THE UNIVERSITY OF CHICAGO

MULTISCALE SIMULATIONS OF PROTON-COUPLED MEMBRANE TRANSPORTERS AND PROTON CHANNELS

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

YU LIU

CHICAGO, ILLINOIS

AUGUST 2023

To my family and friends

Table of Contents

List of Figures

List of Tables

Acknowledgements

I'd like to express my deepest thanks to Professor Gregory A. Voth, my Ph.D. adviser, for his unwavering support and exceptional academic guidance throughout my graduate studies. I cherished the liberty he gave me to explore entirely new protein systems within our research group. I deeply appreciate Greg's perspective that research is fundamentally about communication, and I have gained immensely from interacting with other researchers at various conferences.

My gratitude extends to Professor Robert M. Stroud and Dr. Meghna Gupta at the University of California, San Francisco, along with Professor Douglas Tobias and Dr. Alfredo J. Freites at the University of California, Irvine. Their intellectual contribution and dedication towards our shared projects elevated the quality of our work. I'm also grateful to Prof. Benoit Roux at the University of Chicago, Prof. Thomas E. DeCoursey at Rush University, Prof. Bjørn P. Pedersen at Aarhus University, Prof. Jessica Swanson at University of Utah, Dr. Santanu Roy at Oak Ridge National Laboratory, Dr. Gregory K. Schenter and Dr. Christopher J. Mundy at Pacific Northwest National Laboratory and finally Dr. Trung Nguyen at research computing center of University of Chicago. They all provided a unique perspective that greatly enriched the outcomes of my research.

It is my pleasure to have interactions with both past and present members of the Voth group, including Dr. Chenghan Li, Dr. Paul B. Calio, Dr. Xinyou Ma, Dr. Yuxing Peng, Dr. Laura Watkins, Dr. Joshua Zuchniarz, Dr. Zhi Wang, Mr. Vilmos Zsolnay, Ms. Sijia Chen, Ms. Jeri Beiter and Mr. Da Teng. Engaging with them in enlightening discussions on scientific topics, as well as enjoyable conversations about life and other subjects, has been an enriching and delightful experience.

Beyond my academic journey, I must express profound gratitude to my pillars of strength, my parents: Hong Yang and Haifeng Liu. You have been my unceasing cheerleaders through every

stage of my growth, showering me with an abundance of love and support at every turn. Your unwavering belief in my potential has been the driving force behind all my endeavors. I am truly blessed to have you both in my life.

Abstract

Proton transport (PT), defined by the migration of a delocalized positive charge, is a ubiquitous occurrence that plays an instrumental role in numerous physiological processes in biological systems, especially in proton-coupled phosphate transporters and proton channels. Due to its crucial contribution to the functioning of these membrane proteins, a deeper understanding of how protons couple with ligand transport in the proton-coupled phosphate symporter and how proton transport differs from other ion transports - thus ensuring exceptional selectivity in the proton channel - can illuminate the protein features that may trigger enhanced or impaired transport behaviors, as well as how diseases linked to these proteins can be treated.

Simulating the proton transfer (PT) process in pure solution is already a complex and nontrivial task. This problem becomes even more challenging within the context of protein systems. Molecular Dynamics (MD) simulation has emerged as a powerful tool for studying chemical reactions, especially PT in intricate protein environments. In MD simulations, PT is typically considered implicitly, as seen in simulations with empirical classical forcefields, or explicitly accounted for through hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) and the Multiscale Reactive MD (MS-RMD) method. It's crucial to remember that PT can be associated with changes in a protein's conformation, ranging from brief, localized adjustments taking only several nanoseconds to more extensive, global conformational changes that can span milliseconds to minutes.

The timespan of PT and its related conformational transition usually involve high free energy barriers, and as a result, can easily go beyond the limit of the vanilla unbiased MD approaches. For sufficient sampling of such rare events, enhanced sampling techniques are often required to accelerate key collective motions relevant to PT processes. The free energy profile obtained by

enhanced sampling methods offers the thermodynamics property of the studied system in principle. Additional kinetic information can be inferred based on either a transition state theory assumption, a Markov state model, or a bottom-up kinetic network model.

Here, we employed multiscale simulations in conjunction with enhanced sampling approaches to investigate the functional cycle of a proton-coupled phosphate symporter. We specifically focused on confirming the proposed functional role of a key residue, D324. By developing multiple possible reaction pathways within a bottom-up kinetic network model, we were able to account for different forms of phosphate. This approach allowed us to analyze the interaction between proton and phosphate gradients within each elementary reaction while making a prediction of the optimal pH condition for the phosphate uptake activity of this transporter under various phosphate concentrations.

A critical part of our research involved examining the treatment of transition rate theory in the proton transfer (PT) process. We scrutinized this across multiple protein systems where one-dimensional and two-dimensional free energy profiles were available. Consequently, we derived and analyzed the source of error in each treatment.

Moreover, we used MS-RMD simulations to predict the pH-dependent conductance behavior of a proton channel. The quantitative permeability analysis for proton, tetramethylammonium (TMA) cation and $CH_3SO_3^-$ anion proposed a perspective on the origin of the proton selectivity of the WT channel and the anion selectivity of the mutated channel D112N.

1 Introduction

1.1 Proton transport in biomolecular systems

Proton movement occurs through several pathways in proteins, including via water, titratable ligands, and ionizable residues. This transport process can be defined by the Grotthuss hopping mechanism, which is modified when one or more water molecules are replaced with titratable entities that can either accept or donate protons. Given its capacity to modify electrostatic and hydrogen bonding interactions, proton transport serves as a vital element within the functional cycles of a range of channels, enzymes, and transporters.[1-9]

The PiPT transporter, a proton-coupled phosphate transporter found in fungi, shifts phosphate ions following the direction of the proton gradient. This process allows for the accumulation of inorganic phosphate ions within the organism. While it's established that PiPT utilizes a proton gradient across the membrane to facilitate phosphate uptake, the intricate mechanics of the phosphate import, binding, and release processes are yet to be fully understood. Previous mutagenesis experiments have suggested that the titration of residue D324 has a pivotal role in phosphate transport. Evidence supporting this emerged from the D324N mutation (analogous to D358N in Pho84), which allows phosphate to bind in a typical manner but significantly curtails its transport efficiency. Further insight was gained from the crystal structure of PiPT, which depicted the transporter in a phosphate-bound, inward-facing occluded state. This structure demonstrated a close interaction between D324 and the bound phosphate at the binding site, implying that the phosphate could be released from the binding site upon D324's deprotonation. To thoroughly comprehend the role of specific residues, including the key residue D324, in the coupling of proton transport to phosphate symport, we must undertake multiscale molecular dynamics simulations. These simulations can expose potential protonation states of the protein and phosphate during each

phase of the functional cycle. Further exploration of the proposed sequential PT process that propels phosphate transport, through either hybrid quantum mechanics/molecular mechanics or multiscale reactive molecular dynamics, will be vital for deeper clarification of these mechanisms. Another important aspect to explore is how this transporter adapts its detailed functional pathway to environmental changes. For this, a bottom-up kinetic reaction network framework is necessary. This approach can provide a comprehensive molecular understanding of the diverse dynamic reaction pathways facilitating the transition from phosphate-bound OC states to phosphate-released IO states.

Human voltage-gated proton (hHv1) channels are crucial for regulating essential biological processes such as immune cell respiratory burst, sperm capacitation, neuronal excitability, and cancer cell migration. The central role of hHv1 in these processes is to rapidly extrude proton to perform pH regulation and remain electrical excitability. One of the most fascinating properties of this ion channel is its remarkable proton selectivity[10] when compared to other cations under depolarized membrane potential. The proton selectivity was demonstrated by the experimental measurement of reversal potential (V_{rev}, the membrane potential at zero current), that aligns with the Nernst potential of protons. Considering that protons are a million times less concentrated than other ions under physiological conditions, the remarkable proton selectivity observed necessitates a permeability ratio between proton over other ions ranging from 10^6 to 10^8.[11] Mutagenesis studies revealed D112 to be a crucial residue to the proton selectivity evidenced by the proton-specific selectivity shifting to the anion selective upon a site mutation on D112 to asparagine. There has been a longstanding debate on how this residue performs its selectivity filtration role and to make a conclusive statement on the underlying mechanism of such proton selectivity, yet the explicit

proton transport simulation is needed to make a conclusive remark on the functioning mechanism of selective filter residue D112 and to compare with other ion conduction.

1.2 Modeling explicit proton solvation and transport

In conventional molecular mechanics force fields (FF), the multifaceted interactions between atoms are typically decomposed into bonded and non-bonded potentials. The bonded interactions, which include bonds, angles, and dihedrals, are modeled using straightforward energy functions like springs and cosines. In this modeling framework, the bonding structure is predetermined and stays constant throughout the entire simulation. Consequently, this constrains its capability to simulate reactions that involve bond formation and breaking. While excess protons can be modeled as classical hydronium ions, this ad-hoc approach cannot accurately depict the process of charge delocalization or explicit proton transfer with a Grotthuss mechanism.

The unique Grotthuss nature of proton transfer (PT) makes ab initio molecular dynamics (AIMD) an appropriate choice for simulating this process, as AIMD explicitly calculates the electronic structure while evolving the nuclei on the potential energy surface (PES) shaped by the electronic ground state. Born-Oppenheimer (BO) approximation of this methodology assumes the adiabatic separation between nuclei and electrons. Often, density functional theory (DFT) under the generalized gradient approximation (GGA) framework is selected to balance between accuracy and computational efficiency. However, GGA can lead to an overestimation of intermolecular charge transfer, resulting in overly potent hydrogen bonds and underestimated proton transfer barriers. To rectify these inaccuracies, empirical dispersion corrections are typically incorporated into the functionals, serving to adjust the hydrogen bond strength and more accurately estimate PT barriers. To simulate the PT process in protein systems, hybrid quantum mechanics/molecular

mechanics (QM/MM) method with quantum mechanics treated with dispersion corrected GGA-level of DFT functionals is often employed to lower the computational cost.

Conventional QM/MM requires QM and MM atom identity to be predefined and stay constant during MD simulations, making it difficult to handle the situations in which the reaction center translocate a distance beyond the ranges of the QM box. This is often the case of PT between two distant reactive amino acid residues in a protein channel. The Multiscale Reactive Molecular Dynamics (MS-RMD) method, developed by the Voth group, presents an alternative approach. This method stands out due to its ability to simulate explicit proton transport on a large scale, a feat made possible by its computational efficiency.

The key feature of MS-RMD is to expand the ground state of the system on a basis of valence bond (VB) resonance states (diabatic states), and the Hamiltonian of the system can then be expressed as

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

Each diabatic state $|i\rangle$ corresponds to a distinct bonding topology and the diagonal energy h_{ii} can be directly derived from the employed classical FF. The off-diagonal energy h_{ij} demonstrates how easily can the proton transfer occur between states $|i\rangle$ and $|j\rangle$. In each timestep of the MS-RMD simulation, a state search algorithm is triggered to find all possible diabatic states that contribute non-negligibly to the Hamiltonian. The state search starts from the pivot central hydronium or protonated amino acid and goes up to three solvation shells to rearrange the bonding topology via a chain of proposed proton transfers. The ground state expression of the system in terms of the diabatic state representation is obtained by solving the eigen-value problem

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

where $\boldsymbol{c} = \{c_i\}$ is the ground-state vector associated with the lowest eigenenergy (ground-state energy). The ground-state eigenvector states the relative weights of each diabatic state that contributes to the adiabatic ground state. Thus, the ground state energy of the system can also be expressed as

$$E = \sum_{ij} c_i c_j h_{ij} \tag{1-3}$$

The forces can be calculated by the Hellman-Feynman theorem as

$$\boldsymbol{F} = \sum_{ij} c_i c_j \boldsymbol{F}_{ij} \tag{1-4}$$

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

1.3 Enhanced sampling techniques and collective variables

Due to the presence of substantial thermodynamic barriers (5 to 20 kcal/mol) associated with proton transfer between weak acids/bases and the water molecules that accept them, both QM/MM and MS-RMD necessitate the integration of free energy sampling techniques. This combination enables the creation of an accurate depiction of the proton transfer process, offering statistically significant results within a manageable simulation time frame. Techniques such as umbrella sampling and metadynamics are often used, relying on one or multiple pre-defined reaction coordinates, often termed "collective variables" (CV). The CV offers a mathematical means to measure the progression of the reaction. Typically, molecular dynamics sampling is employed to average the degrees of freedom orthogonal to the reaction coordinate into the free energy profile, otherwise known as the potential of mean force (PMF).

The underlying idea of the enhanced sampling approaches is to add bias either on force or on potential at the chosen degree of freedoms (DOFs) to bring down the free energy barrier to ensure sufficient sampling on the rare cases. The sampled ensemble with the bias in potential, $u(\xi)$, is

thus deviated from the Boltzmann distribution and a simple correction can be done to restore the unbiased observation,

$$\rho^{\text{biased}}(\boldsymbol{r}) = \rho(\boldsymbol{r})\, e^{-\beta u(\xi(r))} \qquad 1\text{-}5$$

To define a collective variable for a water-assisted PT process, the location of the transported Grotthuss hopping excess proton charge defect must be tracked. The concept of tracking an excess proton was initially formulated in the 1990s, focusing on the overall movement of the water chain's dipole. However, as later research revealed, this original proton tracker, which was defined as a linear combination of atomic coordinates, was significantly influenced by the permanent dipole movement of fluctuating water molecules far from the proton transfer active center. This influence led to it being deemed a less suitable proton tracker. A more recent approach to tracking the excess proton, widely referred to as the center of excess charge (CEC), has been proposed. All CEC definitions - MS-RMD CEC, rCEC, mCEC, and the "proton indicator" - utilize a non-linear form of the atomic coordinates. This allows them to capture the delocalized nature of the excess proton charge defect and the chemically reactive (Grotthuss hopping) nature of proton transfer, while reducing contamination from environmental water dipoles.

To detail the progression of proton transfer within a protein, the collective variable (CV) is generally defined as a function of the distance between the CEC position and a site of interest, such as a protein amino acid residue or substrate. Given the curvilinear expression of the CEC CV, all forms of proton transfer CV designed for water-assisted proton transfer do not correspond to a linear combination of the system's atomic coordinates. Other coupled DOF such as substrate transport, the PT pathway hydration process (noted earlier), and local protein conformational changes can also be slow and therefore critical to the PT process.[12-23]. In such cases, additional CVs are necessary to be explicitly expressed in the free energy sampling process to reflect coupling

of PT to those DOFs, along with the CEC describing the PT CV, thereby giving a more complex PMF for the coupled PT process having two dimensions or more.

1.4 Reaction kinetics: Transition state theory

Transition state theory is widely employed for estimating the transition rate of a reaction when combined with free energy sampling techniques. Conventional TST predicts an upper bound to the reaction rate of the system. The upper bound arises from the non-recrossing assumption of a hypersurface that divides the phase space into the reactant and the product regions, i.e. a successful reaction will take place once the system reaches the transition state.

The TST rate constant expression in a general form for the reaction CV, $\boldsymbol{\xi}$, and with a dividing surface described by $\boldsymbol{\xi}=\mathbf{s}(\mathbf{r})$ in the context of a PMF is given by

$$k_{TST} = \frac{1}{Q^R}\int \frac{d\mathbf{r}\, d\mathbf{p}}{h^{3N}}\delta(\boldsymbol{\xi} - \mathbf{s}(\mathbf{r}))\,\Theta(\dot{\xi}_{ver})\dot{\xi}_{ver}e^{-\beta H} \tag{1}$$

where $\mathbf{r}$ are the 3N atomic coordinates of the system, $\mathbf{p}$ are the conjugate linear momentum, N is the total number of the atoms in the system, and $\mathbf{s}(\mathbf{r})$ is the transition state dividing manifold. Q^R is the reactant partition function and can be calculated from the integral of the Boltzmann factor over the reactant phase space. Note that $\dot{\xi}_{ver}$ is the vertical velocity component of $\boldsymbol{\xi}$ to the transition ridge; then according to TST, this factor contributes to the forward reaction flux when found to be a positive value. This relationship is expressed by the step function $\Theta(\dot{\xi}_{ver})$ in the expression.

There are two critical components in the TST formalism: the first is the exponential dependence of the reactant minimum to transition state free energy barrier, while the second is the pre-factor related, in essence, to the velocity of the effective mass of the CV as it crosses the dividing surface.

1.5 Outline of thesis

The work presented in this thesis uses reactive MD approaches, including QM/MM and MS-RMD, combined with enhanced free energy sampling, to study the proton-phosphate coupled

transport in PiPT transporter and the proton conduction and selectivity mechanism in hHv1 channel. The remaining chapters are organized as follows:

Chapter **Error! Reference source not found.** states the classical MD and QM/MM simulations confirming the key role of residue D324 to phosphate release and propose a broad picture on the functional cycle of the phosphate transporter PiPT. Chapters 3 proceeds to examine each elementary reaction for multiple reaction pathways from the phosphate-bound occluded state to the phosphate released inward open state within a bottom-up kinetic network model and analyzes the interaction between proton and phosphate gradients within each elementary reaction while making a prediction of the optimal pH condition for the phosphate uptake activity of this transporter under various phosphate concentrations.

Chapter 4 uses MS-RMD simulations to predict the pH-dependent conductance behavior of a proton channel, hHv1. The quantitative permeability analysis for proton, TMA+ cation and $CH_3SO_3^-$ anion proposed a perspective on the origin of the proton selectivity of the WT channel and the anion selectivity of the mutated channel D112N.

Chapters 5 include advances in examining the treatment of transition state theory in the proton transfer (PT) process. Various implementation of transition state theory were scrutinized across multiple protein systems where one-dimensional and two-dimensional free energy profiles were available.

2 Key computational findings reveal proton transfer as driving the functional cycle in the phosphate transporter PiPT

This chapter was reprinted with permission from *Proceedings of the National Academy of Sciences* 2021, 118 (25), e2101932118.

2.1 Introduction

Inorganic phosphate (Pi) is indispensable to life by virtue of its ubiquitous involvement in metabolic processes. The cell uses phosphate transporters to import Pi with the energy provided by either proton or sodium electrochemical potentials across the membrane[24]. There has been an increasing interest in high-affinity proton-coupled phosphate transporters which play an essential role in protozoan parasites[25, 26], in breast cancer[27] and bone reabsorption[28] in human health. Transporters in the phosphate proton symporter (PHS) family use a proton gradient as the driving force to import phosphate across the plasma membrane of plant cells and fungi. PHS, a family within the major facilitator superfamily[29] (MFS), can provide a necessary perspective on how proton transport (PT) facilitates the substrate transport during the functional cycle of proton-coupled symporters.

The *Piriformospora indica* phosphate transporter (PiPT), purified from a eukaryotic fungus, is the only transporter in the PHS family for which a crystal structure has been determined [9]. The stoichiometry factor H+: Pi of the PHS family varies between 2:1 ~ 4:1[30-33]. PiPT shares high homology with the *Saccharomyces cerevisiae* high-affinity phosphate transporter Pho84 (stoichiometry factor H^+: Pi 3:1) and with plant phosphate transporters inside the PHS family. It is a member of the SLC22 family of human organic anion and cation transporters and is homologous to the SLC2 family of glucose facilitative transporters (GLUTs) and the related synaptic vesicle 2 protein (SV2) family of the MFS[9]. The only structure of the PHS family to date (Figure 2-1A) is of a phosphate-bound, ‘inward-facing occluded’ state and reveals phosphate binding residues

together with a cytosolic proton exit tunnel emanating from the binding site. This structure confirms that PiPT, like other transporters in the MFS, has two homologous domains (N and C domains) each with 6 transmembrane helices (TM) related by a quasi-two-fold symmetry perpendicular to the membrane plane. Based on the structure, the Stroud group[9] postulated that D324 de-protonation alone triggers phosphate cytosolic release while D45 and D149 are part of the proton exit tunnel (shown in Figure 2-1A). Mutagenesis of the highly homologous yeast Pho84 [33-35], where the D324N (D358N in Pho84) mutant binds phosphate with normal affinity but significantly reduces its transport ability, indicates that D324 is essential for both phosphate and proton transport. The D149N/E (D178N/E in Pho84) mutants reduce half of the phosphate uptake activity while keeping normal binding affinity, which indicates that D149 can be a good candidate residue for proton transfer. The crystal structure of PiPT thus provides a snapshot of the transporter.

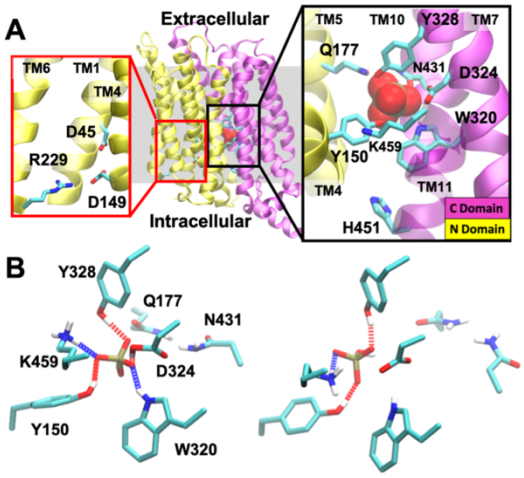


Figure 2-1. PiPT crystal structure and phosphate binding interactions. A) PiPT crystal structure (PDB 4j05) with enclosed phosphate binding (black box) and proton transfer residues (red box). B) Representative phosphate binding interactions in the occluded state (left panel) and the inward open state prior to phosphate exit (right).

In the present work, we have sought to determine how the proton relay occurs and how it is coupled to the phosphate transport. A previous computational analysis[33] of yeast Pho84 starting

from an inward open homology model based on the glycerol 3-phosphate transporter GlpT as template, postulated that the de-protonation of D149 accompanies the conformational transition from the occluded state to the inward-open state while the Y150 (Y179 in Pho84) sidechain acts as its intracellular gate. That computational work protonated only one residue, D149, to form a hydrogen bond with D45 (D76 in Pho84) and found that the docked phosphate remains stable for 30 ns at a putative binding site at least 5 Å further to the cytoplasmic side than in the crystal structure, and that upon protonation of D149, Y150 swings back to its occluded orientation as in the PiPT crystal structure.

It is known that PiPT uses a proton gradient across the membrane to uptake phosphate; however, the finer details of the phosphate import, binding, and release mechanism remain to be defined. Hence with the goal of defining which residues govern the coupling of proton transport to phosphate symport, classical molecular dynamics (MD) simulations were run in the present work starting from the occluded state seen as in the crystal structure of PiPT. Multiple protonation assignments were applied to key residues in order to stabilize the occluded (OC) structure and identify those that produce outward open (OO) and inward open (IO) conformations. These classical MD simulations provided insight into relative key residue titrations that induce transporter conformational change and phosphate release. In conjunction with enhanced free energy sampling (umbrella sampling), hybrid quantum-mechanics molecular-mechanics (QM/MM) was then employed to quantitively define the proton transfer pathways from D324 at the phosphate binding site to D45 both in wild-type PiPT and in the D45N mutant designed to block the pathway through the cytosolic proton exit tunnel. These results demonstrate that the D324 de-protonation event triggers the functional cycle of the transporter through simulation of explicit proton transfer from D324 to the proton exit tunnel QM/MM MD. These simulations yield

the novel result that the phosphate titratable substrate itself participates directly in the PT process in PiPT.

2.2 Methods

2.2.1 Simulation Debrief

Classical MD simulations were run for the PiPT crystal structure [Protein Data Bank ID code 4j05] embedded in a solvated dimyristoylphosphatidylcholine (DMPC) lipid bilayer to simulate the biological environment. The disordered residues from 230 to 296 between N- and C-domains in the crystal structure were not simulated, however we assume this does not affect the function of PiPT within the MFS based on the Pho84 and LacY deletion experimental results [36-39]. The models were built in the CHARMM-GUI web server [40, 41] following standard equilibration protocols. Protonation states were assigned based on PROPKA[42] calculations. The CHARMM-CMAP and CHARMM 36 force field[43-45] were employed to describe the protein, lipid and ligand interactions. Simulations for the choices of the protonation states of the system with phosphate release were replicated (from the same starting structure but with different velocity assigned) three times in GROMACS software [46]. Principal component analysis on the trajectories is shown in Figure 2-13. QM/MM MD simulations for the wild-type PiPT were initiated from a structure with D324 protonated, D45 deprotonated and H451 protonated, equilibrated in classical constant NPT simulations for at least 100 ns. Similarly, the D45N mutant was simulated with the same D324, H451 protonation states. The QM/MM simulations were performed in the CP2K package [47]. The QM region was described by density functional theory (DFT) with the BLYP functional [48, 49] and the D3 dispersion correction [50, 51]. The free energy profile (PMF) for proton transport from D324 to D45/N45 were then calculated with QM/MM umbrella sampling by tracking the excess proton

center of excess charge [52] (CEC) relative to phosphate along the direction from D324 to D45 side-chain oxygens. The proton-transfer rate constant was then estimated using transition state theory.

2.2.2 Classical Molecular Dynamics Simulations

Starting from the crystal structure [Protein Data Bank (PDB) code 4j05], the classical molecular dynamics (MD) simulations were performed for distinct protonation states assigned wild-type or D45 mutant PiPT embedded in a pre-equilibrated and hydrated 1,2-Dimyristoyl-sn-glycero-3-phosphocholine (DMPC) lipid bilayer with approximately 150 mM of NaCl. We capped N- and C- domains of proteins with ACP and CT3 and built phosphate models with Avogadro software version 1.2.0 [53]. TIP3P water model was used [54]. All interactions were described by the CHARMM36 force field with CMAP correction [40, 41, 43-45]. Models were built and equilibrated using a standard protocol in CHARMM-GUI [41] and simulations were performed in GROMACS [46, 55] version 2019.4 with the Velocity Verlet integrator in the isothermal-isobaric (constant particle Number, Pressure, Temperature (NPT)) ensemble using a semi-isotropic Parrinello-Rahman barostat [56] at 1 atm and a velocity rescale thermostat [57] set to 310° K. Electrostatic interactions were calculated using the particle mesh Ewald method with a cutoff of 12.0 Å and a precision of 10^{-5} was used to treat the electrostatic interactions. The Lennard-Jones (LJ) interaction was set to cutoff to 0 at 12.0 Å radius and used a switching function starting at 10.0 Å to smooth the truncation. The interaction energies in Table 2-1 are calculated for stabilized structures over consecutive time frames in each system by GROMACS.

2.2.3 QM/MM MD Umbrella Sampling

The potential of mean force (PMF) for an excess proton passing the phosphate binding region from D324 to D45 was calculated with hybrid QM/MM MD simulations, using the CP2K package [47]. The proton transfer reaction coordinate, also called collective variable (CV), is defined as the

projection of the vector pointing to the center of excess charge (CEC; defined below) from the phosphate heavy atom center of mass (COM) onto the vector connecting D324 and D45 carboxylic COMs. The excess proton CEC coordinate was defined in [52] as follows,

$$\boldsymbol{r}_{\mathrm{CEC}} = \sum_{i=1}^{N} c_i^2 \boldsymbol{r}_i^{\mathrm{COC}} \qquad 2\text{-}1$$

Here, i denotes the index of diabatic states, corresponding to different bonding topologies by assigning the excess proton to a different species, which is the phosphate, the aspartic acids or solvation water molecules. The $\boldsymbol{r}_i^{\mathrm{COC}}$ is the center of charge of the protonated species in diabatic state i, specially the protonated phosphate, the protonated aspartic acids or the hydronium. The c_i^2 measures the population of the system in each diabatic state, i.e. the probability of the excess proton belonging to each possible proton acceptor species. We follow the definition of c_i^2 in ref [52],

$$\frac{c_j^2}{c_i^2} = \exp -k\left(\delta_{ijl} - \delta_0\right) \qquad 2\text{-}2$$

where $\delta_{ijl} = r_{jl} - r_{il}$, is the proton sharing coordinate, and r_{jl} is the distance between the proton acceptor oxygen in state j and the transferring proton K; r_{il} follows the same meaning except of referencing the proton acceptor in state i. The parameter $k = 7.34\ \text{Å}^{-1}$ and $\delta_0 = 0.293$ Å were chosen for aspartic acid; $k = 4.98\ \text{Å}^{-1}$ and $\delta_0 = 0$ were chosen for water and phosphate following the reference[52]. Then the c_i^2's are computed by applying the normalization condition $\sum_i c_i^2 = 1$.

The initial structure for the QM/MM simulations was chosen from representative snapshots after at least 100 ns classical MD equilibration of OC protonation-state system or from nearby windows. The QM region was chosen to contain phosphate, phosphate binding residues (Y150, Q177, W320, D324, Y328, N431 and K459) and a few more nearby residues including D45, F369, F327, F426,

F435. Water molecules within at least 4 Å of all these residues are included as QM atoms. The QM region atoms were updated about every 2 ps to avoid missing any water molecules within 4 Å of the QM described residues stated in the SI. There are always more than 210 QM atoms during the QM/MM simulation and the phosphate is always sufficiently solvated in the water at the binding site. The QM box size was chosen to be 6 Å larger than the range of the QM atoms in all dimensions. The electronic structure of the QM region was treated at the density functional theory (DFT) level using the BLYP functional and the D3 dispersion correction [48-51]. The Goedecker-Teter-Hutter (GTH) pseudopotentials [58] were used to describe the core electrons and the Kohn-Sham orbitals were expanded in the Gaussian TZV2P basis set and plane wave basis set with a cutoff of 360 Ry. The classical region of the system was described by the CHARMM36 forcefield and the QM-MM electrostatic coupling was treated by the Gaussian expansion of the electrostatic potential (GEEP) algorithm with a periodic boundary condition [59-61]. Constant temperature ensembles used the Velocity Verlet integrator and a Nosé–Hoover thermostat at 310 K.

Around forty-five windows with a general spacing of ~0.25 in CV were constructed for all umbrella sampled wild-type PiPT and D45N mutant systems. Additional windows with shorter distance along CV dimension were included to obtain better overlap between nearby windows around the steep-slope region in PMFs. All windows were simulated for at least 20 ps and the collective variable was collected at every time step (0.5 fs). The PMF was constructed by the method in dTRAM [62] instead of that in WHAM [63] to avoid the assumption of a global equilibrium and the independency between time frames in each umbrella window. Block averages were calculated over the last four out of five blocks to give a rough estimation of the statistical error in collective variable distribution in each window. We used transition state theory to estimate the rate

constant of the proton transfer process from D324 to D45(N). The calculation of the rate constant[64] from the 1D PMF is shown as

$$k_{TST} = \left\langle \left(\frac{Z_\xi}{2\pi\beta}\right)^{\frac{1}{2}} \right\rangle_{\xi=s} \frac{e^{-\beta W(s)}}{\int_{-\infty}^{s} d\xi\, e^{-\beta W(\xi)}} \qquad 2\text{-}3$$

where Z_ξ is the inverse of the affective mass of the collective variable ξ at transition state s, β is the inverse value of $k_B T$, and W(ξ) is the PMF value corresponding to ξ.

2.2.4 Cloning, expression, growth and phosphate transport analysis

PiPT was cloned into p112A1NE vector and the mutants (D45E and D45N) were made using a 'Quick Change XL' site-directed mutagenesis kit (Stratagene, La Jolla, CA, U.S.A). Yeast has two high-affinity transporters- Pho84 that is optimally active at pH 4.5 and driven by protons, and Pho89 optimally active at pH 9.5 and driven by Na^+ [65, 66]. We therefore used a *Saccharomyces cerevisiae* strain, MB192 in which Pho84 was deleted (pho84Δ), for heterologous expression of PiPT for functional analysis [67], and performed phosphate transport assays at pH 4.5. The tryptophan auxotroph MB192 strain was transformed with either p112A1NE (Vector only), or PiPTWT, or PiPTD45E or PiPTD45N and the transformed strains used for growth and phosphate transport assays as described in [68]. For growth assays, yeast strains were grown in 5 ml Yeast Nitrogen Base (YNB) media without tryptophan at 30°C overnight. The cells were pelleted down and washed twice with saline (0.9% sodium chloride solution) and re-suspended in 1 ml saline then adjusted to 0.3 $O.D._{600nm}$. 10X serial dilutions were made and 5 µl spots were placed on YNB (without Trp) high- (HPi) and low-phosphate (LPi) agar plates. The cell densities in LPi conditions are most sensitive and proportional to PiPT activity. In general, all the MB192 transformants even with vector alone grow in HPi conditions. The plates were incubated at 30°C for 72 h and imaged.

For transport, PiPT strains were grown overnight in LPi medium at 30°C. Phosphate transport activity was assayed using carrier-free [^{32}P] Pi. The stock concentration of [^{32}P] Pi was 0.75 µM with 5000 Ci/mmole activity when received from BARC, Trombay, India. 9.5 nM of [^{32}P] Pi was mixed with non-radioactive (cold) KH_2PO_4 solution to prepare final concentration of Pi of 20 µM, 50 µM, 100 µM, 150 µM, 200 µM, 400 µM. Cells were washed pelleted and washed with buffer T (25 mM Tris Succinate pH 4.5 and 3% (w/v) glucose) at room temperature. 50 µl of this cell suspension was mixed with 50 µl phosphate solutions (triplicates at each concentration) to achieve 10 µM, 25 µM, 50 µM, 75 µM, 100 µM and 200 µM final concentrations. After 2 min of incubation at room temperature, for each concentration, cells were rapidly washed three times on Whatman GF/F filter paper with ice-cold Buffer T. The [^{32}P] radioactivity retained on the filter papers was counted using liquid scintillation spectrometry in a Packard Tri-Carb 2100 TR Scintillation Counter (Perkin Elmer Life Sciences, Boston, MA, USA). The total counts/min was plotted versus concentration of Pi outside the cells. The readings were measured in 3 different cell growths. The empty vector p112A1NE was transformed into MB192 as the negative control. GraphPad Prism 6 was used to generate the plots from which apparent K_m and V_{max} were determined by regression against the Michaelis-Menten equation.

2.3 Results and Discussion

2.3.1 Classical Simulation Suggests Sequential PT Process

Classical simulations indicate that the D324 de-protonation event triggers phosphate cytosolic release, in which multiple proton transfer events proceed in sequence during the transition from the occluded to the inward conformation.

