry-Us* **2010,** *49* (19), 4138-4146.

128. Fitch, C. A.; Karp, D. A.; Lee, K. K.; Stites, W. E.; Lattman, E. E.; García-Moreno, E. B., Experimental pKa values of buried residues: analysis with continuum methods and role of water penetration. *Biophys. J.* **2002,** *82* (6), 3289-3304.

129. García-Moreno, B. E.; Dwyer, J. J.; Gittis, A. G.; Lattman, E. E.; Spencer, D. S.; Stites, W. E., Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophys. Chem.* **1997,** *64* (1-3), 211-224.

130. Chandler, D., *Introduction to modern statistical mechanics*. Oxford University Press: New York, 1987; p xiii, 274 p.

131. Bowman, G. R.; Pande, V. S.; Noé, F., *An introduction to Markov state models and their application to long timescale molecular simulation*. Springer Science & Business Media: 2013; Vol. 797.

132. Chen, B.; Ivanov, I.; Klein, M. L.; Parrinello, M., Hydrogen bonding in water. *Phys. Rev. Lett.* **2003,** *91* (21), 215503.

133. Medders, G. R.; Babin, V.; Paesani, F., Development of a “first-principles” water potential with flexible monomers. III. Liquid phase properties. *J. Chem. Theory Comput.* **2014,** *10* (8), 2906-2910.

134. Morrone, J. A.; Car, R., Nuclear quantum effects in water. *Phys. Rev. Lett.* **2008,** *101* (1), 017801.

135. Pople, J. A.; Gill, P. M.; Johnson, B. G., Kohn—Sham density-functional theory within a finite basis set. *Chem. Phys. Lett.* **1992,** *199* (6), 557-560.

136. Sun, J.; Ruzsinszky, A.; Perdew, J. P., Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.* **2015,** *115* (3), 036402.

137. VandeVondele, J.; Hutter, J., An efficient orbital transformation method for electronic structure calculations. *J. Chem. Phys.* **2003,** *118* (10), 4365-4369.

138. Chen, M.; Ko, H.-Y.; Remsing, R. C.; Andrade, M. F. C.; Santra, B.; Sun, Z.; Selloni, A.; Car, R.; Klein, M. L.; Perdew, J. P., Ab initio theory and modeling of water. *Proc. Natl. Acad. Sci. U. S. A.* **2017,** *114* (41), 10846-10851.

139. Soper, A.; Benmore, C., Quantum differences between heavy and light water. *Phys. Rev. Lett.* **2008,** *101* (6), 065502.

140. Chau, P.-L.; Hardwick, A., A new order parameter for tetrahedral configurations. *Mol. Phys.* **1998,** *93* (3), 511-518.

141. Skinner, L. B.; Huang, C.; Schlesinger, D.; Pettersson, L. G.; Nilsson, A.; Benmore, C. J., Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide Q-range. *J. Chem. Phys.* **2013,** *138* (7), 074506.

142. Giberti, F.; Hassanali, A. A.; Ceriotti, M.; Parrinello, M., The role of quantum effects on structural and electronic fluctuations in neat and charged water. *J. Phys. Chem. B* **2014,** *118* (46), 13226-13235.

143. Cassone, G., Nuclear quantum effects largely influence molecular dissociation and proton transfer in liquid water under an electric field. *J. Phys. Chem. Lett.* **2020,** *11* (21), 8983-8988.

144. Holmes, M.; Parker, N.; Povey, M. In *Temperature dependence of bulk viscosity in water using acoustic spectroscopy*, Journal of Physics: Conference Series, IOP Publishing: 2011; p 012011.

145. Holz, M.; Heil, S. R.; Sacco, A., Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1H NMR PFG measurements. *Phys. Chem. Chem. Phys.* **2000,** *2* (20), 4740-4742.

146. Cukierman, S., Et tu, Grotthuss! and other unfinished stories. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **2006,** *1757* (8), 876-885.

147. Swanson, J. M. J.; Maupin, C. M.; Chen, H. N.; Petersen, M. K.; Xu, J. C.; Wu, Y. J.; Voth, G. A., Proton solvation and transport in aqueous and biomolecular systems: Insights from computer simulations. *J. Phys. Chem. B* **2007,** *111* (17), 4300-4314.

