THE UNIVERSITY OF CHICAGO

PROTON TRANSPORT IN BIOMOLECULAR SYSTEMS: MODELING AND APPLICATIONS TO MITOCHONDRIAL RESPIRATORY COMPLEX I

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
JOSHUA ZUCHNIARZ

CHICAGO, ILLINOIS

JUNE 2023

To the ones who have helped me along the way

Table of Contents

List of Figures

List of Tables

Acknowledgements

First, I'd like to thank my advisor, Prof. Greg Voth, for his mentorship over these last five years. Greg's deep scientific knowledge and guidance have been of enormous value to me in acquiring the skills and information necessary to perform this work, and I appreciate his encouragement and support. I will look back with fondness on my time in the Voth Group.

I must also offer heartfelt thanks to many group members and former group members. To Dr. Paul Calio and Dr. Laura Watkins, for their guidance and aid, especially when I was a new member of the group. To Dr. Zack Jarin, Dr. Thomas Dannenhoffer-Lafage, and Dr. Morris Sharp, for keeping our corner of the lab lively. To Dr. Yuxing Peng for invaluable assistance with the group's code packages and the cluster. To Dr. Chenghan Li, in respect of whose work in the group I am particularly indebted, since much of the work set forth here builds on his earlier innovations. To Grace Liu, for her friendship, positive attitude, and great efforts in making much of the work presented here possible. To Ian Bongolata, for perseverance and a willingness to help. And finally, to Tim Loose, for both insightful scientific conversations at various times and for his friendship, which has been, to me, one of the most valuable products of this entire experience.

Finally, there are several people who have provided personal support to me, without whom this may not have come to fruition. I am grateful to my parents, Irek and Sherri Zuchniarz, who have always supported me in my education and career goals. I am also thankful for both the hardworking Chemistry Department staff, who are always a huge help, and the Law School administration, who bent over backwards to help me achieve my goal of pursuing a J.D./Ph.D. My friends Christian Goodnow, Madeline Meier, and AJ and Victoria Skiera have provided support and an outlet which has been essential, especially during the pandemic. Last, but certainly not least, I am tremendously grateful to my boyfriend, Zach Baughman, for being a source of joy and

encouragement as I have finished this work. I hope I have not forgotten anyone, but one of the dubious blessings of a Ph.D. program is being confronted with one’s own limitations, so I must admit the possibility that I have crassly failed to thank someone. I hope that they will accept thanks in person as a substitute.

Abstract

Proton transport, the movement of hydrogen ions across membranes, is a pervasive biological process which appears in many contexts and is central to cellular energy transduction, hence life itself. Because it plays a central role in many enzymatic mechanisms, understanding what protein features cause defective or, conversely, facile proton transport can produce insights into how to treat diseases linked to those enzymes, as well as suppress or enhance their function. Proton transport processes in enzymes, though they enable such important events as the synthesis of ATP and co- or anti-transport of small molecules, are often not mechanistically well-understood. In particular, existing proton transport studies have often assumed that transport will only occur readily when coherent, stable hydration structures can be resolved. More recent work has cast doubt on this assumption and produced evidence of relatively facile proton transport even through only transiently hydrated structures, illuminating the need to account for the coupling of proton motion to local water structure when studying this important process.

In this thesis, I present model parameters for reactive molecular dynamics simulations of glutamate, aspartate, and histidine, three titratable amino acids that often mediate proton transport in channels. I show that these models produce physically accurate results and are transferrable to aqueous environments despite being derived from gas-phase data. I then apply some of these models to a real system with important biological implications, mitochondrial respiratory Complex I, a key component of the electron transport chain and one of a few proteins responsible for maintaining the mitochondrial proton gradient that powers ATP synthase. First, I present a study detailing the employment of the fitRMD parameterization procedure to produce a physically accurate reactive molecular dynamics model of aspartate, as well as improvements to the existing methodology that result in a more accurate glutamate model than previously existed. Because

excellent experimental data on the pKas of glutamate and aspartate mutants of a certain residue in staphylococcus nuclease exists, I was also able to show that these models strongly agree with available data for both bulk and protein environments. Then, I present similar work for histidine, discussing the modeling differences from the carboxylates due to the two non-equivalent protonation sites of histidine, and validating the model produced against experimental data.

Finally, I use some of these models in a real system, the ND1-ND4L hypothetical proton pathway in Complex I. This pathway, unlike three other proton pathways in the enzyme, is not well-understood, and the lateral proton transport between several critical titratable carboxylate residues has been hypothesized to enable delivery of the driving force produced by ubiquinone reduction elsewhere in the enzyme to each of the sites of proton translocation. This enzyme is a case study in the coupling of local hydration dynamics to proton transport energetics, as well as the more general phenomenon of proton transport being coupled to processes with much longer and larger time- and length-scales. I calculate the potential of mean force of proton transport with respect to local water connectivity and proton progress through the channel to shed light on how transport through this channel might interact with sequential transport events, and build on the growing body of work showing that transiently dry channels do not prevent a finding of relatively facile proton transport if water structure is adequately accounted for in simulation design.

Taken together, these results represent a step forward toward more accurate and versatile models of reactive residues in protein proton transport processes, as well as an advance in our understanding of the energetics and dynamics of proton transport in one particularly important system, Complex I, with important implications for future modeling efforts as well as further experimental work on the enzyme.

CHAPTER 1

Introduction

1.1 Proton Transport in Biological Systems

Proton transport (PT) is a ubiquitous biological process involving the movement of hydrogen ions (protons or H^+) across membranes. [1-3] Like other ion transport processes, PT may serve to establish a charge gradient[1, 4, 5], it may be active[6] or passive[7], it may even sometimes run in reverse[8], and it is often coupled to changes in the cellular context on larger and longer length- and time-scales than the reactions which comprise the transport process itself[3, 9-11]. PT is critical to the functioning of a wide variety of proteins and other biological systems.[3] PT is essential to produce the mitochondrial proton gradient which drives the synthesis of ATP,[12] to the transport of other ions and molecules,[4, 13] and to cellular homeostasis.[14] The ubiquity of PT in biological systems means that there is a wide variety of potential applications for a better understanding of PT processes generally and in particular instances, from drug design[14-16] to energy storage.[17] But while study of aqueous systems, especially computational study, has made enormous progress, many of the aspects of PT in biological systems, including mechanistic details and energetics, remain obscure.

A proton in aqueous solution interacts strongly with neighboring water molecules, rather than existing isolation, tending to form hydrogen bonds with nearby water molecules.[18] The excess proton is frequently denoted $\mathrm{H_3O^+}$, the hydronium ion, a water molecule bearing an additional proton. This is not a particularly accurate description of the physical reality.[19] The nature of the water structure around such an excess proton can change rapidly but is commonly described as a water cluster comprising two or four water molecules and an excess proton: $\mathrm{H_5O_2^+}$ or $\mathrm{H_9O_4^+}$. These are, respectively, termed the Zundel[20] and Eigen[21] cations. Which, if either, of these two structures dominates under ordinary conditions remains contested,[22] but either serves to illustrate an important point about the excess proton in water.

1.2 The Grotthuss Mechanism

The fact that an excess proton tends to exist as clusters of water molecules sharing the positive charge gives rise to an important phenomenon in the PT context. Unlike other ion transport processes, PT has the unique feature that the identity of the particle being transported changes. In a process known as Grotthuss shuttling, a proton can hop from one water molecule to another.[23] Hops can occur singly and serially, or in concerted fashion.[24] Double or triple hops inescapably result in the charge moving while the atom that originally bore that charge remains behind, a strange concept in the ion transport context;[25] this conceptual difficulty arises from improperly analogizing protons to other ions. When a chloride or similar ion moves through water, its charge moves with it, because its electronegativity makes it extremely unlikely that it will transfer electrons to the surrounding solution. This phenomenon of the charge moving with the particle is called vehicular transport and is the dominant mode by which non-proton ions are transported in biological systems. But hydrogen atoms in water are interchangeable—hydrated protons can and frequently do receive electron transfers ultimately from other hydrogens, which has the effect of moving the positive charge and is the basis of Grotthuss shuttling (**Figure 1.1**). The phenomenon of the charge moving through a comparatively stationary medium is sometimes termed structural transport.[26] It is useful in a structural transport context to think not of a particle being transported, but a charge, and to think of the charge as being located at the center of charge of the local system.[27] PT is thus unique among biological ion transport processes in that mechanistically it presents as a combination of vehicular transport and Grotthuss shuttling modes. The combination of mechanisms makes the diffusion of protons in water much faster than that of other ions,[28, 29] but also has important ramifications for modeling efforts.

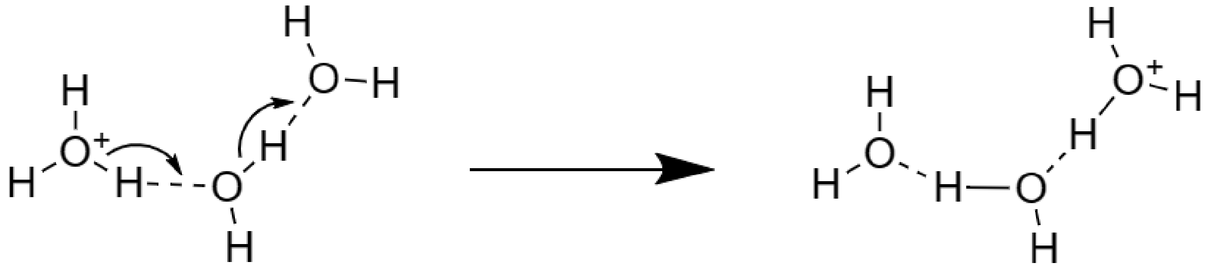

Figure 1.1 A schematic of a Grotthuss shuttling, in this case a concerted two-hop proton transfer in a three-molecule water wire. Notice that the charge has moved to the rightmost water molecule from the leftmost water molecule, while the nuclear positions have remained effectively constant. Bond lengths and angles not necessarily to scale.

Vehicular transport, because it does not involve breaking or forming chemical bonds, is relatively easily modeled. Grotthuss shuttling, conversely, is more complex. The hopping of protons is the breaking and formation of bonds, involving changes in electronic interactions. Classical methods, to avoid having to model explicit electrons, must assume a fixed bonding topology, generally representing bonded interactions as spring potentials fit to high-level data; these methods thus cannot represent Grotthuss shuttling. But to solve the Schrödinger equation for all the electrons in any sizable system would quickly become intractable and seem to foreclose the possibility of studying systems that are both large and reactive. Many methodologies have been devised to address this problem. Some, like *ab initio* molecular dynamics (AIMD)[30] and quantum mechanics / molecular mechanics (QM/MM)[31], retain on-the-fly solution of the Schrodinger equation. Another approach is a bond order formalism method,[32, 33] which computes bond order between given atoms according to some function of the nuclear positions (parameterized against quantum mechanical simulation data) and thereby permits dissociation. Each of these approaches has downsides in the context of PT. The former approach can still become prohibitively expensive when the portion of the system for which explicit QM simulation must be done is large, which is frequently the case with long-range PT processes. Long timescales present a similar problem; *ab initio* methods are inefficient to simulate at the nanosecond timescale and above. The latter approach is less computationally prohibitive once adequate models have been prepared, but the process of creating accurate reactive potentials requires a large amount of QM simulation up front.

A third approach is the multiscale reactive molecular dynamics (MS-RMD) method.[34-36] This method strikes a balance between QM accuracy on the one hand and classical efficiency on the other that is well-suited to the study of reactive processes in large biological systems. MS-RMD treats the system as a linear combination of diabatic states $|i\rangle$, each with a fixed bonding topology. In a given diabatic state, an excess proton in water will thus be covalently bound to one water molecule. Each diabatic state has some potential energy calculated according to the force field used for the computation. These force field energies become the diagonal elements of the MS-RMD Hamiltonian matrix, with the off-diagonal elements representing coupling between diabatic states, corresponding to the transition probability between states in a given configuration. Diagonalization of the MS-RMD Hamiltonian produces a ground state vector which describes the contribution of each diabatic state. Atomistic forces can be computed by applying the Hellman-Feynman theorem allowing the system to be propagated in time. This approach has been successfully applied to a wide array of PT processes in biological systems to describe their energetics and shed light on microscopic PT mechanisms largely inaccessible to experiments.[37-50]

1.3 Complex I

Mitochondrial respiratory complex I (Complex I) is a vital part of the mitochondrial electron transport chain (ETC) by which energy is stored in the body as adenosine triphosphate (ATP).[51-54] Complex I catalytically oxidizes NADH to NAD+, ultimately transferring two electrons to ubiquinone, and captures some of the energy released in that process to transport four protons from the mitochondrial matrix to the intermembrane space.[55-57] The proton gradient thereby established later drives the function of ATP synthase. [52-54, 58] This process is vital and central to energy transduction in biological systems. It is anything but simple, however. Oxidation happens over 100 Å from the location of PT, and the turnover rate of the protein is on the order of milliseconds, orders of magnitude slower than either the rate of NADH oxidation or the time it would take a proton to diffuse a distance equal to the breadth of the membrane.[59] Moreover, the enzyme can function in reverse, under certain circumstances, resulting in flow of protons into the matrix and reduction of NAD+.[60-62] The binding of various cofactors and other considerations have been suggested to influence the overall protein conformation and catalysis.[59, 63, 64] In addition to its core role in cellular energy transduction and storage, Complex I has been implicated in certain neurodegenerative diseases and as a source of oxidative stress.[65]

Understanding the function of this vital enzyme has at least three purposes, then. First, it promises to illuminate the coupling of disparate length- and time-scales in biological systems, a feature of many biological processes that frequently remains poorly understood. Second, it holds insights into PT processes in biological systems in general, as a ubiquitous example of a proton transporting enzyme, with implications for the understanding of other such processes as well as the design of biologically inspired materials. Third, specific knowledge of the factors which influence PT in Complex I may be key to understanding how defects in the protein might cause it to malfunction and aid research into therapies.

1.4 Thesis Summary

This work presents several improvements in MS-RMD models of amino acids as well as the application of the MS-RMD method to an important biological proton transporting protein, Complex I. Chapter II presents the development of reactive models of glutamic acid and aspartic acid derived by diabatic matching of MS-RMD energies to quantum mechanical diabatic state energies. This represents an advance over an older force-matching methodology, as well as refinements and additions to more recent work. Chapter II goes on to present simulations which validate the models in both bulk water and protein contexts, showing that they accurately reproduce the correct physics even in contexts where proton transport is modulated by significant environmental factors. Chapter III presents similar work for a histidine model, with modifications specific to the unique features of histidine. In particular, unlike carboxylic acids, histidine has two non-equivalent protonatable nitrogen atoms, and it was necessary to derive a model which accurately represented that the energetics of proton dissociation from each should be different. Chapter IV presents an application of the models reported in Chapter II to a complex and important biological proton transporter, Complex I. The route of PT through the membrane portion of Complex I is not completely understood, and one suggested pathway involves carboxylate residues. The application of diabatically matched MS-RMD models to this pathway was used to compute the potential of mean force of PT and determine how the hydration of the channel is coupled to the translocation of the proton through the channel. Finally, Chapter V presents some concluding remarks.

CHAPTER 2

Accurate $\mathrm{p}K_\mathrm{a}$ Calculations in Proteins with Reactive Molecular Dynamics Provide Physical Insight Into the Electrostatic Origins of Their Values

Reprinted with permission from: Joshua Zuchniarz, Yu Liu, Chenghan Li, and Gregory A. Voth. Accurate $\mathrm{p}K_\mathrm{a}$ Calculations in Proteins with Reactive Molecular Dynamics Provide Physical Insight Into the Electrostatic Origins of Their Values. *Journal of Physical Chemistry B* **2022**, 126 (38) 7321-7330.

2.1 Introduction

Molecular dynamics (MD) simulations have typically been used to investigate systems and phenomena, such as protein folding or ligand docking, involving force fields with a constant bonding topology. These classical MD methods, so-called because they do not involve a quantum mechanical treatment of the bonds or atoms involved, have been successfully applied to a wide range of biomolecular systems.[66] Despite the versatility of classical MD, alternative methods are required for any process where chemical reactivity plays a central role, such as protonation / deprotonation events by amino acids. Such alternative methods are valuable, indeed essential, for accurately calculating fundamental chemical properties accessible to experiment like the free energy of a reaction or the $\mathrm{p}K_\mathrm{a}$ of an amino acid in a biological context. To that end, a number of modifications of classical MD to permit reactivity have been developed. *Ab initio* molecular dynamics[67] (AIMD) and hybrid quantum mechanics / molecular mechanics[68] (QM/MM) methods permit reactivity by incorporating on-the-fly electronic structure calculations. Naturally, these approaches are much more computationally intensive than classical MD given the need to carry out an electronic structure calculation, and typically they are only feasible for short (< 1 ns) timescales. Another approach is to incorporate information from electronic structure calculations into a reactive MD force field (FF) called “fitRMD”. In such a simulation, electronic structure is treated implicitly, resulting in much greater efficiency than QM methods. Such reactive molecular dynamics (RMD) methods include bond-order formalisms, like ReaxFF,[33] in which the bond order between pairs of atoms is computed as a function of atomic positions and reactivity is the result of changes in bond order over time.[69] RMD also encompasses multistate methods as described below, which compute interatomic potentials and forces from a linear combination of possible bonding topologies, or diabatic states.