In order to analyze the PT events in the PT pathway correlate with the conformational transitions, we ran extensive classical MD simulations (see Methods) of 8 combinations of possible

protonation states of titratable residues around the phosphate binding site and compared these between different conformations. First, we sought to determine protonation states of key residues based on the crystal structure of PiPT by calculation with PROPKA[42], and reran the system for multiple OC conformations. D324 was assigned as protonated due to its close contact with negatively-charged Pi (Figure 2-1 A, B). Residue D45, the closest protonatable residue to the phosphate binding site (Figure 2-1A), was assigned as protonated. This agrees with the Pho84 study[33] where protonated D149 (D178 in Pho84) forms a hydrogen bond with D45, causing the original IO structure to adopt the OC conformation with the gating Y150 orientation restored. Other residues were assigned protonation states based on their calculated pKa values assuming neutral pH conditions. With this protonation state assignment, both mono- and di-basic Pi were found to interact favorably with K459, D324, Y328 and W320, and the OC structure was stable for more than one microsecond in our classical MD simulations (Figure 2-1B left panel, Figure 2-2 A,F).

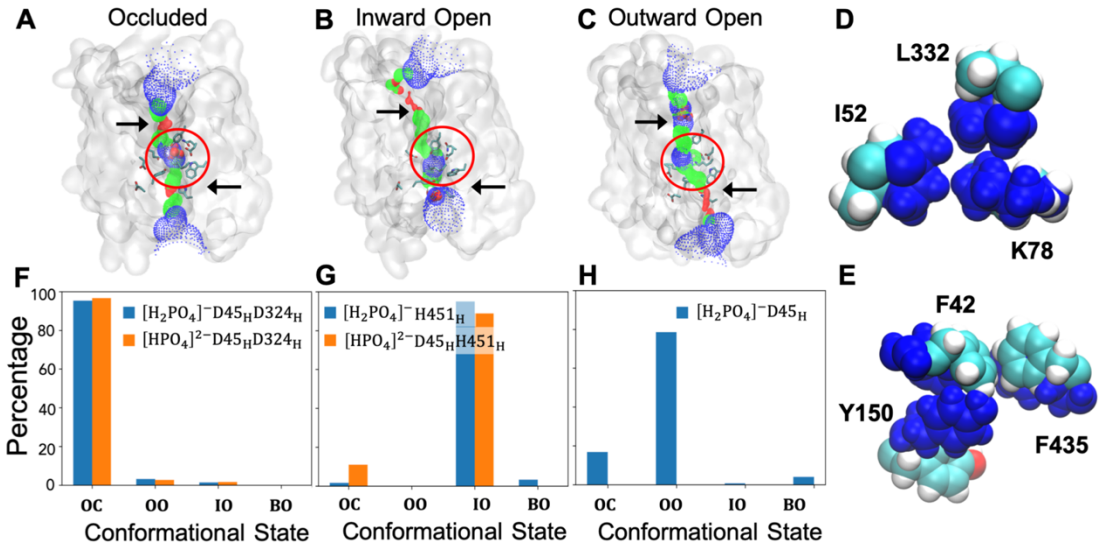


Figure 2-2. Population density of states for given combinations of protonation states. Representative structures in the Occluded (A), Inward-Open (B) and Outward-Open (C) states. The phosphate binding site is circled in red and pore radius along the axis perpendicular to membrane plane is plotted in color red for smaller than the radius of a water molecule (1.15 Å), blue for larger than the ionic radius of phosphate (2.38 Å), and green in between. D) Extracellular gate, and E) Intracellular gate; open in light colors vs. closed in dark blue. Arrows indicate the positions of the gates in the structure. F, G, H show percentage of time spent in each conformation over time period of ~1 microsec total MD simulation, listed as occluded (OC), outward open (OO), inward open (IO) and both gates open (BO). Simulations were for monobasic (blue) or dibasic Pi (orange), for protonation states of residues indicated in each panel. Quantitative measures of Different conformational states are defined in Supplementary Information Fig. S3. All residues shown are fully conserved in PHS family.

Based on mutagenesis of Pho84 [33, 34] where D324N mutants bind normally but cannot release Pi due to the inability of N324 to de-protonate, we assigned D324 as de-protonated and tested whether this change in D324 protonation alone triggers the conformational transition to the IO state. To our surprise, the PiPT structure transformed to the OO state instead (shown in Figure 2-2 D, E). After careful scrutiny of that structure, we found that H451 (assigned as de-protonated in our previous structures) is the other key residue to phosphate cytosolic release; H451 is the only basic residue beneath the phosphate binding site and above the intracellular edge of the protein and is fully

conserved in the entire PHS family. When H451 was assigned as being protonated alongside a de-protonated D324, we observed the expected overall conformational transition from OC to IO state with Pi released into cytosol on a fairly rapid timescale (less than 10 ns, Figure 2-2 B, G).

Consequently, we looked back at the protonation state previously assigned that stabilized the OC structure and reassessed whether it is possible for H451 to be protonated in the OC structure. Further simulations demonstrated that changing only the H451 protonation state does not make a difference – the OC structure is stable for more than one microsecond with Pi remaining inside the binding site (for conformational population distribution, see Figure 2-11 of the Appendix section). This finding correlates with our understanding of the transition from the OF state to the IO state. The OO state requires H451 to be deprotonated for the bottom of the Pi binding gate to close, but during the conformational transition to the IO structure, H451 can abstract a proton from nearby water in the Pi binding site and anchor the direction of Pi release. This eliminates the possibility for the Pi to be released prior to D324 deprotonation event where one may expect Pi to be released with D324, D45 and H451 all assigned as protonated in the wildtype PiPT.

Upon de-protonation of D324 and prior to Pi release, both forms of phosphate lose their previous close contact with D324 and W320 (Figure 2-1B right panel). The contrast between the interaction energies of D324 and W320 with Pi in the OC and IO states shows that the de-protonation of D324 is enough to repel Pi out of the binding site (Table 2-1); the protonated form of D324 is required to hold Pi at the binding site. The titration of H451, in contrast, has negligible interaction energy differences with Pi, because of the relatively large distance (~9 Å) between them. This result confirms the hypothesis[9] that D324 de-protonation is crucial for Pi release. The missing negatively charged residue at the corresponding location of D324 in both of the organic cation and organic

anion SLC22 family transporters[69-71] explains why these transporters cannot utilize proton gradient and use sodium gradient instead.

The missing hydrogen bond between Pi and the N7H of W320 prior to Pi release in the IO state suggests that W320 is another binding residue important to keeping Pi in the OC state. This is probably due to the proximity of W320 to D324 on TM7 and the participation of its bulky side chain in the nearby hydrophobic intracellular gate. The SLC22 family transport anions, cations, and zwitterions. The highly conserved W320 and Y328 residues in the SLC22 family transporters [9] suggests a similar substrate-binding mechanism particularly using W320 TM7.

The dominant conformational form and relative probabilities were determined for each protonation state of residues and of the substrate, by MD simulations for ~1 microsecond each. The open/closed nature of the extracellular, and intracellular gates, as depicted in Figure 2-2, were defined quantitatively (Figure 2-10) and used to quantitate the percentage of time spent in different conformational states. The relative populations show that once D324 is deprotonated but H451 remains neutral, monobasic Pi is repelled by negatively-charged D324 and no longer restrained by H451, and it exits to the extracellular space. This again confirms that the protonated form of D324 is essential to retaining Pi at the binding site. The intracellular gate is composed of bulky hydrophobic residues Y150, F42 and F435 beneath the Pi binding site, that are fully conserved across the PHS family. Among these, Y150 most clearly participates in gating the channel in the MD simulations. In IO states, deprotonated D324 repels Pi and pushes it from the top of Y150, while protonated H451 beneath Y150 pulls Pi from below. Y150 then alters its side-chain orientation to give way to Pi and open the sealed hydrophobic gate (shown in Figure 2-2E). Our findings confirm the role of Y150 as a key part of the gate. Mutations [33] that are similarly large

hydrophobic Y150F or polar Y150S support transport of phosphate, while small hydrophobic Y150A, Y150G mutants do not presumably due to some structural collapse in the latter cases.

2.3.2 QM/MM Simulations and D45N Mutagenesis Suggests Exit PATH of D324 Proton

QM/MM simulations together with D45 mutagenesis find that the exit path of the D324 proton requires D45. During this process, Pi acts as a "transfer station" in between these two aspartic acids.

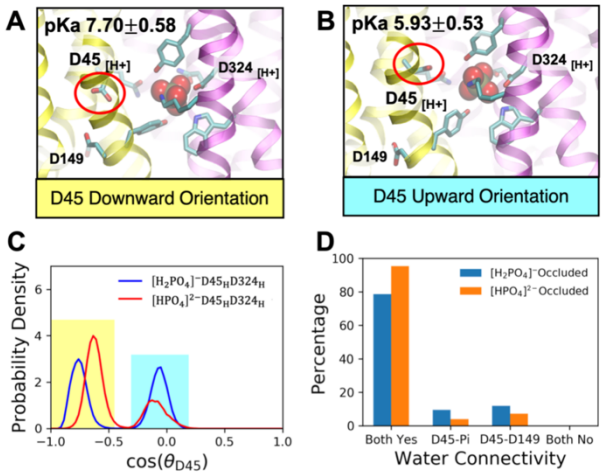


Figure 2-3. D45 as a proton transfer residue. A, B) Typical D45 orientations in the occluded state structure. C) The bimodal distribution of D45 side-chain orientation in the occluded states. cos (θ_{D45}) quantitively describes D45 side-chain orientation where negative value represents for D45 side chain pointing below the xy plane. The angle $\theta_{D45}$is defined as the angle between z-CA-CG. D) Water connectivity probability between D45-phosphate and D45-D149 in OC states.

While the departure of the proton from D324 triggers Pi release and conformational transition (see above), D45 participates in transferring this proton to the cytosol through the PT tunnel as suggested by the pKa calculation of D45 in our OC state classical MD simulations. We observed two typical orientations of protonated D45 in the OC state simulations shown in Figure 2-3 A and B. Protonated D45 can either orient downward to form a hydrogen bond with D149 or orient upward, pointing to Pi. Interestingly, in apo OO state simulations, the same protonated D45 always

hydrogen bonds with D149 and never flips up toward the phosphate binding site. To understand the correlation of proton transfer and D45 side-chain behavior, we calculated pKa of D45 with PROPKA separately on frames of OC state trajectories displaying different orientations of the D45 side chain. The predicted pKa of D45 with either downward- or upward-oriented side chain corresponds to 7.7±0.6 and 5.9±0.5, respectively. These results suggest that the D45 side chain orients downwards in the OO state and once Pi has been imported to the binding site, D45 transfers its proton to cytosol and then its side chain reorients towards Pi for re-protonation from the Pi. As shown in Fig.3D, the water connectivity between D45 side chain and Pi, D45 side chain and D149 illustrates that pathways of hydration are maintained throughout. This supports the proposal of the PT tunnel that allows transport of the proton from Pi to D45 and eventually to cytosol via D149 and R229.

To fully address questions such as where the D324 proton transports and whether Pi titration happens during the process, we explored this problem in a more quantitative manner and calculated the free energy profile (potential of mean force, or PMF) for the PT process from D324 to D45 using umbrella sampling free energy sampling [72] in with QM/MM MD simulations [60, 61]. The reaction coordinate, also called the collective variable (CV), was defined as

$$\mathrm{CV} = \vec{r}_{\mathrm{CEC-Pi}} \cdot \frac{\vec{r}_{\mathrm{D324-D45}}}{|\vec{r}_{\mathrm{D324-D45}}|} \qquad 2\text{-}4$$

where CEC is the excess proton center of excess charge [52] located originally on D324, and Pi is the center of mass of the phosphate. The D324 and D45 subscripts in that CV equation denote the center of mass of the carboxylic three terminal heavy atoms of the side chain. The primary advantage of this reaction coordinate (CV) design is that it does not bias the titration of Pi while still permitting water titration within four solvation shells of the CEC [52].

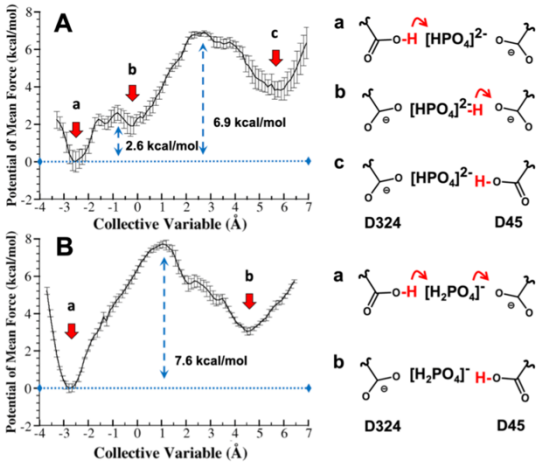


Figure 2-4. Potential of mean force (PMF) from QM/MM MD for the proton transfer process from D324 to D45 mediated by phosphate at the binding site. Vertical red arrows represent thermodynamic stable states depicted on the right column. The activation free energy barrier along the reaction coordinate (Collective Variable, CV) is marked in blue.

For the initial QM/MM umbrella sampling window we chose representative snapshots from the classical MD simulations with protonated D324 and deprotonated D45 assigned. Even though the CV does not drive Pi titration in either direction, we still observed explicit proton transfer from D324 to both forms of Pi and eventually to D45 via Grotthuss proton shuttling through water molecules present in wild-type PiPT after ~100 ns equilibration. This result is fascinating in the sense that, to our knowledge, no previous transporter simulation has observed substrate titration itself as being crucial to enable the functional cycle. It is thus possible that other transporters of titratable substrates might also change their protonation state and therefore their chemical and physical properties midway through the transport process. The excess proton remained on D324 in the lowest free energy state, shown as *state a* for both forms of Pi in Figure 2-4. Then, when the proton is transferred to Pi, the previously dibasic phosphate becomes mono-basic and contributes to a local free energy minimum at *state b* in the upper panel. The previously mono-basic phosphate, on the other hand, transforms to H_3PO_4 and this emerges as the transition state of the PT process. The proton comes to the side chain of D45 in both systems via shuttling through water molecules,

and this forms the corresponding local free energy minimum in both PMFs. The overall proton transfer process in both mono- and di-basic phosphate systems has a kinetic barrier of around 7 kcal/mol, and the time constant (inverse of the rate constant) for both processes is thus less than 5 ns, which matches the classical MD computational observation of the phosphate release time scale in the IO state. The overall barrier to PT through dibasic Pi is slightly lower than that for monobasic Pi. These results therefore support the proposed phosphate release steps depicted in Figure 2-5.

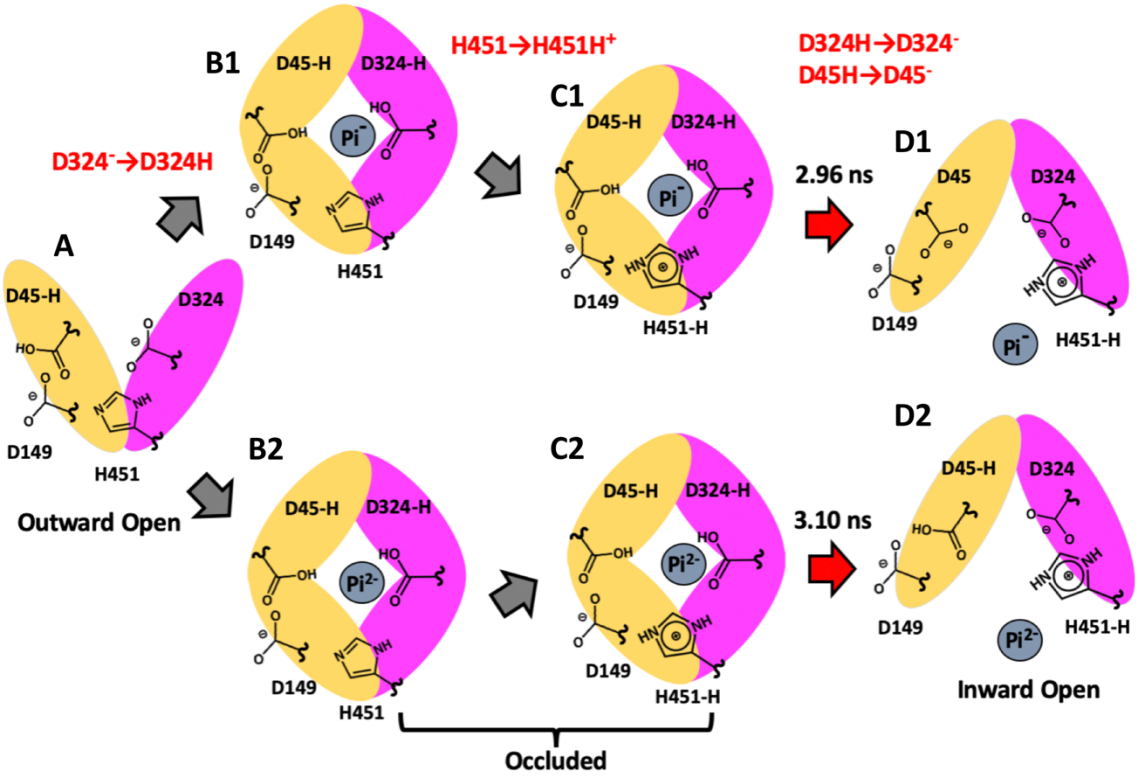


Figure 2-5. Deduced mechanism of how proton transfers couple conformational changes to phosphate transport: A: stabilized outward open state. B1, 2: stabilized occluded states either with monobasic or dibasic phosphate in the primary binding site. C1, 2: intermediate states lead to protonation of H451. D1, 2: stable inward open state accompanies phosphate release (either monobasic or dibasic), and deprotonation of D324. Arrows represent inferred proton transfer processes from classical MD simulations while the ones in red have the proton transfer timescale calculated as the inverse of the transfer rate constant calculated by transition state theory and QM/MM MD PMF calculations.

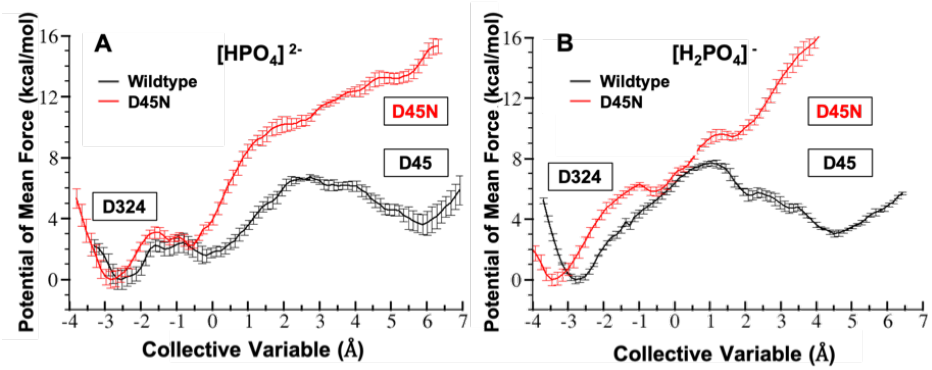
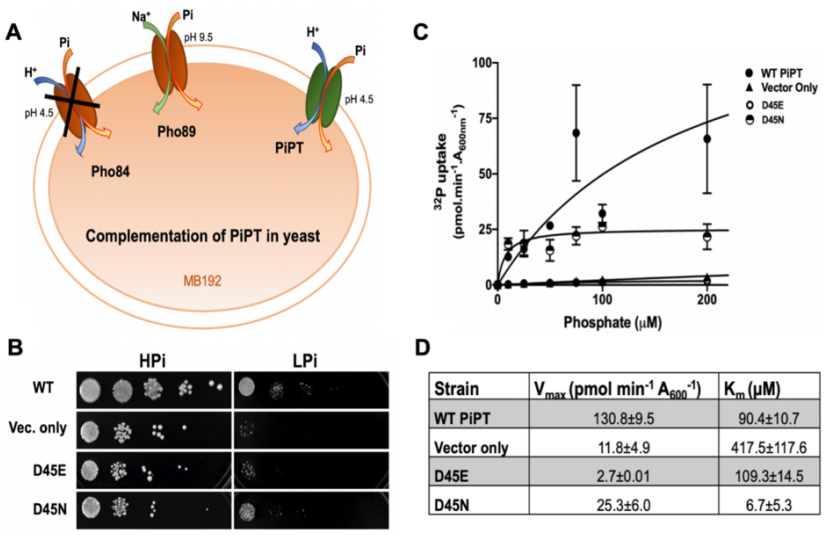


Figure 2-6. QM/MM MD potential mean force (PMF) comparison for the proton transfer process from D324 to D45 in wildtype PiPT (black) and D45N mutant (red). A) Proton transfer PMF for dibasic phosphate at the binding site; B) PMF for mono-basic phosphate at the binding site. Detailed definitions see Appendix section.

In order to further validate the role of titratable D45, we performed another parallel set of QM/MM MD umbrella sampling simulation for the D45N mutant, which mimics the irreversibly protonated D45, and compared the corresponding PMFs with those of the wild-type (see Figure 2-6). In contrast to the barrier height of the WT PMFs in Figure 2-4, those of the D45N mutant have barriers higher than 15 kcal/mol and the mutant PMFs show no apparent local minimum in the vicinity of the D45N side chain (< 2Å). The shapes of the PMFs of both systems are similar during the PT process from D324 to both forms of phosphate, but they show a major difference in the subsequent step of the mechanism where the proton is passed to D45(N) via water proton shuttling. Because of the non-protonatable nature of the N45 residue, there is no local free energy minimum at N45 (contrasted with protonated D45) in the mutant PMF. These two findings support our hypothesis based on classical MD simulations that D45 is the starting point of the PT tunnel and should be deprotonated first to receive the proton from D324. The proton transfer tunnel lining along the direction of the proton gradient across the membrane, with D149 immediately transferring a proton from D45 to bulk, can provide a constant driving force for the excess proton leaving D324 moving towards D45.



Strain	V_{max} (pmol min^{-1} A_{600}^{-1})	K_m (μM)
WT PiPT	130.8±9.5	90.4±10.7
Vector only	11.8±4.9	417.5±117.6
D45E	2.7±0.01	109.3±14.5
D45N	25.3±6.0	6.7±5.3

Figure 2-7. Mutagenesis identifies D45 as a key proton transfer residue in the functional cycle of PiPT. A) A pictorial representation of high-affinity phosphate transporters in *S. cerevisiae*. Pho84 is absent in MB192 and complemented by PiPT WT or the D45 mutants. B) Growth assay showing serial dilutions (from O.D.$_{600nm}$ 0.3 and 10X serial dilutions) of the cells in high- (HPi) and low-phosphate (LPi) condition on minimal media agar plates. C) Michaelis-Menton curves for the phosphate transport assay of PiPT D45E and D45N. D) Values of K_m and V_{max} derived from curve fitting (GraphPad Prism 6) are tabulated.

Experimental mutagenesis followed by growth (Figure 2-7) also demonstrates the essential nature of D45 for phosphate release. The D45N mutant has ~10 fold lower Km, and 5 fold lower Vmax compared with wildtype PiPT showing that the non-titratable D45N mutant binds phosphate, but does not transport it as well. The mutant D45E that preserves the charge allows the same K_m, but lowers the Vmax by a further 10 fold (50 fold below wild type). Classical MD simulations of D45E with corresponding protonation state assignments of the WT occluded states show that it induces a structural collapse (Figure 2-12). A new turn is formed in helix 4 below D149 (in dark blue) that severely disrupts the phosphate binding environment compared to WT and D45N mutants which may explain the diminished phosphate uptake rate in D45E mutant.

Based on our classical and QM/MM MD simulations, here we propose the overall proton transfer pathway through conformational transitions as follows: D45 is initially protonated. During phosphate import from the apo-OO state to the OC state, one proton binds to D324 to hold either

form of Pi at the binding site and another proton protonates H451 that stabilizes the negative charge on Pi (state A to state Cs in Figure 2-5). Phosphate release begins with the deprotonation of D45 to the intracellular side. The proton of D324 moves to the phosphate and eventually is passed to D45 (state Cs to state Ds). Due to the proton gradient across the membrane, this second proton on D45 exits the PT tunnel to reach the cytosol either before the cytosolic departure of mono-basic Pi or after the cytosolic departure of dibasic Pi. The proton on H451 leaves to the intracellular side upon recycling of the transporter from IO to OC, proceeding on to the OO state. As a result, the stoichiometry factor H^+: Pi for PiPT functional cycle is estimated to be 3:1, which fits into the range of stoichiometry factor experimentally observed across the entire PHS family.

Here, we need to clarify that it is, ultimately, the balance of electrochemical potentials at either side of the membrane that drives the direction of the functional cycle. When we claim that proton transfer reactions drive the functional cycle, we are stressing the coupling relationship between proton and phosphate transport during the phosphate cytosolic release. We did assume the reversibility of the phosphate extracellular release during the conformational transition from the OC to the OO state. Thus, we marked the reverse process in Figure 2-6 and inferred the protonation of D45 and D324 upon the import of phosphate. We believe that a co-dependence of proton and phosphate transfer can exist while the electrochemical potential difference from both sides of the membrane shifts sign. A 2D PMF with one CV describing the proton transfer process and the other CV defining the phosphate release could fully address this issue. However, due to the computationally demanding nature of QM/MM, this task is exceedingly hard to achieve.

2.4 Conclusions

In this work, we justify at a detailed molecular level the hypothesis that the deprotonation of the key residue D324 triggers phosphate release. Our work combines results from classical MD

simulations and QM/MM MD free energy sampling to successfully answer the questions of where and how the proton on D324 travels vectorially across the membrane during the phosphate release process. The quantitative QM/MM PMFs confirm D45 as the starting residue of the PT tunnel and offer a new perspective on proton-coupled transporters where titratable substrates such as phosphate may be explicitly involved in the PT process as a key part of driving the functional cycle. Our design of appropriate collective variable together with QM/MM MD simulations proves to be critical to the study of such complicated systems with multiple titratable sites along the transport pathway.

2.5 Appendix

Table 2-1. The calculated phosphate binding interaction energy in occluded (OC) and inward-open (IO) structures prior to phosphate release. These results indicate binding energies that favor release of Pi in the IO conformation. They also show minimal impact on binding energy from the H451 (protonated in OC, unprotonated in IO).

Interaction Energy (kJ/mol)	OC H_2PO4^-	OC HPO_4^{2-}	IO $H_2PO_4^-$	IO HPO_4^{2-}
D324-Pi	- 46 ± 3	- 68 ± 3	- 26 ± 2	- 10 ± 0.8
W320-Pi	- 52 ± 5	- 79 ± 0.4	8 ± 0.8	31 ± 2
H451-Pi	0.01 ± 0.01	0.08 ± 0.02	- 3 ± 0.6	- 2 ± 0.2

Table 2-2. Sequence similarity of PiPT to homologues from the PHS Family.

PHS Homologue	Acc. Num.	E Value	Bit Score	Per. Ident
Pho84(S.cerevisiae)	P25297	9.00E-109	338	38.63
AtPht1;6(A.thaliana)	Q9ZWT3	2.00E-83	271	33.79
LjPht1;4(L.japonicus)	B5RHV8	6.00E-82	267	33.4
OsPht1;4(O.sativa)	Q8H6H2	2.00E-81	266	33.58
OsPht1;2(O.sativa)	Q8GSD9	2.00E-80	263	34.11
OsPht1;13(O.sativa)	Q7XRH8	3.00E-80	262	33.4
PhPht1;5(P.hybrida)	B2CPI7	1.00E-79	261	33.71
OsPht1;12(O.sativa)	Q8H074	2.00E-79	261	33.27
OsPht1;7(O.sativa)	Q8H6G9	3.00E-79	260	33.96
OsPht1;5(O.sativa)	Q7X7V2	8.00E-79	259	33.52
PhPht1;4(P.hybrida)	B2CPI6	9.00E-79	259	32.82
PhPht1;3(P.hybrida)	B2CPI5	2.00E-78	258	34.44
OsPht1;11(O.sativa)	Q94DB8	3.00E-78	258	33.4
OsPht1;3(O.sativa)	Q7XDZ7	6.00E-78	257	32.31
AtPht1;1(A.thaliana)	Q8VYM2	1.00E-77	256	33.52
OsPht1;6(O.sativa)	Q8H6H0	9.00E-77	254	34.03
AtPht1;5(A.thaliana)	Q8GYF4	2.00E-76	253	32.89
OsPht1;8(O.sativa)	Q8H6G8	3.00E-76	253	34.35
AtPht1;3(A.thaliana)	O48639	1.00E-75	250	33.46
AtPht1;4(A.thaliana)	Q96303	4.00E-75	249	32.76
AtPht1;7(A.thaliana)	Q494P0	1.00E-74	248	32.58
PhPht1;1(P.hybrida)	A7KTC5	2.00E-74	248	32.38
AtPht1;8(A.thaliana)	Q9SYQ1	2.00E-74	247	32.16
AtPht1;9(A.thaliana)	Q9S735	7.00E-74	246	32.39
MtPht1;4(M.truncatula)	Q8GSG4	1.00E-73	245	32.95
OsPht1;1(O.sativa)	Q8H6H4	3.00E-69	234	33.86
AtPht1;2(A.thaliana)	Q96243	4.00E-68	231	33.72
MtPht1;8(M.truncatula)	G7KDA1	9.00E-67	228	31.34
OsPht1;10(O.sativa)	Q69T94	9.00E-66	225	30.86
OsPht1;9(O.sativa)	Q8H6G7	3.00E-63	219	29.68

Pairwise sequence similarity as E Value and Scores in Bit are calculated using BLASTP[73, 74] program with PiPT (Acc. Num. A8N031) as the query sequence among the swissprot Non-redundant UniProtKB/SwissProt sequences database of size 476336.

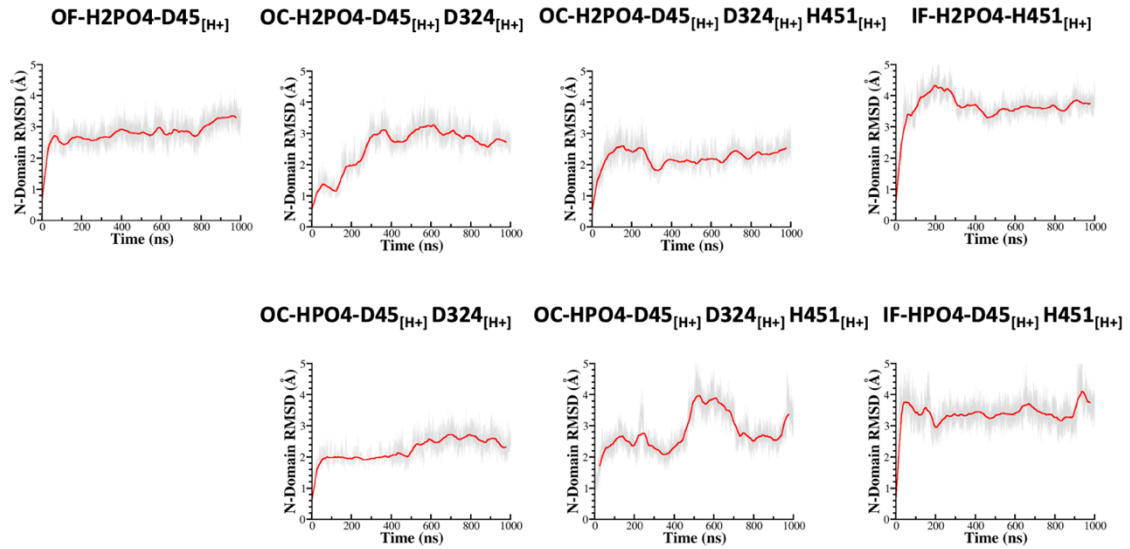


Figure 2-8. Equilibration of various PiPT systems showing the RMSD of the transmembrane backbone of the N-Domain (N, C_α, and carbonyl C), over 1 microsecond equilibration time. The N-Domain backbone RMSD was referenced to the C-Domain backbone as in the starting crystal structure to define domain motion associated with alternate access.

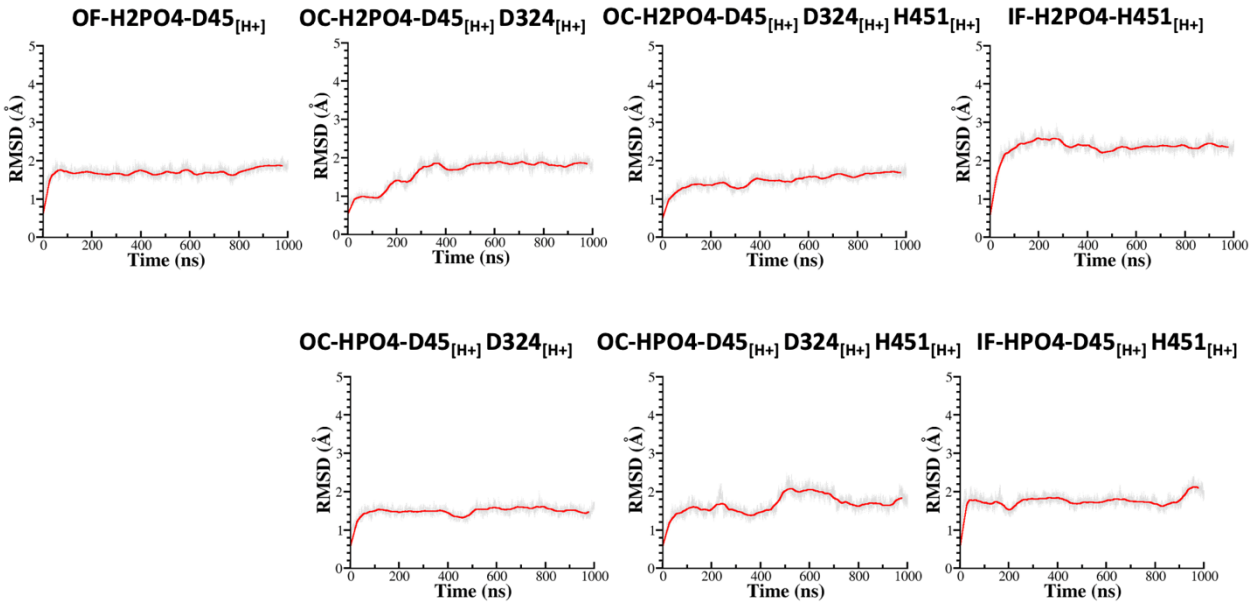


Figure 2-9. Equilibration of various PiPT systems showing the RMSD of the transmembrane backbone of the whole protein (N, alpha C, and carbonyl C) referenced to the crystal structure, over 1 microsecond classical MD simulation time.