148. Kreuer, K.-D.; Paddison, S. J.; Spohr, E.; Schuster, M., Transport in proton conductors for fuel-cell applications: simulations, elementary reactions, and phenomenology. *Chem. Rev.* **2004,** *104* (10), 4637-4678.

149. Savage, J.; Tse, Y.-L. S.; Voth, G. A., Proton transport mechanism of perfluorosulfonic acid membranes. *J. Phys. Chem. C* **2014,** *118* (31), 17436-17445.

150. Arntsen, C.; Savage, J.; Tse, Y. L. S.; Voth, G. A., Simulation of Proton Transport in Proton Exchange Membranes with Reactive Molecular Dynamics. *Fuel Cells* **2016,** *16* (6), 695-703.

151. Bernal, J. D.; Fowler, R. H., A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions. *J. Chem. Phys.* **1933,** *1* (8), 515-548.

152. Knight, C.; Voth, G. A., The curious case of the hydrated proton. *Acc. Chem. Res.* **2012,** *45* (1), 101-109.

153. Agmon, N.; Bakker, H. J.; Campen, R. K.; Henchman, R. H.; Pohl, P.; Roke, S.; Thämer, M.; Hassanali, A., Protons and hydroxide ions in aqueous systems. *Chem. Rev.* **2016,** *116* (13), 7642-7672.

154. Lobaugh, J.; Voth, G. A., The quantum dynamics of an excess proton in water. *J. Chem. Phys.* **1996,** *104* (5), 2056-2069.

155. Vuilleumier, R.; Borgis, D., Transport and spectroscopy of the hydrated proton: a molecular dynamics study. *J. Chem. Phys.* **1999,** *111* (9), 4251-4266.

156. Marx, D.; Tuckerman, M. E.; Hutter, J.; Parrinello, M., The nature of the hydrated excess proton in water. *Nature* **1999,** *397* (6720), 601-604.

157. Marx, D., Proton transfer 200 years after von Grotthuss: Insights from ab initio simulations. *Chemphyschem* **2006,** *7* (9), 1848-1870.

158. Lapid, H.; Agmon, N.; Petersen, M. K.; Voth, G. A., A bond-order analysis of the mechanism for hydrated proton mobility in liquid water. *J. Chem. Phys.* **2005,** *122* (1), 014506.

159. Vuilleumier, R.; Borgis, D., An extended empirical valence bond model for describing proton mobility in water. *Isr. J. Chem.* **1999,** *39* (3 - 4), 457-467.

160. Kornyshev, A.; Kuznetsov, A.; Spohr, E.; Ulstrup, J., Kinetics of proton transport in water. ACS Publications: 2003.

161. Swanson, J. M. J.; Simons, J., Role of Charge Transfer in the Structure and Dynamics of the Hydrated Proton. *J. Phys. Chem. B* **2009,** *113* (15), 5149-5161.

162. Asmis, K. R.; Pivonka, N. L.; Santambrogio, G.; Brümmer, M.; Kaposta, C.; Neumark, D. M.; Wöste, L., Gas-phase infrared spectrum of the protonated water dimer. *Science* **2003,** *299* (5611), 1375-1377.

163. Headrick, J. M.; Diken, E. G.; Walters, R. S.; Hammer, N. I.; Christie, R. A.; Cui, J.; Myshakin, E. M.; Duncan, M. A.; Johnson, M. A.; Jordan, K. D., Spectral signatures of hydrated proton vibrations in water clusters. *Science* **2005,** *308* (5729), 1765-1769.

164. Heine, N.; Fagiani, M. R.; Rossi, M.; Wende, T.; Berden, G.; Blum, V.; Asmis, K. R., Isomer-selective detection of hydrogen-bond vibrations in the protonated water hexamer. *J. Am. Chem. Soc.* **2013,** *135* (22), 8266-8273.

165. Heine, N.; Fagiani, M. R.; Asmis, K. R., Disentangling the contribution of multiple isomers to the infrared spectrum of the protonated water heptamer. *J. Phys. Chem. Lett.* **2015,** *6* (12), 2298-2304.

166. Shellman, S. D.; Lewis, J. P.; Glaesemann, K. R.; Sikorski, K.; Voth, G. A., Massively parallel linear-scaling algorithm in an ab initio local-orbital total-energy method. *J. Comput. Phys.* **2003,** *188* (1), 1-15.