The multiscale reactive molecular dynamics (MS-RMD) method combined with a fitRMD reactive FF follows the latter approach.[34, 70, 71] (It should be noted that in this paper we associate the acronym MS-RMD with the reactive MD algorithm and fitRMD with the reactive MD FF, which has not always been done in past papers.) Because it does not require on-the-fly solutions to the electronic Schrödinger equation, MS-RMD is several orders of magnitude faster than AIMD and QM/MM. This lower cost renders microsecond-scale reactive MD simulations tractable, meaning that MS-RMD can be applied to study complex biomolecular phenomena where fast electronic degrees of freedom are coupled to slower, larger-scale processes like protein conformational changes. At the same time, a recent fitRMD advance[72] has demonstrated that MS-RMD parameters can be trained by diabatic matching (DM) on gas-phase constrained density functional theory (CDFT) data and then used, rather remarkably, without modification in condensed phase simulations in both liquid water and in the staphylococcal nuclease (SNase) protein to produce pK_a predictions in quantitative agreement with experiment. This represents an advantage over other RMD methods, which generally require more extensive or more system-specific FF parameterization to achieve the same level of accuracy.

The ultimate end of any reactive MD simulation is to produce insights into the system in question which are difficult or impossible to probe experimentally or (often) by other computational means. By correctly describing the fundamental, microscopic physics of a reactive process, MS-RMD simulations using DM-derived parameters do not just reproduce macroscopically observable "single value" quantities like pK_a (although in many instances, even determining the pK_a of an amino acid (AA) in a protein can be experimentally challenging[73]). Rather, such MS-RMD simulations can produce an entire multi-dimensional free energy surface for the reactive process at hand, which can then be used to not only compute quantities of interest but also to understand the physics behind them.[74] This degree of insight into the microscopic physics of the system permits elucidation of the mechanism of the process and identification of the features which have the largest effects on its energetics. Put another way, while other methods might possibly correctly predict the value of the pK_a of an amino acid, the MS-RMD approach produces not just *what* the pK_a value is but *why* it is what it is, by describing all of the interactions which collectively determine the physical equilibrium characterized by the pK_a value.

The current fitRMD approach[72] represents a form of physics-constrained supervised machine learning (ML). The training set in our recent study and in this work contains gas-phase diabatic state energies of reactive complex configurations computed by CDFT, but it should be noted that the procedure is agnostic to the QM method chosen; in principle, any diabatic method could be used. The key physical insight is that the information necessary to accurately reproduce a reaction event in a diabatic model must be primarily encoded in the electronic structure of the few diabatic states which dominate the linear combination of diabatic states that describes the entire reactive complex. Although the relationship between these electronic structure calculations and the proper fitRMD parameters is not intuitively obvious, by minimizing the residual of the diabatic state vector (rather than, as in past iterations of fitRMD model development, minimizing the residual of the atomistic forces[34-36, 75]), the fitRMD model learns how to properly reproduce the diagonal elements of the quantum mechanical Hamiltonian (corresponding to the contributions of the diabatic states to the ground state energy) as well as the off-diagonal elements (corresponding to the transition probabilities between states). Once trained, these parameters can be used in environments like bulk water and proteins, which differ greatly from the gas-phase training set conditions, and still achieve quantitatively accurate pK_a predictions evidently without further modification.[72]

In this work, we expand on recent developments of our MS-RMD models in two ways: first, we present newly computed fitRMD parameters for aspartate (Asp; D) from CDFT data, analogous to those previously presented[72] for glutamate (Glu; E) and lysine (Lys; K) and demonstrate quantitative accuracy and transferability similar to our earlier results. Second, we report even better transferability of the fitRMD model to a protein environment by first fitting the sidechain carboxyl group O-H bond force field potential parameters to DFT data. Afterwards, we discuss the notable experimental pK_a shift of both the aspartate and glutamate sidechains between bulk water and SNase in light of the computed potentials of mean force (PMF) of deprotonation. We conclude that a large part of this difference is due sidechain flexibility with regard to rotation and extension, and show how these considerations constrain the minimum free energy path (MFEP) of proton transport from the protonated residue to the bulk water, thus helping to define the physical origins of the pK_a value.

2.2 Methods

2.2.1 MS-RMD.

In a classical MD simulation, unlike MS-RMD, there is a single, fixed bonding topology throughout the simulation, and the dynamical evolution of the system is governed by a force field (FF) which describes the forces between the atoms as a function of their positions. The heart of the MS-RMD approach is to treat a reactive system as a linear combination of different bonding topologies, each with its own potential energy described by a classical FF. These possible bonding topologies, or diabatic states, are evaluated on-the-fly, and allowed to change as the system evolves, so that all those topologies and *only* those topologies (lest the computation become intractably complex) that significantly contribute to the linear combination are considered. Consider the deprotonation of the generic acid HA in water into its conjugate base and a hydrated proton:

$$\mathrm{HA} + \mathrm{H_2O} \rightarrow \mathrm{A^-} + \mathrm{H_3O^+} \tag{2.1}$$

In this representation, the reactant state and product state of the hydrated excess proton shown are two of the diabatic states whose bonding topologies' potential energies enter into the linear combination from which the total system potential energy and interatomic forces are derived. In practice, however, many other diabatic states make contributions to the energy which cannot be neglected. These are taken into account by algorithmically identifying all the reasonable possible bonding topologies out to the third solvation shell of the hydrated excess proton,[76] including topologies representing binding of the proton to nearby titratable amino acids (AAs). That set of states $|i\rangle$ forms the basis for the MS-RMD Hamiltonian as

$$\mathbf{H}^{\mathrm{RMD}} = \boldsymbol{\Sigma}_{ij} \, |i\rangle h_{ij} \langle j| \tag{2.2}$$

where when $i=j$, h_{ii} represents the energy of state $|i\rangle$, typically defined as the classical FF potential energy of the corresponding bonding topology, and when $i\neq j$, h_{ij} represents a coupling term corresponding to the transition probability between the two states. Each of these terms is dependent on the coordinates of the system nuclei. The eigenvalue problem

$$\mathbf{H}^{\mathrm{RMD}}\mathbf{c} = \mathrm{E}\mathbf{c} \tag{2.3}$$

then produces eigenvectors **c** whose components c_i correspond to the contributions of the various diabatic state energies to the eigenvalues of the Hamiltonian. The ground state energy of the system is the lowest eigenvalue of the Hamiltonian (and also a function of the instantaneous set of positions of the nuclei). The system is evolved in time by computing forces using the Hellman-Feynman theorem:

$$\mathbf{F} = \mathbf{\Sigma}_{ij} c_i c_j \mathbf{F}_{ij} = -\mathbf{\Sigma}_{ij} c_i c_j \nabla h_{ij} \tag{2.4}$$

The diabatic state vector **c** is also useful for defining the position of the net positive charge defect associated with the excess proton. This “center of excess charge” (CEC) is defined as[77]

$$\mathbf{r}_{\mathrm{CEC}} = \sum_i c_i^2\, \mathbf{r}_i^{\mathrm{COC}} \tag{2.5}$$

where $\mathbf{r}_i^{\mathrm{COC}}$ is the center of charge of the hydronium or other protonated species in state $|i\rangle$. For a more complete treatment of aspects of the MS-RMD method, please see refs[34, 71].

2.2.2 fitRMD Parameterization.

The electronic structure approach taken in this work to develop fitRMD parameters follows the CDFT-based protocol of ref[72]. We restate the essential relationships here but refer the reader to ref[72] for more thorough derivations and to ref[78] for in-depth discussion of CDFT generally. We also emphasize that this particular choice of electronic structure method is not essential to the broader ML paradigm being described. There is no theoretical obstacle to choosing a different diabatic electronic structure method.

In order to perform an electronic structure calculation on a diabatic basis state of an *adiabatic* ground state system for a given configuration of nuclear coordinates, constraints must be placed on the electron density. Considering again the dissociation of a generic acid HA, the diabatic states of greatest importance are those corresponding to the reactant ($\mathrm{HA} + \mathrm{H_2O}$) and the product ($\mathrm{A^-} + \mathrm{H_3O^+}$). For a single nuclear configuration, these states differ only in the charges of the two molecules, and thus the electron density. So, we can find the energy of a fragment of the system subject to some constraint on the electron density by the minimization

$$E^{CDFT}(N) = \underset{\rho}{\min} \quad \underset{\lambda}{\max}\left(E[\rho(\boldsymbol{r})] + \lambda\left(\int_{\Omega}\rho(\boldsymbol{r})d^3r - N\right)\right) \tag{2.6}$$

where $E[\rho(\boldsymbol{r})]$ is the density functional and $\lambda\left(\int_{\Omega}\rho(\boldsymbol{r})d^3r - N\right)$ is a Lagrange multiplier term which enforces a constraint *N* on the number of electrons in the volume Ω.[78] The Becke partition scheme[79] was used to assign volumes of space to each of the fragments in each diabatic state. Having computed E^{CDFT} for both diabatic states, the ground state energy, analogous to the MS-RMD ground state energy, can be expressed as the lowest eigenvalue of

$$\mathbf{H}^{\mathrm{CDFT}}\mathbf{c} = E\mathbf{S}\mathbf{c} \tag{2.7}$$

where **S** is the overlap matrix between the diabatic states.

With electronic structure data for the diabatic states in hand, parameterization of the MS-RMD model is achieved by minimizing the residual of the diabatic state vectors **c**, given by[72]

$$\chi^2 = \langle \left|\mathbf{c}^{\mathrm{RMD}} - \mathbf{c}^{\mathrm{CDFT}}\right|^2 \rangle \tag{2.8}$$

2.2.3 Density Functional Theory.

DFT simulations were performed on protonated Asp and Glu residues in CP2K[80] using the ωB97X functional[81] and TZV2P basis set. Eleven single-point energy calculations were performed for each residue, varying the length of the sidechain carboxyl O-H bond from 0.9 Å to 1.9 Å in 0.1 Å increments. The resulting bond dissociation energy curves were used to optimize Morse potential parameters for the sidechain carboxyl O-H bond for both residues prior to the more extensive optimization of MS-RMD parameters described below.

For Asp only, DFT simulations at the same level of theory were performed on 63 configurations of a protonated Asp/ water molecule complex, where the distance between the water oxygen and the nearest sidechain carboxyl oxygen varied between 2.2 Å and 2.8 Å in increments of 0.1 Å and the position of the proton shared by the sidechain carboxyl group and the water molecule varied from 1.0 Å from the nearest carboxyl oxygen to 1.0 Å from the water oxygen in nine equally spaced increments. The entire Asp residue was modeled, rather than just its sidechain. This was found to affect the values of **c** computed, unlike earlier work on Glu, due to the smaller size of the Asp sidechain and proximity of the carboxyl group to the backbone. The analogous simulations for Glu were already performed in ref[72] and no further electronic structure data was required.

2.2.4 CDFT.

The CDFT calculations were performed on the same Asp configurations as above in CP2K at the same level of theory. The fragments were defined as the molecules of the diabatic states discussed in Methodology; namely, for the reactant state the fragments were AspH and H_2O and for the product state the fragments were Asp^- and H_3O^+, where the constraint on the number of electrons per fragment was set to achieve the correct overall charge per fragment and the volumes of space corresponding to each fragment were defined by the Becke partition scheme.

2.2.5 Parameter Optimization.

Several sets of parameters were optimized simultaneously for each residue:

$$U_{\mathrm{OX}}^{\mathrm{rep}} = B\left(-b(r_{\mathrm{OX}} - d_{\mathrm{OX}}^{0})\right)\sum_{j=1}^{3} \exp\left(-b'\mathbf{q}_{\mathrm{H}_j\mathrm{X}}^{2}\right) \tag{2.9}$$

$$U_{\mathrm{HX}}^{\mathrm{rep}} = A\exp\left(-a(r_{HX} - d_{HX}^{0})\right) \tag{2.10}$$

Equations 2.9 and 2.10 represent repulsive corrections to the classical FF between the hydronium oxygen and Asp/Glu sidechain carboxyl oxygen ($U_{\mathrm{OX}}^{\mathrm{rep}}$) and hydronium hydrogen and Asp/Glu sidechain carboxyl oxygen ($U_{\mathrm{HX}}^{\mathrm{rep}}$). Parameters $B, b, b', C,$ and c were optimized algorithmically; d_{OX}^{0} and d_{HX}^{0} were 2.4 Å and 1.0 Å, respectively. (A and a correspond, respectively, to parameters C and c of refs[35, 72]) The sum in eq 2.9 runs over the hydronium hydrogens and the exponential term becomes unity if a proton is equidistant between the two oxygens involved in the interaction. Nonbonded interaction parameters between 1) Glu/Asp carboxyl oxygen (OP) and water oxygen (Ow), 2) Glu/Asp carboxyl proton (HP) and Ow, 3) $\mathrm{Glu}^{-}/\mathrm{Asp}^{-}$ carboxylate oxygen (O) and hydronium oxygen (OH), and 4) O and hydronium proton (HH) were fit to a 12-6 Lennard-Jones (LJ) potential (see Supporting Information). The off-diagonal coupling term between diabatic states in the MS-RMD Hamiltonian, h_{ij}, was defined as[35]

$$h_{ij} = g_1\exp(-g_2(r_{\mathrm{HX}} - g_3)^2) \tag{2.11}$$

where r_{HX} is the distance between the nearest Asp/Glu carboxyl oxygen and the proton shared with a water molecule, while parameters $g_1, g_2,$ and g_3 were tunable and correspond, respectively, to $c_1, c_2,$ and c_3 of refs[35, 72]. (Note that in this section all parameters "c" are not the same as the diabatic coefficients in eqs 2.7 and 2.8.) A final tunable term, V_{ii}, which is a constant energy correction for the difference in energy between the protonated and deprotonated forms of Glu/Asp in the classical FF, was also included. In total, 17 parameters were fit for each AA. Optimization was accomplished by a Nelder-Mead[82, 83] minimization followed by Broyden-Fletcher-Golfdfarb-Shanno (BFGS) minimization of the residual in eq 2.8. The initial guess parameter set for Glu was the model of ref[72]while for Asp it was the model in ref[35].

2.2.6 MS-RMD Simulations.

All MS-RMD simulations were performed in LAMMPS MD software[84] using the RAPTOR[34] extension. Enhanced free energy sampling used the PLUMED2 plugin[85, 86] to LAMMPS. Classical FF parameters were from CHARMM36.[87] The SPC/Fw water model[88] was used for classical waters throughout. Except where noted for the AAs, the MS-RMD parameters used for when the dissociated hydrated excess proton was in water molecules were those of MS-EVB 3.2.[89] Prior to any reactive MD simulations, all systems were equilibrated classically with the GROMACS MD package.[90]

The MS-RMD simulations employed a Nose-Hoover chain[91] thermostat using a damping parameter of 100 fs and were performed at 300 K with a timestep of 1 fs. The particle-particle, particle-mesh method was used to compute long-range electrostatic interactions in reciprocal space with a cutoff of 10 Å and a precision of 10^{-4}.[92] Well-tempered metadynamics[93] (WT-MTD) was performed to improve the free energy sampling of the proton dissociation event. The reaction coordinate (RC) of proton dissociation, denoted by ξ_{CEC}, was defined as the distance between the nearest sidechain carboxyl oxygen and the excess proton CEC (eq 2.5), as done previously.[72] The WT-MTD Gaussian height and bias factor were 0.8 kcal/mol and 12, respectively, and Gaussians of width 0.1 Å were deposited every 1 ps. A harmonic potential of 25 $\mathrm{kcal/mol} * \text{Å}^2$ was applied beyond a ξ_{CEC} value of 8 Å to avoid needless sampling of bulk-like excess proton configurations far from the protonatable residue.