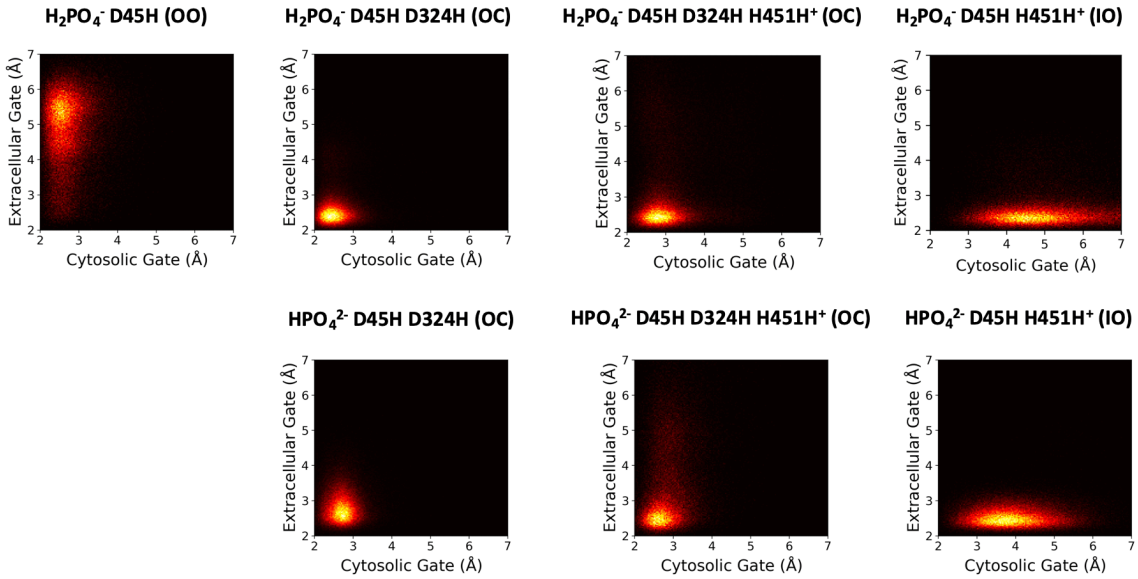


Figure 2-10. The 2D histograms of the extracellular gate and the cytosolic gate measured distance (the shortest distance between any two atoms of different gating residues: extracellular gating residues L49, I52, L332, K78; cytosolic gating residues F42, Y150, F435). The 2D histogram is calculated for the equilibrated structures produced by classical simulations of the different protonation states as indicated above the panels. The hotspots in 2D histograms located at the bottom left correspond to the occluded state (OC), at the upper left to outward open (OO), at the bottom right to the inward open (IO) state. The dominant conformational state classified according to the histogram for each protonation state is specified in the parentheses as OO, OC and IO.

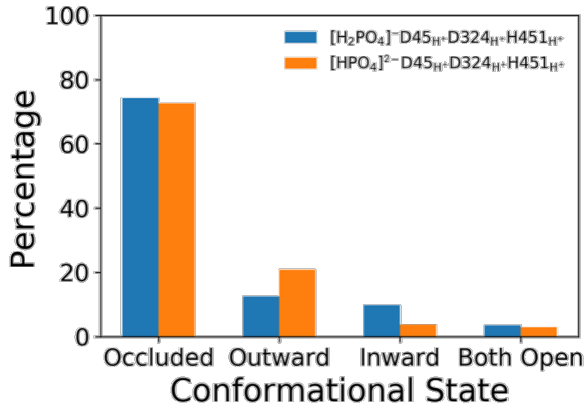


Figure 2-11. Protonation of H451 still favors the occluded state. Percentage of time spent in each of the major states reported in Fig. 2 after equilibration, and over the timeframe of the two classical simulations. The protonation states used in the simulations are indicated in blue for monobasic Pi and in orange for dibasic Pi.

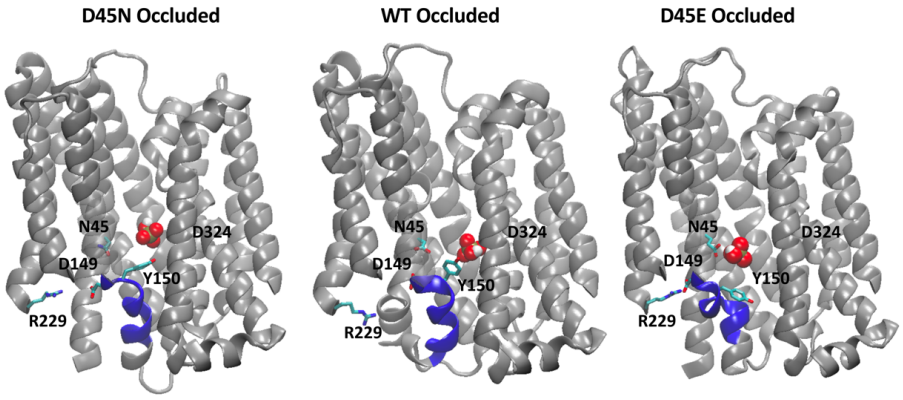


Figure 2-12. Representative configurations of the D45N and the D45E mutants on either side of the wildtype in classical simulations. The three simulations all used the same protonation state that stabilize the occluded conformation in Fig. 2 (HPO_4^{2-} and $D45_{[H+]}$ $D324_{[H+]}$). The residues D45, D149, R229 and D324 are shown as sticks. The protein backbone is rendered in grey while residues 149 to 158, which transit to a special turn in the D45E simulation, are in dark blue. Helix 2 and part of helix 3 are removed for clarity of the proton exit tunnel behind them.

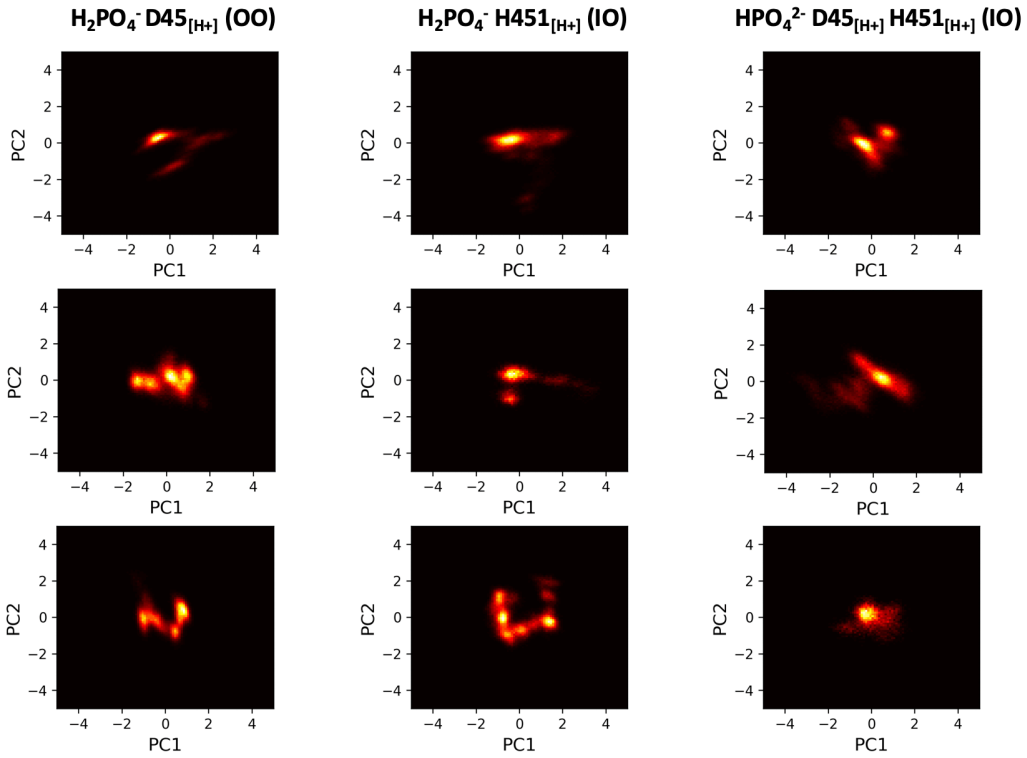


Figure 2-13. Principal component analysis shows the similarities and differences of the conformational distribution along the first two components for phosphate release cases. All trajectories were aligned with the crystal structure according to the C domain alpha carbons of the protein.

3 Kinetic network modeling combined with enhanced free energy sampling unveils a dominant reaction pathway within the proton-coupled phosphate symporter PiPT

3.1 Introduction

Inorganic phosphate (Pi) is essential for all forms of life, given its widespread involvement in the numerous metabolic processes. Cells import Pi using phosphate transporters, leveraging the energy from proton or sodium electrochemical gradients across the membrane[24]. Increasing attention has been given to proton-coupled phosphate transporters, due to their critical roles in protozoan parasites[25, 26], breast cancer[27], and human bone absorption[28]. Transporters in the phosphate-proton symporter (PHS) family, a family within the major facilitator superfamily (MFS), utilize a positive proton gradient across the plasma membranes to drive phosphate import of plant cells and fungi. Examining a transporter from the PHS family can provide significant insights into the molecular mechanisms of how substrate transport is linked with proton transport (PT) during its functional cycle.

The *Piriformospora indica* phosphate transporter (PiPT), purified from a eukaryotic fungus, holds the unique distinction of being the only transporter in the PHS family with an available crystal structure[9]. PiPT shares high homology with all transporters inside the PHS family[9, 75], including the *Saccharomyces cerevisiae* yeast phosphate transporter Pho84. The resolved structure[9] captures PiPT in a phosphate-bound, "inward-facing occluded" conformational state and confirms that PiPT, sharing the quasi-twofold symmetry feature of MFS transporters, has two homologous domains (N and carboxyl domains) with six transmembrane helices each(Figure 1). Combined evidence from mutagenesis studies[33-35] of highly analogous Pho84 transporter and structural data[9] from the phosphate binding site strongly suggest a pivotal role for the aspartate 324 (D324) residue in both phosphate release and proton coupling process.

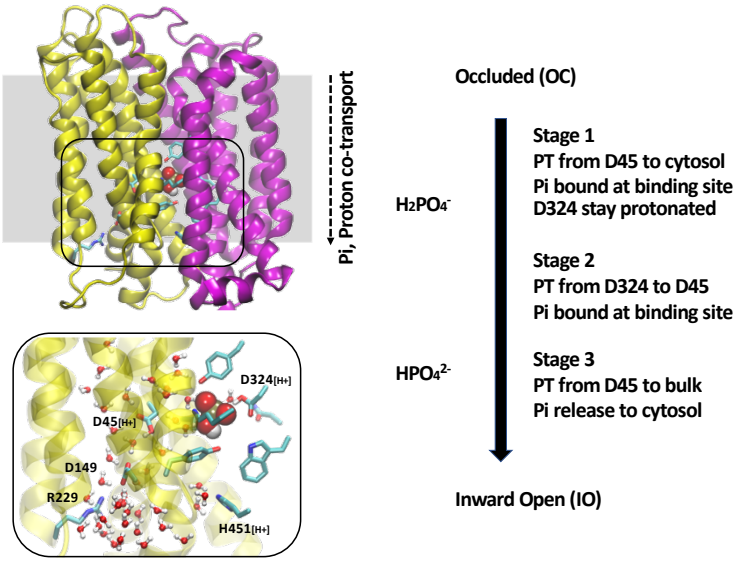


Figure 3-1. PiPT representative occluded structure with phosphate bound at the binding site. The zoom-in box shows the phosphate binding site with the proposed proton exit tunnel containing residue D45 and D149. The right panel depicts the previously proposed three-stage scheme of the transformation from the occluded (OC) state to the inward-open (IO) state.

Previous computational work, complemented by mutagenesis examinations of a key residue D45 inside the proton exit tunnel, suggested a probable proton exit path following its departure from D324 - a trigger for phosphate release from the binding site into the cytosol[75]. This was explored via hybrid quantum mechanics/molecular mechanics (QM/MM) molecular dynamics (MD) combined with enhanced sampling approach to study the proton transport event from D324 to D45. Drawing on insights from classical simulations and the free energy profile calculation of the PT event, our preceding work suggested a comprehensive overview of the PiPT functional cycle. This encompassed a three-step transition (Figure 3-1 right panel) from a phosphate-bound occluded state (OC) to a phosphate-released inward-open (IO) state. Initially, the process starts from the occluded (OC) state, with a proton leaving D45 and dispersing into the cytosol, setting up an attractive trap for another proton. The second stage involves the transportation of a proton from D324 to the now unoccupied D45. This transportation involves the titration of phosphate and water

molecules located at the phosphate binding site. In the final stage, both the phosphate and the newly transported proton on D45 exit into the cytosol.

Nonetheless, several questions and uncertainties persist. For instance, is our presumption concerning the initial transition step—where a proton can depart from D45 and reach the cytosolic bulk—entirely accurate? Additionally, does the phosphate remain bound at the binding site while the proton is being transported to D45 during the second transition step? Given the multiple potential protonation forms of phosphate binding in the OC state, numerous reaction pathways could be established from the OC state to an IO state. So, which is the dominant one, and how does it react to varying pH and phosphate concentration conditions? Arguably, the most challenging question to address is what make pH 4.5 the optimal condition for the phosphate uptake activity of PiPT? Our aim in this work is to answer all these questions with a bottom-up kinetic network model combined with enhanced free energy sampling and multiscale reactive MD (MS-RMD)[76, 77].

3.2 Methods

3.2.1 Classical Molecular Dynamics

Classical MD simulations were run for the PiPT crystal structure (Protein Data Bank ID code 4j05) embedded in a pre-equilibrated and solvated 1,2-Dimyristoyl-sn-glycero-3-phosphocholine lipid bilayer with approximately 150 mM of NaCl to simulate the biological environment. Models were built and equilibrated using a standard protocol in CHARMM-GUI[41] and simulations were performed in GROMACS[46] version 2019.4 with the Velocity Verlet integrator in the isothermal-isobaric (constant particle Number, Pressure, Temperature (NPT)) ensemble using a semi-isotropic Parrinello-Rahman barostat (10) at 1 atm and a velocity rescale thermostat set to 310° K. The

assigned protonation state of each system was based on the previously published observation for an equilibrated OC or IO state. Simulation protocol can be found in the previous literature[75].

3.2.2 Multiscale Reactive Molecular Dynamics

Starting from the equilibrated monobasic and dibasic phosphate-bound occluded structures in classical MD, the MS-RMD simulations were performed to simulate the explicit proton release process through the proton exit tunnel. Both Asp residues, D45 and D149, were modelled as potential protonation sites. All interactions were described by the CHARMM36[45] force field with CMAP correction with the developed Asp MS-RMD model. Simulations were performed by LAMMPS MD package coupled with RAPTOR to enable the PT reactions. MD was integrated with time step set to 1 fs in the canonical (constant particle Number, Volume, Temperature (NVT)) ensemble using a Nose–Hoover chain thermostat at 310 K and with an orthogonal box of previously equilibrated dimension (74.239 Å, 74.239 Å, 112.739 Å) for monobasic phosphate OC structure and (75.485 Å, 75.485 Å, 109.501 Å) for dibasic phosphate OC structure. Electrostatic interactions were calculated using the particle-particle particle-mesh method with the precision of 10-4 in force. The Lennard-Jones (LJ) and coulombic interactions was set to zero at 10.0 Å radius and utilized a switching function starting at 8.0 Å to smooth the truncation.

3.2.3 Umbrella sampling

A total of over 40 umbrella windows were established for the one-dimensional umbrella sampling of the PT process in OC states with a spacing of 0.5 Å in the reaction coordinate value. Each umbrella window was run for at least 1 ns and the total simulation time exceeded 40 ns. To restrain the lateral diffusion of the hydrated proton when completely dissociated from the protein, a harmonic potential was introduced to the reaction coordinate $r_{\perp}$ defined as the radial distance of the transported proton to D149 carboxylate atoms. The force constant of the wall was set to 5

kcal/mol/Å^2 and the restraining potential was switched on once $r_\perp \geq 10$ Å in umbrella windows where $z_H \geq 6$ Å.

A total of over 60 umbrella windows were established for the one-dimensional umbrella sampling of the phosphate process in IO states with a spacing of 0.5 Å in the reaction coordinate value. Each umbrella window was run for at least 5 ns and the total simulation time exceeded 300 ns. To restrain the lateral diffusion of the phosphate when completely dissociated from the protein, a harmonic potential was introduced to the reaction coordinate $r_\perp$ defined as the radial distance of the phosphate to the center of PiPT transmembrane backbone atoms. The force constant of the wall was set to 5 kcal/mol/Å^2 and the restraining potential was switched on once $r_\perp \geq 10$ Å in umbrella windows where $z_H \geq 20$ Å.

The weighted histogram analysis method[63] (WHAM) was employed to combine umbrella sampling trajectories and calculate the PMF for both PT and phosphate release process. PMF error bars were obtained by partitioning the trajectories of all umbrella windows into 6 equally sized blocks and calculating the standard deviation using the last 5 blocks.

3.3 Results

3.3.1 Examination of the PT from D45 to cytosol in the first transition step.

To calculate the free energy profile of the PT process from D45 to cytosol, MS-RMD simulations combined with enhanced sampling approach was employed. The starting two OC states have both D45 and D324 protonated and the only difference between is the protonation form of the bound phosphate at the binding site. The reaction coordinate z_H to describe the proton cytosolic release is defined as $z_H = z_{D149} - z_{H^+}$, where $z_{H^+}$and z_{D149} are the z axial coordinates of the transported proton and the carboxylate group accordingly. Here, D45 and D149 are the only two acidic residues located inside the proton exit tunnel (Figure 3-1). The proton resides on D45 corresponds

to the free energy minimum at around -5Å and as the proton being transported towards the cytosolic pore, the value of reaction coordinate z_{H} increases (Figure 3-2 A, B). The free energy profile (also known as potential of mean force, PMF) reaches a plague around 7Å where the transported proton dissipated fully into the bulk (Figure 3-3 A, C). A cylinder restraint was applied once proton reach to the less restricted bulk region to ensure a well-defined one-dimensional (1D) PMF. The transition rate constant of each process was estimated based on the transition state theory and the calculated rate constants for PT from D45 to bulk with a mono-/di-basic phosphate bound at the binding site are $(6.1 \pm 5.4) \times 10^6\ \mathrm{s}^{-1}$ and $(7.4 \pm 5.1) \times 10^5\ \mathrm{s}^{-1}$ respectively. The pKa of the residue D45 with mono- and di- basic phosphate bound in the OC states are 5.2 ± 0.5 and 5.7 ± 0.4 respectively based on the expression[76-78],

$$\mathrm{p}K_{\mathrm{a}} = \lg\left(c_0 S_u \int_{\mathrm{Reactant}} \mathrm{d}z_{\mathrm{H}}\ e^{-\beta(W(z_{\mathrm{H}})-W(+\infty))}\right) \qquad 3\text{-}1$$

where c_0 is the standard state concentration (1 M) expressed in number density of value $1/1660\ \text{Å}^3$, $W(z_{\mathrm{H}})$ is the 1D PMF value, and $W(+\infty)$ is the value when z_{H} is sufficiently large such that the PMF profile reaches a plateau. S_u term corrects for the effect of introduced cylinder constraint on the proton transport horizontal plane, and can be expressed as

$$S_u = \int_0^{\infty} \mathrm{d}r_{\perp}\, 2\pi r_{\perp} e^{-\beta U_{res}(r_{\perp})} \qquad 3\text{-}2$$

, where $r_{\perp}$ is the radial component of the distance between the transported proton and protein.

The backward reaction rate constant can be obtained by the relationship between the equilibrium constant (K_{a}) expression as follows,

$$K_a = \frac{[\mathrm{H}^+]_i[\mathrm{D45}^-]}{[\mathrm{D45H}]} = \frac{k_f}{k_b} \qquad 3\text{-}3$$

, where k_f is the forward rate constant and k_b is the backward rate constant. The first-order backward reaction rate constant was then estimated to be the product of the second-order backward rate constant and the cytosolic proton concentration. The free energy barrier for the proton to depart from D45 in the OC state is about 9~10 kcal/mol and the forward reaction timescale estimated is on the order of hundred nanoseconds.

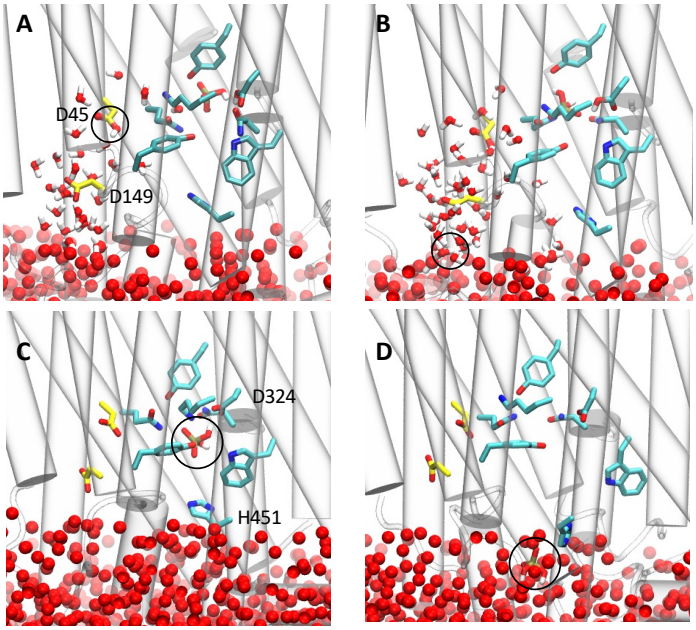


Figure 3-2. Representative Molecular Configurations for Proton and Phosphate Release in PiPT (with Monobasic Phosphate as a Reference). A, B) The PT process from D45 to cytosol with respect to proton residing on D45 (reaction coordinate of value -5Å in Figure 2A) and proton released into cytosolic bulk (reaction coordinate of value 4Å). The transported proton was circled and the water molecules within the proton exit tunnel and in the bulk region were showed in sticks and balls. C, D) The phosphate release process with respect to phosphate residing at the binding site (reaction coordinate of value 8Å in Figure 2B) and phosphate released into cytosolic bulk (reaction coordinate of value 20Å). The transported phosphate was circled and only bulk water molecules were shown to mark the protein-water interface.

3.3.2 Examination of the Phosphate release with classical MD PMF

In our previous work[75], two inward open states were found by assigning protonation state to PiPT protein and equilibrate the system with classical MD where phosphate managed to be released to the cytosol within tens of nanosecond. We inferred based on the classical observation

of the assigned protonation states in this previous study that for the cytosolic release of a dibasic phosphate, right after the PT process from D324 to D45, the dibasic phosphate can leave the binding site and be released into the cytosol. In contrast to the dibasic phosphate, a monobasic phosphate due to its smaller repel force against the negatively charged residue D324, cannot leave the binding site unless D45 transports the proton to cytosol and becomes unoccupied by proton again.

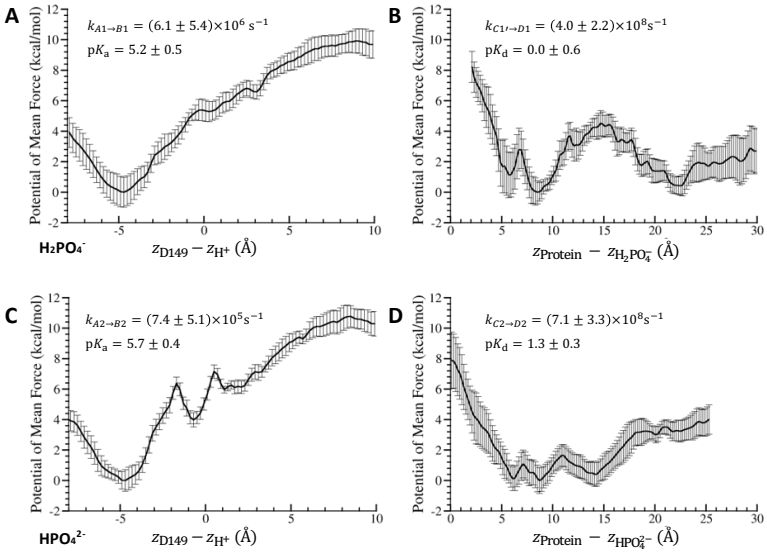


Figure 3-3. PiPT proton and phosphate release PMFs. A, C) The free energy profile of the PT event from D45 to cytosol at the phosphate bound OC states where D324 is protonated. B, D) The free energy profile of the phosphate release event from binding site to cytosol at the two proposed IO states (D324 deprotonated for both cases; D45 deprotonated for monobasic phosphate release and D45 protonated for dibasic phosphate release). The forward reaction rate constants were calculated based on transition state theory and the equilibrium constant were calculated.

Here, we employed classical MD simulations combined with enhanced sampling method to evaluate the three cases mentioned above (D45H [D324]$^-$ $H_2PO_4^-$, [D45]$^-$ [D324]$^-$ $H_2PO_4^-$, and D45H [D324]$^-$ $[HPO_4]^{2-}$) with the 1D PMF calculation. The reaction coordinate z_{P_i}, defined as $z_{P_i} = z_{PO_4} - z_{Prot}$, is used to demonstrate the phosphate cytosolic release process. Here, z_{PO_4} is the z axial coordinate value of the center of mass of the non-hydrogen atoms in the corresponding

protonation form of phosphate ion and the z_{Prot} is the z coordinate value of the geometric center of the transmembrane backbone of PiPT. The free energy basin in the 1D PMFs with $z_{\mathrm{P_i}}$ ranges from 3Å to 11Å in all 1D PMFs corresponds to the phosphate binding at the binding site (Figure 2C) and the free energy basin with $z_{\mathrm{P_i}}$ value at around 13Å corresponds to the phosphate being a close contact with the positive charged residue H451 (Figure 3-3 B, D and Figure 3-6). The phosphate reaches to cytosolic bulk water once fully depart from the attractive residue H451 with $z_{\mathrm{P_i}}$ value greater than 18Å (Figure 3-2D) and the PMF profile comes to a plague after 25Å for monobasic phosphate cases and after 20Å for a di-basic phosphate case. The forward reaction rate constants were calculated based on the transition state theory approach for the three phosphate cytosolic release cases and the equilibrium constant of the phosphate release process pK_{d} was calculated following the procedure of the PT from D45 to bulk. The forward rate constant k and the equilibrium constant pK_{d} were calculated and the timescale of phosphate release for these three cases can be estimated as the inverse value of the forward reaction rate constant calculated from the 1D PMFs. Both the dibasic phosphate release time scale with protonated D45 and monobasic phosphate release time scale with deprotonated D45 are less than 10 ns (Figure 3-3 B, D) and the monobasic phosphate release timescale with protonated D45 exceeds 10 microseconds (Figure 3-6). The calculation result matches with our previous published observation[75] that only the first two cases allow for phosphate release.

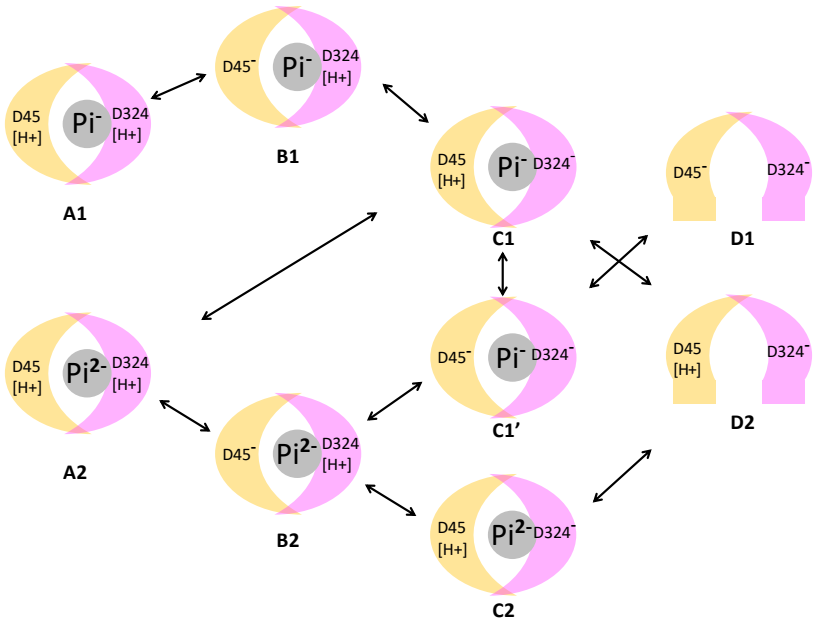


Figure 3-4. Bottom-up kinetic network model. The double-sided arrow represents the elementary reactions calculated. The grey solid circle stands for the mono- and di- basic phosphate bound at the binding site.

3.3.3 Bottom-up kinetic network model scheme

With all PMFs available for each elementary step of the bottom-up kinetic model, first-order reaction forward and backward rate constants were calculated based on transition state theory[64, 79] and detailed balanced assumption. We explored the impact of the three different factors extracellular pH (pH_o), cytosolic phosphate concentration (Pi_i) and extracellular phosphate concentration (Pi_o) on the network reaction flux from the OC to the IO state while the cytosolic pH (pH_i) is fixed to a value of pH 7.0. The two different OC state populations were determined by equilibrating the kinetic network model under the extracellular pH and Pi concentration. The three different IO state populations were determined by equilibrating the kinetic network model under the cytosolic pH and Pi concentration. The populations of the rest of the states were then calculated by solving the transition rate constant matrix with the fixed reactant and product populations.

Figure 3-4 shows the complete kinetic network model for the entire functional cycle of PiPT with an assumption of a fast equilibrium between an outward open phosphate-uncaptured state

(not shown) and the phosphate-bound occluded states. The states A1 and A2 corresponds to the previous mentioned OC states, D1 and D2 corresponds to the phosphate-released IO states. The first stage of the PT process from D45 to cytosol corresponds to elementary reaction from A1 to B1 and from A2 to B2. The second stage of the PT process from D324 to D45 via the titration of both forms of phosphate corresponds to the elementary reaction from B1 to C1, from B2 to C1' and from B2 to C2. These reaction rates were calculated in the previous work with QM/MM PMFs[75] and shown in Table 3-1. The third stage of the phosphate release process involves elementary steps from C1' to D1, from C1 to D2, and from C2 to D2 (Figure 3-4). The related PT from D45 to cytosol after D324 deprotonation event necessitates the monobasic phosphate release is indicated by reaction from C1 to C1'. This PT reaction is assumed to share the same rate constant as the PT reaction from A2 to B2, given that a monobasic phosphate with deprotonated D324 (C1) occupies the same net charge as the one with a dibasic phosphate and a protonated D324 (state A2).

3.3.4 Confirmed D45N mutant unable to release phosphate

The additional elementary reaction that could be possible during the functional cycle is the proton transport from D324 to the bound dibasic phosphate at the OC state (A2 → C1) where D45 stays in its protonated, neutral charged form. We assume that the influence of the charge form of D45 to this PT reaction is rather small and the rate constants can be well simulated by the reaction from B2 to C1' where D45 is negatively charged. This introduced elementary reaction provides a pathway transitioning from the holo-OC state to the apo-IO state all while D45 maintains its protonated form, eliminating the need for titration. This closely mimics the D45N mutant scenario, suggests a possible pathway for D45N mutant of the phosphate release process and raises questions about the experimental mutagenesis results that indicate an inability of the D45N mutant to release

phosphate into the cytosol. The detailed reaction flux analysis of the kinetic network then showed that this specific D45N associated reaction pathway always contribute negligibly to the total reaction flux towards a phosphate released IO state, confirming a consistency with the experimental result where K_{m} of the D45N mutant is 6.7 ± 5.3 μM and $\mathrm{V_{max}}$ is 25.3 ± 6.0 pmol $\mathrm{min^{-1}A_{600}^{-1}}$. The high binding affinity of phosphate to this D45N mutant is indicated by the contrast of the mutant K_{m} value against the wildtype one (90.4 ± 10.7 μM) and the inability of phosphate release in D45N mutant is suggested by the same range of V_{max} between D45N mutant and the vector only control result ($V_{\mathrm{max}}$is 11.8 ± 4.9 pmol $\mathrm{min^{-1}A_{600}^{-1}}$).

3.3.5 Proton and phosphate transport coupling in reaction flux

With the assumption of cytosolic pH being a constant value of pH 7.0, we explored the impact of proton gradient and phosphate gradient across the membrane on the reaction flux of transforming from the OC to the IO state. The proton gradient is explicitly expressed in ΔpH, defined as $\Delta\mathrm{pH} = \mathrm{pH_i} - \mathrm{pH_o}$. The extracellular pH condition ranges from 2.0 to 7.5 in 0.5 increments. The phosphate gradient was explored with cytosolic phosphate in abundant, adequate, and low-level condition with $\mathrm{Pi_i}$ value of $10^{-4}, 10^{-5}, 10^{-6}$ M separately and extracellular phosphate is set to medium and low-level condition with $\mathrm{Pi_o}$ value of $10^{-5}, 10^{-6}$ M.