167. Xu, J.; Zhang, Y.; Voth, G. A., Infrared spectrum of the hydrated proton in water. *J. Phys. Chem. Lett.* **2011,** *2* (2), 81-86.

168. Dahms, F.; Fingerhut, B. P.; Nibbering, E. T.; Pines, E.; Elsaesser, T., Large-amplitude transfer motion of hydrated excess protons mapped by ultrafast 2D IR spectroscopy. *Science* **2017,** *357* (6350), 491-495.

169. Kundu, A.; Dahms, F.; Fingerhut, B. P.; Nibbering, E. T.; Pines, E.; Elsaesser, T., Hydrated excess protons in acetonitrile/water mixtures: Solvation species and ultrafast proton motions. *J. Phys. Chem. Lett.* **2019,** *10* (9), 2287-2294.

170. Day, T. J.; Soudackov, A. V.; Čuma, M.; Schmitt, U. W.; Voth, G. A., A second generation multistate empirical valence bond model for proton transport in aqueous systems. *J. Chem. Phys.* **2002,** *117* (12), 5839-5849.

171. Tse, Y. L. S.; Knight, C.; Voth, G. A., An analysis of hydrated proton diffusion in ab initio molecular dynamics. *J. Chem. Phys.* **2015,** *142* (1).

172. Carpenter, W. B.; Yu, Q.; Hack, J. H.; Dereka, B.; Bowman, J. M.; Tokmakoff, A., Decoding the 2D IR spectrum of the aqueous proton with high-level VSCF/VCI calculations. *J. Chem. Phys.* **2020,** *153* (12), 124506.

173. Park, K.; Lin, W.; Paesani, F., A refined MS-EVB model for proton transport in aqueous environments. *J. Phys. Chem. B* **2012,** *116* (1), 343-352.

174. Roy, S.; Schenter, G. K.; Napoli, J. A.; Baer, M. D.; Markland, T. E.; Mundy, C. J., Resolving Heterogeneous Dynamics of Excess Protons in Aqueous Solution with Rate Theory. *J. Phys. Chem. B* **2020,** *124* (27), 5665-5675.

175. Daly Jr, C. A.; Streacker, L. M.; Sun, Y.; Pattenaude, S. R.; Hassanali, A. A.; Petersen, P. B.; Corcelli, S. A.; Ben-Amotz, D., Decomposition of the experimental Raman and

infrared spectra of acidic water into proton, special pair, and counterion contributions. *J. Phys. Chem. Lett.* **2017,** *8* (21), 5246-5252.

176. Babin, V.; Medders, G. R.; Paesani, F., Development of a "first principles" water potential with flexible monomers. II: Trimer potential energy surface, third virial coefficient, and small clusters. *J. Chem. Theory Comput.* **2014,** *10* (4), 1599-1607.

177. Berkelbach, T. C.; Lee, H.-S.; Tuckerman, M. E., Concerted hydrogen-bond dynamics in the transport mechanism of the hydrated proton: A first-principles molecular dynamics study. *Phys. Rev. Lett.* **2009,** *103* (23), 238302.

178. Forge, V.; Mintz, E.; Guillain, F., Ca2+ binding to sarcoplasmic reticulum ATPase revisited. I. Mechanism of affinity and cooperativity modulation by H+ and Mg2+. *J. Biol. Chem.* **1993,** *268* (15), 10953-10960.

179. Buoninsegni, F. T.; Bartolommei, G.; Moncelli, M. R.; Inesi, G.; Guidelli, R., Time-resolved charge translocation by sarcoplasmic reticulum Ca-ATPase measured on a solid supported membrane. *Biophys. J.* **2004,** *86* (6), 3671-3686.

180. Obara, K.; Miyashita, N.; Xu, C.; Toyoshima, I.; Sugita, Y.; Inesi, G.; Toyoshima, C., Structural role of countertransport revealed in Ca2+ pump crystal structure in the absence of Ca2+. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102* (41), 14489-14496.

181. Floyd, R.; Wray, S., Calcium transporters and signalling in smooth muscles. *Cell calcium* **2007,** *42* (4-5), 467-476.