2.2.7 Bulk Water and Protein pK_{a} Calculations.

Two different types of MS-RMD simulations were performed for each Asp and Glu model. First, each system was simulated as a protonated single residue in a box of bulk water consisting of 238 water molecules for Asp and 241 water molecules for Glu. The pK_{a} in bulk water was calculated from statistical mechanics via the formula[72]

$$\mathrm{p}K_{\mathrm{a}} = \log\left[C^0 \int_0^{\dagger} 4\pi\xi_{\mathrm{CEC}}^2\, e^{-\beta(F(\xi_{\mathrm{CEC}})-F(+\infty))}\ \mathrm{d}\xi_{\mathrm{CEC}}\right] \tag{2.12}$$

where ξ_{CEC} is the distance between the excess proton CEC and the AA sidechain carboxyl group, $F(\xi_{\mathrm{CEC}})$ is the conditional free energy (PMF) as a function of that distance, and $C^0 = \frac{1}{1660}$ Å^{-3} is the standard state concentration expressed as a number density. $F(+\infty)$ is the value of the free energy at infinite distance between the proton and the AA, taken to be the value of the free energy curve when ξ_{CEC} is sufficiently large that $F(\xi_{\mathrm{CEC}})$ has plateaued. The symbol † denotes the position of the transition state. [35, 72]

Second, SNase (PDB: 1U9R[94]) structures were prepared by computationally mutating E66 to D66, if necessary, protonating E/D66, solvating in a cubic box of water 70 Å to a side containing ~10,000 water molecules (9,973 for Glu, 9,911 for Asp), adding NaCl to a concentration of 0.15 M, and equilibrating under constant NPT conditions at 298 K and 1 atm for 200 ns. A Nose-Hoover chain thermostat with a damping parameter of 100fs and Parrinello-Rahman barostat[95] with a damping parameter of 1 ps were used. A timestep of 2 fs was used and the LINCS algorithm[96] constrained bonds involving hydrogen atoms. Subsequent biased MS-RMD simulations used a timestep of 1 fs at a temperature of 298 K in the constant NVT ensemble. For the protein systems, two-dimensional umbrella sampling (US) simulations were carried out with respect to two collective variables (CVs): ξ_{CEC} and d_{SC}.[72] ξ_{CEC} is defined identically to the bulk water case, while d_{SC} is a relative measure of the orientation of the Asp/Glu sidechain with respect to rest of the protein; negative values correspond to buried conformations of the sidechain, interacting with various internal water molecules,[94, 97, 98] while positive values are exposed to bulk (see Supporting Information). For V66E, a total of 900 US windows were used spanning ξ_{CEC}=0.4-11.25 Å and d_{SC}= -1.5-4.2 Å, with starting configurations taken from the US windows of ref[72]. For V66D, 580 windows were used spanning ξ_{CEC}=0.25-8.75 Å and d_{SC}= -1.5-3.0 Å. Each window was simulated for a minimum of 1 ns, with individual simulations extended as necessary to achieve convergence. A radially symmetric constraint $u_{res}(r_{\perp})$ of 10 kcal/mol*$\mathrm{Å}^2$ was placed on the CEC beyond a radius of 7 Å from the mouth of the cavity containing E/D66 to prevent lateral diffusion (see Supporting Information). The weighted histogram analysis method (WHAM)[99] was used to obtain the PMF from the US data per the following.

First, the d_{SC} degree of freedom was integrated out in order to express the free energy only as a function of ξ_{CEC} according to

$$F(\xi_{\mathrm{CEC}}) = -\beta^{-1} \ln \int \exp[-\beta F(\xi_{\mathrm{CEC}}, d_{\mathrm{SC}})] \, \mathrm{d}d_{\mathrm{SC}} \tag{2.13}$$

The pK_{a} can then be calculated from the resulting PMF by:

$$\mathrm{p}K_{\mathrm{a}} = \log\left[C^{0} S_{\mathrm{u}} \int_{0}^{\dagger} 4\pi \xi_{\mathrm{CEC}}^{2}\, e^{-\beta(F(\xi_{\mathrm{CEC}})-F(+\infty))}\ \mathrm{d}\xi_{\mathrm{CEC}}\right] \tag{2.14}$$

$S_{\mathrm{u}} = \int_{o}^{\infty} 2\pi r_{\perp} \exp\left[-\beta u_{res}(r_{\perp})\right] dr_{\perp}$ is a correction factor to account for the presence of the radial restraint $u_{res}(r_{\perp})$ on the excess proton position.

2.3 Results and Discussion

2.3.1 DM-Derived Asp Model.

The DM-optimized Asp MS-RMD parameters are reported in **Table 2.1**. The $\mathrm{p}K_{\mathrm{a}}$ was computed by eq 2.12 as 3.8 ± 0.2. These parameters produce quantitative agreement with the experimental $\mathrm{p}K_{\mathrm{a}}$ of Asp in bulk water of 3.71,[100] in line with our previous results for Glu and Lys. This validation against experimental data is persuasive evidence for the model's accuracy, but it must also be noted that it is encouraging that the parameter sets for Glu and Asp are quite similar to each other. As these AAs differ in their structure only by a methyl group and their sidechain $\mathrm{p}K_{\mathrm{a}}$ by less than one unit, they ought to have substantial similarities in a model which, as our does, aims to encode the proper fundamental physics of the systems. Especially with regard to the LJ parameters, which intuition suggests ought to be almost identical for an Asp sidechain interacting with water compared to a Glu sidechain interacting with water, the two sets of parameters differ very little. This was not a constraint placed on the optimization, nor were the optimizations even conducted from the same initial guess, indicating that the DM paradigm is properly motivated, and that appropriate choice of training set allows the model to learn the physics essential to the reaction being described.

Table 2.1. MS-RMD parameters for Asp and Glu obtained by DM to CDFT data. Definitions of these parameters are provided in Simulation Details. Units of energy are kcal/mol and units of distance are Å.

	Asp	Glu		Asp	Glu
B	0.000928477	3.94487	V_{ii}	-139.912	-153.282
b	1.41581	1.41583	$\epsilon_{\mathrm{OE-HH}}^{\mathrm{LJ}}$	0.231526	0.227986
b'	1.08883	1.09180	$\sigma_{\mathrm{OE-HH}}^{\mathrm{LJ}}$	1.36801	1.37334
A	2.72026	3.85746	$\epsilon_{\mathrm{Ow-HEP}}^{\mathrm{LJ}}$	0.717000	0.730093
a	1.15572	1.15358	$\sigma_{\mathrm{Ow-HEP}}^{\mathrm{LJ}}$	1.22018	1.24711
g_1	-20.2207	-25.0434	$\epsilon_{\mathrm{OE-OH}}^{\mathrm{LJ}}$	0.141951	0.112701
g_2	3.03394	2.99967	$\sigma_{\mathrm{OE-OH}}^{\mathrm{LJ}}$	3.00880	3.00179
g_3	1.43771	1.40739	$\epsilon_{\mathrm{OEP-Ow}}^{\mathrm{LJ}}$	0.150728	0.195512
			$\sigma_{\mathrm{OEP-Ow}}^{\mathrm{LJ}}$	3.08218	3.11138

2.3.2 Improved Glu Model.

In our previous study, to accurately model the bond dissociation energy for the deprotonation of lysine, it was necessary to replace the sidechain NH_2-H^+harmonic bond potential with a Morse potential, as has been done for all reactive species in previous fitRMD models.[35] To determine the appropriate Morse potential parameters, a DFT single point energy scan along different lengths of the NH_2-H^+bond was performed, and the Morse parameters optimized by the BFGS method. It was suggested at the time that a similar procedure might be used to improve the existing models which make use of the Morse potential described in ref[35]. The greater accuracy of the Lys model in SNase compared to Glu lent credence to this proposal.

Table 2.2. Morse potential parameters where $U(r) = D\left(1 - e^{-\alpha(r-r_e)}\right)^2$; $D, \alpha,$ and r_e were fit to a DFT scan along the Asp/Glu sidechain carboxyl OH bond. The first row shows the parameters derived and used in earlier MS-RMD models for comparison.

	D	α	r_e
Nelson[35]	143.003	1.80	0.975
Asp	164.564	1.74	0.960
Glu	163.527	1.73	0.962

In that vein, we report here new Morse potential parameters for the sidechain carboxyl OH bond in both Glu and Asp in **Table 2.2**. Note that these parameters were computed before optimization of the MS-RMD parameters. Two facts are immediately apparent: 1) Both new sets of Morse parameters differ significantly from the old set, especially in the depth of the well D, but also in a somewhat shorter equilibrium bond length r_e. The older parameters were fit to condensed-phase QM/MM forces computed using the B3LYP functional[101] and a double zeta basis set, while these here were computed using gas-phase ωB97X/TZV2P diabatic state energies. At least one study,[102] published after ref[35], concluded that while B3LYP is certainly not a bad choice of functional, others offer better performance across several criteria when dealing with water systems specifically. Another recent publication[81] showed that ωB97X often produced more accurate results than B3LYP for a variety of test sets, and specifically for reaction barrier heights, which is of particular importance to this work. Taken together these indicate that our new Morse parameters rest on more reliable electronic structure data. They also produce better results across the board in terms of convergence of the DM fitting procedure, agreement with experiment, and transferability. A direct comparison can also be made between the parameters of the recently published Glu model[72] and the new model by evaluating the value of eq 2.8 for each. For the older model this is 0.014 while the current model produces a value of 0.0042. **Figure 2.1** represents this same data visually by plotting the values of $\boldsymbol{c}^{\mathrm{CDFT}}$ against $\boldsymbol{c}^{\mathrm{RMD}}$. It is clear from this data that the current model, in which only the Glu carboxyl OH Morse potential was changed before reoptimization, produces fitRMD parameters which better replicate the training data. 2) The two sets of Morse parameters are much alike, especially in comparison with the prior set. This is not unexpected. Asp and Glu, as already noted, have similar structures. Additionally, their experimental pK_{a}'s in bulk water, which are largely determined by the OH bond dissociation energy, differ by only about half a unit. In fact, the earlier force-matched Glu and Asp models used Morse parameters that were actually identical, because they had been fit to a harmonic CHARMM bond potential that was the same for the carboxyl OH of both molecules.[35] Once again, these values were not constrained to be identical in the present work. The underlying DFT energy data, though computed analogously, were different. That they nevertheless produced such similar results reflects well on the choice of electronic structure and optimization methods.

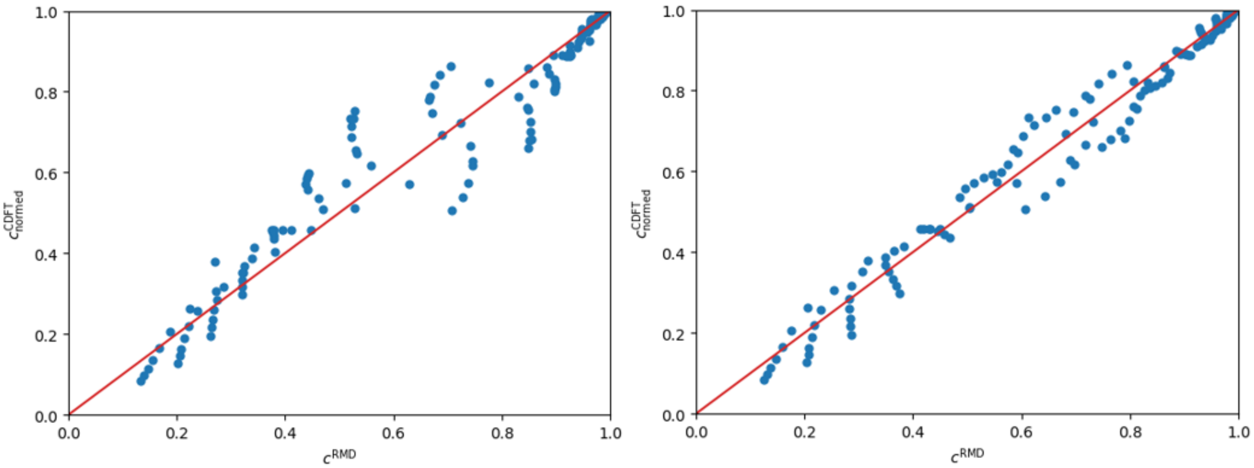

Figure 2.1. Plots of $\boldsymbol{c}^{CDFT}$ versus $\boldsymbol{c}^{RMD}$ for the training set after optimization for Left: the Glu model of ref 9; Right: this work. Points farther from the red line indicate training configurations where the optimized MS-RMD model less accurately reproduced the relative contributions of the diabatic states to the ground state energy of the Glu-water complex. The new model generally reproduces the CDFT training data better.

2.3.3 Model Performance in SNase and pK_a Analysis.

The goal of this model parameterization effort is to perform simulations in biomolecular contexts that have predictive value with quantitative accuracy. To assess the models' ability to perform well in systems different from the training environment, deprotonation simulations in the SNase protein were performed. SNase is ideal for our purposes not just because it allows for direct comparison with our earlier results but because it is a well-studied system for which the pK_a values of several mutated internal residues have been experimentally determined.[103] Notably, the experimental pK_a's of E66 in the V66E mutant and D66 in the V66D mutant are both ~ 9 (see **Table 2.3**), some 5 units more basic than in bulk. Two-dimensional umbrella sampling calculations were performed with respect to the CVs defined in Simulation Details; the resulting 2D PMFs are presented in **Figure 2.2** and the pK_a values of the mutant residues calculated by eq 2.14 are shown in Table 2.3 below.

		SNase	Water
Asp	Computed	8.5 ± 0.2	3.8 ± 0.2
	Experimental	$8.45 - 9.03^b$	3.71^a
Glu	Computed	9.3 ± 0.2	4.1 ± 0.2
	Experimental	$8.73 - 9.28^b$	4.15^a

Table 2.3. pK_a values for deprotonation of Asp and Glu in bulk and SNase derived from PMFs computed according to eq 2.14 by biased MS-RMD simulations.
[a]Values from ref[100]. [b]Values from ref[103] from chemical denaturation.

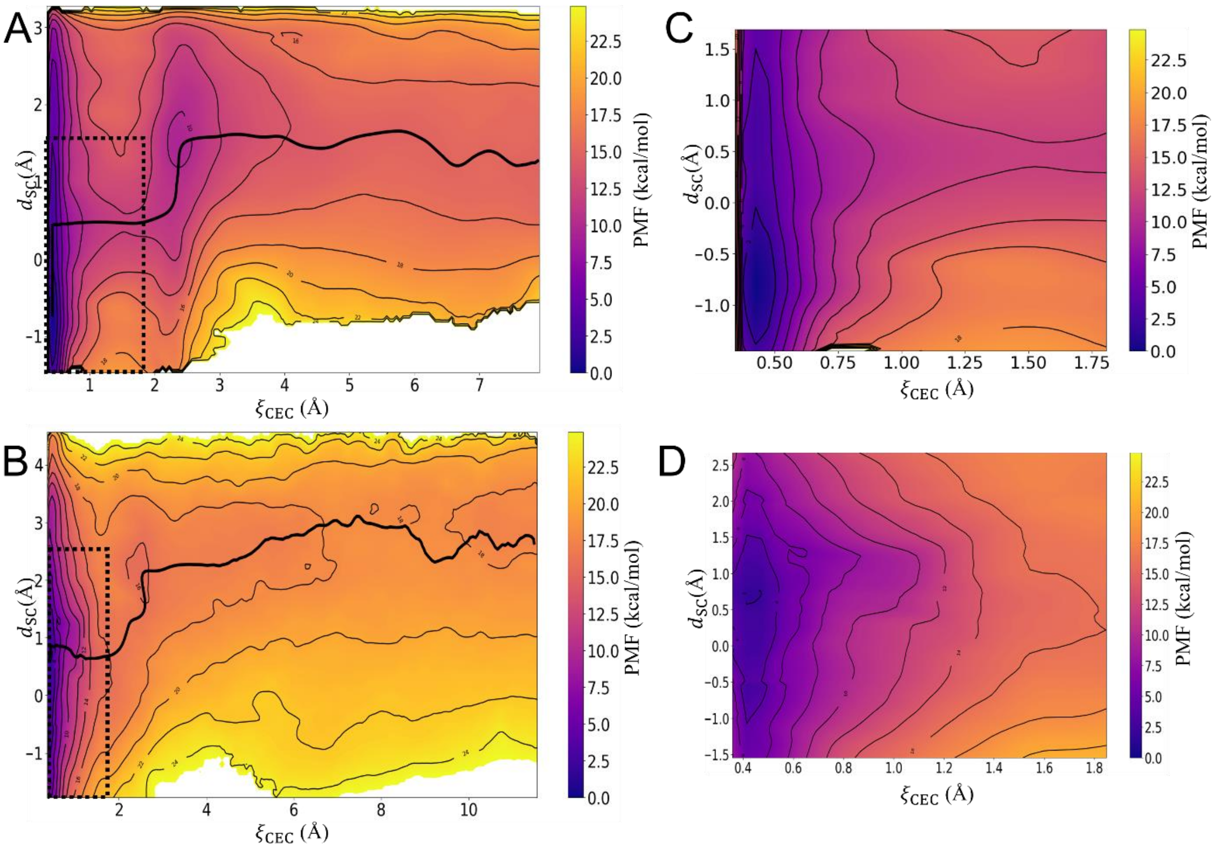

Figure 2.2. 2D PMFs in units of kcal/mol of E/D66 deprotonation in SNase with respect to a distance CV ξ_{CEC} and a sidechain orientation CV d_{SC}. In A/B the black lines represent the MFEP of proton dissociation from the sidechain, corresponding to the likeliest reaction pathway. The dashed boxes indicate the location of the detail shown in C/D. Negative values of d_{SC} indicate a buried E/D66 sidechain, while positive values indicate a solvent-exposed sidechain (see Supporting Information). A) D66. B) E66. C) Detail of the well of (A). D) Detail of the well of (B).

In line with prior results, and buttressing the evidence for the transferability of DM-derived MS-RMD parameters, the pK_a computed for D66 is within the range of values determined by experiment. Note also that the E66 pK_a with the inclusion of the new Morse parameters and subsequent reoptimization has been reduced compared to our earlier study (which found a value of 9.8 ± 0.2)[72] and is now in better agreement with experiment. This transferability of gas-phase-matched models to the biomolecular systems is remarkable and gives us confidence that not just the overall energetics of the reaction but the mechanism suggested by the MFEP across the PMF is correct and has explanatory power for the large pK_a shift for Glu and Asp in the V66E/D SNase mutants. Several features of these PMFs thus merit further discussion as follows.