The reaction flux was calculated as the sum of the elementary net flux from all states directly transfer into each product state. If the calculated overall reaction flux yields a negative value, which implies the system would transition from the IO state to the OC state under a given circumstance, then such datapoints would not be displayed. As shown in Figure 3-5, only a positive pH gradient or a positive Pi gradient necessitates the phosphate uptake by PiPT or else, the phosphate will be transported reversely from cytosol to extracellular bulk. As depicted by the blue, orange, and red lines in the plot, phosphate can be transported against its electrochemical gradient

when there is a positive pH gradient. This demonstrates the capability of PiPT to facilitate Pi accumulation under conditions of low Pi levels. As indicated by the purple line in the plot, in situations where there is a positive phosphate gradient from the extracellular bulk to the cytosol, protons can be transported even against their electrochemical gradient. Hence, the process of proton-coupled phosphate co-transport is demonstrated to be bidirectional; phosphate transport can be propelled by a positive proton gradient, and conversely, proton transport can be driven by a positive phosphate gradient. The latter scenario, however, is seldom encountered as the concentration of extracellular inorganic phosphate is typically less than or equal to the cytosolic concentration.

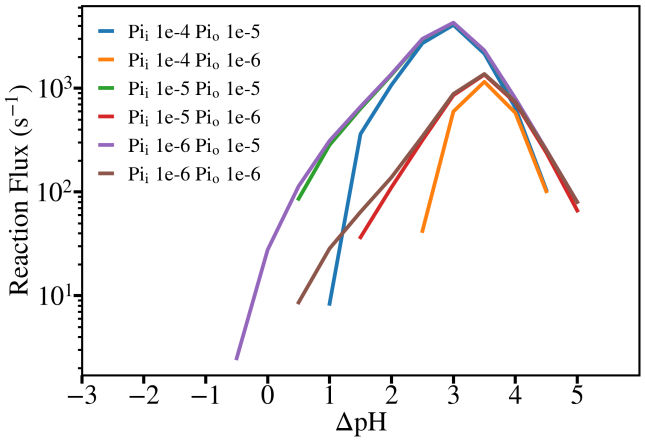


Figure 3-5. Computed Reaction Flux towards the Inward-Open (IO) States under Various pH and Phosphate Concentration Conditions, with a Consistent Cytosolic pH of 7. Pi_i and Pi_o represent the cytosolic and extracellular phosphate concentrations respectively, expressed in molarity (M).

3.3.6 Dominant reaction pathway and rate-limiting step identification

Since multiple reaction pathways exist for the transformation from phosphate bound OC states to phosphate released IO states, we analyzed the contribution of each reaction pathway to the reaction flux and identified a new dominant reaction pathway. The contribution of each pathway was calculated by starting with each ion release state and tracing back through transition rate constants of each elementary net flux. Once multiple ion elementary transport was encountered,

the net flux through each pathway was assigned, and subsequent searches started. The pathway that contains an elementary reaction with a negative net flux was considered to have no contribution to the overall net flux of the reaction towards the IO states. The transporter can adapt to dominantly transport different forms of phosphate under different pH and phosphate conditions and there are only two different kinds of dominant pathways found under all circumstances, each starts with a different form of phosphate bound at the binding site. The first pathway starts from a monobasic phosphate-bound OC state A1 and it follows exactly the previously proposed three-stage pathway, the first proton departs from D45 (A1→ B1), then D324 loses its proton to D45 and eventually to cytosol (B1 → C1 → C1'), and finally the monobasic phosphate released from the binding site and reach to bulk water. The second pathway differs slightly from the three-stage transition. Initially, the transition begins with the transport of a proton from D45 to the bulk, while a dibasic phosphate remains bound in the OC state structure. Following this, the proton from D324 is directly transported to the bound dibasic phosphate. Instead of passing the proton from phosphate to D45 as proposed, the phosphate, carrying the transported proton, departs from the binding site immediately.

For all the conditions illustrated in Figure 3-5,, it's the second pathway that predominates, not the first, except for the scenario where $\Delta\mathrm{pH} = 1.0$, $\mathrm{Pi_i} = 10^{-4}\mathrm{M}$ and $\mathrm{Pi_o} = 10^{-5}\mathrm{M}$. Since the population ratio between the monobasic and dibasic phosphate bound OC states only changes with the extracellular pH condition, as a result, A2 state overpopulates A1 state when $\Delta\mathrm{pH} \leq 1.0$ and A1 state dominates over A2 when $\Delta\mathrm{pH} > 1.0$. This means that for all conditions list in Figure 3-5, with $\Delta\mathrm{pH} > 1.0$, phosphate should be dominantly in the monobasic form, however, it must transform into a dibasic form for the OC state to start the dominant reaction flux towards a phosphate release state. The rate limiting step of this second reaction pathway, defined as the

smallest elementary reaction net flux, is always the PT process from D324 to the bound dibasic phosphate. This is surprising considering that this PT process maintains the largest forward reaction rate constant among all elementary reactions along the given pathway. The rate-determining step for the only scenario where the first reaction pathway dominates is the phosphate release process (C1'→D1).

In brief, even though the dominant reaction pathways share the same elementary phosphate release step (C1'→ D1), the rate limiting step of the two pathways are different in stages – for a monobasic phosphate bound OC state pathway, the rate limiting step is the final phosphate release process whereas for a dibasic phosphate bound OC state pathway, the rate limiting step is the PT event from D324 to phosphate prior to the phosphate release process.

3.3.7 Optimal pH condition and its molecular origin

The molecular underpinnings of the optimal extracellular pH condition for phosphate uptake in PiPT, found to be 4.5 in its close homologue Pho84[65, 75], have scarcely been discussed. Our bottom-up kinetic network model revealed that the optimal pH_o depends on the external phosphate concentration and with extracellular phosphate concentration spanning in a range from 1 μM to 10 μM, the optimal pH increases from pH 3.5 to pH 4.0. This is indicated in Figure 5 where reaction flux profile of blue, green and purple lines with external phosphate concentration as 10 μM all reach a maximum value at $\Delta\mathrm{pH} = 3.0$ and the other three with external phosphate concentration as 1 μM reach to their peak at $\Delta\mathrm{pH} = 3.5$. Our calculation with an external phosphate concentration of 0.1 mM renders an optimal pH value 4.5 (Figure 3-7). Then the pH 4.5 condition of Pho84 experimental work was examined in detail and a 0.1 mM phosphate was guaranteed in the sample preparation step.

Now we have successfully reproduced the optimal pH with the reaction flux calculated from the bottom-up kinetic network model, the next step is to take advantage of this bottom-up kinetic network model and explore the molecular origin of the optimal pH condition. Using $\mathrm{pH}_i = 7.0, \mathrm{Pi}_i = 10^{-5}\mathrm{M}, \mathrm{Pi}_o = 10^{-4}\mathrm{M}$ with different pH_o value ranging from 3.5 to 7.5 at a 0.5 pH unit interval as an example, the net flux that directly produce the IO states goes up and then down as the extracellular pH increases its value. The concentration of state A2, B2 and C1' along the dominant reaction pathway all peak concurrently at the same pH value (Table S2). The A2 state population was solved solely based on the extracellular pH and phosphate condition according to the calculation scheme and as $\mathrm{pH_o}$ increases, the equilibrium between the mono- and di-basic phosphate constantly lifts the concentration of dibasic phosphate (A2 state) and the equilibrium between the protonated and deprotonated D45 constantly suppresses the concentration of a protonated D45 (A2 state). Thus, pH 4.5 becomes the sweet spot for these two counter forces and makes itself the optimal pH to reach the maximum. The optimal pH is the balance between the phosphate protonation state and D45 protonation state.

3.4 Conclusion

We employed multiscale reactive molecular dynamics simulations in conjunction with an enhanced sampling approach to examine the elementary proton transport reactions. Furthermore, classical simulations were performed to explore various phosphate release processes, again coupled with an enhanced sampling technique to ascertain the timescale of the final phosphate release process. Each elementary reaction step in the kinetic network model was examined by the calculated free energy profile and the forward and backward reaction rate constants associated were determined by transition state theory. Utilizing a bottom-up framework, we were able to construct a comprehensive understanding of the myriad dynamical reaction pathways facilitating

the transition from phosphate-bound OC states to phosphate-released IO states. This holistic approach allowed us to examine the impact of environmental factors on reaction flux. We discovered that the PiPT transporter can adaptively handle various forms of phosphate under differing conditions. Interestingly, the primary pathway for dibasic phosphate transport diverges slightly from our initial proposal. In this updated scenario, the deprotonation event at D324 is directly succeeded by a joint transport of proton and phosphate via the newly opened cytosolic gate, rather than a separate proton transport through the previously identified proton exit tunnel. Our kinetic network model revealed the co-transport of proton and phosphate to be bidirectional and can be driven by either a positive phosphate gradient or a positive proton gradient across the membrane. The model's ability to reproduce the optimal pH conditions under experimental setup stems from its control over the equilibrium between D45 and phosphate titrations.

3.5 Appendix

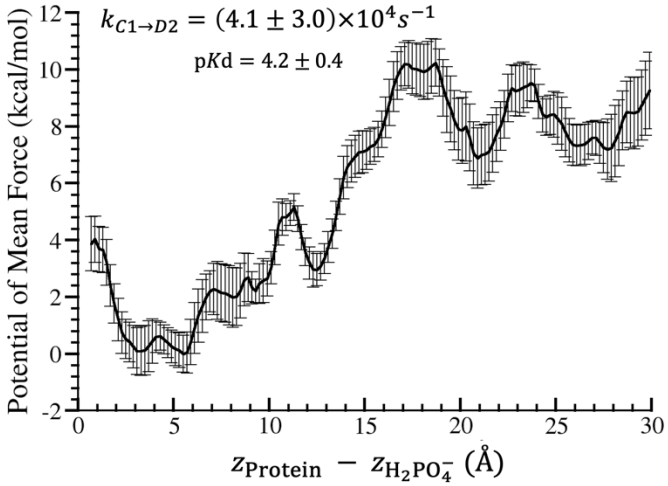


Figure 3-6. The free energy profile of the monobasic phosphate release event from binding site to cytosol where D45 stays protonated and D324 deprotonated.

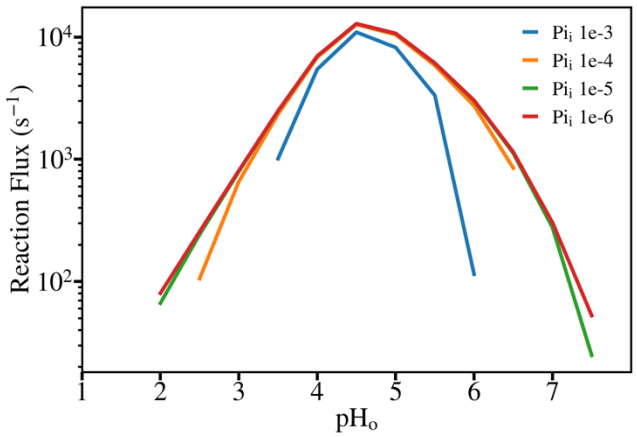


Figure 3-7. Computed Reaction Flux towards the Inward-Open (IO) States under Various pH and Phosphate Concentration Conditions, with a Consistent Cytosolic pH of 7 and extracellular phosphate concentration of 0.1 mM. Pi_i represents the cytosolic phosphate concentrations in molarity (M).

Table 3-1. The calculated forward and backward reaction rate constant for PT process in stage 2 transition, based on the published 1D PMFs calculated from QM/MM simulations.

Elementary reaction	Forward reaction rate constant (s^{-1})	Backward reaction rate constant (s^{-1})
B1→C1	$(3.4 \pm 0.3) \times 10^8$	$(2.7 \pm 0.3) \times 10^{10}$
B2→C1'	$(6.3 \pm 1.6) \times 10^{11}$	$(1.7 \pm 0.8) \times 10^{13}$
B2→C2	$(3.3 \pm 0.7) \times 10^8$	$(1.1 \pm 0.4) \times 10^{11}$

Table 3-2. The probability of state A2, B2 and C1' under different extracellular pH condition with pHi 7.0, Pio 1e-4 M, Pio 1e-5 M.

pHo	A2	B2	C1'	B2/A2	C1'/A2
3.5	0.003	0.0002	8.5*1e-6	0.07	0.003
4.0	0.009	0.0005	1.8*1e-5	0.05	0.002
4.5	0.016	0.0009	3.2*1e-5	0.05	0.002
5.0	0.013	0.0008	2.9*1e-5	0.06	0.002
5.5	0.008	0.0005	1.9*1e-5	0.06	0.002
6.0	0.004	0.0003	1.3*1e-5	0.08	0.003

4 Insights into the quantitative mechanism of ion selectivity of the human voltage-gated proton channel Hv1

4.1 Introduction

Human voltage-gated proton channels (hHv1) are widely acknowledged for their immense clinical significance, as they play pivotal roles in a diverse array of fundamental biological processes[2]. These processes encompass the intricate regulation of respiratory burst in immune cells[80-82], which is essential for effective immune responses against pathogens and tissue damage. Additionally, hHv1 channels contribute to the sperm capacitation process[82, 83], facilitating the acquisition of fertilization competence by spermatozoa. Furthermore, these channels are implicated in the modulation of neuronal excitability[84-86], a fundamental mechanism underlying proper neuronal function and communication in the nervous system. hHv1 channels are also involved in cancer cell migration[87-89], a crucial step in tumor progression and metastasis. The multifaceted involvement of hHv1 channels in such critical biological processes highlights their significance and underscores the need for a comprehensive understanding of their functional properties and underlying mechanisms.

The central role of hHv1 in these processes is to rapidly extrude proton to perform pH regulation and remain electrical excitability. One of the most fascinating properties of this ion channel is its remarkable proton selectivity[10] when compared to other cations under depolarized membrane potential. The proton selectivity was demonstrated by the experimental measurement of reversal potential (V_{rev}, the membrane potential at zero current), that aligns with the Nernst potential of protons. Considering that protons are a million times less concentrated than other ions under physiological conditions, the remarkable proton selectivity observed necessitates a permeability ratio between proton over other ions ranging from 10^6 to 10^8.[11] To further understand the

underlying mechanism of such proton selectivity, numerous experimental and computational attempts have been made to identify key residues and potential proton conduction pathways.

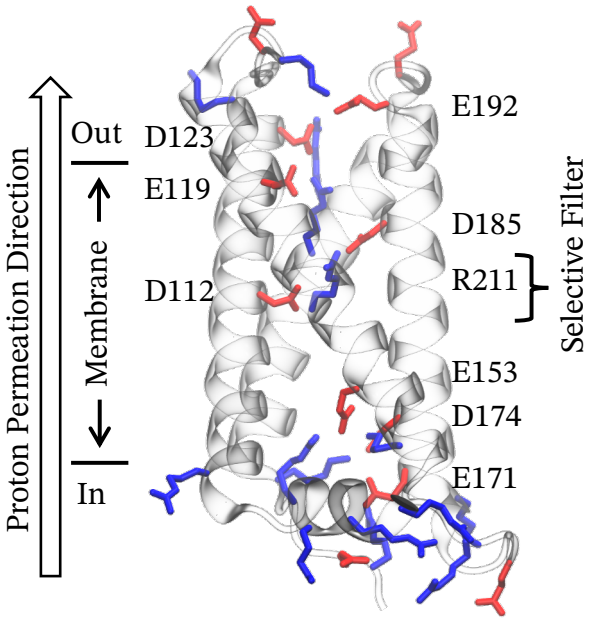


Figure 4-1. Structure of WT conductive Hv1 channel under depolarized membrane potential. Charged pore-lining key residues to proton permeation marked with residue ID (red: acidic; blue: basic). The direction of the proton permeation is from cytoplasmic to extracellular bulk water (marked in arrow).

Musset et al. identified D112 as a crucial residue for proton selectivity via electrophysiological measurements on D112 mutants[90] – D112N mutation was found to shift the channel towards anion selectivity, while the D112E mutant maintains the proton selectivity. These findings highlight the potential role of titratability and electrostatic effect of D112 in the channel proton selectivity. A follow-up quantum mechanical study[91] on a gas-phase model suggested that the D112-R211 salt bridge acts as a gatekeeper, blocking the passage of other cations while allowing protons to permeate by undergoing protonation and deprotonation at D112. Additionally, both hybrid quantum mechanics/molecular mechanics (QM/MM)[92] and empirical valence bond[93] (EVB) molecular dynamics (MD) simulations provided evidence that explicit proton passage occurs via

the titration of D112. However, the formidable computational cost of QM/MM prevented a converged free energy profile and thus lacked a valid conclusion on the proton permeation mechanism. The EVB model by Lee et al. [93] ignored the Grotthuss transport inside the channel but only modeled a shared proton between D112 and its first solvation shell water. Such D112 titration hypothesis implies that the selective filter (SF) consists of D112 and its close residues imposes stronger restrictions on ions other than protons, which resembles a conventional selectivity mechanism in other ion channels.

On the other hand, the alternative water-wire viewpoint argues that the SF region provides sufficient space and hydration, allowing ions of various sizes to pass through. Ramsey et. al. found[94] neutralizing each of the conserved titratable residues (D174A, E153A, K157A, D185A, D112A, D112N) being insufficient to abolish the proton conductance and thus the only possible proton permeation mechanism should be Grotthuss transport via a robust continuous water wire. Evidence of classical MD simulations[94-96] revealed the presence of a persistent water wire within the selective filter region, suggesting its involvement in the proton conduction pathway. However, explicit proton transport with a dynamic hydration environment is needed for further investigation since a continuous water wire observed without the proton can exhibit distinctive thermodynamic properties from one with presence of an explicit proton[19, 23].

The determination of the proton permeation and selectivity mechanism of hHv1 was hindered by the significant challenges to resolve a conductive structure, due to the pronounced conformational changes in response to membrane potential and pH gradients[97-99]. However, advancements in modern-day supercomputers[100], have now opened new avenues for obtaining the physiologically relevant conductive structure of these channels[101]. Previously, we tackled this challenge by employing long-time MD simulations to obtain both the closed and open structure

models of hHv1 and validated the obtained structure models with calculations on the redistribution of charges during channel activation[101]. Using this structure, we made significant progress in the field of hHv1 inhibitor discovery with classical MD simulations[102-104]. However, there is still a need for explicit proton permeation simulations to uncover the precise molecular determinants that govern the selectivity of protons over other ions in hHv1 channels.

Recently, we have made notable progress in the multiscale reactive MD (MS-RMD) method[105, 106], a powerful approach that enables efficient, accurate and transferable descriptions of proton transport (PT) involving amino acids and water chains in complex biomolecular environments[76, 77, 107]. Building upon these achievements, we performed extensive free energy sampling with MS-RMD for the full PT process in the hHv1 channel mediated by pore water and all pore lining Asp and Glu residues, without any prior bias towards a specific PT mechanism. Moreover, our investigations extend beyond proton transport and include the examination of free energy profiles of cation and anion permeation through the channel. Remarkably, our calculations on proton and other ion permeability exhibit excellent agreement with experimental measurements on the ion selectivity and proton conductance, providing a quantitative understanding of the proton selectivity mechanism and marks a significant step towards unraveling the fundamental functional mechanism of voltage-gated proton channels. Methods

4.2 Methods

4.2.1 Classical molecular dynamics equilibration (classical MD)

Starting from the conductive wildtype Hv1 structure[101], we performed classical MD equilibration for 2 microseconds with protonation states of the protein assigned as in neutral pH condition according to pK_a predictions by PROPKA[42] on the starting structure. The protein was embedded in a pre-equilibrated and hydrated 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC)

lipid bilayer with approximately 150 mM of NaCl under a constant external electric field of 0.0156 V/nm on z axis to simulate the conductive Hv1 channel under the +150mV depolarized membrane potential. All interactions were described by the CHARMM36 force field[45] with CMAP correction and the TIP3P water model. Models were built and equilibrated using a standard protocol in CHARMM-GUI[41] and simulations were performed in GROMACS[46, 55] version 2020.4 with the Velocity Verlet integrator using a time step of 2 fs in the isothermal-isobaric (constant particle Number, Pressure, Temperature (NPT)) ensemble using a semi-isotropic Parrinello-Rahman barostat[56] at 1 atm) and a velocity rescale thermostat[57] (300 K). The same equilibration procedure was conducted in parallel for the D112N mutant. Electrostatic interactions were calculated using the particle mesh Ewald method with a cutoff of 12.0 Å and a precision of 10-5 was used to treat the electrostatic interactions. The Lennard-Jones (LJ) interaction was set to cutoff to 0 at 12.0 Å radius and used a switching function starting at 10.0 Å to smooth the truncation. All bonds involving hydrogens were constrained using the LINCS algorithm[108].

4.2.2 Multiscale reactive molecular dynamics simulations (MS-RMD)

Starting from the equilibrated conductive Hv1 structure of both wildtype and D112N mutant in classical MD, the MS-RMD simulations were performed to simulate the explicit proton permeation process with all pore-lining Glu/Asp residues modelled as potential protonation sites and one additional proton was added to a pore-water replacing a Na^+ in the bulk. All interactions were described by the CHARMM36 force field with CMAP correction with the developed Asp and Glu MS-RMD models[76, 77]. Simulations were performed by LAMMPS MD package[109] coupled with RAPTOR[106] for proton reactions. MD was integrated with time step set to 1 fs in the canonical (constant particle Number, Volume, Temperature (NVT)) ensemble using a Nose–Hoover chain thermostat at 300 K and with an orthogonal box of previously equilibrated dimension (77.493 Å,

77.483 Å, 99.581 Å) for WT Hv1 and (77.471 Å, 77.471 Å, 99.587 Å) for D112N mutant. Electrostatic interactions were calculated using the particle-particle particle-mesh method with the precision of 10^{-4} in force. The Lennard-Jones (LJ) and coulombic interactions was set to cutoff to 0 at 10.0 Å radius and used a switching function starting at 8.0 Å to smooth the truncation.

4.2.3 2D umbrella sampling with MS-RMD simulations

To enhance the sampling of free energy landscape of proton permeation through the channel, two-dimensional umbrella sampling on the two reaction coordinates were performed by PLUMED2 package[110] incorporated with LAMMPS/RAPTOR. The principal reaction coordinate is defined as

$$z_{\mathrm{H}} = z_{\mathrm{CEC}} - z_{\mathrm{Prot}} \tag{4-1}$$

where CEC is the abbreviation of center of excess charge, a virtual particle mathematically defined to track the position of the transported excess proton. Detailed definition of CEC involved in the MS-RMD theoretical framework could be found in the MS-RMD literature[76]. Prot stands for the center of alpha carbons of transmembrane Hv1 helices S1, S2 and S3, defined as residues 103 to 122, 138 to 160 and 172 to 188. The second reaction coordinate is defined as a smooth version of minimum distance,

$$r_{\min} = \frac{\beta}{\log \sum_i \exp\left(\frac{\beta}{s_i}\right)} \tag{4-2}$$

where β parameter is set to be 90 Å and s_i is the distance between the *i*-th carboxyl oxygen of all pore lining Asp/Glu residues and the transported proton (CEC).

To restrain the lateral diffusion of the hydrated proton when completely dissociated from the protein, a flat-bottom harmonic-shaped potential was introduced to the reaction coordinate $r_{\perp}$ defined as the radial distance of the transported proton to the center of protein.

$$r_{\perp} = \sqrt{(\mathrm{x_{CEC}} - \mathrm{x_{Prot}})^2 + (\mathrm{y_{CEC}} - \mathrm{y_{Prot}})^2} \quad 4\text{-}3$$

The force constant of the wall was set to 5 kcal/mol/Å^2 and the restraining potential was switched on once $r_{\perp} \geq 7$ Å in umbrella windows where $|z_{\mathrm{H}}| \geq 26$ Å.

A total of over 1500 umbrella windows were established for 2D umbrella sampling in WT and D112N Hv1 systems, with a spacing of 0.5 Å in both the z_{H} and r_{min} dimension. Each umbrella window was run for ~1 ns and the total simulation time exceeded 1 μs. The weighted histogram analysis method[63] (WHAM) was employed to combine the 2D umbrella sampling data and calculate the 2D PMF. The 1D PMF as a function of z_{H} was obtained from integration of the 2D PMF over the r_{min} dimension. PMF error bars were obtained by partitioning the trajectories of all umbrella windows into 6 equally sized blocks and calculating the standard deviation using the last 5 blocks.

4.2.4 Replica exchange umbrella sampling with classical simulations

To enhance the sampling of free energy landscape of TMA^+ and $CH_3SO_3^-$ ion permeation through the channel, replica exchange umbrella sampling on the reaction coordinate z_{ion} were performed by PLUMED2[110] package incorporated with GROMACS[46, 55]. The reaction coordinate was defined as

$$z_{\mathrm{ion}} = \mathrm{z_{Ion}} - \mathrm{z_{Prot}} \quad 4\text{-}4$$

where Ion represents the center nitrogen atom of the TMA^+ cation or the center sulfur atom of the $CH_3SO_3^-$ anion. Prot adopts the same definition of protein center when defining z_{H} in the 2D proton umbrella sampling. To restrain the lateral diffusion of the transported ion when completely dissociated from protein, we employed a similar cylinder restraint on $r_{\perp}$ with a force constant of 10 kcal/mol/Å^2. The timestep for each classical simulation was set to 2 fs and adjacent umbrella windows attempted to exchange every 10 ps. Umbrella windows were set for a spacing of 0.5 Å

and the force constant of each umbrella potential was set to 5 kcal/mol/Å^2. More than 125 umbrella windows were set, and each umbrella window was run for ~10 ns with a total simulation time exceeding 1 μs for both WT Hv1 and D112N mutant systems. We utilized the SPC/E water model enable a direct comparison with the MS-RMD simulations, which use SPC/Fw as the water model. Figure 4-7 shows how results are dependent on the water model. WHAM was employed to combine the umbrella sampling data and calculate the 1D PMF. PMF error bars were obtained by partitioning the trajectories of all replica-exchanged umbrella windows into 6 equally sized blocks and calculating the standard deviation using the last 5 blocks.

4.2.5 Permeability calculation

The ion dissociation constant can be calculated from the expression[111] below.

$$\mathrm{p}K_\mathrm{d} = \lg c_0\, xV = \lg c_0 \int_{\mathrm{Protein}} \mathrm{d}z\, e^{-\beta\left(W(z)-W(z_0)\right)} \int_0^{+\infty} 2\pi r e^{-\beta U^{\mathrm{bias}}(r_\perp, z_0)}\, \mathrm{d}r_\perp \qquad 4\text{-}5$$

Here c_0 = 1/1660 Å^{-3} is the standard state concentration (1M) expressed in number density. $W(z)$ is the 1D PMF of ion permeation and $W(z_0)$ is its value when reaching a plateau in the bulk region. $U^{\mathrm{bias}}(r_\perp, z_0)$ is the cylinder restraint potential used in umbrella windows. The integral was performed over the transmembrane protein region defined as $-22\ \text{Å} \le z \le +22\ \text{Å}$. The binding constant is related to the dissociation constant via $K_\mathrm{b} = 10^{\mathrm{p}K_\mathrm{d}}$. The permeability were obtained separately from the PMFs of each block and calculating the standard error using the last 5 blocks out of 6 blocks.

The transport rate constant can be calculated from the Nernst-Planck equation[112] in terms of a 1D PMF,

$$k = \frac{1}{\left(\int_{\mathrm{Protein}} \mathrm{d}z \exp -\beta W(z)\right)\left(\int_{\mathrm{Protein}} \mathrm{d}z \frac{\exp \beta W(z)}{D(z)}\right)} \qquad 4\text{-}6$$

where $D(z)$ is the position-dependent diffusion constant. A comparison to an alternative, Markov state model (MSM) approach to compute the transport rate can be found in supplementary computational details.

4.3 Results and Discussion

4.3.1 Proton permeation mechanism in WT Hv1

We explored the unitary proton transport mechanism via simulating explicit proton transport with two-dimensional (2D) umbrella sampling (US). The reaction coordinates (RCs) used for the US includes (1) the excess proton z coordinate (the cell membrane defines the xy-plane) to the protein center, denoted as z_{H}, representing the PT process through the channel from cytoplasm to the extracellular bulk, and (2) the distance between the transported proton and its closest pore-lining Asp/Glu residue, denoted as r_{min} . The second RC is crucial to differentiate whether the proton permeates the channel via titrating a protonatable residue (Figure 4-1) or traveling through the water wires.

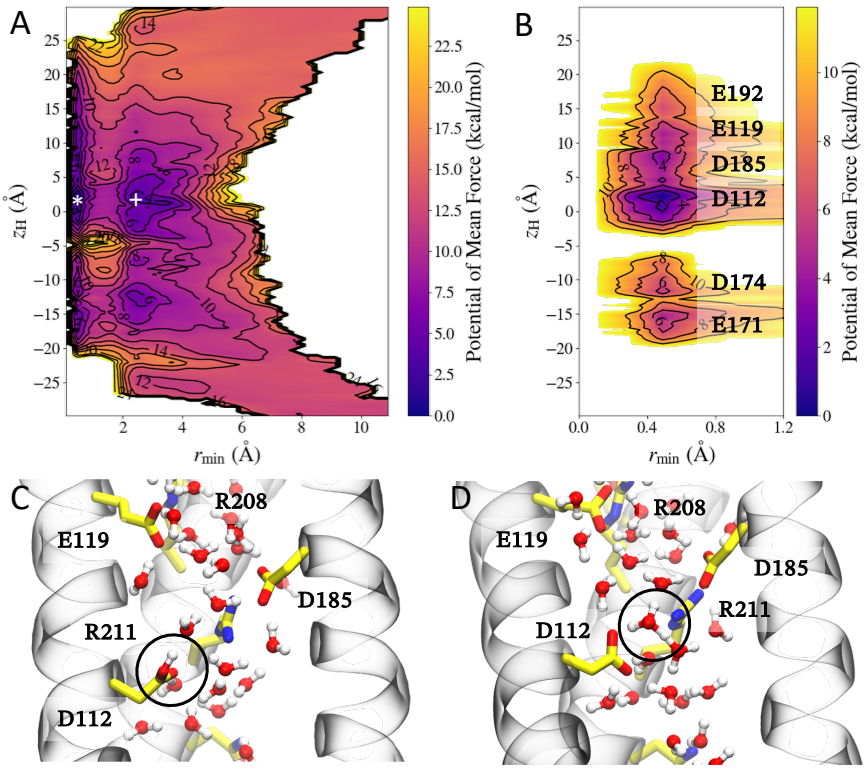


Figure 4-2. Free energy profile of proton permeation process in wildtype hHv1 channel and representative molecular configurations. A The full free energy profile as a function of proton transport progress (z_H) and the proton distance to the closest protonatable residue (r_{min}). B A zoom-in of the free energy profile focusing on the states where proton resides on the pore-lining Asp/Glu residues. C A molecular configuration of the selective filter in D112 protonated state corresponding to position of the "star" symbol in panel A. D A molecular configuration of the selective filter with solvated hydronium forming a contact ion pair with the deprotonated D112, corresponding to the "plus" symbol in panel A. The transported proton has been circled for clarity.

Figure 4-2 A and B show the potential of mean force (PMF), also referred to as the free energy profile, of the excess proton as it traverses the inner pore of the hHv1 channel. The 2D PMF exhibits multiple deep wells situated around $r_{min} \approx 0.5$ Å, (Figure 2A, more clearly shown in Figure 4-2B) corresponding to proton staying on the protonatable residues (Asp/Glu) lining the pore. This observation aligns with previous QM/MM metadynamics study[94] that observed spontaneous protonation events on pore-lining Asp/Glu residues. The free energy wells in the

range of $r_{\min} = 2$ Å to 4 Å correspond to instances where the proton resides in the first solvation shell of these titratable residues. Notably, these minimum wells are less energetically favorable than the protonated wells of the corresponding residues. This observation clearly indicates that the most probable proton transfer pathway involves sequential proton hopping between the titratable Asp and Glu residues.

The selective filter region spanning from below residue D112 to above residue D185 corresponds to a vertical distance, z_{H}, ranging from -5 Å to +10 Å. Notably, the global free energy minimum is associated with the proton residing on D112 (Figure 4-2C), whereas the free energy minimum for the proton and D112 forming a contact ion pair (CIP, Figure 4-2D) exhibits a higher free energy of ~3 kcal/mol. The 1D free energy profile (Figure 4-3A), obtained by integrating 2D PMF over the second reaction coordinate, $r_{\min}$, shows the rate-limiting step of the full proton permeation process to be the dissociation of the proton from D112 to the extracellular bulk water, and the rate-limiting free energy barrier corresponds to the difference between a proton residing in bulk water and a proton in the D112 horizontal plane. Consequently, the rate-limiting barrier is ~3 kcal/mol lower if the proton never hops onto the D112 residue, but instead passes through the SF region via water wires. Thus, despite the titration of residue D112 being the most probable pathway for proton permeation according to the free energies, it presents a slower proton permeation pathway, compared to the water wire mechanism. This finding contradicts to the D112 titration mechanism, where the proton permeation is either assumed to occur only through the D112 titration[91, 92, 113, 114], or is faster and more feasible due to barrier lowering as a result of D112 titration[93]. The detailed proton selectivity mechanism associated with this proton permeation mechanism will be discussed in the next sections along with the free energy profiles for other cation permeation.

4.3.2 Proton Permeation Mechanism in D112N hHv1 Mutant

Electrophysiological experiments revealed that the D112N channel abolishes proton selectivity but becomes an anion channel[90], but how a single site mutation can shift a proton channel into anion selective is still elusive. As a first step of understanding the anion versus proton selectivity, we examined the PT mechanism in the mutant channel by calculating the 2D PMF adopting the same reaction coordinates as were used in the WT free energy sampling. As shown in Figure 4-5, the dominant PT pathway remains to be the proton hopping between protonatable residues mediated by water wires, while D112, being mutated to an asparagine, loses its reactive capability to bind a proton, and thus the proton can only permeate through the SF via water wires. In addition, the mutation switches the total charge of selective filter residues (N112, D185 and R221) from negative to be neutral, and as a result, a huge free energy barrier is now seen in the proton 1D PMF (Figure 3A), even before the proton reaches the selective filter region ($z_{\mathrm{H}} = 0$ Å). Notably, the global free energy minimum is shifted to the position of D174. This implies that in the case of a low cytoplasmic pH, a two-proton concerted permeation mechanism may occur in WT hHv1 if we regard the D112N channel mimics a state of the WT channel where a proton has traveled to D112. Given this assumption, the partially reduced proton conductance in the D174A channel under neutral pH[114] may be understood by considering the impaired two-proton concerted transport, which is a contributing but insignificant pathway unless in low pH conditions.