182. Inesi, G.; Tadini-Buoninsegni, F., Ca 2+/H+ exchange, lumenal Ca 2+ release and Ca 2+/ATP coupling ratios in the sarcoplasmic reticulum ATPase. *J. Cell Commun. Signal.* **2014,** *8* (1), 5-11.

183. Meis, L. d.; Vianna, A. L., Energy interconversion by the Ca2+-dependent ATPase of the sarcoplasmic reticulum. *Annu. Rev. Biochem* **1979,** *48* (1), 275-292.

184. Møller, J. V.; Juul, B.; le Maire, M., Structural organization, ion transport, and energy transduction of P-type ATPases. *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes* **1996,** *1286* (1), 1-51.

185. Toyoshima, C., How Ca2+-ATPase pumps ions across the sarcoplasmic reticulum membrane. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **2009,** *1793* (6), 941-946.

186. Kuum, M.; Veksler, V.; Liiv, J.; Ventura-Clapier, R.; Kaasik, A., Endoplasmic reticulum potassium–hydrogen exchanger and small conductance calcium-activated potassium channel activities are essential for ER calcium uptake in neurons and cardiomyocytes. *J. Cell Sci.* **2012,** *125* (3), 625-633.

187. Pollock, N.; Kargacin, M.; Kargacin, G., Chloride channel blockers inhibit Ca2+ uptake by the smooth muscle sarcoplasmic reticulum. *Biophys. J.* **1998,** *75* (4), 1759-1766.

188. Hirota, S.; Trimble, N.; Pertens, E.; Janssen, L. J., Intracellular Cl– fluxes play a novel role in Ca2+ handling in airway smooth muscle. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **2006,** *290* (6), L1146-L1153.

189. Kim, J. H.; Johannes, L.; Goud, B.; Antony, C.; Lingwood, C. A.; Daneman, R.; Grinstein, S., Noninvasive measurement of the pH of the endoplasmic reticulum at rest and during calcium release. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (6), 2997-3002.

190. Kneen, M.; Farinas, J.; Li, Y.; Verkman, A., Green fluorescent protein as a noninvasive intracellular pH indicator. *Biophys. J.* **1998,** *74* (3), 1591-1599.

191. Pape, P.; Konishi, M.; Hollingworth, S.; Baylor, S., Perturbation of sarcoplasmic reticulum calcium release and phenol red absorbance transients by large concentrations of fura-2 injected into frog skeletal muscle fibers. *J. Gen. Physiol.* **1990,** *96* (3), 493-516.

192. Kamp, F.; Donoso, P.; Hidalgo, C., Changes in luminal pH caused by calcium release in sarcoplasmic reticulum vesicles. *Biophys. J.* **1998,** *74* (1), 290-296.

193. Takeshima, H.; Venturi, E.; Sitsapesan, R., New and notable ion - channels in the sarcoplasmic/endoplasmic reticulum: do they support the process of intracellular Ca2+ release? *J. Physiol.* **2015,** *593* (15), 3241-3251.

194. Espinoza-Fonseca, L. M., The Ca2+-ATPase pump facilitates bidirectional proton transport across the sarco/endoplasmic reticulum. *Mol Biosyst* **2017,** *13* (4), 633-637.

195. MacLennan, D. H.; Kranias, E. G., Phospholamban: a crucial regulator of cardiac contractility. *Nat. Rev. Mol. Cell Biol.* **2003,** *4* (7), 566-577.

196. Toyoshima, C.; Iwasawa, S.; Ogawa, H.; Hirata, A.; Tsueda, J.; Inesi, G., Crystal structures of the calcium pump and sarcolipin in the Mg 2+-bound E1 state. *Nature* **2013,** *495* (7440), 260-264.

197. Winther, A.-M. L.; Bublitz, M.; Karlsen, J. L.; Møller, J. V.; Hansen, J. B.; Nissen, P.; Buch-Pedersen, M. J., The sarcolipin-bound calcium pump stabilizes calcium sites exposed to the cytoplasm. *Nature* **2013,** *495* (7440), 265-269.

198. Chen, H.; Ilan, B.; Wu, Y.; Zhu, F.; Schulten, K.; Voth, G. A., Charge delocalization in proton channels, I: the aquaporin channels and proton blockage. *Biophys. J.* **2007,** *92* (1), 46-60.