The most obvious feature of each of the 2D PMFs is the narrow (in the ξ_CEC dimension) and deep free energy well. Each is centered around $\xi_\mathrm{CEC} \approx 0.5$ Å with a depth of ~15 kcal/mol relative to the bulk, but perhaps more conspicuously a depth of ~14 kcal/mol relative to the position of the contact ion pair (CIP) local minimum. The prominence of the CIP well in both plots comports with the suggestion of a prior study that ns-scale persistent CIP interactions modulate the local electrostatic environment.[104] Compare this with a well depth in bulk of 9-11 kcal/mol for both Glu and Asp.[72] This difference accounts almost entirely for the large pK_a shift – even if the integral in eq 2.14 is only performed out to ξ_CEC=3 Å for V66E (approximately the position of the CIP), the resulting pK_a is 8.2. This indicates that the most important factor in the pK_a shift compared to bulk is the difference in the initial bond dissociation event energetics, modulated by the local protein environment, with only a small contribution to the overall change in free energy from the remainder of the proton translocation process out to the bulk. It should be noted that, while such an approximate viewpoint is useful for understanding the origins of the differing behavior of Glu/Asp in SNase compared to bulk, quantitatively accurate, physics-motivated pK_a predictions must include the entire free energy curve of the dissociation event. Because pK_a is a logarithmic scale, an approximation which only includes the first few Å of the curve and neglects the remainder of the process, while appearing to result in a fairly accurate pK_a (differing from experiment by "just" one unit), overestimates the degree of dissociation by a factor of 10. For example, a popular pK_a prediction tool, PROPKA 3[105, 106], yields a pK_a value of 8.02 for E66, similar to the value approximately obtained here by considering only the first few Å of the dissociation curve. While this is useful for quickly gauging the direction of a pK_a shift in a given environment, it indicates that for reliably accurate predictions, the full proton dissociation curve based on a statistical mechanics formulation and an explicitly reactive MD model is essential.

Several differences between the E66 and D66 dissociation curves are also worth discussing. Note the difference in shape of the MFEP between the two plots (**Figure 2.2 A,B**). Where the MFEP is vertical or horizontal, the reaction is proceeding along only one of the CVs, while the other remains constant. Where the MFEP is sloped, the two CVs are said to be coupled, or change concurrently as the reaction proceeds. The D66 MFEP suggests an almost purely stepwise mechanism; that is, the MFEP is either vertical or horizontal for most of the dissociation event, demonstrating significant coupling between the CVs only beyond $\xi_{\mathrm{CEC}} > 4$ Å, where the free energy surface is basically bulk-like and the relevance of the MFEP diminishes. The dissociation takes place in three distinct steps: first, a partial exposure of the Asp sidechain; next, movement of the proton away from the Asp sidechain to a distance of $\xi_{\mathrm{CEC}} \approx 2.5$ Å; and finally, another reorientation of the Asp carboxylate group to establish the CIP local minimum. The E66 MFEP, by comparison, is not so starkly stepwise, but shows greater coupling between the two CVs. The likeliest explanation for this seems the greater length, and therefore flexibility, of the Glu sidechain compared to Asp. In a study of ligand binding in proteins, binding pocket Glu residues were found to be flexible about twice as often as Asp residues, owing to one more rotatable bond (3 compared to 2) in the sidechain.[107] An analogous effect is at work here. This difference in sidechain flexibility manifests itself in another way: compare the values of d_{SC} at the end of the MFEP for E66 and D66. The minimum for E66 at high values of ξ_{CEC} is broad but generally between d_{SC} values of 2 and 3 Å. For D66, the equivalent minimum is closer to 1 to 2 Å. This is a subtle difference but corresponds to greater exposure of deprotonated Glu to the bulk compared to deprotonated Asp. **Figure 2.3** sheds some light on the causes. In the first panel, corresponding to the global minimum, the Asp sidechain is buried, far from bulk water molecules and participating in a hydrogen bond with an internal water molecule. In the second panel, corresponding to the saddle point in the PMF, the sidechain has both rotated to interact strongly with the bulk and translocated, while the relative position of the excess proton CEC to the carboxyl group remains largely constant (notice the large relative distance between the sidechain and the internal water molecules.) In the third panel, corresponding to the hill above the saddle point, the Asp sidechain is still interacting strongly with bulk water molecules, but the alpha helical structure of the nearby backbone has been significantly disrupted to achieve the desired orientation. The energetic penalty associated with that disruption appears to explain the different behavior of E/D66 with respect to their sidechain orientations during deprotonation. This D66 behavior, threading a needle between significant protein secondary structure disruption on the one side and the need to establish contact with the solvent on the other, sheds light on prior studies which describe greater local protein disorganization in V66D.[97, 98, 103, 108] We also note that our simulations support the hypothesis of α-helix unfolding,[98, 103, 108] rather than that of ref[97] of β-sheet melting. The more flexible Glu sidechain can rotate toward the bulk more easily, without destabilizing the protein secondary structure, and so the minimum of the PMF is at higher values of d_{SC}, despite otherwise very similar deprotonation curves and pK_{a} values for the two residues. We also note that the consistently high value of d_{SC} at large values of ξ_{CEC} in the MFEP of both structures indicates that the reorientation necessary for deprotonation is persistent, allowing the negatively charged carboxylate sidechain to be well-hydrated, in agreement with a number of prior works suggesting that deprotonation of E/D66 is coupled to conformational reorganization.[98, 103, 104, 108]

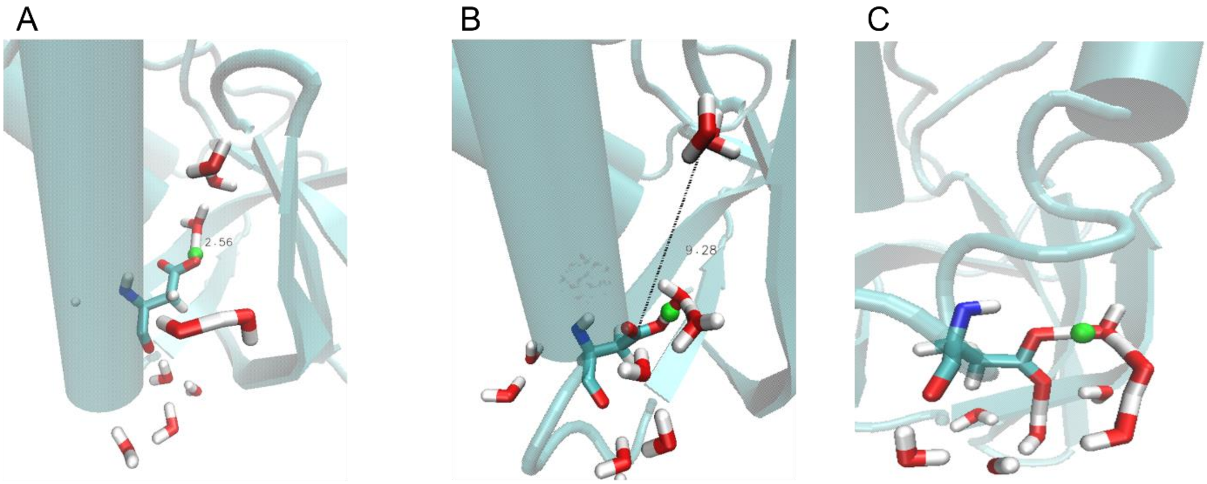

Figure 2.3. Representative structures of SNase D66 taken from three different umbrella sampling windows. The CEC is shown in green. A) $\xi_{CEC} = 0.5$, $d_{SC} = -0.75$, the global minimum of the PMF. The distance between the D66 proton and a nearby internal water molecule oxygen is shown in Å. B) $\xi_{CEC} = 1.5$, $d_{SC} = 0.5$, the saddle point between the global minimum and the CIP well. The same distance as in (A) is shown for comparison. c) $\xi_{CEC} = 1.5$, $d_{SC} = 2.00$, atop the energetic barrier associated with entering the CIP local minimum along a constant value of $d_{SC} = 2.00$ (see **Figure 2.2B**).

2.4 Conclusions

We have presented a new DM-based fitRMD model for Asp and an updated model for Glu which both produce quantitative agreement with experimental values of pK_a in bulk water and in the SNase protein. Comparison of the MFEPs of 2D PMFs of deprotonation of E/D66 in SNase reveals subtle differences in mechanism not apparent from the pK_a values alone, and supported by prior experimental and computational studies. Specifically, the deprotonation event for Asp appears to take place in a nearly perfectly stepwise fashion, with little coupling between the chosen CVs, while more significant coupling is apparent in the PMF of Glu deprotonation; additionally, the Asp sidechain rotates to a lesser degree into the bulk. The transferability of the Asp model to a protein context is further evidence that the DM approach to fitRMD model parameterization correctly captures the fundamental physics of AA deprotonation reactions. The updated DM paradigm proposed here, of first refitting classical potentials for reactive bonds to electronic structure data and then optimizing the MS-RMD parameters by minimizing the residual of the diabatic state vector, yields powerful tools for investigating the mechanistic features of reactive events. We anticipate that we will be able to extend this paradigm to other titratable AAs and to other types of reactions to broaden the set of systems that can currently be investigated by our combined fitRMD/MS-RMD methodology. We also expect that these mechanistic insights into protonation equilibria and other reactions in biomolecular contexts will advance our understanding of proton coupling in proteins in general and will aid in the design of new experiments to study such systems.

2.5 Supporting Information

$$U_{\mathrm{LJ}}(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{2.S1}$$

Equation 2.S1. The 12-6 Lennard-Jones potential, a function of interatomic distance r. Parameters σ and ϵ were fit in the present model for four pairs of atoms, described in the main text.

$$\mathbf{n}_{\mathrm{prot}} = \frac{\mathbf{r}_{C2} - \mathbf{r}_{C1}}{|\mathbf{r}_{C2} - \mathbf{r}_{C1}|} \tag{2.S2}$$

$$r_{\perp} = |\mathbf{v}_{\mathrm{CEC}} - (\mathbf{v}_{\mathrm{CEC}} \cdot \mathbf{n}_{\mathrm{prot}})\mathbf{n}_{\mathrm{prot}}| \tag{2.S3}$$

$$d_{\mathrm{SC}} = \mathbf{v}_{\mathrm{SC}} \cdot \mathbf{n}_{\mathrm{prot}} \tag{2.S4}$$

Equations 2.S2-S4. Several relationships first defined in ref[9] and used for the enhanced sampling performed in this work. $\mathbf{r}_{C1}$ and $\mathbf{r}_{C2}$ are the centroids of the backbone atoms of residues 15-19 and residues 61-65 of SNase, respectively. $\mathbf{v}_{\mathrm{CEC}}$ is the vector pointing from the D/E66 sidechain carboxyl centroid to the excess proton CEC. $\mathbf{v}_{\mathrm{SC}}$ is the vector pointing from the D/E66 α-carbon to the sidechain carboxyl centroid.

Table 2.S1. fitRMD Parameters Used to Compute Potentials of Mean Force

[a]Parameters of MS-EVB 3.2, published in ref[28] of the main text. Definitions of all these parameters can be found in ref[15]. [b]Parameters for water-hydronium interactions only. Similarly-named parameters in the main text and the top half of this table are the equivalent for amino acid-water interactions.

	Asp	Glu		Asp	Glu
B	0.000928477	3.94487	$\mathrm{V_{ii}}$	-139.912	-153.282
b	1.41581	1.41583	$\epsilon_{\mathrm{OE-HH}}^{\mathrm{LJ}}$	0.231526	0.227986
b'	1.08883	1.09180	$\sigma_{\mathrm{OE-HH}}^{\mathrm{LJ}}$	1.36801	1.37334
A	2.72026	3.85746	$\epsilon_{\mathrm{Ow-HEP}}^{\mathrm{LJ}}$	0.717000	0.730093
a	1.15572	1.15358	$\sigma_{\mathrm{Ow-HEP}}^{\mathrm{LJ}}$	1.22018	1.24711
g_1	-20.2207	-25.0434	$\epsilon_{\mathrm{OE-OH}}^{\mathrm{LJ}}$	0.141951	0.112701
g_2	3.03394	2.99967	$\sigma_{\mathrm{OE-OH}}^{\mathrm{LJ}}$	3.00880	3.00179
g_3	1.43771	1.40739	$\epsilon_{\mathrm{OEP-Ow}}^{\mathrm{LJ}}$	0.150728	0.195512
			$\sigma_{\mathrm{OEP-Ow}}^{\mathrm{LJ}}$	3.08218	3.11138

Table 2.S1 (Cont.)

C	5.0917317 [a, b]	γ	1.783170[a]
c	8.9920023 [a, b]	P	0.1559053 [a]
B	9.9178410 [a, b]	k	5.0664471 [a]
b	1.1021518 [a, b]	D_{OO}	2.8621690 [a]
b'	2.0066249 [a, b]		5.2394128 [a]
q_O^{ex}	-0.0895456 [a]	R_{OO}^0	2.9425969 [a]
q_H^{ex}	0.0252683 [a]	P'	7.6147672 [a]
$q_{H^*}^{ex}$	0.0780180 [a]	α	7.4062624 [a]
$q_O^{H_3O^+}$	-0.32 [a]	r_{OO}^0	1.8 [a]
$q_H^{H_3O^+}$	0.44 [a]	V_{const}^{ij}	-21.064268 [a]
d_{OO}^0	2.4 [a]	d_{OH}^0	1.0 [a]
$\epsilon_{O^*O^*}$	0.098609686 [a]	$\sigma_{O^*O^*}$	3.118508 [a]
$\epsilon_{H^*H^*}$	0.000040458 [a]	$\sigma_{H^*H^*}$	0.0 [a]
$\epsilon_{O^*H_w}$	3.0 [a]	$\sigma_{O^*H_w}$	1.6 [a]
D_{OH}	136.2026 [a]	α_{OH}	2.0834 [a]
r_{OH}^0	0.98 [a]	k_α	77.4868 [a]
		α_0	111.7269 [a]

CHAPTER 3

Diabatically Matched Multiscale Reactive Molecular Dynamics Models for Histidine Produce Physically Correct Behavior

3.1 Introduction

The amino acid building blocks which comprise all proteins are ubiquitous and each is, in its own way, critical to life. They are not, however, interchangeable. Each has its own unique chemical features, and by virtue of these features will tend to play certain roles in a protein. Proline, for example, is commonly found in the β-hairpin turns of antiparallel β-sheets, because its sidechain rigidity imparts a local twist in the protein structure which tends to reverse the backbone direction.[109] Tryptophan, the largest amino acid by weight, is frequently found at the interface of transmembrane α-helices, and is thought to promote correct folding of integral membrane proteins.[110] Cysteine is famous for its propensity to form disulfide bridges which stabilize protein tertiary structure.[111] It should not come as a surprise, then, that certain amino acids are more frequently found in proton channels.

Of particular importance to proton transport (PT) processes are the amino acids whose sidechains tend to be charged at physiological pH. Amino acids that can accept a proton from or donate a proton to solution may potentially participate directly in the hydrogen bonding network along which the center of excess charge (CEC) travels during PT.[59, 71] Two of these, glutamate and aspartate, are acidic and usually negatively charged. Two more, arginine and lysine, are basic and usually positively charged. But one, histidine, has a sidechain pKa close enough to physiological pH that it is difficult to generalize about its protonation state. The His sidechain has a pKa of 6.04,[100] while physiological pH is around 7.4. It is thus frequently found in both its protonated (cationic) and deprotonated (neutral) states in proteins, and changes in the local environment can lead to changes in His protonation.[49, 112] Furthermore, the imidazole moiety of the His sidechain has two protonatable nitrogen atoms, N_ϵ and N_δ, located on opposite sides of the ring. When both are protonated, the residue is in its positively-charged, +1 state, denoted HIP. When only one nitrogen is protonated, the residue is in its neutral state, denoted HID or HIE, depending on which nitrogen is protonated (**Figure 3.1**). Doubly-deprotonated histidine, which has a negatively charged imidazolate sidechain, is not usually biologically relevant[113] (but see ref[114] for an important counterexample).