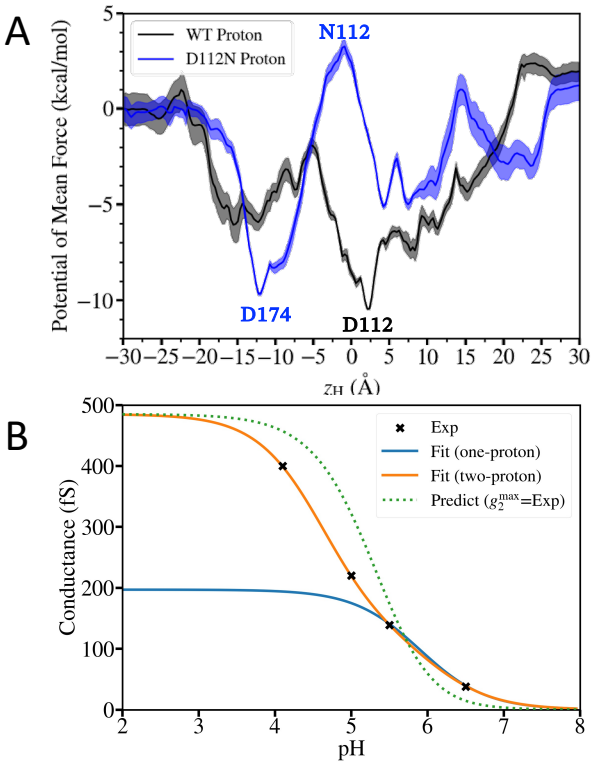


Figure 4-3. Proton permeation in wildtype (WT) and mutant Hv1 and validation against experimental conductance data. A One-dimensional free energy profile for proton permeation in WT (black) and D112N mutant (blue); B The validation against measured conductance under different pH conditions. The experimental data was fitted with a one-proton model (blue solid) and a two-proton model (orange solid). The predicted conductance curve with g_2^{max} obtained from the experimental fit was plotted in dotted lines.

4.3.3 pH-dependent Proton Conductance of hHv1

Cherny et al.[115] measured the unitary proton conductance under various pH conditions, and the data was summarized in Figure 4-3B. To understand the origin of the pH-dependency, we first fit the data with a two-proton conducting model of equation (1)

$$g_2(c) = \frac{cK_1 g_1^{\max}}{1 + cK_1 + c^2K_1K_2} + \frac{c^2K_1K_2 g_2^{\max}}{1 + cK_1 + c^2K_1K_2} \qquad 4\text{-}7$$

where $c = 10^{-\mathrm{pH}}$ is the proton concentration, $g_2(c)$ is the conductance value as the function of proton concentration, K_1 and K_2 are the binding constants of a single proton and a second proton to the protein respectively. $g_1^{\max}$ and $g_2^{\max}$ are the saturated maximum conductance for a one-

proton and for a two-proton mechanism accordingly. In equation 4-7 and 4-8, we assume that the pH-dependency of conductance arises from the pH-dependent probabilities of the channel being in a single-proton-bound or a two-proton-bound state, and the total conductance is thus the weighted average of the two conducting mechanisms. The fitted curve is shown in Figure 4-3B and the fitted parameters are listed in Table 4-1. When the two-proton mechanism is disregarded, the model simplifies to a single-proton model,

$$g_1(c) = \frac{cK_1 g_1^{max}}{1 + cK_1} \qquad 4\text{-}8$$

, which coincides with Levitt's derivation[112] using the Nernst-Planck equation, assuming a channel only allows one single ion to permeate at the same time. Such a single-proton model, however, can only explain the pH-dependency at high pH but fails to adequately fit the experimental data at low pH (blue solid curve in Figure 4-3B. This inconsistency is in line with the hypothetical two-proton mechanism, which was inferred from the PT PMF of the D112N channel.

Table 4-1. Conductance model parameters

	$g_1^{\max}$ (fS)	pK_{d_1}	$g_2^{\max}$ (fS)	pK_{d_2}
Comp.	21.7 ± 8.6	6.4 ± 0.2	--	5.3 ± 0.2
Exp. Fit	131	6.1	485	4.6

$g_1^{\max}$ and $g_2^{\max}$ stand for the maximum conductance of one-proton and two-proton mechanisms respectively. K_{d_1} and K_{d_2} stand for the proton disassociation constant from the protein. The fitted parameters of the one-proton and two-proton conducting model were obtained by fitting equation 4-7 and 4-8 against the experimental conductance measurement. The computational parameters were obtained from proton permeation PMFs and diffusion constants. Details can be found in Methods and Supplementary computational details.

The proton permeation PMFs allow us to directly compute the one-proton binding constant K_1 and, combined with the Nernst-Planck equation, also the one-proton maximum conductance $g_1^{\max}$ (see Supplementary computational details). The second proton binding constant K_2 represents the binding affinity of hHv1 to a second proton given that a proton already resides in the channel.

Since D112 is the global minimum in WT PMF, it is the most probable protonation site for the first incoming proton. Presumably the D112N mutant mimics the one-proton-bound state of WT, and K_2 is estimated as the single proton affinity to the D112N channel. Table 4-1 presents the agreement between our calculated K_1, K_2 and $g_1^{\max}$ and the experimentally fitted data, providing validation for our proton transport free energy calculations. The second-proton maximum conductance $g_2^{\max}$ cannot be directly obtained from our calculations, but it can be approximated by considering the deprotonation timescale of D112 to the bulk when D174 remains protonated. This timescale represents the transition of the channel from a state where two protons occupy the most probable protonation sites to a state where one proton is already transported into the extracellular bulk. Considering that D174, located beneath D112, is charge neutral when it is protonated, it is expected that the deprotonated barrier for proton to depart from D112 would be lower compared to the case in WT with a negatively charged D174. Consequently, the transport rate of the second proton is anticipated to be faster than that of the first proton. This expectation is consistent with the observation that the fitted $g_2^{\max}$ is larger than the fitted $g_1^{\max}$. Given our computed K_1, K_2 and $g_1^{\max}$, together with the experimentally fitted $g_2^{\max}$, we plotted the pH-dependent conductance in Figure 3B (green dotted curve), which shows reasonable agreement with the experimentally determined one.

Table 4-2. The selectivity of proton over other ions when pH = 7 and other ion concentrations are 0.15 M.

Selectivity	WT	D112N mutant
Proton over TMA^+	9.3	12.7
Proton over $CH_3SO_3^-$	2.0×10^4	4.7×10^{-3}

Selectivity was quantified by the product of the proton permeability and proton concentration, divided by the ion permeability times its concentration. The permeability was calculated according to equation (4).

4.3.4 Proton selectivity to TMA+ cation in WT Hv1

The central quantity governing selectivity is the ion permeability through the channel. The selectivity can be reflected from the comparison between ions' Nernst potential and the experimentally measured V_{rev} based on the Goldman-Hodgkin-Katz (GHK) voltage equation[116, 117],

$$V_{\mathrm{rev}} = \frac{RT}{zF}\log\frac{P_{CH_3SO_3^-}[CH_3SO_3^-]_{\mathrm{i}} + P_{H^+}[H^+]_{\mathrm{o}} + P_{TMA^+}[TMA^+]_{\mathrm{o}}}{P_{CH_3SO_3^-}[CH_3SO_3^-]_{\mathrm{o}} + P_{H^+}[H^+]_{\mathrm{i}} + P_{TMA^+}[TMA^+]_{\mathrm{i}}} \qquad 4\text{-}9$$

where R is the ideal gas constant, T is the temperature in Kelvins, F is the Faraday's constant, z is the charge of the ions (here $z = 1$ for monovalent ion composition). The GHK equation establishes a framework for the competition among potentially permeant ions to determine V_{rev}. Each ion strives to shift V_{rev} towards its respective Nernst potential, and the observed $V_{\mathrm{rev}} = V_{\mathrm{H}}$ for WT hHv1 can be attributed to the dominant contribution from proton, depicted as the product of its permeability and concentration largely surpassing the ones of other ions.

Several studies[90, 118-121] have conducted electrophysiological measurements to investigate the selectivity of protons over the tetramethylammonium cation (TMA^+). The proton-to-TMA^+ permeability ratio determined from the measured V_{rev} using the GHK equation consistently yielded $\frac{P_{H^+}}{P_{TMA^+}} > 1.0 \times 10^7$. Similar to proton permeation PMF calculations, we ran umbrella sampling using the z-coordinate of nitrogen atom of TMA^+, z_{TMA} as the reaction coordinate to quantify the cation permeation free energy profile. In contrast to proton, TMA^+ cations are unable to bind to D112 and their larger radius makes it unfavorable for them to reside within the narrow region of the selectivity filter. Consequently, TMA^+ cations tend to be more stable below the selective filter region rather than inside it, as indicated by the presence of a free energy well below $z_{\mathrm{TMA}} = 0$ Å (Figure 4-4A).

The ion permeability can be calculated as[122-124]

$$P_{\text{ion}} = k_1 \int_{\text{Protein}} \mathrm{d}z\, e^{-\beta(W(z)-W(z_0))} \propto k_1 K_1 \qquad 4\text{-}10$$

, where $W(z)$ is the 1D PMF of ion permeation and $W(z_0)$ is its value when reaching a plateau in the bulk region. The integral was performed over the transmembrane protein region. k_1 is the rate constant of a single ion permeating through the channel, and K_1 is its first order binding constant of the given ion to the protein (see Methods for a detailed expression of k_1 and K_1). The calculated permeability ratio between a proton and a TMA cation is 1.4×10^7, in good agreement with the experimentally estimated lower bound (1×10^7).

By expressing permeability as the product of the rate constant and the binding constant, we can interpret it as the combination of the ion's probability of being inside the channel (binding constant, determined by the depth of the free energy well) and the speed at which the ion can pass through the channel (transport rate constant, influenced by the free energy barrier and its diffusion constant). Notably, the origin of proton selectivity, i.e., the large permeability ratio between the proton and TMA comes from the tight binding of protons ($K_1(\mathrm{H^+}) = 10^{6.4\pm0.2} >> K_1(\mathrm{TMA^+}) = 10^{-0.4\pm0.3}$), instead of the transport rate constant ($k_1(\mathrm{H^+}) = 3.5 \pm 0.7 \times 10^3\ \mathrm{s^{-1}} \approx k_1(\mathrm{TMA^+}) = 1.2 \pm 0.6 \times 10^3\ \mathrm{s^{-1}}$). This finding contradicts the conventional notion of a selective filter, which is typically believed to restrict the movement of ions other than the selected ion, resulting in a noticeable distinction in the transport rate and/or the rate-limiting free energy barrier. The voltage-gated potassium channel from *Streptomyces lividans*[125] (KcsA) is a classic example that contains such a selective filter. In this channel, the interaction between potassium ions and specific residues lining the pore reduces the free energy barrier by 4-5 kcal/mol compared to sodium ions[126-128]. Another example is the Outer Membrane Protein F from *Escherichia Coli*[129] (OmpF) whose wide

inner pore favors conduction of a broad range of cations to anions. The comparison of free energy and diffusion constant profiles of K^+ and Cl^- permeation[129] suggested transport rate constant being the root cause of its cation selectivity. The distinctive behavior of hHv1 can be explained by the unique properties of proton permeation involving the titration events of pore-lining residues. The association and dissociation of protons require a much larger free energy change compared to ion binding, as a result, the free energy profile for proton permeation has a much deeper free energy well compared to ion transport. This deep free energy well is reminiscent of the observed well at the titratable residues[13] H37 in the A M2 channel, suggesting a similar proton selectivity mechanism in A M2, but indeed further calculations on the transport free energy of other ions are needed to confirm this hypothesis. In contrast, a de novo designed proton channel[130] may exhibit a more "conventional" selectivity mechanism, where the difference in permeation rate plays a dominating role in determining selectivity, due to the absence of proton binding residues.

The significant influence of binding affinity on proton permeability emphasizes the unique nature of D112 among all pore-lining reactive residues. The protonated D112 serves as the global minimum in the free energy landscape (Figure 2A), contributing the most to the binding affinity of the proton. The distinctive function of D112 helps elucidate the experimental observations that neutralizing single-site mutations of residue D185 and D174 still maintain proton selectivity[90, 114], as they exhibit weaker proton-binding properties compared to D112 in hHv1.

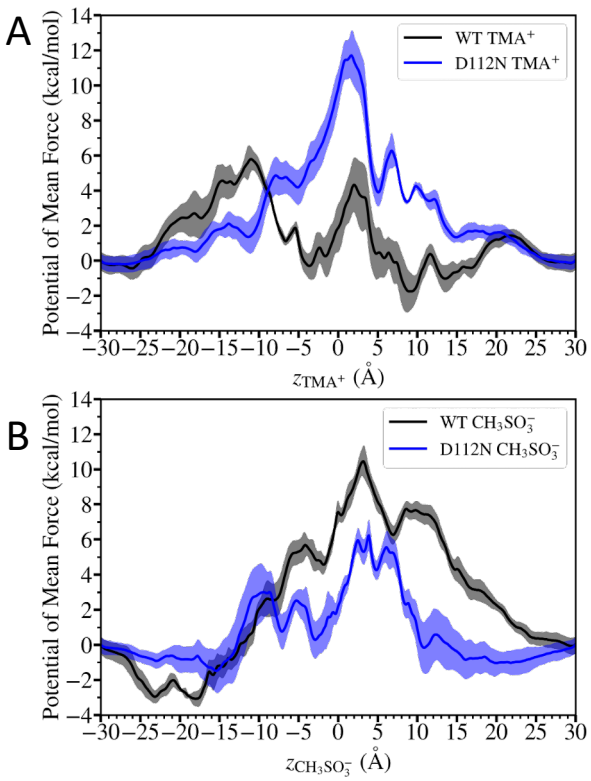


Figure 4-4. The permeation PMFs. A Permeation PMFs of TMA+ cation in wildtype (WT, black) and D112N mutant (blue). B Permeation PMFs of $CH_3SO_3^-$.

4.3.5 Anion selectivity in D112N Hv1 mutant

As discussed earlier, electrophysiological experiments revealed D112N mutant as an anion channel, transporting both $CH_3SO_3^-$ and Cl^-. This was evidenced by the positive shift in the V_{rev} value when all types of ions (except H^+ and OH^-) were diluted in the extracellular solution. A negative shift in the V_{rev} value, however, was observed when the extracellular pH was adjusted from 5.5 to 7.0, suggesting considerable permeation of either protons, hydroxides, or both, at a comparable level to anions. Considering the anion selectivity of the D112N mutant and the pH dependent shift of the V_{rev} value upon dilution, it was postulated that OH^- is the likely permeating ion, though direct evidence is lacking to distinguish the two species.

To address these uncertainties, we conducted umbrella sampling to examine the permeation free energies of TMA^+ and $CH_3SO_3^-$ in the D112N mutant. Figure 4-4 indicates that the anion could be

stabilized below the selective filter (D185 and R211), evidenced by the free energy well around $z_{CH_3SO_3^-} = -3$ Å, whereas the cation could not. Compared to the permeation free energies in the WT channel, the barrier for TMA^+ transport was increased by the mutation and the barrier for $CH_3SO_3^-$ was greatly reduced, clearly shifting the selectivity to anions. This discrepancy can be attributed to the electrostatic effect in the D112N mutant, where the arginine residue is positioned closer to the base of the selective filter.

Permeability calculations revealed that the permeability ratio of protons to $CH_3SO_3^-$, $\frac{P_{H^+}}{P_{CH_3SO_3^-}} = 7.0 \times 10^3$, and the ratio of TMA^+ to $CH_3SO_3^-$ is $\frac{P_{TMA^+}}{P_{CH_3SO_3^-}} = 3.7 \times 10^{-4}$. These suggest a dominant contribution from the anion to the GHK equation, when considering the physiological pH and anion concentration, confirming the anion selective nature of the mutant channel. The anion-to-proton selectivity rules out the possibility of H^+ being the permeating species but supports the hypothesis that considerable OH^- permeation is responsible for the pH-dependent shift in V_{rev}. Although the D112N mutant loses its proton selectivity over anions, it is worth noting that the channel continues to exhibit proton selectivity over the TMA^+ cation, as indicated by a permeability ratio of $\frac{P_{H^+}}{P_{TMA^+}} = 2.0 \times 10^7$.

In summary, the anion-selectivity behavior of the D112N mutant results from substantial anion leakage through the channel. This leakage is attributed to the markedly enhanced anion permeability along with reduced proton/cation permeability compared to the WT. Interestingly, proton selectivity over cations is retained, although this preference is underneath the dominant anion permeability.

4.4 Appendix

Transport rate constant calculations. The transport rate constant can be derived from the Nernst-Planck equation[112] in terms of the 1D free energy profile,

$$k = \frac{1}{\left(\int_{-z_0}^{z_0} \mathrm{d}z \exp -\beta W(z)\right)\left(\int_{-z_0}^{z_0} \mathrm{d}z \ \frac{\exp \beta W(z)}{D(z)}\ \right)} \qquad 4\text{-}11$$

where $W(z)$ is the 1D free energy profile (potential of mean force; PMF) of reaction coordinate at value z, $D(z)$ is the ion position-dependent diffusion constant at position z. The integral was performed over the transmembrane protein region defined as $-22\ \text{Å} \leq z \leq +22\ \text{Å}$.

Due to the exponential dependency on $W(z)$, the expression can be simplified as[131]

$$k = \frac{D(z_{TS})}{\left(\int_{-z_0}^{z_0} \mathrm{d}z \exp -\beta W(z)\right)\left(\int_{-z_0}^{z_0} \mathrm{d}z \exp \beta W(z)\right)} \qquad 4\text{-}12$$

where $D(z_{TS})$ is the diffusion constant calculated at the transition state of the system, corresponding to the top of the free energy barrier of the 1D PMF.

The diffusion constant can be calculated as follows, according to Hummer et. al.[132]

$$D(z) = \frac{\mathrm{Var}(z)^2}{\int_0^{\infty} \mathrm{d}t\ C_{zz}(t)} \qquad 4\text{-}13$$

, where $\mathrm{Var}(z)$ stands for the variance of z, and $C_{zz}(t)$ is the time correlation function. $\mathrm{Var}(z)$ and $C_{zz}(t)$ were computed from the z time series in the corresponding umbrella sampling window. And the time integral of $C_{zz}(t)$ was performed by firstly fitting to a triple exponential function, $a_1 \exp\left(-\frac{t}{t_1}\right) + a_2 \exp\left(-\frac{t}{t_2}\right) + a_3 \exp\left(-\frac{t}{t_2}\right)$, and integrated analytically.

In the case of proton permeation in the WT channel, the highest free energy point corresponds to the stage of proton freely diffusing in the bulk water. Thus, we took the experimental proton diffusion constant[133] (0.94 Å^2/ps) in bulk water in the transport rate constant calculation for WT

proton permeation process. We note that the proton diffusion constant in bulk directly obtained from MS-RMD[134] is 0.37 Å^2/ps, smaller than the experimental value. The discrepancy can be attributed to the missing nuclear quantum effects (NQE) in simulations with classical nuclei[135]. In D112N mutant, the transition state of the proton permeation corresponds to proton staying on the first solvation shell of residue N112. We ran a separate MS-RMD simulation where an umbrella potential acted only on the reaction coordinate z_{H} with window center of value ~ -1.0 Å. The proton diffusion constant obtained from the simulation for D112N mutant at the transition ridge was 0.01Å^2/ps. The diffusion constant was corrected for the missing NQE by the ratio between experimental and the MS-RMD proton diffusion in bulk water and the corrected value was 0.03Å^2/ps.

The transport rate constant can also be computed from the transition matrix in the reaction coordinate space using a Markov state model (MSM) approach[136]. The specific flavor of MSM we employed was the dynamic histogram analysis method (DHAM) framework[137], which relates the transition dynamics in biased simulations, such as umbrella sampling, to the unbiased transition probability via a reweighting factor as a function of the bias potential. The proton permeation rate constant was computed to be the conditional reaction flux per unit time τ given that the proton had resided in the cytosolic bulk water last,

$$k^{\mathrm{MSM}} = \frac{F}{\tau \sum_i \pi_i q_i^-} \tag{4-14}$$

The reaction flux for the entire proton permeation process was computed from the committor via

$$F = \sum_{i \in \mathrm{Cytosol}} \sum_{j \notin \mathrm{Extracellular}} \pi_i P_{ij} q_j^+ \tag{4-15}$$

, where π_i is the stationary probability of state i. The reactant state was defined as the transported proton inside cytosolic bulk water ($z_{\mathrm{H}} \leq -22\text{Å}$) and the product state was defined as the one inside extracellular bulk water ($z_{\mathrm{H}} \geq 22\text{Å}$).

The forward committor q_i^+ demonstrates[138-140] the probability for the proton when currently being at state i will reach the extracellular bulk water next rather than back to cytosol. The committor gradually increases its value from 0 to 1 from cytosolic bulk region to extracellular space and its value for intermediate states i ($-22\ \text{Å} < z_{\mathrm{H}} < 22\ \text{Å}$) can be calculated from the transition matrix by solving the equations below[141],

$$q_i^+ = \sum_{i \in \mathrm{Intermediate}} P_{ik} q_k^+ + \sum_{k \in \mathrm{Extracellular}} P_{ik} \qquad 4\text{-}16$$

where P_{ik} is the transition matrix element, indicating the transition probability from state I to state k within lag time τ. At equilibrium, the backward committor can be calculated as $q_i^- = 1 - q_i^+$.

The umbrella sampling trajectories were partitioned into 6 equally sized blocks and the MSM rate standard error was calculated using the last 5 blocks out of 6. Results are in Figure 4-6. The resulting MSM was checked by the saturated rate constant as a function of increasing lag time (Figure 4-8).

Maximum conductance calculation. The unitary maximum conductance $g_1^{\max}$ was calculated from transport rate constant k_1 via[142, 143]

$$g_1^{\max} = \frac{k_1 q^2}{k_B T} \qquad 4\text{-}17$$

where q is the charge of the transported ion, k_B is the Boltzmann constant and T is the temperature in Kelvin (300 K for our system).

Table 4-3. Diffusion constant calculated from classical ion permeation.

WT hHv1	Diffusion constant (Å^2/ps)	D112N mutant	Diffusion constant (Å^2/ps)
TMA^+	0.004 (0.002 to 0.004)	TMA^+	0.004 (0.003 to 0.006)
$CH_3SO_3^-$	0.002 (0.001 to 0.002)	$CH_3SO_3^-$	0.002 (0.002 to 0.006)

The diffusion constants were calculated from umbrella window trajectories corresponding to the transition state position in the 1D PMF of ion permeation process. The experimental TMA^+ diffusion constant[144] in bulk water is 0.093 Å^2/ps. The value in the parenthesis represents the range of the diffusion constant of the reaction coordinate in $\pm$0.5 Å adjacency of the transition state.

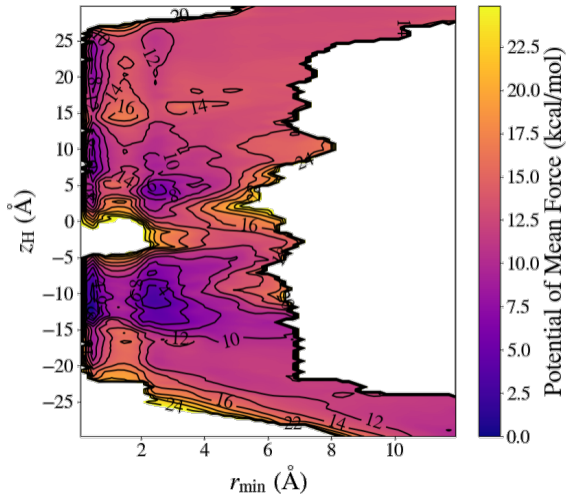


Figure 4-5. Proton 2D PMF in D112N mutant.

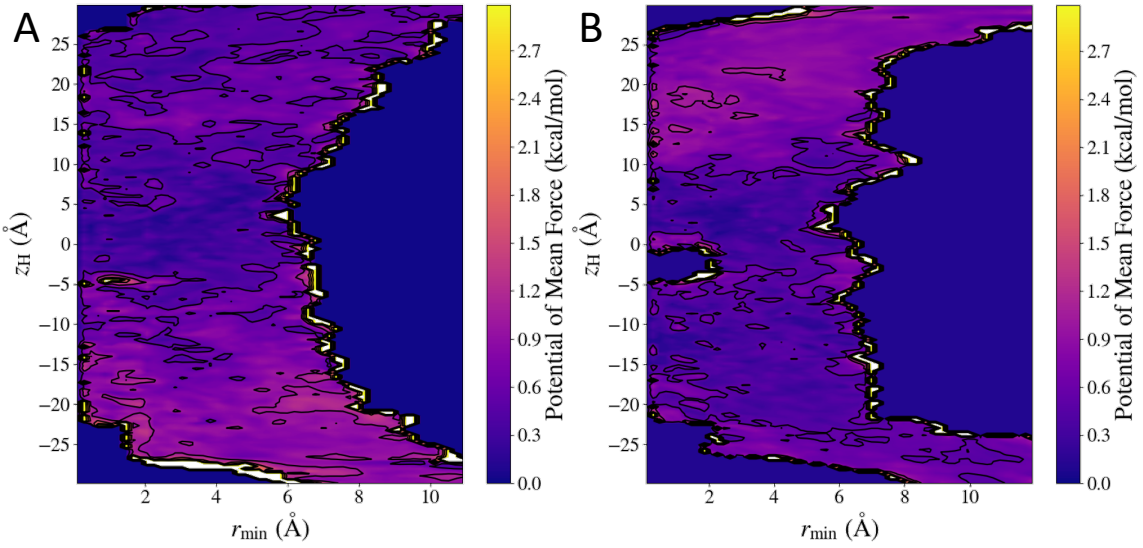


Figure 4-6. Proton 2D PMF error profiles. A) wildtype, B) D112N mutant. The standard deviation was calculated by aligning the 2d PMF of each block by setting the total probability of 2D PMF to 1.

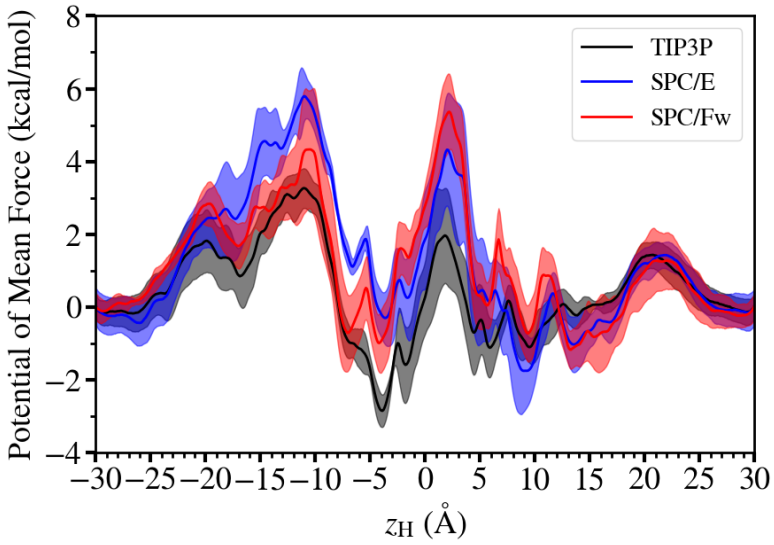


Figure 4-7. TMA+ PMF with different water model: TIP3P, SPC/E SPC/Fw.

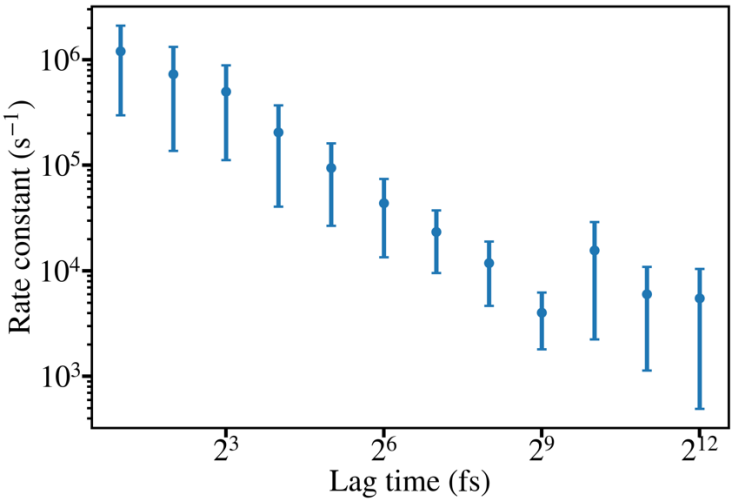


Figure 4-8. Rate constant calculated from MSM with different lag time for proton permeation in WT hHv1. The rate constant value calculated from Nernst-Planck equation with diffusion constant is $(3.5 \pm 0.7) \times 10^3 s^{-1}$, which is in good agreement with the MSM estimated rate constant $(4.0 \pm 2.2) \times 10^3 s^{-1}$ of lag time 0.5 ps.

5 A generalized transition state theory treatment of water-assisted proton transport processes in proteins

5.1 Introduction

Water is the key solvent in all biological and many chemical reactions and it can play an important role in mediating proton transport (PT). The PT process occurs across a broad range of molecular systems in aqueous solution, from simple acid solutions to complex proteins. The mechanisms associated with PT can be exceedingly intricate and are known to be crucial for proton-coupled transporters, enzymes, and channels,[2, 145] wherein the explicit process of PT shows complicated coupling mechanisms to hydration,[19, 23] substrate transport,[75] and protein conformational change.[12, 107]

Molecular dynamics (MD) simulation can represent a powerful approach for investigating chemical reactions such as PT. In MD, PT is usually taken into account implicitly, such as in MD simulations with empirical classical forcefields, or it is treated explicitly via hybrid Quantum Mechanics/Molecular Mechanics[146, 147] (QM/MM) or the Multiscale Reactive MD (MS-RMD) method.[76, 105, 106, 148] In classical MD simulations, the PT process is considered to occur implicitly between two titratable residues within a protein, or between one titratable residue and water when assigning different protonation states to the protein for separate MD simulations. PT is typically inferred based on analyzing the water network within a confined region of the protein. However, due to the fixed bonding topology of classical MD simulation forcefields, the key features of the water-assisted PT, notably the Grotthuss proton hopping[149, 150] component of the PT and the explicit correlation[12-23] between water hydration and PT, cannot be described. In contrast, although the most accurate QM/MM MD describes the electronic structure of the QM region explicitly[146]

possibly with some treatment to account for nuclear quantum effects[151-154] (NQE), its high computational cost limits the degree of conformational (free energy) sampling and hence the ability to investigate the correlation between water hydration and PT with sufficient statistical sampling.

To overcome these shortcomings of both classical MD and QM/MM, the MS-RMD[76, 77, 105, 106, 148] reactive MD approach has been developed which can provide both high accuracy and computational speed that is on average 10^3 or more times faster than direct QM/MM. At the same time, MS-RMD also describes the net positive charge defect delocalization associated with the hydrated excess proton[155-157] as well as the Grotthuss hopping nature of the protonic charge defect.[135]

Due to the initial high thermodynamic barrier (5 to 20 kcal/mol) for proton transfer between weak acids/bases and the accepting water molecule(s), both QM/MM and MS-RMD require a combination of free energy sampling techniques to draw an accurate picture of the PT process with statistically meaningful results and within accessible simulation time. Free energy sampling approaches such as umbrella sampling[72] and metadynamics[158-163] require a pre-defined reaction coordinate, more recently called a "collective variable" (CV), that provides a mathematical measure of the reaction progress. Most commonly, molecular dynamics sampling is used to average out the orthogonal degrees of freedom (DOF) into the free energy profile, the so-called potential of mean force (PMF). As a result, to begin to describe a water-assisted PT process within a protein, the location of the transported Grotthuss hopping excess proton charge defect must be tracked. As will be discussed later, in many cases one must also sample a water hydration CV, which ideally also describes the functional connectivity of the proton transporting water chain.[23]

An initial concept of an excess proton tracker was established in the 1990s as the overall dipole movement of the water chain.[164, 165] As later reported,[166] this initial proton tracker, which is defined as a linear combination of atomic coordinates, is greatly influenced by the permanent dipole movement of the fluctuating water molecules far from the PT active center, making it a less appropriate proton tracker. A more recent concept for the excess proton tracker is commonly known as the center of excess charge (CEC). All CEC definitions – MS-RMD CEC,[167] rCEC,[52, 76] mCEC,[166] and the "proton indicator"[168] – adopt a non-linear form of the atomic coordinates to capture the delocalized nature of the excess proton charge defect and the chemically reactive (Grotthuss hopping) nature of PT, while minimizing the environmental water dipole contamination described earlier. To describe the progress of PT within a protein, the CV is usually defined as a function of the distance between the CEC position and the site of interest (e.g., protein amino acid residue or substrate). Due to the curvilinear CEC CV expression, all forms of PT CV designed for a water-assisted PT process do not, therefore, correspond to a linear combination of the atomic coordinates of the system. As we will discuss in more detail, a curvilinear CV employed in TST produces a nontrivial prefactor in the form of a statistical ensemble average.

Other coupled DOF such as substrate transport, the PT pathway hydration process (noted earlier), and local protein conformational changes can also be slow and therefore critical to the PT process.[12-23]. In such cases, additional CVs are necessary to be explicitly expressed in the free energy sampling process to reflect coupling of PT to those DOFs, along with the CEC describing the PT CV, thereby giving a more complex PMF for the coupled PT process having two dimensions or more. However, if such a PMF can be obtained (or if 1D for just the CEC is sufficient), then with the application of transition state theory (TST),[64, 169] a transition rate can be directly estimated for the PT process from PMF data.

Conventional TST predicts an upper bound to the reaction rate of the system. The upper bound arises from the non-recrossing assumption of a hypersurface that divides the phase space into the reactant and the product regions, i.e. a successful reaction will take place once the system reaches the transition state. There are two critical components in the TST formalism: the first is the exponential dependence of the reactant minimum to transition state free energy barrier,[170] while the second is the pre-factor related, in essence, to the velocity of the effective mass of the CV as it crosses the dividing surface.