199. Wu, Y.; Ilan, B.; Voth, G. A., Charge delocalization in proton channels, II: the synthetic LS2 channel and proton selectivity. *Biophys. J.* **2007,** *92* (1), 61-69.

200. Watkins, L. C.; Liang, R.; Swanson, J. M.; DeGrado, W. F.; Voth, G. A., Proton-induced conformational and hydration dynamics in the influenza A M2 channel. *J. Am. Chem. Soc.* **2019,** *141* (29), 11667-11676.

201. Liang, R.; Li, H.; Swanson, J. M.; Voth, G. A., Multiscale simulation reveals a multifaceted mechanism of proton permeation through the influenza A M2 proton channel. *Proc. Natl. Acad. Sci. U. S. A.* **2014,** *111* (26), 9396-9401.

202. Liang, R.; Swanson, J. M.; Madsen, J. J.; Hong, M.; DeGrado, W. F.; Voth, G. A., Acid activation mechanism of the influenza A M2 proton channel. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113* (45), E6955-E6964.

203. Vineyard, G. H., Frequency factors and isotope effects in solid state rate processes. *J. Phys. Chem. Solids* **1957,** *3* (1-2), 121-127.

204. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007,** *126* (1).

205. Berendsen, H. J.; Postma, J. v.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984,** *81* (8), 3684-3690.

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

207. Maseras, F.; Morokuma, K., IMOMM: A new integrated ab initio+ molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.* **1995,** *16* (9), 1170-1179.

208. Hockney, R. W.; Eastwood, J. W., *Computer simulation using particles*. Special student ed.; A. Hilger: Bristol England ; Philadelphia, 1988; p xxi, 540 p.

209. Laio, A.; Gervasio, F. L., Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics* **2008,** *71* (12).

210. Hastie, T.; Stuetzle, W., Principal curves. *J. Am. Stat. Assoc.* **1989,** *84* (406), 502-516.

211. Zhang, Y.; Voth, G. A., Combined metadynamics and umbrella sampling method for the calculation of ion permeation free energy profiles. *J. Chem. Theory Comput.* **2011,** *7* (7), 2277-2283.

212. Casey, J. R.; Grinstein, S.; Orlowski, J., Sensors and regulators of intracellular pH. *Nat. Rev. Mol. Cell Biol.* **2010,** *11* (1), 50-61.

213. Tunuguntla, R. H.; Allen, F. I.; Kim, K.; Belliveau, A.; Noy, A., Ultrafast proton transport in sub-1-nm diameter carbon nanotube porins. *Nat. Nanotechnol.* **2016,** *11* (7), 639-+.

214. Miyake, T.; Rolandi, M., Grotthuss mechanisms: from proton transport in proton wires to bioprotonic devices. *J. Phys.: Condens. Matter* **2016,** *28* (2), 023001.

215. Mayes, H. B.; Lee, S.; White, A. D.; Voth, G. A.; Swanson, J. M. J., Multiscale Kinetic Modeling Reveals an Ensemble of Cl-/H+ Exchange Pathways in ClC-ec1 Antiporter. *J. Am. Chem. Soc.* **2018,** *140* (5), 1793-1804.

216. Li, C.; Yue, Z.; Espinoza-Fonseca, L. M.; Voth, G. A., Multiscale Simulation Reveals Passive Proton Transport Through SERCA on the Microsecond Timescale. *Biophys. J.* **2020,** *119* (5), 1033-1040.

217. Ma, X.; Li, C.; Martinson, A. B.; Voth, G. A., Water-Assisted Proton Transport in Confined Nanochannels. *J. Phys. Chem. C* **2020,** *124* (29), 16186-16201.

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

219. Jentsch, T. J.; Stein, V.; Weinreich, F.; Zdebik, A. A., Molecular structure and physiological function of chloride channels. *Physiol. Rev.* **2002,** *82* (2), 503-568.

220. Picollo, A.; Pusch, M., Chloride/proton antiporter activity of mammalian CLC proteins ClC-4 and ClC-5. *Nature* **2005,** *436* (7049), 420-423.

221. Accardi, A.; Miller, C., Secondary active transport mediated by a prokaryotic homologue of ClC Cl-channels. *Nature* **2004,** *427* (6977), 803-807.