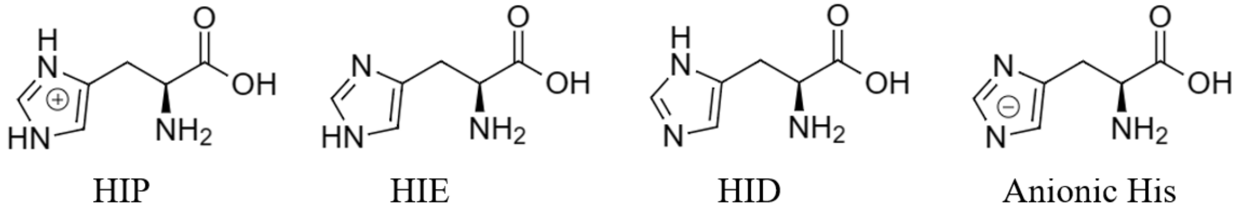

Figure 3.1 Shown, from left to right, are: cationic (doubly-protonated) histidine, HIP; neutral histidine protonated at N_ϵ, HIE; neutral histidine protonated at N_δ, HID; and anionic (doubly-deprotonated) histidine. This latter form is not typically relevant in biological contexts.[113] "His" in this work generically refers to the neutral form of the residue, either HIE or HID. Of the two forms, HIE is the more biologically common.[115]

The fact that neutral His can protonate from solution and deprotonate at the other nitrogen position gives rise to the possibility that proton transport can occur "through" the His sidechain (**Figure 3.2**). HIE, for example, could protonate to HIP, then deprotonate to HID, resulting in the translocation of the CEC from one side of the imidazole ring to the other without the requirement of any intervening water molecules. Indeed, due to the ability of His to participate in proton shuttling in this way, and the relatively facile protonation owing to its pKa, critical His residues have been discovered to mediate PT in various biological contexts.[21, 116, 117] His has also been found to facilitate PT in more complex ways, as in Influenza A M2, where the energetics of proton conduction through a channel are governed by the total charge of a His tetrad coupled to local hydration.[49]

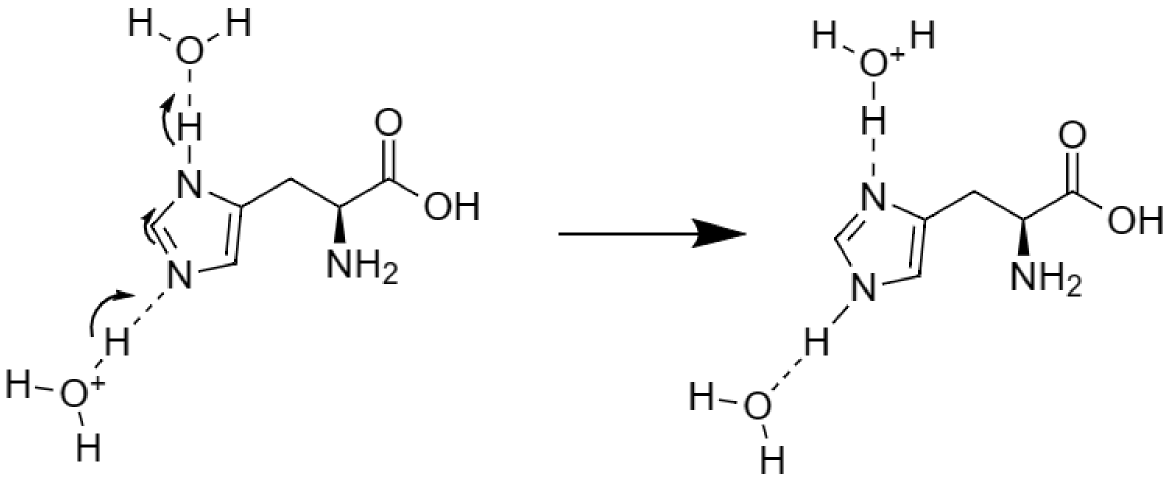

Figure 3.2 A reaction scheme for Grotthuss-type shuttling involving a central histidine imidazole sidechain. Note that in the process of charge translocation, the histidine is converted from the HID form (left) to the HIE form (right).

Clearly, classical models of His cannot capture this important behavior. To accurately describe the role of His residues in PT, reactive models are needed. Quantum-mechanical techniques, such as QM/MM, have been used successfully in some systems.[49, 118] These are, however, expensive, and perhaps totally infeasible for very large systems, systems where the PT path is long, or systems where simulations longer than the nanosecond scale are desired. A previous solution has involved force-matching MS-RMD potentials to QM forces.[36, 47] This approach, however, is system-specific; QM forces are computed for a large basis of structures consisting of various CEC positions with respect to the amino acid in the appropriate local environment, including several solvation shells of water and electrostatically important nearby residues. The MS-RMD parameters are then iteratively derived by minimizing the residual of force. [36] The large number of heavy atoms that must be modeled and the large basis of structures typically used can make parameterization expensive. The advantage of this procedure is that the model produced is likely to be quite accurate for the specific residue, in the specific environment, under study; the disadvantages are that it requires very significant QM computation before any MS-RMD can be done, and that the models must be reparameterized for each new system. They are not, in other words, typically transferable. A model that can be parameterized using a minimal basis and that is reasonably transferable is thus desirable.

In the following work, I apply and extend the diabatic matching physics-constrained machine learning methodology of Chapter 2 and ref[72] to histidine to enable future studies on histidine-containing channels at the level of accuracy that the fitRMD approach produces. I report physically accurate bulk pKa values for the deprotonation of the resulting models, including slightly different pKas for each imidazole nitrogen, as expected based on the molecule's electronic structure. I also discuss the effects of certain choices of training data on the resulting models and the prospect of extending the methodology to more complex reactions.

3.2 Methodology

The methodology of this work was substantially similar to that of Chapter 2. Only the most important relationships and any differences from the approach of Chapter 2 are detailed here.

3.2.1 MS-RMD.

The MS-RMD approach treats the system as a linear combination of diabatic states, or bonding topologies. That set of states $|i\rangle$ determines the MS-RMD Hamiltonian:

$$\mathbf{H}^{\mathrm{RMD}} = \mathbf{\Sigma}_{ij} |i\rangle h_{ij} \langle j| \tag{3.2}$$

where when i=j, h_{ii} is the energy of state $|i\rangle$ according to the FF used. When i≠j, h_{ij} corresponds to the transition probability between states $|i\rangle$ and $|j\rangle$ as a function of nuclear positions.

$$\mathbf{H}^{\mathrm{RMD}}\mathbf{c} = \mathrm{E}\mathbf{c} \tag{3.3}$$

can be solved for eigenvectors **c** describing the contributions of the various diabatic state energies to the eigenvalues of the Hamiltonian. Forces are computed by the Hellman-Feynman theorem:

$$\mathbf{F} = \mathbf{\Sigma}_{ij} c_i c_j \mathbf{F}_{ij} = -\mathbf{\Sigma}_{ij} c_i c_j \nabla h_{ij} \tag{3.4}$$

The CEC, as usual, is defined as[77]

$$\mathbf{r}_{\mathrm{CEC}} = \sum_i c_i^2 \, \mathbf{r}_i^{\mathrm{COC}} \tag{3.5}$$

where $\mathbf{r}_i^{\mathrm{COC}}$ is the center of charge of state $|i\rangle$.

3.2.2 fitRMD Parameterization

The energy of a portion of a system subject to an arbitrary constraint and $\lambda\left(\int_\Omega \rho(\boldsymbol{r}) d^3 r - N\right)$ on the electron density in volume Ω is

$$E^{CDFT}(N) = \min_{\rho} \max_{\lambda} \left(E[\rho(\boldsymbol{r})] + \lambda \left(\int_\Omega \rho(\boldsymbol{r}) d^3 r - N \right) \right) \tag{3.6}$$

where $E[\rho(\boldsymbol{r})]$ is the density functional.[78] Here, the Becke partition scheme[79] was used to define the volumes Ω. The ground-state CDFT energy is then the lowest eigenvalue of

$$\mathbf{H}^{\mathrm{CDFT}}\mathbf{c} = E\mathbf{S}\mathbf{c} \tag{3.7}$$

Parameterization of the MS-RMD model proceeds by iteratively minimizing the residual of the diabatic state vector **c**:[72]

$$\chi^2 = \langle \left| \mathbf{c}^{\mathrm{RMD}} - \mathbf{c}^{\mathrm{CDFT}} \right|^2 \rangle \tag{3.8}$$

3.2.3. Density Functional Theory.

DFT simulations were performed on protonated His (HIP) in CP2K[80] using the ωB97X functional[81] and TZV2P basis set. Forty-eight single-point energy calculations were performed, varying the length of each $\mathrm{N}_{\{\delta,\epsilon\}}$-H bond from 0.5 Å to 2.8 Å in 0.1 Å increments, though only the most physically relevant portion of this curve ended up being used in the production models. Morse potential parameters for the imidazole N-H bonds were fit to the bond dissociation energy scans prior to fitRMD parameterization, according to the potential

$$U^{Morse}(r) = D[1 - \exp\left(-\alpha\,(r - r_e)\right)]\text{^}2 \tag{3.9}$$

Additionally, 126 configurations of HIP in the presence of two flanking water molecules were prepared. The entire HIP residue was included, rather than just the sidechain, on the expectation that the polarizability of the ring and the proximity of flanking waters to the backbone groups might have important effects on electronic structure. The water molecules were placed so that they were interacting with the imidazole nitrogen protons, $\mathrm{H}_{\{\delta,\epsilon\}}$. For clarity, water oxygen $\mathrm{O}_{\mathrm{w}\{\delta,\epsilon\}}$is the oxygen nearest $\mathrm{N}_{\{\delta,\epsilon\}}$. In half (63 structures), the positions of H_δ and $\mathrm{O}_{\mathrm{w}\delta}$ were varied: the $\mathrm{O}_{\mathrm{w}\delta}$ positions were varied from a distance of 2.2 Å from N_δ to a distance of 2.8 Å in increments of 0.1 Å, and for each $\mathrm{O}_{\mathrm{w}\delta}$ position, 9 H_δ positions equally spaced between a N_δ-H_δ distance of 1.0 Å and a H_δ-$\mathrm{O}_{\mathrm{w}\delta}$ distance of 1.0 Å were generated (see **Figure 3.3**). While the δ positions were varied, the corresponding ϵ positions were held at their equilibrium values. In the other half (the remaining 63 structures) the same procedure was followed, but for $\mathrm{N}_\epsilon, \mathrm{H}_\epsilon, \text{and } \mathrm{O}_{\mathrm{w}\epsilon}$. DFT simulations at the same level of theory as the dissociation scan were performed on all 126 structures.

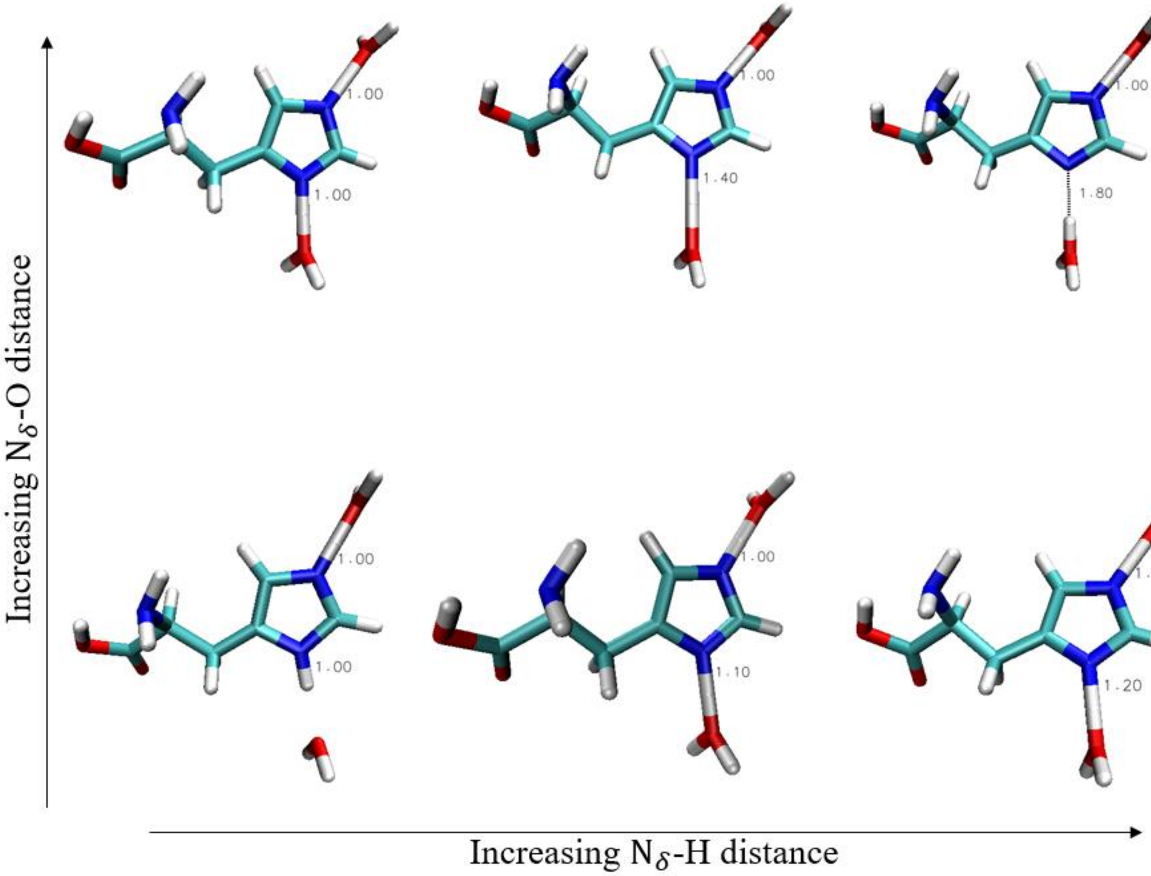

Figure 3.3 A representative sampling of the basis of structures used to collect QM data on the dissociation of the N_{δ}-H bond of histidine (HIP) in the presence of coordinating water molecules. Note that as the N_{δ}-O distance increases (up on the plot), the maximum reasonable value of N_{δ}-H increases as well, such that the rightmost structures do not have the same N_{δ}-H value (but each does correspond to the maximum computed such value for the given value of N_{δ}-O distance. Only six structures are shown here, all corresponding to the deprotonation of HIP to HIE. 63 total structures for that process were used, in total, as well as the equivalent 63 structures for the HIP to HID deprotonation.

3.2.4 CDFT

CDFT calculations were performed His configurations just described in CP2K at the same level of theory. For the reactant state the fragments were (HIP, $\mathrm{H_2O}$) and $\mathrm{H_2O}$ and for the product state the fragments were either (HID, $\mathrm{H_2O}$) or (HIE, $\mathrm{H_2O}$) and $\mathrm{H_3O^+}$, with the identity of the product state neutral histidine fragment corresponded to the N-H bond that was *not* being varied (i.e. the proton that remained on the ring). Constraints on the number of electrons per fragment were imposed to total the correct overall charge per fragment, with volumes per fragment determined by the Becke partition scheme.

3.2.5 Parameter Optimization

Several sets of parameters were optimized simultaneously, following the approach to histidine parameterization of ref[47]. 11 parameters were fit for each imidazole nitrogen, as well as a V_{ij} term common to both that represents a constant correction between the two potential energy surfaces. Nelder-Mead[82, 83] minimization followed by Broyden-Fletcher-Golfdfarb-Shanno (BFGS) minimization of the residual in eq 3.8 was performed, first by restraining the equivalent parameters for $\mathrm{N}_{\{\delta,\epsilon\}}$ to identity with each other, and then refining by removing the restraint.

3.2.6 MS-RMD Simulations

MS-RMD simulations were performed in the RAPTOR modification to LAMMPS[34, 84]. The PLUMED2 plugin[85, 86] was used for all enhanced sampling. Classical FF parameters were CHARMM36.[87] Classical waters were SPC/Fw[88]; EVB waters were MS-EVB 3.2 except where otherwise noted.[89] All systems were equilibrated in GROMACS in NVT at 300K.[90]

MS-RMD equations of motion were integrated every 1 fs. They were performed at 300K using the Nose-Hoover chain[91] thermostat with a damping parameter of 100. Electrostatics were treated by PPPM with a cutoff of 10 Å and a precision of 10^{-4}.[92] Well-tempered metadynamics[93] (WT-MTD) was performed to sample proton dissociation, where the reaction coordinate of proton dissociation was the distance between the appropriate N-H pair. The WT-MTD Gaussian height and bias factor were 0.8 kcal/mol and 12, respectively, and Gaussians of width 0.1 Å were deposited every 1 ps. A harmonic potential of 25 kcal/mol $*$ $Å^2$ was applied beyond a distance of 8 Å.

3.2.7 Bulk Water pK_a Calculations.

His was simulated as HIP in a box of 272 water molecules. The pK_a in bulk water for each of $N_{\{\delta,\epsilon\}}$ was calculated as[72]

$$\mathrm{p}K_\mathrm{a} = \log\left[C^0 \int_0^{\dagger} 4\pi\xi_{\mathrm{CEC}}^2\, e^{-\beta(F(\xi_{\mathrm{CEC}})-F(+\infty))}\, \mathrm{d}\xi_{\mathrm{CEC}}\right] \tag{3.12}$$

where ξ_{CEC} is the distance between the CEC and the relevant imidazole N atom, $F(\xi_{\mathrm{CEC}})$ is the PMF as a function of that distance, and $C^0 = \frac{1}{1660}$ $Å^{-3}$ is the standard state concentration. † denotes the transition state.[35, 72]

3.3 Results and Discussion

3.3.1 Histidine Model Parameters

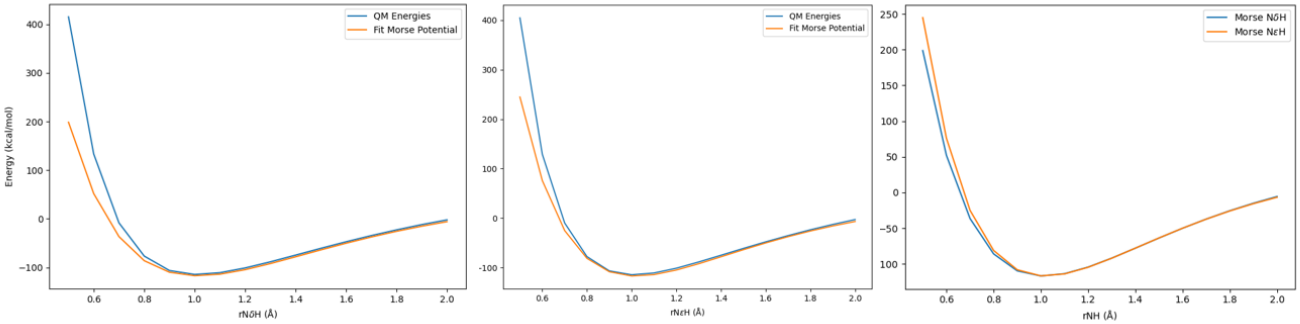

Figure 3.4 Morse potentials for $N_{\{\delta,\epsilon\}}$-H bond dissociation plotted against QM energies of bond dissociation scan and against each other. Left: N_δ-H potential compared to QM energies; Center: N_ϵ-H potential compared to QM energies; Right: Both Morse potentials compared against each other. Note that both fit potentials slightly underestimate the QM energy in the hard-wall region, but hard-sphere repulsion is much less relevant to PT than the equilibrium positions and energies, which are more accurate. At right, note that the N_δ-H bond is in slightly stronger than its N_ϵ counterpart.