Previous research has presented derivations of the mass-dependent pre-factor expression for 1D rate constants for general curvilinear CVs,[64, 171] and applications for both rectilinear[172, 173] and curvilinear[174] CVs with associated PMFs. However, the definition of the water-assisted PT CV, with its effective mass intrinsically dependent on the sampling position of the PMF, requires a more detailed and specific study with respect to the impact of the mass pre-factor on the rate expression. Building on prior research,[64] the present study extends the mass pre-factor expression in terms of a dividing surface on a 2D PMF as a function of both a general curvilinear PT CV and a hydration-related CV. As discussed herein, we develop an exact form of the expression for the mass pre-factor for a general *n*-dimensional case and then study various approximations to the TST transition-rate constant for both 1D and 2D examples of water-assisted PT processes in proteins. Our results[79] confirm the importance of the phase-space position-dependence of the mass pre-factor for water-assisted PT processes and establishes a standard protocol for applying this methodology in future research.

5.2 Methods

5.2.1 Rate Constant Expressions in Generalized Transition State Theory

The TST rate constant expression in a general form for the reaction CV, $\boldsymbol{\xi}$, and with a dividing surface described by $\boldsymbol{\xi}=\mathbf{s}(\mathbf{r})$ in the context of a PMF is given by

$$k_{TST} = \frac{1}{Q^R}\int \frac{d\mathbf{r}\, d\mathbf{p}}{h^{3N}}\delta(\boldsymbol{\xi} - \mathbf{s}(\mathbf{r}))\,\Theta\left(\dot{\xi}_{ver}\right)\dot{\xi}_{ver}e^{-\beta H} \tag{5-1}$$

where $\mathbf{r}$ are the 3N atomic coordinates of the system, $\mathbf{p}$ are the conjugate linear momentum, N is the total number of the atoms in the system, and $\mathbf{s}(\mathbf{r})$ is the transition state dividing manifold. Q^R is the reactant partition function and can be calculated from the integral of the Boltzmann factor over the reactant phase space. Note that $\dot{\xi}_{ver}$ is the vertical velocity component of $\boldsymbol{\xi}$ to the transition ridge; then according to TST, this factor contributes to the forward reaction flux when found to be a positive value. This relationship is expressed by the step function $\Theta\left(\dot{\xi}_{ver}\right)$ in the expression.

Once one transforms the 3N dimensional vector $\mathbf{r}$ to $\boldsymbol{\xi}$, the 3N-*n* dimensional vector $\mathbf{q}$ can be chosen such that no cross terms in the momentum expression between $\mathbf{p}_\xi$ and $\mathbf{p}_q$ remains. The Hamiltonian of the system is then given by

$$H = \frac{1}{2}\mathbf{p}_\xi^T\mathbf{Z}_\xi\mathbf{p}_\xi + \frac{1}{2}\mathbf{p}_q^T\mathbf{Z}_q\mathbf{p}_q + V(\boldsymbol{\xi}, \mathbf{q}) \tag{5-2}$$

where $\mathbf{Z}_\xi$ is the inverse of the effective mass of the CV ξ such that

$$(\mathbf{Z}_\xi)_{jk} \coloneqq \sum_{i=1}^{3N}\frac{1}{M_i}\left(\frac{\partial\xi_j}{\partial r_i}\right)\left(\frac{\partial\xi_k}{\partial r_i}\right) \tag{5-3}$$

and $V(\boldsymbol{\xi}, \mathbf{q})$ is the potential energy expressed as a function of transformed coordinates $\boldsymbol{\xi}$ and $\mathbf{q}$.

Previously, Schenter, et al.[64] derived the TST rate constant expression on the 1D PMF $W(\xi)$, as a function of a general curvilinear CV, given by

$$k_{TST} = \frac{\left\langle \sqrt{Z_\xi} \right\rangle_{\xi=\xi^{TS}}}{\sqrt{2\pi\beta}} \frac{e^{-\beta W(\xi^{TS})}}{Q^R} \quad 5\text{-}4$$

This equation provides an accurate expression for the transition-rate pre-factor with an effective mass factor expression $\left\langle \sqrt{Z_\xi} \right\rangle_{\xi=\xi^{TS}}$, which is the ensemble average value of $\sqrt{Z_\xi}$ sampled at the transition state.

Inspired by the work of Schenter et al, we have derived the TST rate constant expression for an *n*-dimensional PMF with a set of general curvilinear CVs, $\boldsymbol{\xi}$, here we assume the dividing surface can be described as $\xi_1 = \xi_{TS}$. This may be always achieved by a non-linear transformation of $\boldsymbol{\xi}$ such that the generally non-linear $\mathbf{s}(\mathbf{r})$ manifold becomes a hyper-plane and choosing its normal direction as ξ_1. Practically, we found such a transformation is not needed for our case studies. The PT CV was found to already satisfy this assumption as evidenced by the fact that a vertical line approximates well the transition ridge on the 2D-PMFs as shown in the next section. The rate expression in terms of a *n*-dimensional PMF $W(\boldsymbol{\xi})$ is then given by

$$k_{TST} = \frac{\left\langle \sqrt{Z_{11}} \right\rangle_{\xi_1=\xi_{TS}}}{\sqrt{2\pi\beta}} \frac{\int d\boldsymbol{\xi}\, e^{-\beta W(\xi)} \delta(\xi_1 - \xi_{TS})}{Q^R}) \quad 5\text{-}5$$

in which the mass factor $\left\langle \sqrt{Z_{11}} \right\rangle_{\xi_1=\xi_{TS}}$ is the ensemble average of $\sqrt{Z_{11}}$ along the rectilinear transition ridge of the PMF $\xi_1 = \xi_{TS}$ and the expression of Z is given in equation 5-3. Note that the mass factor expression in equation 5-5 is only dependent on ξ_1 (regardless of the other ξ's expression) due to the transition ridge approximation. This derivation is provided in Appendix.

5.2.2 Approximation for Mass Factor in Transition State Theory

Approximations in prior literature have often employed the sampled mass factor from the simulation trajectories corresponding to the reactant state, instead of at the transition state on either

1D or 2D PMFs. An example of this can be found in studies of the water-assisted PT process within a protein[12, 16-18, 21, 175]. This approximation simply amounts to the following expression,

$$\left\langle \sqrt{Z_{11}} \right\rangle_{\xi=\xi_{TS}} \approx \left\langle \sqrt{Z_{11}} \right\rangle_{\xi=\xi_{R}} \qquad 5\text{-}6$$

The approximation stated in equation 5-6 becomes exact if ξ_1 is a linear function of atomic coordinates, such that $\frac{\partial \xi_1}{\partial r_i}$ is constant and does not depend on the atom positions. An example of such a CV is the center of mass of a protein. The approximation could still be valid for some curvilinear CVs, if the value of Z_{11} is constant or insensitive along the minimum free energy path. Examples of such a CV are a pairwise distance or a simple proton transfer coordinate, $r_{HA} - r_{HB}$ where A/B are proton donor/acceptor atoms. We note that, as long as the ensemble average of $\sqrt{Z_{11}}$ is evaluated correctly, this approximation does not rely on the linearity in CVs other than ξ_1.

Moreover, this prior approximation has sometimes utilized the equipartition theorem, by stating that

$$\left\langle \sqrt{Z_{11}} \right\rangle \approx \sqrt{\langle Z_{11} \rangle} \approx \sqrt{\beta \langle \dot{\xi}_1^2 \rangle} \qquad 5\text{-}7$$

The first part of the second approximation stated in equation 5-7, $\left\langle \sqrt{Z_{11}} \right\rangle \approx \sqrt{\langle Z_{11} \rangle}$, is valid when Z_{11} approximates a constant, and thus the conditions we discussed about the first approximation also apply here. Although those conditions are sufficient for $\left\langle \sqrt{Z_{11}} \right\rangle \approx \sqrt{\langle Z_{11} \rangle}$ to be true, they are usually not necessary conditions. It can be seen, from $\mathrm{var}\left(\sqrt{Z_{11}}\right) = \langle Z_{11} \rangle - \left\langle \sqrt{Z_{11}} \right\rangle^2$, that the quality of this approximation is related to the intrinsic fluctuation of $\sqrt{Z_{11}}$. The equipartition theorem used in the second part of the approximation, $\langle Z_{11} \rangle = \beta \langle \dot{\xi}_1^2 \rangle$ is exact (see Appendix). Due to the greater sampling of $\dot{\xi}_1$ (since it can be computed from finite difference upon ξ_1 time series), and due to $\beta \langle \dot{\xi}_1^2 \rangle$ being used in the previous literature, we only report here $\beta \langle \dot{\xi}_1^2 \rangle$ values.

5.2.3 Ab Initio MD and MS-RMD CV Definitions for Water-Assisted Proton Transport

Here, we briefly explain the mCEC[166] and rCEC[52] definitions that have been used as a PT CV in Ab Initio MD (AIMD), and QM/MM simulations, as well as the MS-RMD CEC CV in MS-RMD simulations. This is followed by a discussion of the non-linear properties of the CEC definition function itself with respect to atomic coordinates. The mCEC is defined as

$$\mathrm{r_{mCEC}} = \sum_{i\in\{\mathrm{H}\}} \mathrm{r}_i - \sum_{j\in\{\mathrm{X}\}} \mathrm{w}_j \mathrm{r}_j - \sum_{i\in\{\mathrm{H}\}} \sum_{j\in\{\mathrm{X}\}} \mathrm{f_{sw}}(\mathrm{r}_{ij})(\mathrm{r}_i - \mathrm{r}_j)$$

$$\mathrm{f_{sw}}(\mathrm{r}_{ij}) = \frac{1}{1+\exp\frac{\mathrm{r}_{ij}-\mathrm{r}_0}{\mathrm{d}_0}} \qquad 5\text{-}8$$

where {H} is the collection of hydrogen atoms, {X} is the collection of heavy atoms in the AIMD (or QM/MM) description of the system. The weighting factor w_j is assigned as the number of protons originally bonded to the heavy atom j prior to the acceptance of the proton. For example, the weighting factor of Asp and Glu sidechain oxygens is 0, and the weighting factor of the water oxygen is 2. The switching function $\mathrm{f_{sw}}(\mathrm{r_{ij}})$ employed a Fermi function form with parameter r_0 and d_0 to allow a smooth transition between 0 and 1, which corresponds to the non-bonding and bonding state of each proton-heavy atom pair ij. Taking a Zundel cation $H_5O_2^+$ pair as an example, when proton stays between two water molecules, $\mathrm{f_{sw}}(\mathrm{r_{ij}})$ of each hydrogen-oxygen pair becomes 1 (except for the atomic pair of shared hydrogen with oxygen, switching function becomes 0.5). The water "dipole moment" $(\mathrm{r_{H1}} + \mathrm{r_{H2}} - 2\mathrm{r_{O1}})$ from the first two terms of equation 5-8 was cancelled by the third term due to the pairwise switching function corresponding to $\mathrm{H_1} - \mathrm{O_1}$ and $\mathrm{H_2} - \mathrm{O_1}$ pairs. The switching function does not require any additional normalization factor since this expression is normalized upon construction. More explicitly, the coefficient of all atoms

should be exactly 1. The non-linear property of the mCEC function with respect to the atomic coordinates of the system arises merely from the switching function $f_{sw}(r_{ij})$.

The rCEC employs the diabatic state concept in MS-RMD theory, where it is defined as

$$r_{rCEC} = \sum_{i \in \{I\}} c_i^2 r_i^{COC} \tag{5-9}$$

and {I} denotes the collection of all possible diabatic states of the system, and r_i^{COC} is the position of the center of excess charge of the protonated species in the fixed bonding topology of the *i*th RMD diabatic state

$$r_i^{COC} = \frac{\sum_j |q_j| r_j}{\sum_j |q_j|} \tag{5-10}$$

where j runs over all atoms in the protonated molecule (e.g., all carboxylic hydroxyl atoms in the protonated aspartate or glutamate acid, or all four atoms in the hydronium).

The probability of the system remaining in the diabatic state I is denoted as c_i^2. The approximation made for c_i^2 is as follows,

$$\frac{c_j^2}{c_i^2} = \exp\left(-k\left(\delta_{ijk} - \delta_0\right)\right) \tag{5-11}$$

$$\sum_{i \in \{I\}} c_i^2 = 1 \tag{5-12}$$

where the proton-sharing indicator, $\delta_{ijk} = r_{jk} - r_{ik}$, corresponds to the extent of proton sharing (indicated as k) between the two diabatic state protonated species. The positive δ_{ijk} indicates the proton to be predominantly held by the protonated molecule in diabatic state i, while negative δ_{ijk} indicates the proton to be predominantly held by the protonated molecule in diabatic state j. The term r_{jk} represents the distance between the proton and the oxygen of the proton acceptor in

diabatic state i. Like mCEC, the rCEC function shows a non-linear dependence with respect to the atomic coordinates of the system due to the non-linear dependence of the probability c_i^2 of diabatic state $|\mathrm{i}\rangle$ to the atomic coordinates of the system.

The MS-RMD CEC, noted just as CEC here, shares the same expression as the rCEC, except for the definition of c_i. In the MS-RMD framework, the ground state of the system $|\psi\rangle$ can be expressed as a linear combination of distinct bonding topologies $|\mathrm{i}\rangle$ or diabatic states. The expansion coefficient of the ground state in relation to the diabatic state $|\mathrm{i}\rangle$ is defined as c_I, so the total linear combination is given by

$$|\psi\rangle = \sum_i \mathrm{c}_i \, |i\rangle \qquad 5\text{-}13$$

The c_i's are calculated on the fly during the RMD simulation by solving the eigenvalue problem with the Hamiltonian of the system, expressed as $\mathbf{H} = \sum_{ij} h_{ij} \, |i\rangle\langle j|$, such that

$$\mathbf{H}\,\mathbf{c} = \mathrm{E}_0\mathbf{c} \qquad 5\text{-}14$$

Here, h_{ii} is the potential energy of the system within the diabatic state $|i\rangle$, and h_{ij} is the coupling energy between diabatic states $|i\rangle$ and $|j\rangle$. E_0 is the solved eigenvalue corresponding to the ground state energy of the system as a function of all coordinates, and $\boldsymbol{c}$ is the ground state eigenvector from the diabatic-state basis set. It should be noted that the MS-RMD CEC function is not a linear-combination of atomic coordinates of the system as the on-the-fly solved ground state eigenvector $\boldsymbol{c}$ depends on those coordinates nonlinearly.

5.2.4 Effective TST Mass Factor for Water-Assisted Proton Transport

The mass factor in the TST rate expression for both 1D and 2D PMFs is based on the derivative of the proton-transfer CV to the atomic coordinates within, e.g., a protein-water-substrate complex. Since the CEC incorporated within the PT CV is already a curvilinear function of atomic

coordinates, the derivative value of the PT CV as a function of the atomic coordinates will vary based on the value of the CV. Thus, one can expect the ensemble average of $\sqrt{Z_\xi}$ sampled at a given reactant state to display a significant difference in magnitude in comparison to the (correct) one sampled at the transition state for a PT process, e.g., in proteins.

5.3 Results and Discussion

5.3.1 1D TST Results

The exact form of the TST expression for the effective mass factor was employed here for 1D PMFs of water assisted PT in two proton-coupled transporters.[75, 175] The impact of the approximations on the value of the mass factor was studied and compared to the exact value.

The *Piriformospora indica* phosphate transporter[9] (PiPT) is a proton-coupled phosphate transporter within the major facilitator superfamily[176] (MFS). The coupling relationship between PT and phosphate transport was studied in our recent work[75] via extensive classical simulations and QM/MM MD. The collective variable ξ that describes the PT from D324 to D45 in the QM/MM MD PMF (Figure 5-1A) is defined as a curvilinear function of multiple distances between AIMD rCEC, phosphate, D324 and D45 sidechain oxygens, such that

$$\xi = \vec{r}_{\mathrm{CEC-Pi}} \cdot \frac{\vec{r}_{\mathrm{D324-D45}}}{|\vec{r}_{\mathrm{D324-D45}}|} \qquad 5\text{-}15$$

The calculated mass factor using the exact expression $\left\langle\sqrt{Z_\xi}\right\rangle_\xi = \left\langle\sum_{i=1}^{3N}\frac{1}{M_i}\left(\frac{\partial\xi}{\partial r_i}\right)^2\right\rangle_\xi$ and the equipartition expression $\left\langle\sqrt{Z_\xi}\right\rangle_\xi = \sqrt{\beta\langle\dot{\xi}\rangle_\xi}$ sampled at both the reactant state and transition states in the PMF are provided in the chart of Figure 1B. Note that the orange and blue bars, which correspond to the exact and approximate equipartition approach, respectively, do not show evidence of a marked difference in the calculated value of the mass factor. However, there is a

non-negligible dependence of the mass factor and the position in the PMF when sampling the ensemble average. As shown in the chart, the calculated TST rate constant with these effective mass factors can be different by a factor as large as 4, which borders on a serious difference.

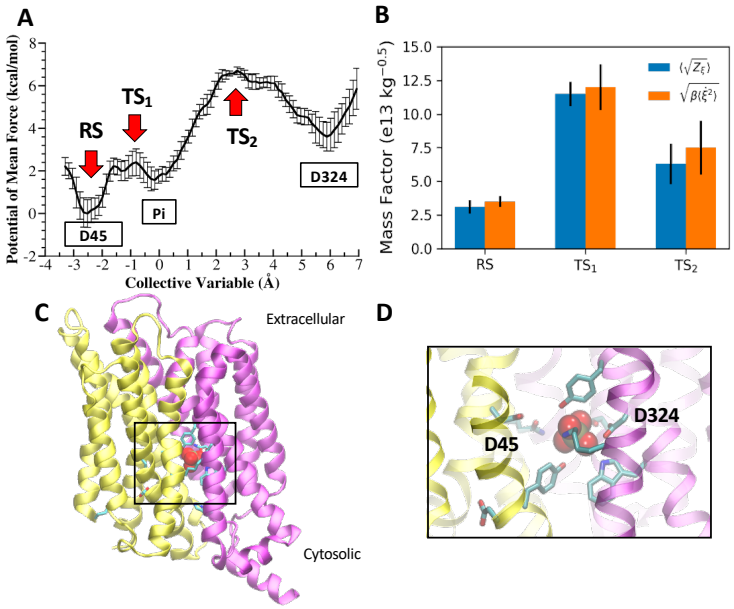


Figure 5-1. QM/MM MD proton transfer PMF in PiPT. A) The potential of mean force (PMF) of the PT process staring from D324 to D45 via waters and phosphate titration events. RS indicates the reactant state, TS_1 and TS_2 correspond to the transition states for the PT process from D45 to Pi and from D45 to D324; B) The calculated mass factor at different locations of the PMF stated in A) with the exact approach and the equipartition approach from the trajectories of the earlier work[75] (error bars are solid black lines); C) The side view of the PiPT protein crystal structure with phosphate bonded, the N domain helices are marked in yellow and C domain marked in purple; D) An enlarged view of the PT local environment with nearby residues represented as sticks and phosphate represented in VDW style.

We then compared the mass factor values from calculated from the different approaches for a proton-coupled peptide transporter, $PepT_{xc}$, in the peptide oligomer transporter family.[177] Our earlier computational work[178] incorporated QM/MM MD combined with enhanced free energy sampling techniques to determine the PT event that effects the inward-open to outward-open conformational transition of the protein. The PMF shown in Figure 5-2A is for the PT process from D322 to E425, which was suggested to trigger the conformational transition of the transporter.

The collective variable ξ that describes the PT from D322 to E425 in the QM/MM MD PMF (Figure 5-2A) is defined as the ratio of distances between the projected value of distance between the AIMD mCEC and D322 and the distance between D322 to E425 sidechain oxygens.

$$\xi = \frac{\mathrm{r_{CEC-D}} \cos \angle(\mathrm{CEC - D - E})}{\mathrm{r_{DE}}} \qquad 5\text{-}16$$

The mCEC marks the position of the proton, r_{DE} and $r_{\mathrm{CEC-D}}$ denote the distance between D322 and E425 and the distance between D322 and the excess proton position.

The calculated mass factor using the exact expression $\left\langle \sqrt{\mathrm{Z}_{\xi}} \right\rangle_{\xi} = \left\langle \sum_{i=1}^{3N} \frac{1}{M_i} \left(\frac{\partial \xi}{\partial r_i} \right)^2 \right\rangle_{\xi}$ and the equipartition expression $\left\langle \sqrt{\mathrm{Z}_{\xi}} \right\rangle_{\xi} = \sqrt{\beta \left\langle \dot{\xi} \right\rangle_{\xi}}$ sampled in the reactant state and at the transition states in the PMF are provided in Figure 2B. Note that the calculated mass factors exhibit a strong contrast between different sampling positions in the PMF, with the transition state mass factor being 7 times greater than the analogous reactant state value. Observed differences in the mass factor values calculated using the two methods were found to be marginally statistically significant; however, the difference of the effective mass factor calculated (correctly) at the transition state versus in the reactant well is still quite large, on the order of a factor of 5-8.

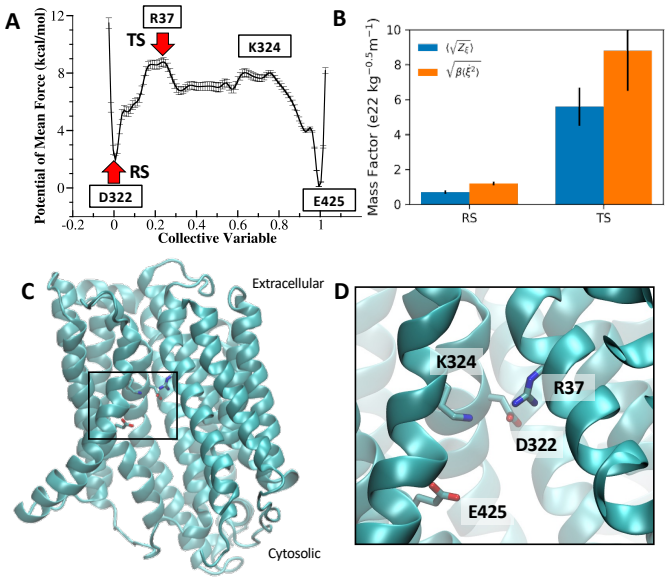


Figure 5-2. QM/MM MD proton transfer PMF in PepT. A) The potential of mean force (PMF) of the PT process stating from D322 to E425 via water wires. B) The calculated mass factor at reactant state and the transition state of the PMF stated in A) with the exact approach and the equipartition approach from the trajectories of the earlier work.[178] C) The side view of the PepT structure inside the membrane; D) An enlarged view of the PT local environment with nearby residues represented as sticks.

5.3.2 2D TST Results and the Role of Hydration as a Second CV

When the PT occurs throughout a pore or an otherwise narrow pathway within the protein, it is likely that the dynamics of the hydration process are strongly coupled with different stages of the PT charge translocation event.[23] Such cases can be identified from post-analysis of trajectories involving explicit PT, as each will show a distinct excess proton CEC position along its transport pathway. When there is a clear dependency between the trend of some water hydration indicator and different stages of the PT CV, either the water connectivity or the water density of the PT pathway needs to be additionally sampled as a CV to accurately capture the PT process.

Below, we have applied a derived 2D TST expression (equation 5-5, derivation see Appendix) to calculate the effective TST mass factor at different positions along the 2D PT PMF of the two protein systems. The ratio between this quantity in the reactant state and the transition ridge

indicates the extent to which prior approximations will impact the mass factor of the TST constant as the correct answer is the one involving the calculation at the transition ridge.

ClC-ec1 is a chloride-proton antiporter that also transports multiple anions with a distinct stoichiometry factor for each transported anion. Extensive experiments have been done in the past to interpret the co-transport mechanisms for each type of anion. Additionally, our prior computational studies[16-18, 20, 179, 180] have employed a range of approaches to analyze the molecular-level anion modulation mechanism with respect to PT. In general, the MS-RMD approach was proven to be a most suitable method for studying the explicit PT mechanisms as affected by different stages of the anion transport, with a crucial consideration of the proton-induced hydration effect.

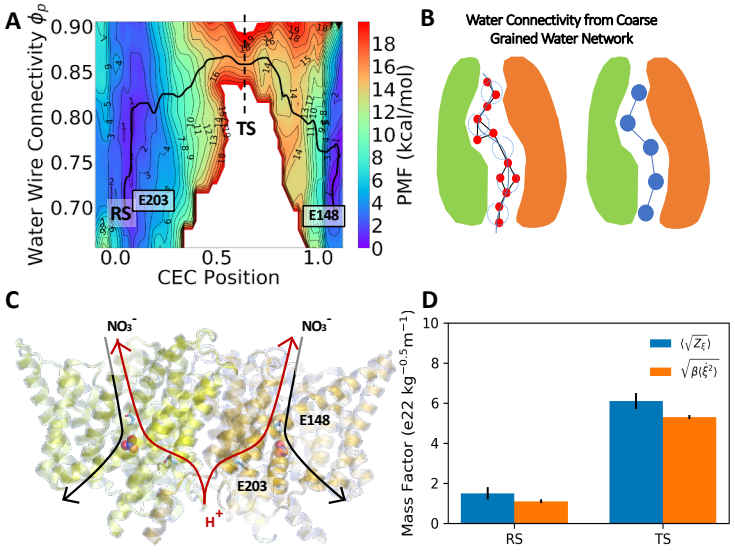


Figure 5-3. MS-RMD PT PMF in ClC-ec1 dimer. A) The potential of mean force (PMF) of the PT process starting from E203 to E148 with a nitrate anion bound in the vicinity. B) A schematic of the water wire connectivity CV. C) A sideview of the ClC-ec1 dimeric structure within the membrane with nitrate in a VDW representation. D) The calculated mass factor at the reactant state and the transition state of the PMF shown in A) with different calculated effective mass approaches based on trajectories from earlier work.[23]

Figure 5-3B depicts a novel hydration indicator CV, the water connectivity, combining graph theory and a coarse-grained representation of the water network within the protein, which was employed to depict the proton transport progress in the free energy landscape (Figure 5-3A).[23] Based on specifying the starting and end points of the water network, this water connectivity CV quantifies the extent of connectivity of the most-connected water wire associated with the explicit PT process. Here, the MS-RMD CEC was also employed to track the location of the transported excess proton charge defect on the fly, while the horizontal PT CV on the PMF, denoted by ξ, is denoted by the relative position of the excess proton CEC compared to the projected distance of the two Glu residues E203 and E148 along the direction perpendicular to the membrane plane. The expression of ξ is as follows,

$$\xi = \frac{\min_{i\in\{1,2\}} (\mathbf{r}_{\mathrm{CEC}} - \mathbf{r}_{\mathrm{E203},i}) \cdot \widehat{\mathrm{n_{PT}}}}{\min_{i,j\in\{1,2\}} (\mathbf{r}_{\mathrm{E148},j} - \mathbf{r}_{\mathrm{E203},i}) \cdot \widehat{\mathrm{n_{PT}}}} \qquad 5\text{-}17$$

where $\widehat{\mathrm{n_{PT}}}$ is a pre-defined unit vector pointing from E203 to E148, $\mathbf{r}_{\mathrm{CEC}}, \mathbf{r}_{\mathrm{E203,i}}, \mathbf{r}_{\mathrm{E148,j}}$ are the coordinates of the MS-RMD CEC position, from the ith and the jth carboxyl oxygen of E148 carboxyl oxygen of E203. From the derived expression in SI and with an assumption of a rectilinear transition ridge parallel to the second CV of the PMF, the mass factor for the transition-rate constant is independent of the gradient of the water connectivity CV to atomic coordinates in the system. The calculated mass factor from the derived exact form of the TST expression at the transition ridge ensemble in Figure 3D presents a 4:1 ratio to the one calculated in the reactant basin, which again highlights the significant flaw in the assumption that the PT CV will adopt the same effective mass between the reactant basin and the transition ridge in the PMF.

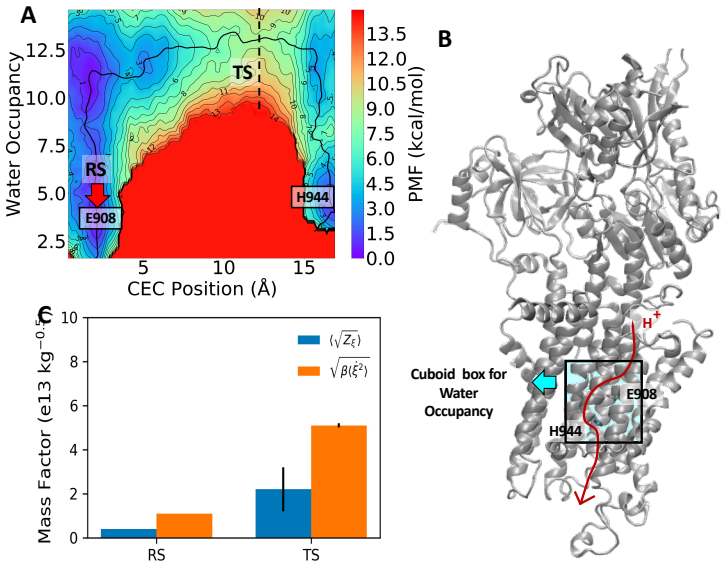


Figure 5-4. MS-RMD PT PMF in SERCA. A) The potential of mean force (PMF) of the PT process staring from E908 to H944. B) A sideview of the protein within the membrane; the red arrow points to the direction of PT within the protein across the membrane. The black box highlights the cuboid box boundary for the water occupancy CV. C) The calculated mass factor at the reactant state and the transition state of the PMF as seen in A) with different calculation approaches from the trajectories taken from earlier work.[21]

The sarcoplasmic reticulum Ca^{2+} -ATPase pump (SERCA) belongs to the P-type ATPase family, which carries out calcium-cation transport in cells to enable muscle contraction and maintain intracellular Ca^{2+} homeostasis. Additionally, SERCA transports multiple protons in the opposite direction of the outflow of Ca^{2+} in the endoplasmic reticulum / sarcoplasmic reticulum domain and is found to co-localize with several chloride-channel (ClC) family proteins. Our recent computational study[21] of explicit PT process in SERCA via the MS-RMD approach along with free energy sampling techniques confirmed the feasibility of PT from E908 to the luminal side.

The 2D MS-RMD PMF obtained via umbrella sampling (Figure 5-4A) illustrates PT from E908 to the H944 residue inside SERCA. The horizontal PT CV is defined as the projection of the MS-RMD CEC position along a curvilinear path. The discrete CV points that define the path were obtained from the sampled configurations from a Metadynamics[158-161] run. The water occupancy

within the cuboid box (shown in Figure 5-4B) was employed as the second CV for free energy sampling to capture the couple water hydration response to the excess proton position. The calculated TST effective mass factor calculated from either the exact form of the expression or the equipartition assumption expression sampled at either the reactant and (correctly) at the transition ridge on the PMF once again displays a difference of about a factor of 4 (Figure 5-4C). We note that the discrepancy in the mass factor value calculated by the two different methods was found to be a result from the post-analysis of the velocity of the first CV and the assumption of $\left\langle \sqrt{Z_\xi} \right\rangle_\xi = \sqrt{\langle Z_\xi \rangle_\xi}$.

5.4 Conclusions

A prefactor ratio of order 3-10 is considered to be significantly important in both a practical and a fundamental sense as the TST rate is proportional to the pre-factor value. The statistical error of the reaction PMF calculation can also be important for $e^{-\beta W(\xi^{TS})}$ in equation 5-4 which constitutes the overall TST rate constant value. The 1D PMF and 2D PMF error profiles are all around $k_B T$ in this work and this statistical error contribution from the free energy barrier can be estimated as $\exp -\beta \Delta W = \exp(\pm 1)$, thus ranging from 0.4 to 2.7 times to the overall TST rate value. The impact of the statistical error in the PMF can therefore be as important as the impact of any prefactor approximation in the overall TST rate calculation, but of course the PMF statistical error can always be better converged numerically whereas a systematic error in the calculation of the TST prefactor cannot be eliminated. Moreover, all approximations to the prefactor calculation will share the same calculated PMF. The correct manner of calculating the prefactor therefore matters for obtaining a correct overall TST rate value given a PMF of a given statistical significance.

In this work, we have derived the exact form of TST for the general *n*-dimensional case with curvilinear CVs and applied this approach to calculate the effective mass factor for PT reactions (with and without coupled hydration) in several biomolecular systems. These four applications to water-assisted PT within selected proteins exhibit up to nearly an order of magnitude difference in value between the actual (correct) effective mass factor calculated at the transition state dividing surface compared to one sampled, as an approximation, in the reactant state for 1D and 2D PMFs. Interestingly, the mass factor calculation was relatively insensitive to using an equipartition approximation. Our results thus highlight the significant impact of the position dependence of the effective mass factor in TST for PT processes and one that should not be taken lightly if quantitative accuracy is the goal in the calculated TST rate constant. The results described herein are therefore expected to provide a more accurate basis for the prediction of the PT rate in future studies of PT in biomolecular and other systems.

5.5 Appendix

5.5.1 Derivation of Transition State Rate Constant in Terms of an n-Dimensional PMF

The general expression of the rate constant when assuming the dividing surface of an N-dimensional PMF is defined as $\xi_1 = c$, is given by

$$k_{TST} = \frac{1}{Q^R}\int \frac{d\mathbf{q}\, d\mathbf{p}_q}{h^{3N}} \int d\boldsymbol{\xi} \int d\mathbf{p}_{\boldsymbol{\xi}}\, \delta(\xi_1 - c) e^{-\beta H}\, \Theta(\dot{\xi}_1)\dot{\xi}_1 \qquad 5\text{-}18$$

Here, $\boldsymbol{\xi} = (\xi)_i$ is the CV vector with dimension *n*, and ξ_1 is the PT CV related to the excess proton CEC., while $\mathbf{p}_\xi$ is the conjugate momentum vector of $\boldsymbol{\xi}$.