222. Scheel, O.; Zdebik, A. A.; Lourdel, S.; Jentsch, T. J., Voltage-dependent electrogenic chloride/proton exchange by endosomal CLC proteins. *Nature* **2005,** *436* (7049), 424-427.

223. Rychkov, G. Y.; Pusch, M.; Roberts, M. L.; Jentsch, T.; Bretag, A. H., Permeation and block of the skeletal muscle chloride channel, ClC-1, by foreign anions. *J. Gen. Physiol.* **1998,** *111* (5), 653-665.

224. Fahlke, C.; Dürr, C.; George Jr, A. L., Mechanism of ion permeation in skeletal muscle chloride channels. *J. Gen. Physiol.* **1997,** *110* (5), 551-564.

225. Fahlke, C.; Henry, T. Y.; Beck, C. L.; Rhodes, T. H.; George, A. L., Pore-forming segments in voltage-gated chloride channels. *Nature* **1997,** *390* (6659), 529-532.

226. Ludewig, U.; Jentsch, T. J.; Pusch, M., Analysis of a protein region involved in permeation and gating of the voltage - gated Torpedo chloride channel ClC - 0. *J. Physiol.* **1997,** *498* (3), 691-702.

227. Hebeisen, S.; Heidtmann, H.; Cosmelli, D.; Gonzalez, C.; Poser, B.; Latorre, R.; Alvarez, O.; Fahlke, C., Anion permeation in human ClC-4 channels. *Biophys. J.* **2003,** *84* (4), 2306-2318.

228. Nguitragool, W.; Miller, C., Uncoupling of a CLC Cl–/H+ exchange transporter by polyatomic anions. *J. Mol. Biol.* **2006,** *362* (4), 682-690.

229. Jiang, T.; Han, W.; Maduke, M.; Tajkhorshid, E., Molecular basis for differential anion binding and proton coupling in the Cl–/H+ exchanger ClC-ec1. *J. Am. Chem. Soc.* **2016,** *138* (9), 3066-3075.

230. Kuang, Z.; Mahankali, U.; Beck, T. L., Proton pathways and H+/Cl– stoichiometry in bacterial chloride transporters. *Proteins* **2007,** *68* (1), 26-33.

231. Yen, J. Y., Finding the k shortest loopless paths in a network. *Manag. Sci.* **1971,** *17* (11), 712-716.

232. Siek, J.; Lee, L.-Q.; Lumsdaine, A., *The boost graph library*. Pearson India: 2002.

233. Grossfield, A. *WHAM: the weighted histogram analysis method*, 2.0.9.

234. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996,** *14* (1), 33–38.

235. Hunter, J. D., Matplotlib: A 2D graphics environment. *Comput Sci Eng* **2007,** *9* (03), 90-95.

236. Rubio-Aliaga, I.; Daniel, H., Mammalian peptide transporters as targets for drug delivery. *Trends Pharmacol. Sci.* **2002,** *23* (9), 434-440.

237. Giacomini, K. M.; Huang, S. M.; Tweedie, D. J.; Benet, L. Z.; Brouwer, K. L. R.; Chu, X. Y.; Dahlin, A.; Evers, R.; Fischer, V.; Hillgren, K. M.; Hoffmaster, K. A.; Ishikawa, T.; Keppler, D.; Kim, R. B.; Lee, C. A.; Niemi, M.; Polli, J. W.; Sugiyama, Y.; Swaan, P. W.; Ware, J. A.; Wright, S. H.; Yee, S. W.; Zamek-Gliszczynski, M. J.; Zhang, L.; Transporter, I., Membrane transporters in drug development. *Nat. Rev. Drug Discov.* **2010,** *9* (3), 215-236.

238. Lin, L.; Yee, S. W.; Kim, R. B.; Giacomini, K. M., SLC transporters as therapeutic targets: emerging opportunities. *Nat. Rev. Drug Discov.* **2015,** *14* (8), 543-560.

239. Quistgaard, E. M.; Löw, C.; Guettou, F.; Nordlund, P., Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nat. Rev. Mol. Cell Biol.* **2016,** *17* (2), 123-132.