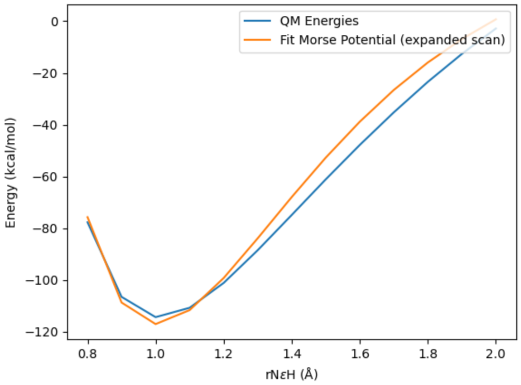

Figure 3.5 Detail of the 0.8 – 2.0 Å region of a Morse potential for the N_ϵ-H bond fit to an expanded set of QM bond dissociation energy scan data, from 0.5 – 2.5 Å. The analogous result for the N_δ-H bond is similar. Notice that including the additional data reduces the quality of the fit in the most physically relevant region, tending to overestimate the bond strength at equilibrium and overestimate the potential energy of the bonded pair beyond ~1.2 Å.

	D	α	r_e
Previous[47]	135.070	2.06	1.01
N_ϵ-H	163.158	1.76	1.018
N_δ-H	169.487	1.68	1.011

Table 3.1 Morse potential parameters for imidazole N-H bonds fit to DFT bond dissociation energy scans.

The Morse potential parameters fit to the DFT bond dissociation energy scans are presented in **Table 3.1**. Graphical representations of the quality of the fit to the DFT data are presented in **Figure 3.4** and an alternate fit to an expanded set of QM data is shown in **Figure 3.5**. The MS-RMD model parameters derived by diabatic matching to CDFT data are presented in **Table 3.2.**

	N_ϵ-H	N_δ-H
V_{ij}^{const}	-23.663	-23.663
r_{sc}^0	1.420	1.419
R_{DA}^0	2.548	2.548
a_{DA}	1.90	1.88
b_{DA}	1.55	1.55
c_{DA}	1.06	1.05
C	0.382	0.382
α	1.30	1.31
β	0.423	0.423
λ	0.455	0.454
ϵ	1.221	1.220
γ	2.179	2.179
V_{ii}	0 (def'n)	-3.12

Table 3.2 MS-RMD model parameters for each nitrogen of histidine.

3.3.2 His Bulk pKa and Model Validation

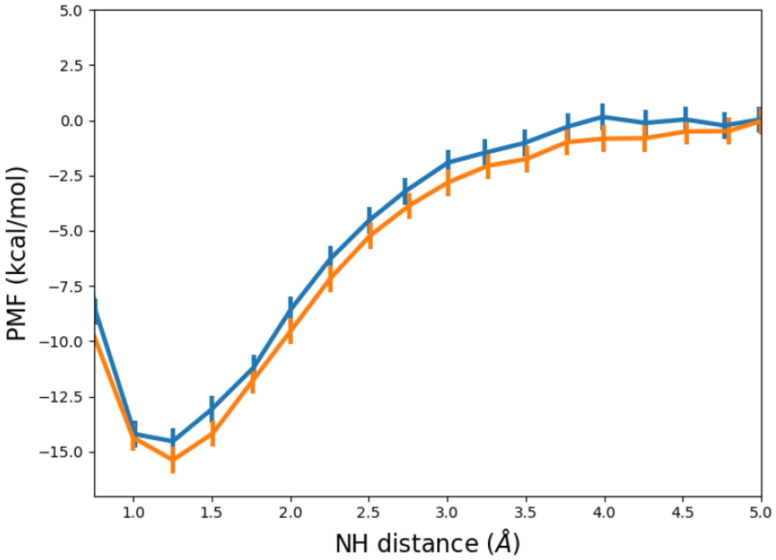

Figure 3.6 PMF of HIP deprotonation in bulk water using the models reported in 3.3.2 from (blue) N_δ and (orange) N_ϵ. The modeled PMF is in good agreement with experimental values for the pKa of the histidine sidechain in bulk. The blue curve represents a pKa of 6.32±0.3 and the orange curve a pka of 6.44±0.3 compared to an experimental value of 6.04.

The value of eq 3.8 for the set of parameters reported was 0.012, indicating quite good agreement between the $\mathbf{c}^{\mathrm{CDFT}}$ and $\mathbf{c}^{\mathrm{fitRMD}}$. Part of the modeling effort was to capture the physics of the system by restricting the twelve equivalent parameters for each nitrogen to be identical in the first iteration of the minimization of the residual of the diabatic state vector, followed by a further unrestrained refinement in order to take some of the pressure for producing the difference in energetics between the two nitrogen off of the V_{ii} term. This choice was physically motivated – if fitting 24 parameters simultaneously, there is a risk that nearly-degenerate solutions will be found for the two different nitrogens which represent more discrepant physics far from the conditions of the training set. Since the two deprotonation events ought to be very similar in most contexts, the choice of initial constraint seemed physically justified. It only appears to have been partially successful, however; most of the parameter values even after the unrestrained minimization remained identical or nearly identical. The models were further validated by computing the pKa in bulk of HIP deprotonation from each imidazole nitrogen; the results of that calculation are shown in **Figure 3.6** Integrating the curves according to eq 3.12 produces a pKa of 6.32±0.3 for N_δ and a pKa of 6.44±0.3 for N_ϵ, compared to an experimentally known value of 6.04.[100] These results are positive for the model, then, though a few caveats should be mentioned. First, the model is probably slightly too basic. The known pKa is only within the error bar for N_δ (which is the more salient comparison, since N_δ is more frequently deprotonated), and then only just. Additionally, the energetics of the two deprotonation events are probably too similar. It is difficult to find quantitative data regarding the relative prevalence of HID and HIE, so the qualitative result here that HIE has a higher pKa is correct, as far as it goes. But the two pKa values are very similar, indicating that all else being equal, the two tautomers ought to be present in bulk in nearly equal proportions.

There is one quite encouraging result as well, pertaining to the ultimate model result after incorporating the QM-derived Morse potentials. Notice that the Morse potential for N_δ actually has a *lower* well depth than N_ϵ, by a small amount but larger, for example, than the difference in well depth between the Glu and Asp Morse potentials in Chapter 2. If our expectations for the deprotonation of histidine were based entirely on scans of bond dissociation energy, we would expect that N_δ would be the more frequently protonated species, or the opposite of the result just described. Instead, however, by also learning the physics of the system associated with *deprotonation to water*, rather than the dissociation of a free proton, the model has compensated for the greater binding strength of N_δ in the other parameters.

3.3.3 Implications for Future Models

It was stated in Chapter 2 that the diabatic matching method is agnostic to choice of QM method. Here, as in Chapter 2, CDFT was again used. While these simulations formed an adequate basis from which to compute physical model parameters, it is not difficult to imagine a more complex reaction requiring a more sophisticated approach than the two-molecule method of this work. In the context of histidine, it is easy to see that a second solvating water molecule, as compared to only one for the carboxylate work, is reasonably intuitively necessary for accurate parameterization. Certainly it is essential to accurately modelling a Grotthuss shuttling mode across the imidazole ring to have a water molecule to either side in the underlying QM data; otherwise, the electronic structure around the unprotonated nitrogen would be significantly different. It would have been a natural choice, then, to describe the molecule as *three* fragments, then, and the possible diabatic states (that is, bonding topologies) for the single excess proton system as (hydronium, HID, water), (water, HIP, water), and (water, HIE, hydronium), for example. This turns out not to have been necessary, but it seems almost certain that for some more complicated reaction mechanism, such a three or more molecule description of the system, with all the associated diabatic state combinations, will be required to adequately sample the important underlying energetics of the reaction.

By the same token, there is nothing that restricts the application of the MS-RMD methodology to proton transfer reactions only. The underlying empirical valence bond formalism is generic. [119] [120] As the available library of MS-RMD models grows, then, and more complex processes become accessible, it may become necessary in the near future to create models for non-proton transfer reactions involving several complex mechanistic steps, and such work will not be feasible without a QM method that accommodates an arbitrary number of molecules and diabatic state definitions.

3.4 Conclusions

The results presented in this section extend the method of Chapter 2 and ref[72] to histidine, a residue important to a number of biological proton transport processes, with a unique sidechain structure that permits Grotthuss shuttling across the imidazole ring. The model presented in this chapter was validated by computing the PMF of deprotonation in bulk of both imidazole nitrogens, and pKa values in good agreement with available experimental data were found. Additionally, the machine learning protocol employed adapted to a discrepancy between the behavior of histidine by itself and when interacting with water to produce the correct qualitative result of a higher N_ϵpKa. These results hold promise for future extensions of the method to more sophisticated QM methods and more complex reactive systems.

CHAPTER 4

MS-RMD Simulations using fitRMD Models of Carboxylates Suggest Relatively Facile Proton Transport Coupled to Local Hydration in Mitochondrial Respiratory Complex I

4. 1 Introduction

In eukaryotic organisms, chemical potential energy is stored in adenosine triphosphate (ATP) through oxidative phosphorylation in the mitochondrion.[58, 121] The mitochondrion is an organelle which has a double-membrane structure: an outer membrane, outside of which is the cellular cytoplasm; an inner membrane, inside of which is the mitochondrial matrix; and between them an inter-membrane space. The reduction of ATP to adenosine diphosphate (ADP) is the principal reaction by which the cell harnesses chemical potential energy to maintain homeostasis.[63] ATP is synthesized by ATP synthase; the mechanism of ATP synthase is powered by the flow of protons across the inner mitochondrial membrane, and in order for the protons to flow, there must first be a proton gradient across the membrane.[58] Thus, the mitochondrial intermembrane space is commonly described as having a negatively charged side (N-side), on the interior of the inner membrane, and a positively charged side (P-side), in the intermembrane space.[122] The proton gradient must be established by the endergonic catalytic pumping of protons from the N-side to the P-side by a group of four protein complexes collectively, along with ATP synthase, called the electron transport chain (ETC).[123] The first four complexes in the ETC are called Complex I, II, IIII, and IV, and have related functions – each catalyzes an exergonic redox reaction, and harnesses the energy released to pump protons into the intermembrane space.[124]

Complex I, also called NADH:ubiquinone oxidoreductase, comprises the NADH dehydrogenase enzyme with flavin mononucleotide (FMN) and iron-sulfide cofactors (Fe-S).[63]

Mammalian Complex I has 45 distinct subunits and a mass of ~980 kDa, making it one of the largest known protein complexes.[125] (Prokaryotes have an analogous structure only about half the size of eukaryotic, mitochondrial Complex I.[126] This work deals specifically with the mitochondrial variant, but the insights produced have obvious applications to the prokaryotic system as well.) The complex is L-shaped, consisting of an integral membrane portion, the hydrophobic "arm," and an N-side perpendicular hydrophilic "arm."[51, 63] Complex I catalyzes the oxidation of NADH to NAD+ in its peripheral arm, resulting in the transfer of two electrons to FMN, from FMN through each of the Fe-S clusters, and finally to coenzyme Q, also called ubiquinone.[52, 53] For each NADH oxidation, four protons are translocated across the membrane portion of Complex I from the N-side to the P-side, contributing to the maintenance of the mitochondrial proton gradient.[127, 128] (Although see ref[129] for a competing view of the enzyme stoichiometry). The resulting proton-motive force (*pmf*, not to be confused with PMF, the potential of mean force) drives the formation of ATP.

The mechanism of Complex I is therefore vital to eukaryotic life and exhibits a number of interesting and important features that are poorly understood. First, the oxidation of NADH, the ultimate reduction of ubiquinone, and the associated free energy change is coupled to four proton translocation events over 100 Å away. Furthermore, while the timescale of the electron transfers from NADH ultimately to ubiquinone is thought to be ~100 μs,[55] the turnover rate of the enzyme as a whole is several orders of magnitude slower, on the order of milliseconds.[52] Study of Complex I thus presents opportunities to better understand the coupling of a large span of length- and time-scales in biological systems. Working to understand this system will shed light on biological energy transduction in general, where the coupling of extremely divergent scales is ubiquitous. (Compare, for example, the femtosecond timescale of chemical reactions with the hour-timescale of the circadian rhythm, and the Å-lengthscale of reactions with the meter lengthscale of hormone signaling). Additionally, Complex I's centrality to biological energy transduction suggests that understanding it may inspire synthetic materials that manipulate energy flows in similar ways. This work may thus have implications for design of materials for the purpose of energy storage or molecular machines. Additionally, mutations in the subunits of Complex I have been associated with increased production of reactive oxygen species, which are linked to a host of neurodegenerative diseases, like Parkinson's, and other maladies.[61, 130] Understanding the enzyme's mechanism in greater detail might, then, provide insights into reducing oxidative stress with a variety of positive health implications.

Most immediately, however, study of this enzyme promises to shed light on an open question as to the nature of PT across the membrane arm. The structure of the mammalian enzyme is shown in **Figure 4.1.** The enzyme adopts one of two conformations, termed open and closed, differing in the angle between the peripheral and membrane arms.[131, 132] It has been suggested that the open state, in particular, represents a catalytic intermediate structure, and is thus relevant to understanding of the catalytic mechanism.[63, 133] Three homologous domains, ND2, ND4, and ND5, contain obvious proton channels, with either persistent coherent water wires or ample transient hydration to facilitate PT.[59] One authoritative work has proposed that all four protons transported by the enzyme travel through the domain of those three farthest from the peripheral arm, ND5, based on more significant hydration compared to the other domains.[63] In particular, that theory regards transport through any of the other membrane arm domains, (ND1, ND3, and ND4L, those closest to the peripheral arm) as unlikely, based on the presence of a dry region between two residues, Asp66 of ND3 and Glu34 of ND4L (here always called $\text{Glu34}_{\text{ND4L}}$ to avoid confusion with another Glu34 elsewhere in the protein), when the enzyme is in its open conformation.[63] A competing proposal has posited that the absence of crystallographic waters does not totally foreclose the possibility of PT between ND1 and ND4L, and proposed that transient hydration between titratable residues in these domains may be feasible, given the ~120 mV potential available to drive PT.[51] This suggestion was later supported by quantum mechanical calculations showing that the barrier to transport between $\mathrm{Glu34}_{\mathrm{ND4L}}$ and Asp66 is lower than the available redox potential.[134]

Here we have used the MS-RMD methodology to study PT in the ND1-ND4L putative proton channel of Complex I and shed light on the energetics of the contested mechanism. Umbrella sampling simulations with respect to both principle curve connectivity of water molecules and titratable residues through the channel and proton progress along the path were performed and the PMF of PT through the channel computed. Our results show that formation of a coherent water wire between the first three residues in the pathway is low-barrier and PT is actually slightly exergonic. In contrast, hydrating the dry region at the end of the channel is higher-barrier and the resulting PT step significantly endergonic, but still feasible given the amount of energy available from reduction of ubiquinone elsewhere in the enzyme and the energy released by PT in the first portion of the channel. In sum, this work suggests that the absence of crystallographic waters in a certain region of the closed Complex I structure does not conclusively determine that PT does not occur, and in fact our results suggest that PT through the ordinarily dry region may be a critical part of the enzyme’s function.