The Hamiltonian of the system can be expressed as a function of the $\boldsymbol{\xi}$ and $\mathbf{p}_\xi$ vectors as

$$H = \frac{1}{2}\mathbf{p}_\xi^T \mathbf{Z}_\xi \mathbf{p}_\xi + \frac{1}{2}\mathbf{p}_q^T \mathbf{Z}_q \mathbf{p}_q + V(\boldsymbol{\xi}, \mathbf{q}) \qquad 5\text{-}19$$

where $\mathbf{Z}_\xi$ and $\mathbf{Z}_q$ are the mass matrices for the CVs from the degrees of freedom. The definition of $\mathbf{Z}_\xi$ is given by equation 5-3.

The rate expression now becomes

$$k_{TST} = \frac{1}{Q^R}\int \frac{d\mathbf{q}\, d\mathbf{p}_q}{h^{3N}} \int d\boldsymbol{\xi}\ \delta(\xi_1 - c) e^{-\beta(\frac{1}{2}\mathbf{p}_q^T \mathbf{Z}_q \mathbf{p}_q + V(\boldsymbol{\xi},\mathbf{q}))} \int d\mathbf{p}_\xi\ \Theta(\dot{\xi}_1)\dot{\xi}_1 e^{-\beta\frac{1}{2}\mathbf{p}_\xi^T \mathbf{Z}_\xi \mathbf{p}_\xi} \qquad 5\text{-}20$$

Note that by incorporating

$$\dot{\boldsymbol{\xi}} = \frac{\partial H}{\partial \mathbf{p}_\xi} = \mathbf{Z}_\xi \mathbf{p}_\xi \qquad 5\text{-}21$$

one can transform the representation from $\mathbf{p}_\xi$ to $\dot{\boldsymbol{\xi}}$,

The transition matrix $\mathbf{A}_{\boldsymbol{\xi}}$ is the inverse matrix of $\mathbf{Z}_{\boldsymbol{\xi}}$, and is expressed in the following form

$$\mathbf{A}_n = \mathbf{Z}_\xi^{-1} = \begin{pmatrix} a & \mathbf{b}^T \\ \mathbf{b} & \mathbf{A}_{n-1} \end{pmatrix} \qquad 5\text{-}22$$

The kinetic term in the exponential now becomes

$$\frac{1}{2}\mathbf{p}_\xi^T \mathbf{Z}_\xi \mathbf{p}_\xi = \frac{1}{2}\dot{\boldsymbol{\xi}}^T \mathbf{Z}_\xi^{-1} \dot{\boldsymbol{\xi}} \qquad 5\text{-}23$$

From equation 5-21, 5-22 and 5-23, we have

$$\int d\mathbf{p}_\xi\ \Theta(\dot{\xi}_1)\dot{\xi}_1 e^{-\beta\frac{1}{2}\mathbf{p}_\xi^T \mathbf{Z}_\xi \mathbf{p}_\xi} = |\mathbf{A}_{\boldsymbol{\xi}}| \int d\dot{\xi}_n \dots \int d\dot{\xi}_2 \int_0^\infty d\dot{\xi}_1\, \dot{\xi}_1 e^{-\frac{\beta}{2}\dot{\boldsymbol{\xi}}^T \mathbf{Z}_\xi^{-1} \dot{\boldsymbol{\xi}}} \qquad 5\text{-}24$$

Using the Gaussian integral formula,

$$\int dx_n \dots \int dx_2 \int_0^\infty dx_1 x_1 \exp -\frac{1}{2}(\mathbf{x}^T A \mathbf{x}) = \left(\sqrt{2\pi}\right)^{n-1} \frac{\sqrt{|\mathbf{A}_{n-1}|}}{|\mathbf{A}_n|} \qquad 5\text{-}25$$

the integration over the velocity representation in equation 5-24 becomes

$$\int d\mathbf{p}_\xi\ \Theta(\dot{\xi}_1)\dot{\xi}_1 e^{-\beta\frac{1}{2}\mathbf{p}_\xi^T \mathbf{Z}_\xi \mathbf{p}_\xi} = |\mathbf{A}_n|\left(\sqrt{2\pi}\right)^{n-1} \frac{\sqrt{\beta^{n-1}|\mathbf{A}_{n-1}|}}{\beta^n |\mathbf{A}_n|} \qquad 5\text{-}26$$

$$= \left(\sqrt{\frac{2\pi}{\beta}}\right)^{n-1} \frac{1}{\beta}\sqrt{|\mathbf{A}_{n-1}|}$$

By then incorporating the Gaussian formula

$$\int d\mathbf{p}_\xi \; e^{-\beta\frac{1}{2}\mathbf{p}_\xi^T \mathbf{Z}_\xi \mathbf{p}_\xi} = \sqrt{\frac{(2\pi)^n |\mathbf{A}_n|}{\beta^n}} \tag{5-27}$$

the integral over the momentum on the left-hand side of equation 5-25 can be expressed as

$$\begin{aligned}\int d\mathbf{p}_\xi \; \Theta(\dot{\xi}_1)\dot{\xi}_1 \, e^{-\beta\frac{1}{2}\mathbf{p}_\xi^T \mathbf{Z}_\xi \mathbf{p}_\xi} &= \frac{\left(\sqrt{\frac{2\pi}{\beta}}\right)^{n-1} \frac{1}{\beta}\sqrt{\beta^n}\sqrt{|\mathbf{A}_{n-1}|}}{\left(\sqrt{2\pi}\right)^n \sqrt{|\mathbf{A}_n|}} \int d\mathbf{p}_\xi \; e^{-\beta\frac{1}{2}\mathbf{p}_\xi^T \mathbf{Z}_\xi \mathbf{p}_\xi} \\ &= \frac{1}{\sqrt{2\pi\beta}} \frac{\sqrt{|\mathbf{A}_{n-1}|}}{\sqrt{|\mathbf{A}_n|}} \int d\mathbf{p}_\xi \; e^{-\beta\frac{1}{2}\mathbf{p}_\xi^T \mathbf{Z}_\xi \mathbf{p}_\xi}\end{aligned} \tag{5-28}$$

By applying the inverse operator on both sides of equation 5-22,

$$\mathbf{Z}_\xi = \mathbf{A}_n^{-1} \tag{5-29}$$

and due to the fact that the specific element in an inverse matrix can be explicitly expressed with the matrix

$$\left(\mathbf{Z}_\xi\right)_{11} = \mathbf{Z}_{11} = \frac{|\mathbf{A}_{n-1}|}{|\mathbf{A}_n|} \tag{5-30}$$

we can substitute the integrand to obtain the same integration results in the rate expression as

$$\begin{aligned}k_{TST} = \frac{1}{Q^R} \int \frac{d\mathbf{q}\, d\mathbf{p}_q}{h^{3N}} \int d\boldsymbol{\xi} \; \delta(\xi_1 \\ - c) e^{-\beta(\frac{1}{2}\mathbf{p}_\xi^T \mathbf{Z}_\xi \mathbf{p}_\xi + V(\boldsymbol{\xi},\mathbf{q}))} \frac{1}{\sqrt{2\pi\beta}} \sqrt{Z_{11}} \int d\mathbf{p}_\xi \; e^{-\beta\frac{1}{2}\mathbf{p}_\xi^T \mathbf{Z}_\xi \mathbf{p}_\xi} \\ = \frac{1}{Q^R \sqrt{2\pi\beta}} \int \frac{d\mathbf{q}\, d\mathbf{p}_q d\boldsymbol{\xi} d\mathbf{p}_\xi}{h^{3N}} \; \delta(\xi_1 - c)\sqrt{Z_{11}} e^{-\beta H}\end{aligned} \tag{5-31}$$

$$= \frac{1}{\mathrm{Q^R}\sqrt{2\pi\beta}} \frac{\int \frac{\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}_\mathrm{q}\mathrm{d}\boldsymbol{\xi}\mathrm{d}\mathbf{p}_\xi}{h^{3N}}\,\delta(\xi_1-\mathrm{c})\sqrt{\mathrm{Z}_{11}}\mathrm{e}^{-\beta\mathrm{H}}}{\int \frac{\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}_\mathrm{q}\mathrm{d}\boldsymbol{\xi}\mathrm{d}\mathbf{p}_\xi}{h^{3N}}\,\delta(\xi_1-\mathrm{c})\mathrm{e}^{-\beta\mathrm{H}}}$$

$$\times \int \frac{\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}_\mathrm{q}\mathrm{d}\boldsymbol{\xi}\mathrm{d}\mathbf{p}_\xi}{h^{3N}}\,\delta(\xi_1-\mathrm{c})\mathrm{e}^{-\beta\mathrm{H}}$$

$$= \frac{1}{\mathrm{Q^R}\sqrt{2\pi\beta}}\left\langle\sqrt{\mathrm{Z}_{11}}\right\rangle_{\xi_1=\mathrm{c}} \int \frac{\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}_\mathrm{q}\mathrm{d}\boldsymbol{\xi}\mathrm{d}\mathbf{p}_\xi}{\mathrm{h}^{3\mathrm{N}}}\,\delta(\xi_1-\mathrm{c})\,\mathrm{e}^{-\beta\mathrm{H}}$$

$$= \frac{1}{\sqrt{2\pi\beta}}\left\langle\sqrt{\mathrm{Z}_{11}}\right\rangle_{\xi_1=\mathrm{c}} \frac{\int \mathrm{d}\boldsymbol{\xi}\,\mathrm{e}^{-\beta\mathrm{W}(\xi_1=\mathrm{c})}}{\mathrm{Q^R}}$$

where $\mathrm{Z}_{11} = \sum_{\mathrm{k}=1}^{3\mathrm{N}} \frac{1}{\mathrm{M_k}}\left(\frac{\partial\xi_1}{\partial \mathrm{r_k}}\right)^2$ is the inverse value of the effective mass of the first CV.

5.5.2 Calculation of the ensemble average of $\langle\sqrt{\mathrm{Z}_\xi}\rangle$ among transition states in terms of a 2D PMF

First, let's review WHAM scheme to obtain a 1D PMF for the transition ridge $\xi_1 = \mathrm{c}$ in the 2D PMF. WHAM here wants to maximize the likelihood of probability summing over window k and bin i of ξ_2 axis at transition ridge $\xi_1 = \mathrm{c}$.

Statement 1:

Minimize $-\ln L_{WHAM} = -\sum_k \sum_i \ln\left(\pi_i^k\right)^{N_i^k} = -\sum_k \sum_i N_i^k \ln\left(f^k \gamma_i^k \pi_i\right)$ so that $\sum_i \pi_i = 1$.

Here, $\pi_i = P(\xi_2 = (\xi_2)_i \mid \xi_1 = \mathrm{c})$ is the conditional probability of ξ_2 falling into bin i out of whole transition ridge $\xi_1 = c$. π_i^k is the biased probability at simulation condition k that falls into bin i. π_i and π_i^k are related by the known umbrella potential added, shown as $\pi_i^k = f^k \gamma_i^k \pi_i$. Here, γ_i^k is the constant bias factor caused by adding umbrella window k.$\gamma_i^k = e^{-\kappa\left((\xi_2)_i - (\mathrm{center}^k)\right)^2}$. $f^k = \frac{1}{\sum_l \pi_l \gamma_l^k}$ is the normalization factor that ensures $\sum_i \pi_i^k = 1$.

Statement 2:

$$\max_{v}\left[\min_{\pi_i}\left(-\ln L_{\mathrm{WHAM}}+v*\left(\sum_i \pi_i-1\right)\right)\right]$$

Statement 1 is equal to statement 2 due to Lagrange Duality theory.

To solve $\min_{\pi_i}(-\ln L_{\mathrm{WHAM}}+v\cdot(\sum_i \pi_i-1))$, we do derivative over π_i for all possible i.

$$\frac{\partial}{\partial \pi_i}\left(-\sum_k\sum_j N_j^k \ln\left(\frac{\gamma_j^k\pi_j}{\sum_l \pi_l\gamma_l^k}\right)+v\cdot\left(\sum_j \pi_j-1\right)\right)=0 \quad \forall\, i \tag{5-32}$$

$$\frac{\partial}{\partial \pi_i}\left(-\sum_k\sum_j\left[N_j^k\cdot\ln\gamma_j^k\pi_j-N_j^k\cdot\ln\sum_l \pi_l\gamma_l^k\right]+v\cdot\sum_j \pi_j-v\right)=0 \quad \forall\, i \tag{5-33}$$

$$-\sum_k\frac{N_i^k}{\pi_i}+\sum_k\left(\left(\sum_j N_j^k\right)\cdot\frac{\gamma_i^k}{\sum_l \pi_l\gamma_l^k}\right)+v=0 \quad \forall\, i \tag{5-34}$$

$$v=\frac{1}{\pi_i}\sum_k N_i^k-\sum_k\left(\left(\sum_j N_j^k\right)\cdot\gamma_i^k f^k\right) \tag{5-35}$$

This equation stands for all π_i. Then we need to get v that maximize the following expression

$$\max_{v}\left[-\sum_k\sum_j N_j^k\ln\left(\frac{\gamma_j^k\pi_j}{\sum_l \pi_l\gamma_l^k}\right)+v\cdot\left(\sum_j \pi_j-1\right)\right] \tag{5-36}$$

It's obvious to get that $v=0$. The solution is shown below,

$$\pi_i=\frac{\sum_k N_i^k}{\sum_k\left(\left(\sum_j N_j^k\right)\cdot\gamma_i^k f^k\right)}$$
$$f^k=\frac{1}{\sum_l \pi_l\gamma_l^k} \tag{5-37}$$

Now we want to calculate the ensemble average for $\langle A\rangle_{\xi_1=c}$ in the 2D PMF, where the physical quantity happens to be $\sqrt{Z_{\xi_{11}}}$.

$$\langle A\rangle_{\xi_1=c} = \sum_i \left(P(\xi_2 = (\xi_2)_i \mid \xi_1 = \mathrm{c}) \cdot \langle A\rangle_{i,\xi_1=\mathrm{c}}\right) \tag{5-38}$$

Here, $P(\xi_2 = (\xi_2)_i \mid \xi_1 = \mathrm{c}) = \pi_i = \frac{e^{-\beta E_i}}{\sum_j e^{-\beta E_j}} = \frac{\sum_k N_i^k}{\sum_k \left(\left(\sum_j N_j^k\right)\cdot \gamma_i^k f^k\right)}$.

Notice that $\langle A\rangle_i^l(\mathrm{unbiased})$ is the unbiased, reweighted ensemble average derived from quantity A sampled at window l and falls into bin i, $\langle A\rangle_i^l(\mathrm{biased})$ is the biased ensemble average of quantity A directly sampled at window l and falls into bin i

$$\langle A\rangle_{i,\xi_1=\mathrm{c}} = \sum_l \left(w_{il} \cdot \langle A\rangle_i^l(\mathrm{unbiased})\right) \tag{5-39}$$

$$\langle A\rangle_i^l(\mathrm{unbiased}) = \frac{\sum_t A_i^j(t) e^{+\beta V_i^l(t)}}{\sum_t e^{+\beta V_i^l(t)}} = \frac{\int \mathrm{d}\xi_1 \int \mathrm{d}\xi_2 \int \mathrm{d}\boldsymbol{r}^{3N-2} A_i^j(t) e^{-\beta\left(\mathrm{H}+V_i^j\right)} e^{+\beta V_i^j}}{\int \mathrm{d}\xi_1 \int \mathrm{d}\xi_2 \int \mathrm{d}\boldsymbol{r}^{3N-2} e^{-\beta\left(\mathrm{H}+V_i^j\right)} e^{+\beta V_i^j}} \tag{5-40}$$

$$\langle A\rangle_i^l(\mathrm{biased}) = \frac{1}{T}\sum_t A_i^j(t) = \frac{\int \mathrm{d}\xi_1 \int \mathrm{d}\xi_2 \int \mathrm{d}\boldsymbol{r}^{3N-2} A_i^j(t) e^{-\beta\left(\mathrm{H}+V_i^j\right)}}{\int \mathrm{d}\xi_1 \int \mathrm{d}\xi_2 \int \mathrm{d}\,\boldsymbol{r}^{3N-2} e^{-\beta\left(\mathrm{H}+V_i^j\right)}} \tag{5-41}$$

Since $V_i^l(t)$ is a time-independent potential, we have

$$\langle A\rangle_i^l(\mathrm{unbiased}) = \langle A\rangle_i^l(\mathrm{biased}) \tag{5-42}$$

Hence,

$$\langle A\rangle_{i,\xi_1=\mathrm{c}} = \sum_l \left(w_{il} \cdot \langle A\rangle_i^l(\mathrm{unbiased})\right) = \sum_l \left(w_{il} \cdot \frac{1}{T}\sum_t A_i^l(t)\right) \tag{5-43}$$

To solve w_l, we look at the probability of bin i and its decomposition to the probability of this bin contributing from window l.

$$\pi_i = \sum_l \left(w_{il} \cdot \pi_i^l(\text{unbiased}) \right) = \sum_l \left(w_{il} \cdot \frac{\frac{N_i^l}{\gamma_i^l}}{\sum_j \frac{N_j^l}{\gamma_j^l}} \right) \quad 5\text{-}44$$

From WHAM derivation in the first section, we know that

$$\pi_i = \frac{\sum_l N_i^l}{\sum_k \left(\left(\sum_j N_j^k \right) \cdot \gamma_i^k f^k \right)} = \sum_l \frac{N_i^l}{\sum_k \left(\left(\sum_j N_j^k \right) \cdot \gamma_i^k f^k \right)} \quad 5\text{-}45$$

Hence,

$$w_{il} \cdot \frac{\frac{N_i^l}{\gamma_i^l}}{\sum_j \frac{N_j^l}{\gamma_j^l}} = \frac{N_i^l}{\sum_k \left(\left(\sum_j N_j^k \right) \cdot \gamma_i^k f^k \right)} \quad 5\text{-}46$$

$$w_{il} = \frac{\gamma_i^l \cdot \sum_j \frac{N_j^l}{\gamma_j^l}}{\sum_k \left(\left(\sum_j N_j^k \right) \cdot \gamma_i^k f^k \right)} = \frac{\pi_i}{\sum_l N_i^l} \cdot \gamma_i^l \cdot \sum_j \frac{N_j^l}{\gamma_j^l} \quad 5\text{-}47$$

To sum up,

$$\begin{aligned} \left\langle \sqrt{Z_\xi} \right\rangle_{\xi_1 = \mathrm{c}} &= \sum_i \left(P(\xi_2 = (\xi_2)_i \mid \xi_1 = \mathrm{c}) \cdot \langle A \rangle_{i, \xi_1 = \mathrm{c}} \right) \\ &= \sum_i \left(\pi_i \cdot \sum_l \left(w_{il} \cdot \frac{1}{T} \sum_t A_i^l(t) \right) \right) \\ &= \sum_i \left[\frac{\pi_i^2}{\sum_l N_i^l} \cdot \sum_l \left(\gamma_i^l \cdot \left(\sum_j \frac{N_j^l}{\gamma_j^l} \right) \cdot \frac{1}{T} \sum_t \left(\sqrt{Z_\xi} \right)_i^l (t) \right) \right] \end{aligned} \quad 5\text{-}48$$

Let us clarify the calculation of $\sqrt{Z_\xi}$ on a given frame of the trajectory mentioned previously

The inverse value of effective mass of ξ is defined as

$$Z_\xi \coloneqq \sum_{i=1}^{3N} \frac{1}{M_i}\left(\frac{\partial \xi}{\partial r_i}\right)^2 \tag{5-49}$$

where N is the number of atoms in the system, M_i is the mass of atom i, r_i is the Cartesian coordinate (either x, or y, or z dimension) of atom i.

$$\frac{\partial \xi}{\partial r_i} = \left(\frac{\partial \xi}{\partial V_{\text{bias}}}\right)\left(\frac{\partial V_{\text{bias}}}{\partial r_i}\right) \tag{5-50}$$

Since

$$\frac{\partial \xi}{\partial V_{\text{bias}}} = -\left(\frac{\partial V_{\text{bias}}(\xi)}{\partial \xi}\right)^{-1} = -\left(\frac{\partial\left[\frac{1}{2}\kappa(\xi - s)^2\right]}{\partial \xi}\right)^{-1} = -\frac{1}{\kappa(\xi - s)}$$

$$\frac{\partial V_{\text{bias}}}{\partial r_i} = -F_{r_i}^{\text{bias}} \tag{5-51}$$

Here, $F_{r_i}^{\text{bias}}$, the bias force on each atom along r_i direction, is provided by PUMED package, dump-force command.

So, we have

$$\frac{\partial \xi}{\partial r_i} = \left(\frac{\partial \xi}{\partial V_{\text{bias}}}\right)\left(\frac{\partial V_{\text{bias}}}{\partial r_i}\right) = \frac{F_{r_i}^{\text{bias}}}{\kappa(\zeta - s)} \tag{5-52}$$

$$Z_\xi = \sum_{i=1}^{3N} \frac{1}{M_i}\left(\frac{\partial \xi}{\partial r_i}\right)^2 = \sum_{i=1}^{3N} \frac{1}{M_i}\left(\frac{F_{r_i}^{\text{bias}}}{\kappa(\xi - s)}\right)^2 \tag{5-53}$$

$$\sqrt{Z_\xi} = \sqrt{\sum_{i=1}^{3N} \frac{1}{M_i}\left(\frac{F_{r_i}^{\text{bias}}}{\kappa(\xi - s)}\right)^2} \tag{5-54}$$

5.5.3 The Proof of equipartition theorem $\langle \mathbf{Z_{11}} \rangle = \boldsymbol{\beta}\langle \dot{\boldsymbol{\xi}}_1^2 \rangle$

$$\beta\langle \dot{\xi}_1^2 \rangle = \int \frac{\mathrm{d}\mathbf{q}\, \mathrm{d}\mathbf{p}_\mathrm{q}}{\mathrm{h}^{3\mathrm{N}}} \int \mathrm{d}\boldsymbol{\xi}\ \beta\, \mathrm{e}^{-\beta\left(\frac{1}{2}\mathbf{p}_\mathrm{q}^\mathrm{T}\mathbf{Z}_\mathrm{q}\mathbf{p}_\mathrm{q} + \mathrm{V}(\boldsymbol{\xi},\mathbf{q})\right)} \int \mathrm{d}\mathbf{p}_\xi\, \dot{\xi}_1^2\, \mathrm{e}^{-\beta\frac{1}{2}\dot{\boldsymbol{\xi}}^\mathrm{T}\mathbf{Z}_\xi^{-1}\dot{\boldsymbol{\xi}}} \tag{5-55}$$

$$= \int \frac{\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}_\mathrm{q}}{\mathrm{h}^{3\mathrm{N}}} \int \mathrm{d}\boldsymbol{\xi}\ \beta\, \mathrm{e}^{-\beta\left(\frac{1}{2}\mathbf{p}_\mathrm{q}^\mathrm{T}\mathbf{Z}_\mathrm{q}\mathbf{p}_\mathrm{q}+\mathrm{V}(\boldsymbol{\xi},\mathbf{q})\right)} \int \mathrm{d}\dot{\boldsymbol{\xi}}\,|\mathbf{A}|\,\dot{\xi}_1^2\ \mathrm{e}^{-\beta\frac{1}{2}\dot{\boldsymbol{\xi}}^\mathrm{T}\mathbf{Z}_\xi^{-1}\dot{\boldsymbol{\xi}}}$$

$$= \int \frac{\mathrm{d}\mathbf{q}}{\mathrm{h}^{3\mathrm{N}}} \int \mathrm{d}\boldsymbol{\xi}\ \ \mathrm{e}^{-\beta\mathrm{V}(\boldsymbol{\xi},\mathbf{q})} \sqrt{\frac{(2\pi)^n}{\beta|\mathbf{A}|}}\,|\mathbf{A}_{n-1}|$$

The last line of equation 5-55 employs the integral formula

$$\int \mathrm{d}\mathbf{x}\,\mathrm{x}_1^2\ \mathrm{e}^{-\frac{\beta}{2}\mathbf{x}^\mathrm{T}\mathbf{A}\mathbf{x}} = \sqrt{\frac{(2\pi)^n}{\beta^3|\mathbf{A}|^3}}\,|\mathbf{A}_{n-1}| \tag{5-56}$$

Plugging in equation 5-30, we get

$$\beta\langle\dot{\xi}_1^2\rangle = \int \frac{\mathrm{d}\mathbf{q}}{\mathrm{h}^{3\mathrm{N}}} \int \mathrm{d}\boldsymbol{\xi}\ \ \mathrm{e}^{-\beta\mathrm{V}(\boldsymbol{\xi},\mathbf{q})} \sqrt{\frac{(2\pi)^n|\mathbf{A}|}{\beta}}\,\mathbf{Z}_{11} \tag{5-57}$$

Since $\mathbf{A} = \mathbf{Z}^{-1}$ and $\left|\mathbf{Z}^{-1}\right| = |\mathbf{Z}|^{-1}$

$$\beta\langle\dot{\xi}_1^2\rangle = \int \frac{\mathrm{d}\mathbf{q}}{\mathrm{h}^{3\mathrm{N}}} \int \mathrm{d}\boldsymbol{\xi}\ \ \mathrm{e}^{-\beta\mathrm{V}(\boldsymbol{\xi},\mathbf{q})} \sqrt{\frac{(2\pi)^n}{\beta|\mathbf{Z}_\xi|}}\,\mathbf{Z}_{11} \tag{5-58}$$

Now, we look at the derivation of $\langle\mathbf{Z}_{11}\rangle$

$$\begin{aligned}\langle\mathbf{Z}_{11}\rangle &= \int \frac{\mathrm{d}\mathbf{q}}{\mathrm{h}^{3\mathrm{N}}} \int \mathrm{d}\boldsymbol{\xi}\ \mathbf{Z}_{11}\mathrm{e}^{-\beta(\frac{1}{2}\mathbf{p}_\mathrm{q}^\mathrm{T}\mathbf{Z}_\mathrm{q}\mathbf{p}_\mathrm{q}+\mathrm{V}(\boldsymbol{\xi},\mathbf{q}))} \int \mathrm{d}\mathbf{p}_\xi\,\mathrm{e}^{-\beta\frac{1}{2}\mathbf{p}_\xi^\mathrm{T}\mathbf{Z}_\xi\mathbf{p}_\xi} \\ &= \int \frac{\mathrm{d}\mathbf{q}}{\mathrm{h}^{3\mathrm{N}}} \int \mathrm{d}\boldsymbol{\xi}\ \mathbf{Z}_{11}\left(\sqrt{\frac{(2\pi)^n}{\beta|\mathbf{Z}_\xi|}}\right)\mathrm{e}^{-\beta\mathrm{V}(\boldsymbol{\xi},\mathbf{q})}\end{aligned} \tag{5-59}$$

Hence,

$$\langle\mathbf{Z}_{11}\rangle = \beta\langle\dot{\xi}_1^2\rangle \tag{5-60}$$

6 Concluding remark

A series of studies have been conducted to unravel the intricate mechanisms of the proton-coupled phosphate transporter and the voltage-gated proton channel at a molecular level.

In Chapter 2, extensive classical MD simulations with different protonation state assigned to protein and phosphate were conducted to infer different stages of the functional cycle. Extensive free energy sampling using QM/MM MD simulations has been performed to justify at a detailed molecular level the hypothesis that the deprotonation of the key residue D324 triggers phosphate release. This work successfully answers the questions of where and how the proton on D324 travels vectorially across the membrane during the phosphate release process. The quantitative QM/MM PMFs confirm D45 as the starting residue of the PT tunnel and offer a perspective on proton-coupled trans- porters in which titratable substrates such as phosphate may be explicitly involved in the PT process as a key part of driving the functional cycle. Our design of an appropriate CV together with QM/MM MD simulations proves to be critical to the study of such complicated systems with multiple titratable sites along the transport pathway.

In Chapter 3, multiscale reactive molecular dynamics simulations in conjunction with an enhanced sampling approach were employed to examine the elementary proton transport reactions. Furthermore, classical simulations were performed to explore various phosphate release processes, again coupled with an enhanced sampling technique to ascertain the timescale of the final phosphate release process. Each elementary reaction step in the kinetic network model was examined by the calculated free energy profile and the forward and backward reaction rate constants associated were deter-mined by transition state theory. Utilizing a bottom-up framework, this work helps construct a comprehensive understanding of the myriad potential reaction pathways facilitating the transition from phosphate-bound OC states to phosphate-released IO

states. This holistic approach allowed us to examine the impact of environmental factors on reaction flux. The PiPT transporter was found to be able to adaptively handle various forms of phosphate under differing conditions. Interestingly, the primary pathway for dibasic phosphate transport diverges slightly from our initial proposal stated in Chapter 2. In this updated scenario, the deprotonation event at D324 is directly succeeded by a joint transport of proton and phosphate via the newly opened cytosolic gate, rather than a separate proton transport through the previously identified proton exit tunnel. The kinetic network model revealed the co-transport of proton and phosphate to be bidirectional and can be driven by either a positive phosphate gradient or a positive proton gradient across the membrane. The model's ability to reproduce the optimal pH conditions under experimental setup stems from its control over the equilibrium between D45 and phosphate titrations.

With detailed molecular mechanism examined for the functional transition from the phosphate-bound occluded state to phosphate-release inward open state, the next step could be exploring the transformation from outward open state to occluded state and identify the sequential step of PT related to phosphate binding.

In Chapter 4, extensive free energy sampling with MS-RMD was performed for the full PT process in the hHv1 channel mediated by pore water and all pore lining Asp and Glu residues, without any prior bias towards a specific PT mechanism. Moreover, the permeability investigations extend beyond proton transport and include the examination of free energy profiles of cation and anion permeation through the channel. Remarkably, the calculations on proton and other ion permeability exhibit excellent agreement with experimental measurements on ion selectivity and proton conductance, providing a quantitative understanding of the proton selectivity

mechanism and marks a significant step towards unraveling the fundamental functional mechanism of voltage-gated proton channels.

In Chapter 5, we have derived the exact form of TST for the general n-dimensional case with curvilinear CVs and applied this approach to calculate the effective mass factor for PT reactions (with and without coupled hydration) in several biomolecular systems. These four applications to water-assisted PT within selected proteins exhibit up to nearly an order of magnitude difference in value between the actual (correct) effective mass factor calculated at the transition state dividing surface compared to one sampled, as an approximation, in the reactant state for 1D and 2D PMFs. Interestingly, the mass factor calculation was relatively insensitive to using an equipartition approximation. Our results thus highlight the significant impact of the position dependence of the effective mass factor in TST for PT processes and one that should not be taken lightly if quantitative accuracy is the goal in the calculated TST rate constant. The results described herein are therefore expected to provide a more accurate basis for the prediction of the PT rate in future studies of PT in biomolecular and other systems.

Moving forward, an intriguing research topic could involve exploring the mechanism by which the hHv1 proton channel inhibits ion conduction, including its highly selective protons, under hyperpolarized membrane potential conditions. With an activated (depolarized) membrane potential, the hHv1 channel undergoes a significant conformational shift, where its 4th helix moves up to 10 Å along the membrane's normal direction. To better understand this channel's proton selectivity and conduction properties, it would be beneficial to carry out extensive free energy sampling on proton permeation under different stages of gate-opening structures. This could shed light on the dynamic behaviors of the channel.

References

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

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

3. 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.

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

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

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

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

8. 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.

9. Pedersen, B. P.; Kumar, H.; Waight, A. B.; Risenmay, A. J.; Roe-Zurz, Z.; Chau, B. H.; Schlessinger, A.; Bonomi, M.; Harries, W.; Sali, A.; Johri, A. K.; Stroud, R. M., Crystal structure of a eukaryotic phosphate transporter. *Nature* **2013,** *496* (7446), 533-536.

10. DeCoursey, T. E., The Voltage-Gated Proton Channel: A Riddle, Wrapped in a Mystery, inside an Enigma. *Biochemistry* **2015,** *54* (21), 3250-3268.

11. DeCoursey, T. E., Voltage-Gated Proton Channels: Molecular Biology, Physiology, and Pathophysiology of the HV Family. *Physiological Reviews* **2013,** *93* (2), 599-652.

12. 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. *Proceedings of the National Academy of Sciences* **2014,** *111* (26), 9396-9401.

13. Liang, R.; Swanson, J. M. J.; Madsen, J. J.; Hong, M.; DeGrado, W. F.; Voth, G. A., Acid activation mechanism of the influenza A M2 proton channel. *Proceedings of the National Academy of Sciences* **2016,** *113* (45), E6955.

14. Liang, R.; Swanson, J. M.; Peng, Y.; Wikström, M.; Voth, G. A., Multiscale simulations reveal key features of the proton-pumping mechanism in cytochrome c oxidase. *Proceedings of the National Academy of Sciences* **2016,** *113* (27), 7420-7425.

15. Liang, R.; Swanson, J. M. J.; Wikström, M.; Voth, G. A., Understanding the essential proton-pumping kinetic gates and decoupling mutations in cytochrome c oxidase. *Proceedings of the National Academy of Sciences* **2017,** *114* (23), 5924.

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

17. 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. *Journal of the American Chemical Society* **2016,** *138* (45), 14923-14930.

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

19. Peng, Y.; Swanson, J. M. J.; Kang, S.-g.; Zhou, R.; Voth, G. A., Hydrated Excess Protons Can Create Their Own Water Wires. *The Journal of Physical Chemistry B* **2015,** *119* (29), 9212-9218.

20. 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. *Journal of the American Chemical Society* **2018,** *140* (5), 1793-1804.

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

22. Ma, X.; Li, C.; Martinson, A. B.; Voth, G. A., Water-Assisted Proton Transport in Confined Nanochannels. *The Journal of Physical Chemistry C* **2020,** *124* (29), 16186-16201.

23. Li, C.; Voth, G. A., A quantitative paradigm for water-assisted proton transport through proteins and other confined spaces. *Proceedings of the National Academy of Sciences* **2021,** *118* (49), e2113141118.

24. Gu, M.; Chen, A.; Sun, S.; Xu, G., Complex Regulation of Plant Phosphate Transporters and the Gap between Molecular Mechanisms and Practical Application: What Is Missing? *Molecular Plant* **2016,** *9* (3), 396-416.