240. Newstead, S.; Drew, D.; Cameron, A. D.; Postis, V. L.; Xia, X.; Fowler, P. W.; Ingram, J. C.; Carpenter, E. P.; Sansom, M. S.; McPherson, M. J., Crystal structure of a prokaryotic homologue of the mammalian oligopeptide–proton symporters, PepT1 and PepT2. *EMBO J.* **2011,** *30* (2), 417-426.

241. Drew, D.; Boudker, O., Shared molecular mechanisms of membrane transporters. *Annu. Rev. Biochem* **2016,** *85*, 543-572.

242. Newstead, S., Molecular insights into proton coupled peptide transport in the PTR family of oligopeptide transporters. *Bba-Gen Subjects* **2015,** *1850* (3), 488-499.

243. Huang, Y. F.; Lemieux, M. J.; Song, J. M.; Auer, M.; Wang, D. N., Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. *Science* **2003,** *301* (5633), 616-620.

244. Doki, S.; Kato, H. E.; Solcan, N.; Iwaki, M.; Koyama, M.; Hattori, M.; Iwase, N.; Tsukazaki, T.; Sugita, Y.; Kandori, H.; Newstead, S.; Ishitani, R.; Nureki, O., Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (28), 11343-11348.

245. Immadisetty, K.; Hettige, J.; Moradi, M., What can and cannot be learned from molecular dynamics simulations of bacterial proton-coupled oligopeptide transporter GkPOT? *J. Phys. Chem. B* **2017,** *121* (15), 3644-3656.

246. Selvam, B.; Mittal, S.; Shukla, D., Free energy landscape of the complete transport cycle in a key bacterial transporter. *ACS Cent. Sci.* **2018,** *4* (9), 1146-1154.

247. Batista, M. R.; Watts, A.; José Costa-Filho, A., Exploring Conformational Transitions and Free-Energy Profiles of Proton-Coupled Oligopeptide Transporters. *J. Chem. Theory Comput.* **2019,** *15* (11), 6433-6443.

248. Aduri, N. G.; Montefiori, M.; Khalil, R.; Gajhede, M.; Jørgensen, F. S.; Mirza, O., Molecular Dynamics Simulations Reveal the Proton: Peptide Coupling Mechanism in the Bacterial Proton-Coupled Oligopeptide Transporter YbgH. *ACS Omega* **2019,** *4* (1), 2040-2046.

249. Jo, S.; Kim, T.; Im, W., Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *Plos One* **2007,** *2* (9), e880.

250. Jo, S.; Kim, T.; Iyer, V. G.; Im, W., CHARNIM-GUI: A Web-Based Grraphical User Interface for CHARMM. *J. Comput. Chem.* **2008,** *29* (11), 1859–1865.

251. Durell, S. R.; Brooks, B. R.; Ben-Naim, A., Solvent-Induced Forces between Two Hydrophilic Groups. *J. Phys. Chem.* **1994,** *98* (8), 2198–2202.

252. MacKerell Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr., R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998,** *102* (18), 3586–3616.

253. Mackerell Jr., A. D.; Feig, M.; Brooks III, C. L., Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004,** *25* (11), 1400–1415.

254. MacKerell Jr., A. D.; Feig, M.; Brooks III, C. L., Improved Treatment of the Protein Backbone in Empirical Force Fields. *J. Am .Chem. Soc.* **2004,** *126* (3), 698–699.

255. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I., CHARMM general force field: A force field for drug ‐ like molecules compatible with the CHARMM all ‐ atom additive biological force fields. *J. Comput. Chem.* **2010,** *31* (4), 671-690.

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

257. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103* (19), 8577–8593.

258. Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W., CHARMM-GUI Membrane Builder Toward Realistic Biological Membrane Simulations. *J. Comput. Chem.* **2014,** *35* (27), 1997–2004.

259. Torrie, G. M.; Valleau, J. P., Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *J. Comput. Phys.* **1977,** *23* (2), 187-199.

260. Rosta, E.; Hummer, G., Free energies from dynamic weighted histogram analysis using unbiased Markov state model. *J. Chem. Theory Comput.* **2015,** *11* (1), 276-285.

261. Sicard, F.; Koskin, V.; Annibale, A.; Rosta, E., Position-Dependent Diffusion from Biased Simulations and Markov State Model Analysis. *J. Chem. Theory Comput.* **2021,** *17* (4), 2022-2033.