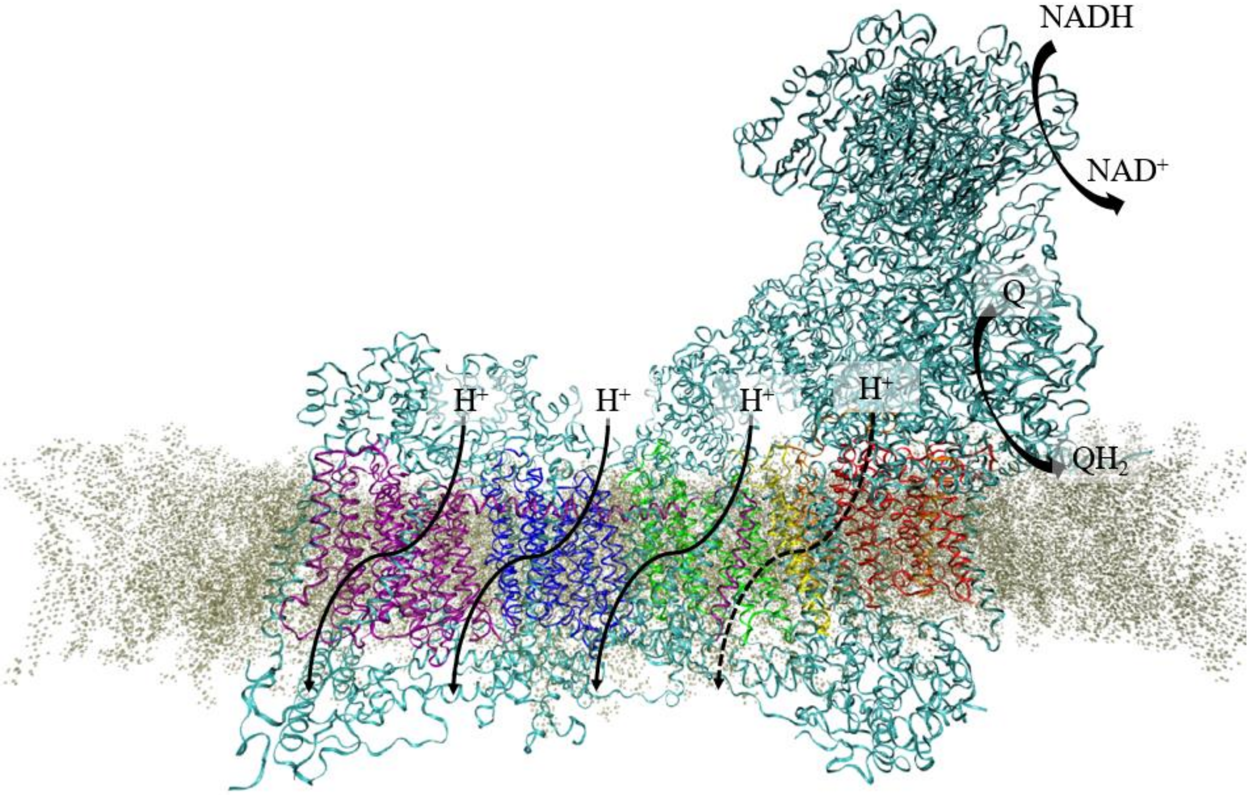

Figure 4.1 Mitochondrial respiratory complex I. The lipid bilayer of the mitochondrial inner membrane is shown in transparent, tan beads. The peripheral, hydrophilic arm is at right, showing approximate locations of NADH oxidation and ubiquinone reduction. Several domains of the membrane arm are color-coded, from right to left: ND1 (red), ND3 (orange), ND4L (yellow), ND2 (green), ND4 (blue), and ND5 (purple). Three well-characterized PT pathways in domains ND5, ND4, and ND2 are depicted as solid black arrows. The approximate location of a PT pathway running from ND1 to ND4L is shown as a dashed black arrow. The arrows point from the N-side (mitochondrial interior) to the P-side (intermembrane space).

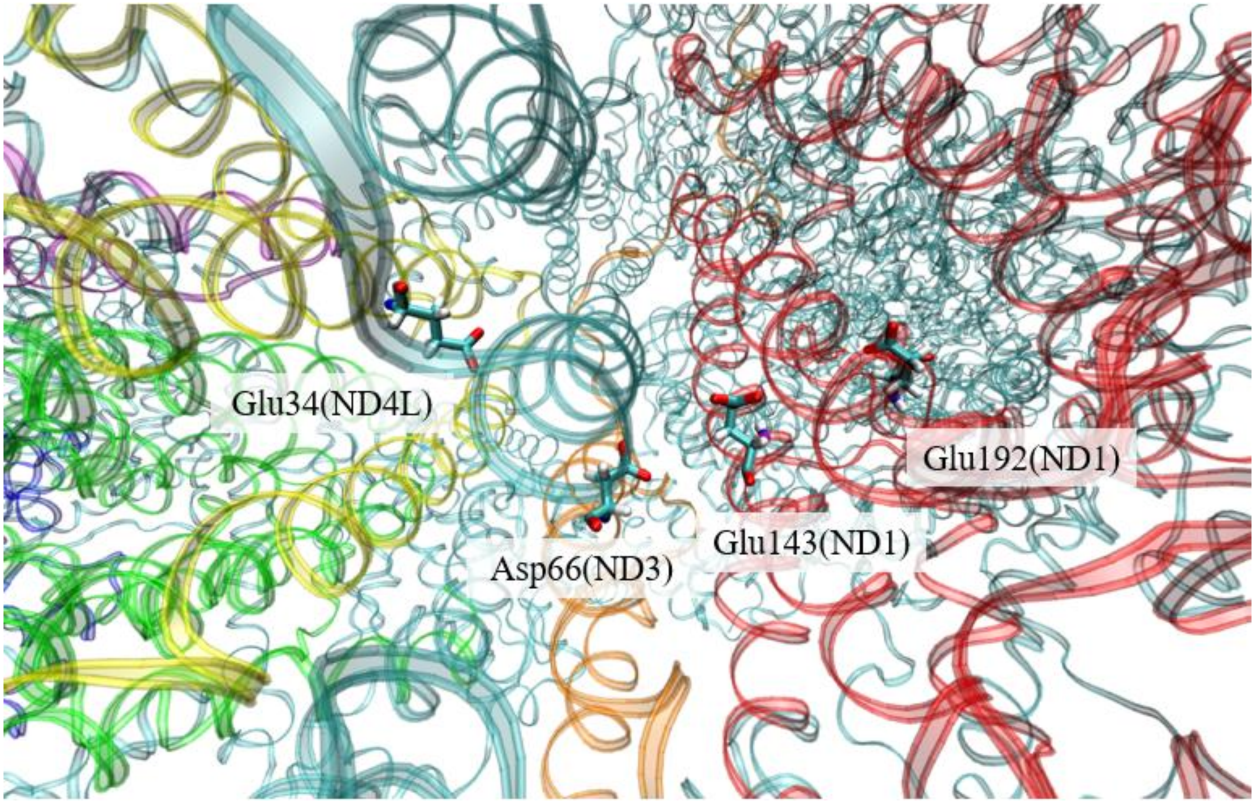

Figure 4.2 A detailed view of the ND1-ND4L putative pathway, internal waters and most sidechains omitted, showing only the key residues. Color coding is identical to Figure 4.1. The perspective of the image is looking from the P-side toward the N-side; PT is from Glu192 (right) to Glu34 (left).

4.2 Methodology

4.2.1 Classical and MS-RMD Modeling Details

The initial full 45-subunit mammalian mitochondrial Complex I structure was PDB ID 6ZTQ.[125 30] placed in a 1:2:2: cardiolipin:POPC:POPE membrane and solvated by ~250k TIP3P water molecules and 150mM NaCl. Bound FMN cofactor was present in the peripheral arm, and a ubiquinone molecule was modeled in the active site of Q_{10} reduction. The CHARMM36 force field was used for classical potentials,[87] except any replaced by MS-RMD potentials. The DFT-derived force field parameters derived by Gamiz-Hernandez et al. were used for the Fe-S clusters in the peripheral arm.[135] Residues Glu192 (ND1), Glu143(ND1), Asp66(ND3) and Glu34(ND4L) were modeled as EVB-active, using the parameters reported in Chapter 2. MS-EVB 3.2 parameters were used to treat the hydrated excess proton.[89] 200 ns classical equilibration in the NVT ensemble at 310 K was performed in LAMMPS.[84]

4.2.2 Collective Variables

As usual, it was necessary to bias the position of the CEC rather than any particular nucleus; the CEC was defined as[77]

$$\mathbf{r}_{\mathrm{CEC}} = \sum_i c_i^2 \, \mathbf{r}_i^{\mathrm{COC}} \tag{4.1}$$

In order to capture dynamic wetting of the channel during PT, it is necessary to include some bias on the hydration of the channel. This work uses the principal curve connectivity of the water network, ϕ, in the channel as described in ref[136]. To briefly restate the most important relationships,

$$\phi = \left(\prod_{i=1}^{N-1} f_{i,i+1} \right)^{\frac{1}{N-1}} \tag{4.2}$$

where ϕ runs from 0 to 1 and where the f_i are two-body connectivity variables defined as

$$f_{i,i+1} = \frac{I_i + I_{i+1}}{2} \tag{4.3}$$

and I_i is the Fermi function

$$I_i = \frac{1}{1 + \exp\left(-\frac{s_i - s_w}{\sigma}\right)} \tag{4.4}$$

Parameter s_i represents the water occupancy of the ith region along the curve whose connectivity is being determined, as measured by locations of water/hydronium oxygens and titratable sidechain atoms. s_w and σ are parameters which reflect when a region is fully occupied by water.

Biasing the motion of the proton along the pathway required first defining a curvilinear geometric CV for progress from Glu192 to $\mathrm{Glu34_{ND4L}}$. This was accomplished by first solvating the channel by biasing the water connectivity CV to 1 between the residues at the ends of the path, ensuring that there would be no dry regions in the channel. Then, a steered MD simulation was performed with a piecewise biasing potential, first between Glu192 and Glu143, then Glu143 and Asp66, and finally Asp66 and $\mathrm{Glu34_{ND4L}}$ such that the CEC was biased the length of the putative channel in linear segments between adjacent titratable residues. The biasing force was a relatively weak 5 kcal/mol$\mathrm{Å}^2$ to avoid forcing the CEC into energetically unfavorable positions, in advance of the next step. The adaptive path functionality of PLUMED[85, 86, 137, 138] was then used to relax the geometric CV, while keeping the channel fully hydrated and produce a curvilinear path CV along which to bias the CEC for the production runs.

4.2.4 MFEP Determination

Once the geometric CV had been determined, it was possible to estimate the MFEP of PT with respect to both the water connectivity and geometric CVs. First, starting structures were prepared corresponding to a reasonable guess path. Based on prior results in a hydrophobic synthetic system[139], and a hydrophobic biological system,[49] and on the fact that protons are rarely transported vehicularly in confined channels, it was assumed that a coherent water wire would form between the nth and n+1th residue along the pathway *before* the proton had entered the segment of the pathway between residues n and n+1. The guess path thus consisted of hydration of the Glu192-Glu143 segment, followed by transport to Glu143, then hydration of Glu143-Asp66, followed by PT to Asp66, and finally hydration of the Asp66-Glu34$_{\mathrm{ND4L}}$ segment, and the final leg of PT.

Then, the swarm of trajectories string method was employed to refine the guess path toward the true MFEP.[140-142] Swarms of thirty-two, 5 ps unbiased trajectories were run for each of fifty images along the guess MFEP. The water connectivity and geometric CV of each trajectory in the swarm was computed at the final step, used to define a vector indicating each trajectory's motion through CV space, and the vectors added together to describe the average motion of the entire swarm. This vector was decomposed into its projection along the vector between the nth bead and either the (n+1)th or (n-1)th bead, depending on whether the net motion was forward or backward along the path, and component perpendicular to that projection. The motion along the guess path was ignored and the perpendicular component was used to update the node positions. The positions of the nodes at the ends of the path were not changed. Convergence was assessed by measuring the RMSD between successive iterations of the string image positions.

4.2.5 Umbrella Sampling and WHAM

Umbrella sampling (US) simulations were performed using the RAPTOR modification[34] to LAMMPS[84] with the PLUMED plugin.[85, 86] 80 US windows were placed evenly along the length of the path. Initial window structures were taken from the prior steered MD simulations through the solvated channel so that no window was starting from an unphysically dry starting point. The system was integrated with a 1-fs timestep and the Nose-Hoover chain thermostat[91] in the NVT ensemble at 310 K. A harmonic restraint of 10 kcal/(mol* $\mathrm{\AA}^2$) was applied to each CV. Each window was equilibrated for 100ps, which was not used in the following free energy calculations;

windows were run a further 500 ps – 2 ns as required to converge. The PMF of PT along the path was computed using the weighted histogram analysis method (WHAM)[143].

4.3 Results and Discussion

4.3.1 Water Connectivity CV Degeneracy and Slow Degrees of Freedom

There are important effects of the choices made in the guess MFEP that bear further discussion. The guess MFEP chosen roughly corresponded to hydrating the entire first half of the channel from N-side bulk, concomitant with PT between the first several residues in the channel, until the proton reached Asp66, then hydrating the second half of the channel from Asp66 to the $\mathrm{Glu34_{ND4L}}$followed by PT from Asp66 to $\mathrm{Glu34_{ND4L}}$. This is a reasonable mechanism, given the result of ref[139], which showed a charge defect at the entrance to a hydrophobic cavity will tend to create a coherent water wire across the cavity before transporting. It is not, however, the only conceivable mechanism. One alternative mechanism, which can probably be ruled out, is that proton translocation and water wire formation are totally concurrent; that is, that the proton travels more or less continuously along the water wire as the water wire forms. (Imagine, as an analogy, an icebreaker ship, clearing a path as it goes.) Such a result would not be totally unintuitive, perhaps, but appears unlikely based on prior work, showing barriers ~25% higher for structures corresponding to incomplete hydration or water connectivity in two-dimensional PMFs of PT relative to completely hydrated/connected pathways.[49, 139]

However, one can also easily imagine a PT mechanism where, instead of the whole channel or any large part of it being fully connected by a water wire at any given time, instead the excess proton creates its own water wire to the first titratable residue along the path, transports down the chain, then the waters in the already existing water wire flow into the next segment of the channel, between the CEC and the next titratable residue, then the proton transports along the new water wire to the next titratable residue, and so on (**Figure 4.3**).

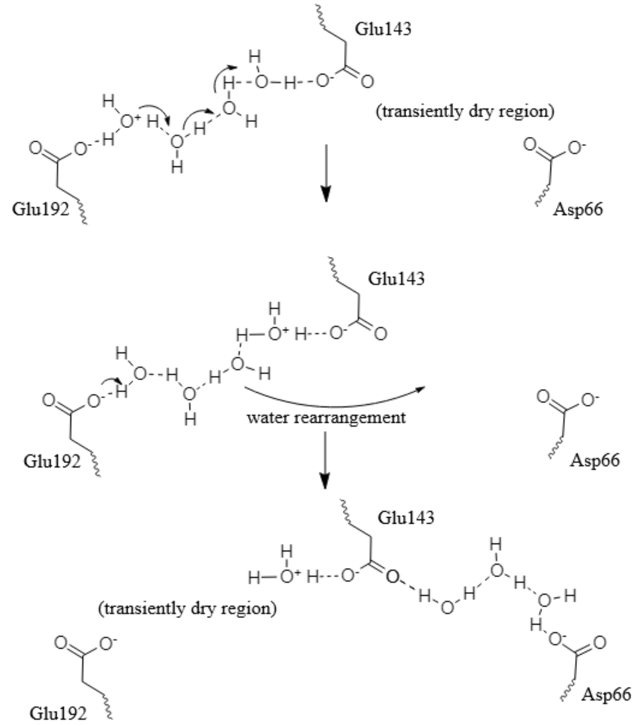

Figure 4.3 A hypothetical hydration mechanism using the ND1-ND4L channel as an example, where rather than a coherent water wire to bulk existing across the entire chain of titratable residues at any one time, the path hydrates in segments, leaving dry regions behind as the proton moves. This "leapfrog" mechanism is a plausible interpretation of the result that an excess proton can create its own water wires in a hydrophobic cavity, but presents certain modeling difficulties (see main text).

Because the principal curve connectivity CV ϕ essentially measures the number of adjacent water-connected beads along a certain path (see generally ref[74]), low values of connectivity like those arrangements shown in **Figure 4.3** may be degenerate or nearly degenerate with each other in their ϕ values. This problem would be particularly acute if biasing water connectivity meant increasing channel hydration in areas far from the CEC, and so likely irrelevant to the PMF of PT; that problem is solved in the CV by the use of a screening function which in effect filters beads far from the CEC out of the computation. However, when the relevant occupied beads are all relatively close to the CEC, and where the value of ϕ on either side of a transition like that hypothesized above is very similar due to a similar number of water molecules in a similar amount of space, there is a hidden slow degree of freedom present in a simulation that biases only the total ϕ across the entire channel rather than jointly biasing the individual ϕ_i in each segment of the channel. (There is an obvious reason not to take this latter step: it would defeat the entire purpose of attempting to reduce the dimensionality of the system down to one or two manageable dimensions. Simultaneously biasing numerous ϕ_i between, e.g. bulk and Glu192, Glu192 and Glu143, Glu143 and Asp66, and so on would be intractable. At best, perhaps, each separate ϕ_i could be biased *sequentially*, recovering a manageable two-dimensional problem, but this would in effect impose on the system the type of mechanism supposed above).