25. Russo-Abrahão, T.; Alves-Bezerra, M.; Majerowicz, D.; Freitas-Mesquita, A. L.; Dick, C. F.; Gondim, K. C.; Meyer-Fernandes, J. R., Transport of inorganic phosphate in Leishmania infantum and compensatory regulation at low inorganic phosphate concentration. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2013,** *1830* (3), 2683-2689.

26. Russo-Abrahão, T.; Koeller, C. M.; Steinmann, M. E.; Silva-Rito, S.; Marins-Lucena, T.; Alves-Bezerra, M.; Lima-Giarola, N. L.; de-Paula, I. F.; Gonzalez-Salgado, A.; Sigel, E.; Bütikofer, P.; Gondim, K. C.; Heise, N.; Meyer-Fernandes, J. R., H+-dependent inorganic

phosphate uptake in *Trypanosoma brucei* is influenced by *myo*-inositol transporter. *Journal of Bioenergetics and Biomembranes* **2017,** *49* (2), 183-194.

27. Lacerda-Abreu, M. A.; Russo-Abrahão, T.; Cosentino-Gomes, D.; Nascimento, M. T. C.; Carvalho-Kelly, L. F.; Gomes, T.; Rodrigues, M. F.; König, S.; Rumjanek, F. D.; Monteiro, R. Q.; Meyer-Fernandes, J. R., H+-dependent inorganic phosphate transporter in breast cancer cells: Possible functions in the tumor microenvironment. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **2019,** *1865* (9), 2180-2188.

28. Ito, M.; Matsuka, N.; Izuka, M.; Haito, S.; Sakai, Y.; Nakamura, R.; Segawa, H.; Kuwahata, M.; Yamamoto, H.; Pike, W. J.; Miyamoto, K.-i., Characterization of inorganic phosphate transport in osteoclast-like cells. *American Journal of Physiology-Cell Physiology* **2005,** *288* (4), C921-C931.

29. Pao, S. S.; Paulsen, I. T.; Saier, M. H., Major facilitator superfamily. *Microbiology and Molecular Biology Reviews* **1998,** *62* (1), 1.

30. Ullrich-Eberius, C. I.; Novacky, A.; Fischer, E.; Lüttge, U., Relationship between energy-dependent phosphate uptake and the electrical membrane potential in *Lemna gibba* G1. *Plant Physiol* **1981,** *67* (4), 797-801.

31. Ullrich-Eberius, C. I.; Novacky, A.; van Bel, A. J. E., Phosphate uptake in *Lemna gibba* G1: Energetics and kinetics. *Planta* **1984,** *161* (1), 46-52.

32. Sakano, K., Proton/Phosphate stoichiometry in uptake of inorganic phosphate by cultured cells of *Catharanthus roseus* (L.) G. Don. *Plant Physiol* **1990,** *93* (2), 479.

33. Samyn, D. R.; Van der Veken, J.; Van Zeebroeck, G.; Persson, B. L.; Karlsson, B. C. G., Key residues and phosphate release routes in the *Saccharomyces cerevisiae* Pho84 transceptor: The role of Tyr179 in functional regulation. *Journal of Biological Chemistry* **2016,** *291* (51), 26388-26398.

34. Samyn, D. R.; Ruiz-Pavon, L.; Andersson, M. R.; Popova, Y.; Thevelein, J. M.; Persson, B. L., Mutational analysis of putative phosphate- and proton-binding sites in the *Saccharomyces cerevisiae* Pho84 phosphate:H+ transceptor and its effect on signalling to the PKA and PHO pathways. *Biochemical Journal* **2012,** *445*, 413-422.

35. Ceasar, S. A.; Baker, A.; Muench, S. P.; Ignacimuthu, S.; Baldwin, S. A., The conservation of phosphate-binding residues among PHT1 transporters suggests that distinct transport affinities are unlikely to result from differences in the phosphate-binding site. *Biochemical Society Transactions* **2016,** *44*, 1541-1548.

36. Bibi, E.; Kaback, H. R., *In vivo* expression of the *lacY* gene in two segments leads to functional *lac* permease. *Proc Natl Acad Sci U.S.A* **1990,** *87* (11), 4325-4329.

37. Zen, K. H.; McKenna, E.; Bibi, E.; Hardy, D.; Kaback, H. R., Expression of lactose permease in contiguous fragments as a probe for membrane-spanning domains. *Biochemistry* **1994,** *33* (27), 8198-8206.

38. Weinglass, A. B.; Kaback, H. R., The central cytoplasmic loop of the major facilitator superfamily of transport proteins governs efficient membrane insertion. *Proc Natl Acad Sci U.S.A* **2000,** *97* (16), 8938.

39. Lundh, F.; Mouillon, J.-M.; Samyn, D.; Stadler, K.; Popova, Y.; Lagerstedt, J. O.; Thevelein, J. M.; Persson, B. L., Molecular mechanisms controlling phosphate-induced downregulation of the yeast Pho84 phosphate transporter. *Biochemistry* **2009,** *48* (21), 4497-4505.

40. Jo, S.; Kim, T.; Iyer, V. G.; Im, W., CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **2008,** *29* (11), 1859-1865.

41. Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L.; MacKerell, A. D.; Klauda, J. B.; Im, W., CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *Journal of Chemical Theory and Computation* **2016,** *12* (1), 405-413.

42. Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H., PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *Journal of Chemical Theory and Computation* **2011,** *7* (2), 525-537.

43. Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W., Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *The Journal of Physical Chemistry B* **2010,** *114* (23), 7830-7843.

44. 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 ϕ, ψ and Side-Chain $\chi 1$ and $\chi 2$ Dihedral Angles. *Journal of Chemical Theory and Computation* **2012,** *8* (9), 3257-3273.

45. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2017,** *14* (1), 71-73.

46. 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.

47. Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; Golze, D.; Wilhelm, J.; Chulkov, S.; Bani-Hashemian, M. H.; Weber, V.; Borštnik, U.; Taillefumier, M.; Jakobovits, A. S.; Lazzaro, A.; Pabst, H.; Müller, T.; Schade, R.; Guidon, M.; Andermatt, S.; Holmberg, N.; Schenter, G. K.; Hehn, A.; Bussy, A.; Belleflamme, F.; Tabacchi, G.; Glöß, 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. *The Journal of Chemical Physics* **2020,** *152* (19), 194103.

48. Lee, C. T.; Yang, W. T.; Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density. *Physical Review B* **1988,** *37* (2), 785-789.

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

50. 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).

51. Klimes, J.; Michaelides, A., Perspective: Advances and challenges in treating van der Waals dispersion forces in density functional theory. *J Chem Phys* **2012,** *137* (12).

52. Li, C.; Swanson, J. M. J., Understanding and tracking the excess proton in *Ab Initio* simulations; Insights from IR spectra. *The Journal of Physical Chemistry B* **2020,** *124* (27), 5696-5708.

53. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R., Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics* **2012,** *4* (1), 17.

54. Jorgensen, W. L.; Jenson, C., Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. *Journal of Computational Chemistry* **1998,** *19* (10), 1179-1186.

55. 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. *Journal of Computational Chemistry* **2015,** *36* (26), 1990-2008.

56. Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981,** *52* (12), 7182-7190.

57. Bussi, G.; Donadio, D.; Parrinello, M., Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007,** *126* (1), 014101.

58. Hartwigsen, C.; Goedecker, S.; Hutter, J., Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Physical Review B* **1998,** *58* (7), 3641-3662.

59. Blöchl, P. E., Electrostatic decoupling of periodic images of plane‐wave‐expanded densities and derived atomic point charges. *The Journal of Chemical Physics* **1995,** *103* (17), 7422-7428.

60. Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M., An Efficient Real Space Multigrid QM/MM Electrostatic Coupling. *Journal of Chemical Theory and Computation* **2005,** *1* (6), 1176-1184.

61. Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M., An efficient linear-scaling electrostatic coupling for treating periodic boundary conditions in QM/MM simulations. *Journal of Chemical Theory and Computation* **2006,** *2* (5), 1370-1378.

62. Wu, H.; Mey, A. S. J. S.; Rosta, E.; Noé, F., Statistically optimal analysis of state-discretized trajectory data from multiple thermodynamic states. *The Journal of Chemical Physics* **2014,** *141* (21), 214106.

63. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **1992,** *13* (8), 1011-1021.

64. Schenter, G. K.; Garrett, B. C.; Truhlar, D. G., Generalized transition state theory in terms of the potential of mean force. *The Journal of Chemical Physics* **2003,** *119* (12), 5828-5833.

65. Martinez, P.; Zvyagilskaya, R.; Allard, P.; Persson, B. L., Physiological Regulation of the Derepressible Phosphate Transporter in *Saccharomyces cerevisiae*. *Journal of Bacteriology* **1998,** *180* (8), 2253.

66. Zvyagilskaya, R. A.; Lundh, F.; Samyn, D.; Pattison-Granberg, J.; Mouillon, J.-M.; Popova, Y.; Thevelein, J. M.; Persson, B. L., Characterization of the Pho89 phosphate transporter by functional hyperexpression in Saccharomyces cerevisiae. *FEMS Yeast Research* **2008,** *8* (5), 685-696.

67. Bun-Ya, M.; Nishimura, M.; Harashima, S.; Oshima, Y., The PHO84 gene of Saccharomyces cerevisiae encodes an inorganic phosphate transporter. *Molecular and Cellular Biology* **1991,** *11* (6), 3229.

68. Yadav, V.; Kumar, M.; Deep, D. K.; Kumar, H.; Sharma, R.; Tripathi, T.; Tuteja, N.; Saxena, A. K.; Johri, A. K., A Phosphate Transporter from the Root Endophytic Fungus Piriformospora indica Plays a Role in Phosphate Transport to the Host Plant. *Journal of Biological Chemistry* **2010,** *285* (34), 26532-26544.

69. Hediger, M. A.; Romero, M. F.; Peng, J. B.; Rolfs, A.; Takanaga, H.; Bruford, E. A., The ABCs of solute carriers: physiological, pathological and therapeutic implications of human membrane transport proteins - Introduction. *Pflugers Archiv-European Journal of Physiology* **2004,** *447* (5), 465-468.

70. Fredriksson, R.; Nordstrom, K. J. V.; Stephansson, O.; Hagglund, M. G. A.; Schioth, H. B., The solute carrier (SLC) complement of the human genome: Phylogenetic classification reveals four major families. *Federation of European Biochemical Societies Letters* **2008,** *582* (27), 3811-3816.

71. Schlessinger, A.; Matsson, P.; Shima, J. E.; Pieper, U.; Yee, S. W.; Kelly, L.; Apeltsin, L.; Stroud, R. M.; Ferrin, T. E.; Giacomini, K. M.; Sali, A., Comparison of human solute carriers. *Protein Sci* **2010,** *19* (3), 412-428.

72. Torrie, G. M.; Valleau, J. P., Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977,** *23* (2), 187-199.

73. Altschul, S. F.; Madden Tl Fau - Schäffer, A. A.; Schäffer Aa Fau - Zhang, J.; Zhang J Fau - Zhang, Z.; Zhang Z Fau - Miller, W.; Miller W Fau - Lipman, D. J.; Lipman, D. J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997,** *25* (0305-1048 (Print)).

74. Altschul, S. F.; Wootton, J. C.; Gertz, E. M.; Agarwala, R.; Morgulis, A.; Schäffer, A. A.; Yu, Y.-K., Protein database searches using compositionally adjusted substitution matrices. *FEBS J* **2005,** *272* (20), 5101-5109.

75. 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. *Proceedings of the National Academy of Sciences* **2021,** *118* (25), e2101932118.

76. Li, C.; Voth, G. A., Accurate and Transferable Reactive Molecular Dynamics Models from Constrained Density Functional Theory. *The Journal of Physical Chemistry B* **2021,** *125* (37), 10471-10480.

77. Zuchniarz, J.; Liu, Y.; Li, C.; Voth, G. A., Accurate pKa Calculations in Proteins with Reactive Molecular Dynamics Provide Physical Insight Into the Electrostatic Origins of Their Values. *The Journal of Physical Chemistry B* **2022,** *126* (38), 7321-7330.

78. Maupin, C. M.; Wong, K. F.; Soudackov, A. V.; Kim, S.; Voth, G. A., A multistate empirical valence bond description of protonatable amino acids. *The Journal of Physical Chemistry A* **2006,** *110* (2), 631-639.

79. Liu, Y.; Li, C.; Voth, G. A., Generalized Transition State Theory Treatment of Water-Assisted Proton Transport Processes in Proteins. *The Journal of Physical Chemistry B* **2022,** *126* (49), 10452-10459.

80. Schilling, T.; Gratopp, A.; DeCoursey, T. E.; Eder, C., Voltage-activated proton currents in human lymphocytes. *The Journal of Physiology* **2002,** *545* (1), 93-105.

81. Wu, L.-J., Voltage-gated proton channel HV1 in microglia. *Neuroscientist* **2014,** *20* (6), 599-609.

82. Zhao, R.; Kennedy, K.; De Blas, G. A.; Orta, G.; Pavarotti, M. A.; Arias, R. J.; de la Vega-Beltrán, J. L.; Li, Q.; Dai, H.; Perozo, E., Role of human Hv1 channels in sperm capacitation and white blood cell respiratory burst established by a designed peptide inhibitor. *Proceedings of the National Academy of Sciences* **2018,** *115* (50), E11847-E11856.

83. Zhao, R.; Dai, H.; Arias, R. J.; De Blas, G. A.; Orta, G.; Pavarotti, M. A.; Shen, R.; Perozo, E.; Mayorga, L. S.; Darszon, A.; Goldstein, S. A. N., Direct activation of the proton channel by albumin leads to human sperm capacitation and sustained release of inflammatory mediators by neutrophils. *Nature Communications* **2021,** *12* (1), 3855.

84. Wu, L.-J.; Wu, G.; Sharif, M. R. A.; Baker, A.; Jia, Y.; Fahey, F. H.; Luo, H. R.; Feener, E. P.; Clapham, D. E., The voltage-gated proton channel Hv1 enhances brain damage from ischemic stroke. *Nature Neuroscience* **2012,** *15* (4), 565-573.

85. Murugan, M.; Zheng, J.; Wu, G.; Mogilevsky, R.; Zheng, X.; Hu, P.; Wu, J.; Wu, L.-J., The voltage-gated proton channel Hv1 contributes to neuronal injury and motor deficits in a mouse model of spinal cord injury. *Molecular Brain* **2020,** *13* (1), 143.

86. He, J.; Ritzel, R. M.; Wu, J., Functions and Mechanisms of the Voltage-Gated Proton Channel Hv1 in Brain and Spinal Cord Injury. *Frontiers in Cellular Neuroscience* **2021,** *15*.

87. Wang, Y.; Li, S. J.; Pan, J.; Che, Y.; Yin, J.; Zhao, Q., Specific expression of the human voltage-gated proton channel Hv1 in highly metastatic breast cancer cells, promotes tumor progression and metastasis. *Biochemical and biophysical research communications* **2011,** *412* (2), 353-359.

88. Wang, Y.; Wu, X.; Li, Q.; Zhang, S.; Li, S. J., Human Voltage-Gated Proton Channel Hv1: A New Potential Biomarker for Diagnosis and Prognosis of Colorectal Cancer. *PLOS ONE* **2013,** *8* (8), e70550.

89. Ribeiro-Silva, L.; Queiroz, F. O.; da Silva, A. M. B.; Hirata, A. E.; Arcisio-Miranda, M., Voltage-Gated Proton Channel in Human Glioblastoma Multiforme Cells. *ACS Chemical Neuroscience* **2016,** *7* (7), 864-869.

90. Musset, B.; Smith, S. M. E.; Rajan, S.; Morgan, D.; Cherny, V. V.; DeCoursey, T. E., Aspartate 112 is the selectivity filter of the human voltage-gated proton channel. *Nature* **2011,** *480* (7376), 273-277.

91. Dudev, T.; Musset, B.; Morgan, D.; Cherny, V. V.; Smith, S. M. E.; Mazmanian, K.; DeCoursey, T. E.; Lim, C., Selectivity Mechanism of the Voltage-gated Proton Channel, HV1. *Scientific Reports* **2015,** *5* (1), 10320.

92. van Keulen, S. C.; Gianti, E.; Carnevale, V.; Klein, M. L.; Rothlisberger, U.; Delemotte, L., Does proton conduction in the voltage-gated H+ channel hHv1 involve Grotthuss-like hopping via acidic residues? *The Journal of Physical Chemistry B* **2017,** *121* (15), 3340-3351.

93. Lee, M.; Bai, C.; Feliks, M.; Alhadeff, R.; Warshel, A., On the control of the proton current in the voltage-gated proton channel Hv1. *Proceedings of the National Academy of Sciences* **2018,** *115* (41), 10321-10326.

94. Ramsey, I. S.; Mokrab, Y.; Carvacho, I.; Sands, Z. A.; Sansom, M. S. P.; Clapham, D. E., An aqueous H+ permeation pathway in the voltage-gated proton channel Hv1. *Nature Structural & Molecular Biology* **2010,** *17* (7), 869-875.

95. Wood, M. L.; Schow, E. V.; Freites, J. A.; White, S. H.; Tombola, F.; Tobias, D. J., Water wires in atomistic models of the Hv1 proton channel. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2012,** *1818* (2), 286-293.

96. Bennett, A. L.; Ramsey, I. S., CrossTalk opposing view: Proton transfer in Hv1 utilizes a water wire, and does not require transient protonation of a conserved aspartate in the S1 transmembrane helix. *The Journal of physiology* **2017,** *595* (22), 6797.

97. Takeshita, K.; Sakata, S.; Yamashita, E.; Fujiwara, Y.; Kawanabe, A.; Kurokawa, T.; Okochi, Y.; Matsuda, M.; Narita, H.; Okamura, Y., X-ray crystal structure of voltage-gated proton channel. *Nature structural & molecular biology* **2014,** *21* (4), 352-357.

98. Li, Q.; Shen, R.; Treger, J. S.; Wanderling, S. S.; Milewski, W.; Siwowska, K.; Bezanilla, F.; Perozo, E., Resting state of the human proton channel dimer in a lipid bilayer. *Proceedings of the National Academy of Sciences* **2015,** *112* (44), E5926-E5935.

99. Bayrhuber, M.; Maslennikov, I.; Kwiatkowski, W.; Sobol, A.; Wierschem, C.; Eichmann, C. d.; Frey, L.; Riek, R., Nuclear magnetic resonance solution structure and functional behavior of the human proton channel. *Biochemistry* **2019,** *58* (39), 4017-4027.

100. Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C., Anton, a special-purpose machine for molecular dynamics simulation. *Communications of the ACM* **2008,** *51* (7), 91-97.

101. Geragotelis, A. D.; Wood, M. L.; Göddeke, H.; Hong, L.; Webster, P. D.; Wong, E. K.; Freites, J. A.; Tombola, F.; Tobias, D. J., Voltage-dependent structural models of the human Hv1 proton channel from long-timescale molecular dynamics simulations. *Proceedings of the National Academy of Sciences* **2020,** *117* (24), 13490.

102. Lim, V. T.; Freites, J. A.; Tombola, F.; Tobias, D. J., Thermodynamics and Mechanism of the Membrane Permeation of Hv1 Channel Blockers. *The Journal of Membrane Biology* **2021,** *254* (1), 5-16.

103. Lim, V. T.; Geragotelis, A. D.; Lim, N. M.; Freites, J. A.; Tombola, F.; Mobley, D. L.; Tobias, D. J., Insights on small molecule binding to the Hv1 proton channel from free energy calculations with molecular dynamics simulations. *Scientific Reports* **2020,** *10* (1), 13587.

104. Zhao, C.; Hong, L.; Riahi, S.; Lim, V. T.; Tobias, D. J.; Tombola, F., A novel Hv1 inhibitor reveals a new mechanism of inhibition of a voltage-sensing domain. *Journal of General Physiology* **2021,** *153* (9), e202012833.

105. Lee, S.; Liang, R.; Voth, G. A.; Swanson, J. M. J., Computationally Efficient Multiscale Reactive Molecular Dynamics to Describe Amino Acid Deprotonation in Proteins. *Journal of Chemical Theory and Computation* **2016,** *12* (2), 879-891.

106. Yamashita, T.; Peng, Y.; Knight, C.; Voth, G. A., Computationally Efficient Multiconfigurational Reactive Molecular Dynamics. *Journal of Chemical Theory and Computation* **2012,** *8* (12), 4863-4875.

107. Li, C.; Yue, Z.; Newstead, S.; Voth, G. A., Proton coupling and the multiscale kinetic mechanism of a peptide transporter. *Biophysical Journal* **2022,** *121* (12), 2266-2278.

108. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., LINCS: A linear constraint solver for molecular simulations. *Journal of computational chemistry* **1997,** *18* (12), 1463-1472.

109. Plimpton, S., Fast parallel algorithms for short-range molecular dynamics. *Journal of computational physics* **1995,** *117* (1), 1-19.

110. Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G., PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **2014,** *185* (2), 604-613.

111. 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)]. *The Journal of chemical physics* **2008,** *128* (22), 105103.

112. Levitt, D. G., Interpretation of biological ion channel flux data—reaction-rate versus continuum theory. *Annual review of biophysics and biophysical chemistry* **1986,** *15* (1), 29-57.

113. Chamberlin, A.; Qiu, F.; Rebolledo, S.; Wang, Y.; Noskov, S. Y.; Larsson, H. P., Hydrophobic plug functions as a gate in voltage-gated proton channels. *Proceedings of the National Academy of Sciences* **2014,** *111* (2), E273-E282.

114. Pohl, P.; Boytsov, D.; Brescia, S.; Chaves, G.; Koefler, S.; Hannesschläger, C.; Siligan, C.; Goessweiner-Mohr, N.; Musset, B., Trapped Pore Waters in the Open Proton Channel HV1. *Small* **2023**.

115. Cherny, V. V.; Murphy, R.; Sokolov, V.; Levis, R. A.; DeCoursey, T. E., Properties of single voltage-gated proton channels in human eosinophils estimated by noise analysis and by direct measurement. *The Journal of general physiology* **2003,** *121* (6), 615-628.

116. Sunderland, M. A., Ion Channels of Excitable Membranes. 2001.

117. Goldman, D. E., Potential, impedance, and rectification in membranes. *The Journal of general physiology* **1943,** *27* (1), 37-60.

118. Cherny, V. V.; Markin, V. S.; DeCoursey, T. E., The voltage-activated hydrogen ion conductance in rat alveolar epithelial cells is determined by the pH gradient. *The Journal of general physiology* **1995,** *105* (6), 861-896.

119. DeCoursey, T. E.; Cherny, V. V., II. Voltage-activated proton currents in human THP-1 monocytes. *The Journal of membrane biology* **1996,** *152*, 131-140.

120. Smith, S. M. E.; Morgan, D.; Musset, B.; Cherny, V. V.; Place, A. R.; Hastings, J. W.; DeCoursey, T. E., Voltage-gated proton channel in a dinoflagellate. *Proceedings of the National Academy of Sciences* **2011,** *108* (44), 18162-18167.

121. DeCoursey, T. E.; Cherny, V. V., Deuterium isotope effects on permeation and gating of proton channels in rat alveolar epithelium. *The Journal of general physiology* **1997,** *109* (4), 415-434.

122. Marrink, S.-J.; Berendsen, H. J. C., Simulation of water transport through a lipid membrane. *The Journal of Physical Chemistry* **1994,** *98* (15), 4155-4168.

123. Marrink, S. J.; Berendsen, H. J. C., Permeation process of small molecules across lipid membranes studied by molecular dynamics simulations. *The Journal of Physical Chemistry* **1996,** *100* (41), 16729-16738.

124. Parisio, G.; Stocchero, M.; Ferrarini, A., Passive membrane permeability: beyond the standard solubility-diffusion model. *Journal of Chemical Theory and Computation* **2013,** *9* (12), 5236-5246.

125. Doyle, D. A.; Cabral, J. M.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R., The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *science* **1998,** *280* (5360), 69-77.

126. Roux, B.; MacKinnon, R., The cavity and pore helices in the KcsA K+ channel: electrostatic stabilization of monovalent cations. *Science* **1999,** *285* (5424), 100-102.

127. Noskov, S. Y.; Roux, B., Importance of hydration and dynamics on the selectivity of the KcsA and NaK channels. *The Journal of general physiology* **2007,** *129* (2), 135-143.

128. Noskov, S. Y.; Berneche, S.; Roux, B., Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands. *Nature* **2004,** *431* (7010), 830-834.

129. Im, W.; Roux, B., Ion permeation and selectivity of OmpF porin: a theoretical study based on molecular dynamics, Brownian dynamics, and continuum electrodiffusion theory. *Journal of molecular biology* **2002,** *322* (4), 851-869.

130. Kratochvil, H. T.; Watkins, L. C.; Mravic, M.; Liu, L.; Voth, G. A.; DeGrado, W. F.; Lawrence, K. S., Transient Water Wires Mediate Selective Proton Conduction in Designed Channel Proteins.

131. Woolf, T. B.; Roux, B., Conformational flexibility of o-phosphorylcholine and o-phosphorylethanolamine: a molecular dynamics study of solvation effects. *Journal of the American Chemical Society* **1994,** *116* (13), 5916-5926.

132. Hummer, G., Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations. *New Journal of Physics* **2005,** *7* (1), 34.

133. Roberts, N. K.; Northey, H. L., Proton and deuteron mobility in normal and heavy water solutions of electrolytes. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases* **1974,** *70* (0), 253-262.

134. Biswas, R.; Tse, Y.-L. S.; Tokmakoff, A.; Voth, G. A., Role of presolvation and anharmonicity in aqueous phase hydrated proton solvation and transport. *The Journal of Physical Chemistry B* **2016,** *120* (8), 1793-1804.

135. Calio, P. B.; Li, C.; Voth, G. A., Resolving the Structural Debate for the Hydrated Excess Proton in Water. *Journal of the American Chemical Society* **2021,** *143* (44), 18672-18683.

136. Suárez, E.; Wiewiora, R. P.; Wehmeyer, C.; Noé, F.; Chodera, J. D.; Zuckerman, D. M., What Markov state models can and cannot do: Correlation versus path-based observables in protein-folding models. *Journal of chemical theory and computation* **2021,** *17* (5), 3119-3133.

137. Rosta, E.; Hummer, G., Free energies from dynamic weighted histogram analysis using unbiased Markov state model. *Journal of chemical theory and computation* **2015,** *11* (1), 276-285.

138. Vanden-Eijnden, E., Towards a theory of transition paths. *Journal of statistical physics* **2006,** *123* (3), 503-523.

139. Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. S., On the transition coordinate for protein folding. *The Journal of chemical physics* **1998,** *108* (1), 334-350.

140. Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L., Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annual review of physical chemistry* **2002,** *53* (1), 291-318.

141. Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R., Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proceedings of the National Academy of Sciences* **2009,** *106* (45), 19011-19016.

142. Roux, B.; Allen, T.; Berneche, S.; Im, W., Theoretical and computational models of biological ion channels. *Quarterly reviews of biophysics* **2004,** *37* (1), 15-103.

143. Zhekova, H. R.; Ngo, V.; da Silva, M. C.; Salahub, D.; Noskov, S., Selective ion binding and transport by membrane proteins–A computational perspective. *Coordination Chemistry Reviews* **2017,** *345*, 108-136.

144. Eriksson, P.-O.; Lindblom, G.; Burnell, E. E.; Tiddy, G. J. T., Influence of organic solutes on the self-diffusion of water as studied by nuclear magnetic resonance spectroscopy. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases* **1988,** *84* (9), 3129-3139.

145. Wraight, C. A., Chance and design—Proton transfer in water, channels and bioenergetic proteins. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2006,** *1757* (8), 886-912.

146. Cui, Q.; Pal, T.; Xie, L., Biomolecular QM/MM Simulations: What Are Some of the "Burning Issues"? *The Journal of Physical Chemistry B* **2021,** *125* (3), 689-702.

147. Hu, H.; Yang, W., Free Energies of Chemical Reactions in Solution and in Enzymes with Ab Initio Quantum Mechanics/Molecular Mechanics Methods. *Annual Review of Physical Chemistry* **2008,** *59* (1), 573-601.

148. Knight, C.; Lindberg, G. E.; Voth, G. A., Multiscale reactive molecular dynamics. *The Journal of Chemical Physics* **2012,** *137* (22), 22A525.

149. Agmon, N., The Grotthuss mechanism. *Chemical Physics Letters* **1995,** *244* (5), 456-462.

150. Voth, G. A., Computer Simulation of Proton Solvation and Transport in Aqueous and Biomolecular Systems. *Accounts of Chemical Research* **2006,** *39* (2), 143-150.

151. Galib, M.; Duignan, T. T.; Misteli, Y.; Baer, M. D.; Schenter, G. K.; Hutter, J.; Mundy, C. J., Mass density fluctuations in quantum and classical descriptions of liquid water. *The Journal of Chemical Physics* **2017,** *146* (24), 244501.

152. Marsalek, O.; Markland, T. E., Quantum Dynamics and Spectroscopy of Ab Initio Liquid Water: The Interplay of Nuclear and Electronic Quantum Effects. *The Journal of Physical Chemistry Letters* **2017,** *8* (7), 1545-1551.

153. 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. *Chemical Reviews* **2016,** *116* (13), 7529-7550.

154. Markland, T. E.; Ceriotti, M., Nuclear quantum effects enter the mainstream. *Nature Reviews Chemistry* **2018,** *2* (3), 0109.

155. 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. *Biophysical journal* **2007,** *92* (1), 46-60.

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

157. Knight, C.; Voth, G. A., The Curious Case of the Hydrated Proton. *Accounts of Chemical Research* **2012,** *45* (1), 101-109.

158. Laio, A.; Parrinello, M., Escaping free-energy minima. *Proceedings of the National Academy of Sciences* **2002,** *99* (20), 12562-12566.

159. 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), 126601.

160. Invernizzi, M.; Parrinello, M., Rethinking Metadynamics: From Bias Potentials to Probability Distributions. *The Journal of Physical Chemistry Letters* **2020,** *11* (7), 2731-2736.

161. Invernizzi, M.; Piaggi, P. M.; Parrinello, M., Unified Approach to Enhanced Sampling. *Physical Review X* **2020,** *10* (4), 041034.

162. Fu, H.; Shao, X.; Chipot, C.; Cai, W., Extended Adaptive Biasing Force Algorithm. An On-the-Fly Implementation for Accurate Free-Energy Calculations. *Journal of Chemical Theory and Computation* **2016,** *12* (8), 3506-3513.

163. Dama, J. F.; Parrinello, M.; Voth, G. A., Well-Tempered Metadynamics Converges Asymptotically. *Physical Review Letters* **2014,** *112* (24), 240602.

164. Pomès, R.; Roux, B., Free energy profiles for H+ conduction along hydrogen-bonded chains of water molecules. *Biophysical journal* **1998,** *75* (1), 33-40.

165. Chakrabarti, N.; Tajkhorshid, E.; Roux, B. t.; Pomès, R., Molecular basis of proton blockage in aquaporins. *Structure* **2004,** *12* (1), 65-74.

166. König, P.; Ghosh, N.; Hoffmann, M.; Elstner, M.; Tajkhorshid, E.; Frauenheim, T.; Cui, Q., Toward theoretical analyis of long-range proton transfer kinetics in biomolecular pumps. *The Journal of Physical Chemistry A* **2006,** *110* (2), 548-563.

167. Brewer, M. L.; Schmitt Uw Fau - Voth, G. A.; Voth, G. A., The formation and dynamics of proton wires in channel environments. (0006-3495 (Print)).

168. Pezeshki, S.; Lin, H., Adaptive-partitioning QM/MM for molecular dynamics simulations: 4. Proton hopping in bulk water. *Journal of chemical theory and computation* **2015,** *11* (6), 2398-2411.

169. Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J., Current status of transition-state theory. *The Journal of physical chemistry* **1996,** *100* (31), 12771-12800.

170. Dietschreit, J. C. B.; Diestler, D. J.; Hulm, A.; Ochsenfeld, C.; Gómez-Bombarelli, R., From free-energy profiles to activation free energies. *The Journal of Chemical Physics* **2022,** *157* (8), 084113.

171. Roux, B., Transition rate theory, spectral analysis, and reactive paths. *The Journal of Chemical Physics* **2022,** *156* (13), 134111.

172. Zinovjev, K.; Tuñón, I., Quantifying the limits of transition state theory in enzymatic catalysis. *Proceedings of the National Academy of Sciences* **2017,** *114* (47), 12390.

173. 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. *The Journal of Physical Chemistry B* **2020,** *124* (27), 5665-5675.

174. Mazzuca, J. W.; Schultz, C. P., Quantum Mechanical Enhancement of Rate Constants and Kinetic Isotope Effects for Water-Mediated Proton Transfer in a Model Biological System. *The Journal of Physical Chemistry A* **2017,** *121* (4), 819-826.

175. 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. *Proceedings of the National Academy of Sciences* **2017,** *114* (50), 13182.

176. Pao, S. S.; Paulsen, I. T.; Saier Jr, M. H., Major facilitator superfamily. *Microbiology and molecular biology reviews* **1998,** *62* (1), 1-34.

177. Nakajima, H.; Hagting, A.; Kunji, E.; Poolman, B.; Konings, W. N., Cloning and functional expression in Escherichia coli of the gene encoding the di-and tripeptide transport protein of Lactobacillus helveticus. *Applied and environmental microbiology* **1997,** *63* (6), 2213-2217.

178. Parker, J. L.; Li, C.; Brinth, A.; Wang, Z.; Vogeley, L.; Solcan, N.; Ledderboge-Vucinic, G.; Swanson, J. M.; Caffrey, M.; Voth, G. A., Proton movement and coupling in the POT family of peptide transporters. *Proceedings of the National Academy of Sciences* **2017,** *114* (50), 13182-13187.

179. Wang, Z.; Swanson, J. M. J.; Voth, G. A., Local conformational dynamics regulating transport properties of a Cl−/H+ antiporter. *Journal of Computational Chemistry* **2020,** *41* (6), 513-519.

180. Wang, D.; Voth, G. A., Proton Transport Pathway in the ClC Cl−/H+ Antiporter. *Biophysical Journal* **2009,** *97* (1), 121-131.