262. Szabo, A.; Schulten, K.; Schulten, Z., First passage time approach to diffusion controlled reactions. *J. Chem. Phys.* **1980,** *72* (8), 4350-4357.

263. Chen, W.; Huang, Y.; Shen, J., Conformational Activation of a Transmembrane Proton Channel from Constant pH Molecular Dynamics. *J. Phys. Chem. Lett.* **2016,** *7* (19), 3961–3966.

264. Wallace, J. A.; Shen, J. K., Continuous Constant pH Molecular Dynamics in Explicit Solvent with pH-Based Replica Exchange. *J. Chem. Theory Comput.* **2011,** *7* (8), 2617–2629.

265. Brooks, B. R.; Brooks III, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009,** *30* (10), 1545–1614.

266. Lee, M. S.; Salsbury Jr., F. R.; Brooks III, C. L., Constant-pH Molecular Dynamics Using Continuous Titration Coordinates. *Proteins* **2004,** *56* (4), 738–752.

267. Khandogin, J.; Brooks III, C. L., Constant pH Molecular Dynamics with Proton Tautomerism. *Biophys. J.* **2005,** *89* (1), 141–157.

268. Steinbach, P. J.; Brooks, B. R., New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation. *J. Comput. Chem.* **1994,** *15* (7), 667–683.

269. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* **1977,** *23* (3), 327–341.

270. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R., Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995,** *103* (11), 4613–4621.

271. Loncharich, R. J.; Brooks, B. R.; Pastor, R. W., Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of *N*-Acetylalanyl-*N*'-Methylamide. *Biopolymers* **1992,** *32* (5), 523–535.

272. Hockney, R. W., The Potential Calculation and Some Applications. *Methods Comput. Phys.* **1970,** *9*, 136–211.

273. Im, W.; Lee, M. S.; Brooks III, C. L., Generalized Born Model with a Simple Smoothing Function. *J. Comput. Chem.* **2003,** *24* (14), 1691–1702.

274. Im, W.; Feig, M.; Brooks III, C. L., An Implicit Membrane Generalized Born Theory for the Study of Structure, Stability, and Interactions of Membrane Proteins. *Biophys. J.* **2003,** *85* (5), 2900–2918.

275. Chen, J.; Im, W.; Brooks III, C. L., Balancing Solvation and Intramolecular Interactions: Toward a Consistent Generalized Born Force Field. *J. Am. Chem. Soc.* **2006,** *128* (11), 3728–3736.

276. Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A., Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* **1999,** *101* (6), 426–434.

277. Metropolis, N.; Ulam, S., The Monte Carlo Method. *J. Am. Stat. Assoc.* **1949,** *44* (247), 335–341.

278. Hill, A. V., The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol.* **1910,** *40* (Suppl), iv–vii.

279. Xu, L.; Haworth, I. S.; Kulkarni, A. A.; Bolger, M. B.; Davies, D. L., Mutagenesis and cysteine scanning of transmembrane domain 10 of the human dipeptide transporter. *Pharm. Res.* **2009,** *26* (10), 2358-2366.

280. Smart, O. S.; Neduvelil, J. G.; Wang, X.; Wallace, B. A.; Sansom, M. S. P., HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **1996,** *14* (6), 354–360.

281. Yue, Z.; Bernardi, A.; Li, C.; Mironenko, A. V.; Swanson, J. M., Toward a Multipathway Perspective: pH-Dependent Kinetic Selection of Competing Pathways and the Role of the Internal Glutamate in Cl–/H+ Antiporters. *J. Phys. Chem. B* **2021**.

282. Liu, Y.; Li, C.; Gupta, M.; Verma, N.; Johri, A. K.; Stroud, R. M.; Voth, G. A., Key computational findings reveal proton transfer as driving the functional cycle in the phosphate transporter PiPT. *Proc. Natl. Acad. Sci. U. S. A.* **2021,** *118* (25).

283. Zhang, L.; Wang, H.; E, W., Reinforced dynamics for enhanced sampling in large atomic and molecular systems. *J. Chem. Phys.* **2018,** *148* (12), 124113.

284. Zhang, L.; Han, J.; Wang, H.; Saidi, W. A.; Car, R., End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems. *arXiv preprint arXiv:1805.09003* **2018**.