This limitation only becomes apparent when there are more than two reasonable choices of segments of the water channel, so two or more sufficiently distant locations titratable residues along the path. This is because in a system with a water channel composed of at most two reasonable, independently-hydrated segments, the only reasonable choices of mechanism are transport through a totally coherent water wire (corresponding to ϕ approaching 1, followed by PT along the relevant geometric CV) and transport through a coherent water wire to the location of the titratable residue(s), followed by formation of a coherent water wire in the second half and further transport. (Precise values of ϕ for such a mechanism would vary depending on how nearly each half actually was half the *distance* of the channel, but would be roughly 0.5 the whole time). The "leapfrog" mechanism supposed above is, in terms of water connectivity, *identical to* the mechanism that would be supposed from the carbon nanotube result of ref[139] in a two-segment system. Note a real example of the phenomenon described: in a forthcoming paper,[144] a designed protein with a hydrophobic channel containing a central Gln residue (a "hydrophobic gasket" system) was simulated. PT in the first half required a ϕ value of ~0.6 (starting from an equilibrium value of ~0.4). PT in the second half required a $\phi \approx 0.4$. Importantly, $\phi \approx 0.4$ in the second half did *not* correspond to the same water structure as the first-half equilibrium starting position. In a system of smaller channel segments separated regularly by titratable residues, the prospect of near-degeneracy of ϕ in adjacent segments becomes more acute.

4.3.2 PMF of PT in Complex I ND1-ND4L Channel

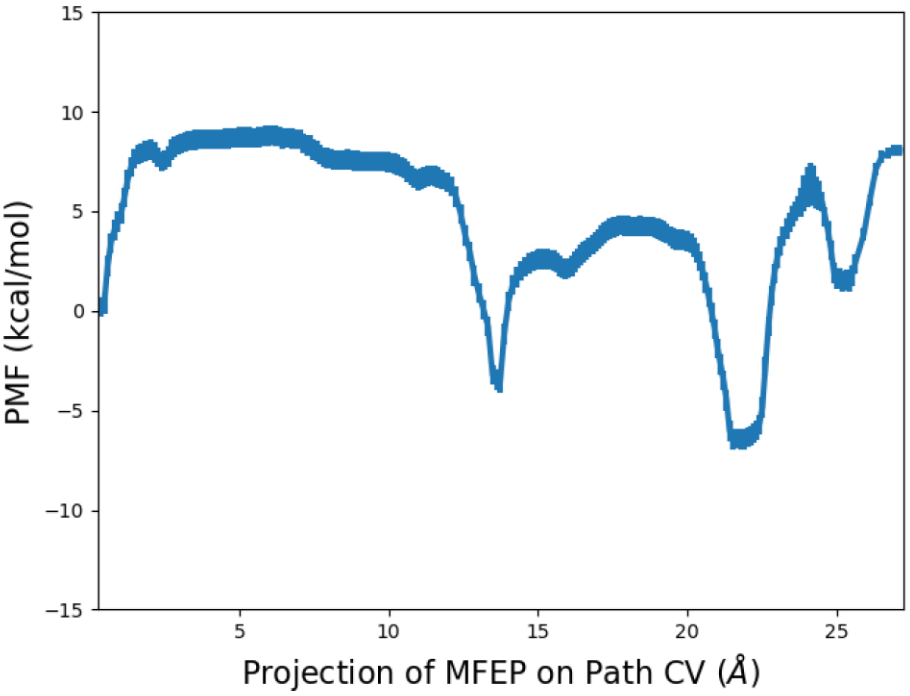

Figure 4.4 The PMF of PT along the minimum free energy path between Glu192 (ND1) and Glu34_{ND4L} computed by WHAM analysis of US simulations along the path. The wells, from left to right, correspond to: Glu192, Glu143, Asp66, and Glu34_{ND4L}. Four contact ion pair interactions are also visible, flanking the first three wells. PT for the first ~22 Å along the path is exergonic, with a maximum barrier of ~8 kcal/mol, and a $\Delta G \approx$-7 kcal/mol between Glu192 and Asp66. PT between Asp66 and Glu34_{ND4L}, conversely, is endergonic by ~9 kcal/mol, with a barrier of ~13 kcal/mol corresponding largely to the energetic cost of hydrating the previously-dry channel. PT progress along coherent water wires is largely isoenergetic, as evidenced by the approximately flat plateaus.

4.3.2.1 The exergonic portion

The PMF of PT along the MFEP of the ND1-ND4L channel with respect to water connectivity in the channel and proton position, projected on to the proton position dimension, is presented in **Figure 4.4**. The PMF reveals that PT is exergonic between Glu192 and Asp66, with a $\Delta\mathrm{G} \approx -6.5\,\mathrm{kcal/mol}$. This suggests that Asp66 is likely to be protonated from bulk spontaneously. This result is in excellent agreement with the suggestion of prior work that proton uptake from the N-side to titratable residues in ND1 is exergonic by about 4 kcal/mol. [51, 145] It also suggests that this initial transfer step, through segments of the channel that are ordinarily partially hydrated, is not the rate-limiting step of the process. Using the rough rate estimation of ref[51] motivated by transition state theory,[146]

$$k_{ij} = 6.4\ \mathrm{ps}^{-1} * \exp\left(-\frac{\Delta G^{TST}}{RT}\right) \tag{4.5}$$

the rate of PT between Glu192 and Asp66 suggested by the ~8kcal/mol barrier is about 16 $\mu\mathrm{s}^{-1}$, far more rapid that the millisecond turnover rate of the enzyme. Furthermore, this barrier itself is in excellent agreement with hypotheses that PT from Glu192 is mediated principally by the energy associated with hydrating the channel, previously estimated at 8-10 kcal/mol.[51]

4.3.2.2 The endergonic portion

PT from Asp66 to $\mathrm{Glu34_{ND4L}}$ is endergonic, with a $\Delta G \approx 7\mathrm{kcal/mol}$ and a barrier height of 13 kcal/mol. This barrier is large, corresponding to a pKa for Asp66 deprotonation to $\mathrm{Glu34_{ND4L}}$ of >8. The base-shift of Asp66 compared to its bulk pKa of ~4 is extreme, representing deprotonation frequency reduced by a factor of over ten thousand relative to bulk. Performing a rate calculation again using eq 4.5 for the barrier involved suggests a rate of about 4 per millisecond, on the order of the turnover rate of the enzyme.

This produces a number of insights into the method and the system. One methodological insight is that the Asp FitRMD models first reported in Chapter 2 remain robust and transferrable to large changes in local environment. Note that the pKa here for Asp-Glu transfer is base-shifted to a degree comparable to SNase Asp66 (see Chapter 2). Second, this high barrier indicates that PT from Asp66 to $\mathrm{Glu34_{ND4L}}$ (and afterwards to P-side bulk) must, absent some sort of large protein conformational change, be powered by the driving force produced ultimately by the redox reaction in the peripheral arm of the protein. That, in itself, is no surprise: Complex I is a pump, and so at least on average the four proton conduction events will be endergonic. It is, however, an encouraging result for the prospect that the fourth proton transport event happens through the ND1-ND4L channel identified in this and prior works, and that this channel, which links to the other three proton channels, is the likely pathway by which the driving force is initially transmitted down the membrane arm. It is further worth noting that the barrier computed in this work, ~13 kcal/mol, is significantly smaller than the available free energy from reduction of ubiquinone ($\Delta\mathrm{G} \approx -18.5$ kcal/mol); furthermore, because, per these calculations, the initial PT between Glu192 and Asp66 is *exergonic*, by about 6 kcal/mol, that energy may also be harnessed by the enzyme to do some of the work. These results also suggest that this proton channel, as compared to the other three, might require a disproportionate share of the driving force available, but this is not entirely unexpected. Given the hydration dynamics at equilibrium, where the three more distant proton channels all display coherent water wires from N-side to P-side bulk, the work required to transport a proton would be expected to correspond almost entirely to the relative pH of the N-side and P-side; essentially, if a coherent water wire already exists, there should be no work associated with first hydrating the channel, and the energy required for PT ought to be proportional to the *pmf* under the chosen conditions. It is natural that when there are dry portions of the channel at equilibrium, on the other hand, that hydrating them comes with some energetic cost, and that therefore, all other things being equal, the work put into transporting a proton through the ordinarily partially dry channel, as here, would consume a disproportionate share of the available potential.

It is worth taking a moment to discuss the overall energetics suggested by the PMF as well. The result in **Figure 4.4** shows that PT from Glu192 to $\mathrm{Glu34}_{\mathrm{ND4L}}$is endergonic by about 2 kcal/mol, and that there is a further barrier of about ~6 kcal/mol associated with the start of proton release to the P-side. This result, that lateral transport is nearly energetically degenerate (though weakly endergonic), while the full N-side-lateral transport-P-side path is likely more strongly endergonic, agrees well with existing literature and prior suggestions of the enzyme's mechanism. In particular, the nearly isoenergetic lateral pathway is a positive result in line with prior predictions given the capacity of the enzyme to run in reverse under conditions of high external *pmf*.[122] On the other hand, the qualitative result that the overall pathway is likely endergonic is also a positive one for the validity of the models; Complex I being a pump, proton translocation ought to require energy. The magnitudes of the energy required are highly correlated to the precise value of the external *pmf*, though under physiologically reasonable conditions of an external 200 mV potential, the overall average energy requirement to translocate a proton is 4.6 kcal/mol; in light of this figure, our results here seem quite reasonable.[122]

4.3.2.3 P-side proton uptake simulations indicate importance of interaction of Glu residues

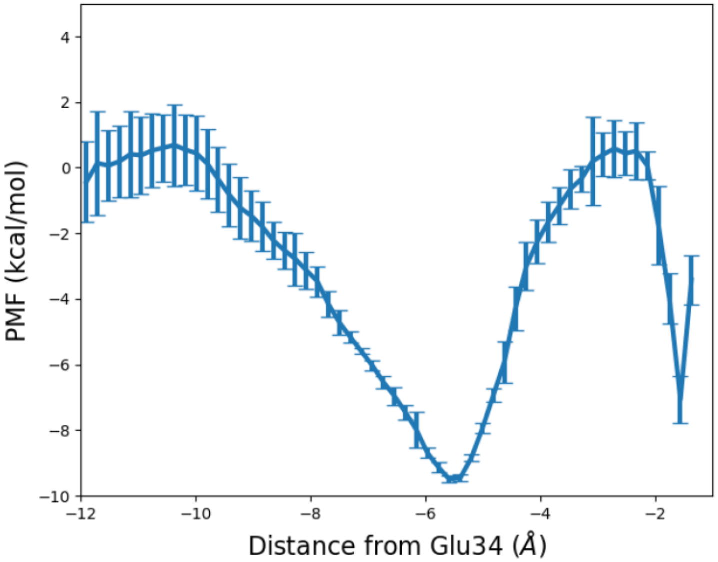

Figure 4.5 A PMF of proton uptake from bulk to Glu34$_{ND4L}$. The sharp well at right corresponds to proton residence on Glu34$_{ND4L}$ and was expected, but the broad well which appears to represent a global minimum was unexpected. Further inspection of the trajectories revealed that this well corresponds to the CEC interacting strongly with three nearby Glu residues at the gateway between the ND1-ND4L and ND2 lateral proton channels. Error bars computed by block averaging over four blocks.

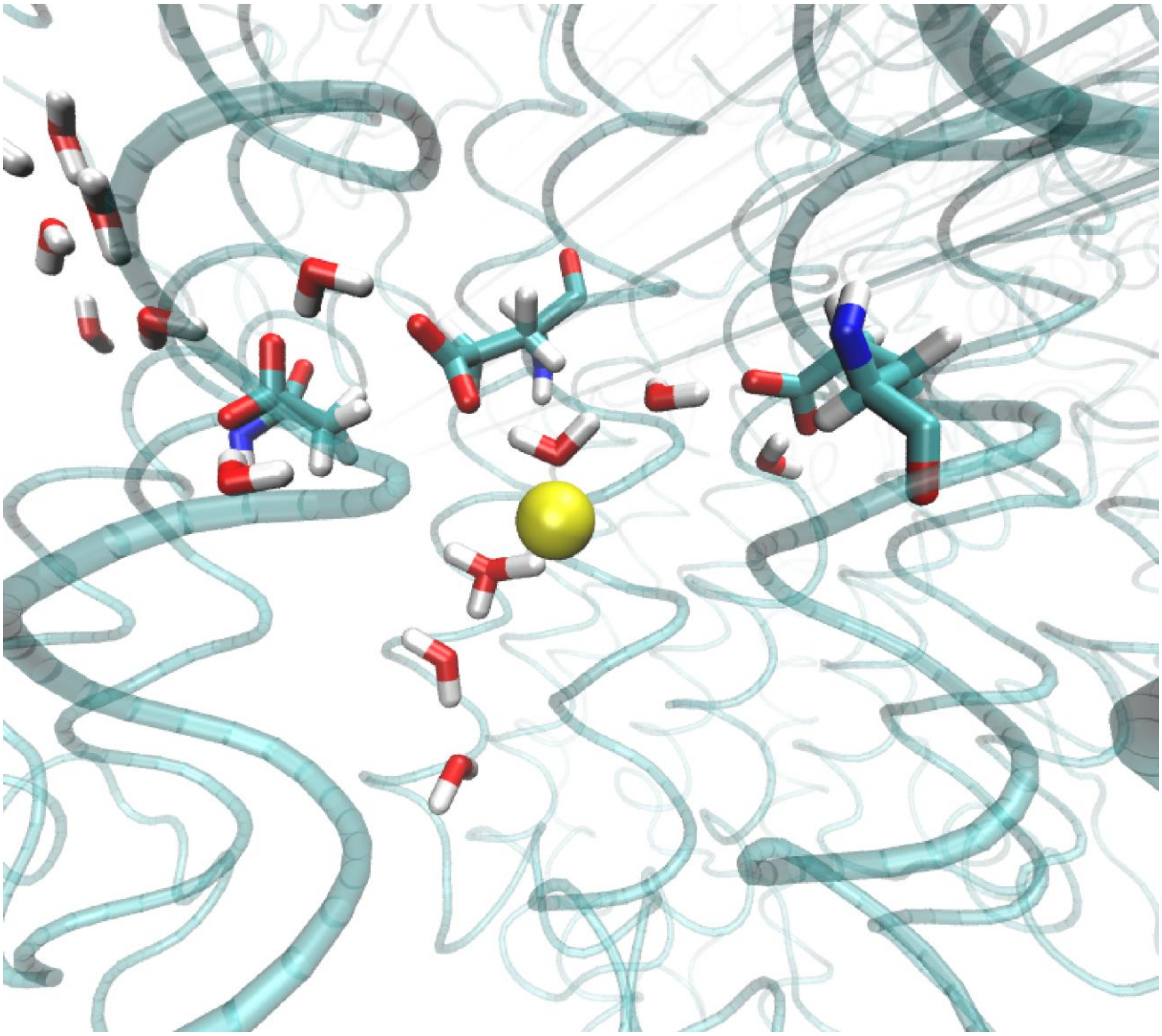

Figure 4.6 A snapshot of PT simulations from the P-side bulk to Glu34$_{ND4L}$ corresponding to the broad well in the PMF of proton uptake (Figure 4.5). The perspective is looking from the P-side toward the N-side. Domain ND4L is at right, ND2 is at left and center. The CEC position is shown as a yellow sphere. Three residues, Glu34$_{ND4L}$ (right) and Glu34$_{ND2}$ (left), and Glu70$_{ND2}$ (center) are also shown. The broad well at ~-5 Å in Figure 4.5 may represent a global minimum in the PMF of proton uptake to ND4L/ND2 from bulk; if that is the case, it appears to be due that the fact that the three Glu residues are just far enough apart that the CEC can participate at least transiently in contact ion pair interactions with all three, and potentially with multiple at once.

Umbrella sampling simulations along the channel axis from Glu34$_{\mathrm{ND4L}}$ to the bulk were performed to produce insights into the energetics of PT to and from the P-side bulk. No further residues were modeled as reactive than those previously described. The PMF computed from these simulations is shown in **Figure 4.5**. Note the broad, ~9 kcal/mol well at spanning ~4-8 Å from the Glu sidechain. This well appears to correspond to proton interaction with another pair of carboxylate residues, Glu34 and Glu70 of the ND2 domain. The position of the well roughly corresponds to the distance at which favorable contact ion pair interactions with each of the neighboring Glu residues could form (see Chapter 2 and ref [72]). $\mathrm{Glu34_{ND2}}$ is known to be part of an ion pair, with Lys105 of ND2, that gates proton transfer between ND4L and ND2.[134] Interestingly, prior modeling work on these residues in their deprotonated states has indicated that, at equilibrium, no water wire exists between ND4L and ND2.[134] Here, as the aim was simply to sample $\mathrm{Glu34_{ND4L}}$-bulk uptake, the water connectivity between Glu70 and the two Glu34 residues was not biased. Nevertheless, as is obvious from **Figure 4.6**, a coherent water has formed without any apparent high barrier resulting from failure to bias a slow degree of freedom. This result has two important implications.

First, PT between ND4L and ND2 may be more rapid than previously thought, given the ease with which the supposedly dry region was solvated. While a follow-on study to determine the energetics of what is now seen to be a bulk-Glu triplet transport process rather than a bulk-$\mathrm{Glu34_{ND4L}}$ transport process, involving explicit reactivity of the two ND2 Glu residues and an explicit bias on the water connectivity between the residues, would be appropriate to be able to be confident of the quantitative accuracy of the calculations, it would be strange if the result of that study were to show a higher barrier, since it would by supposition account for more of the slow degrees of freedom. Thus, we can be fairly confident that the data here represent at worst an upper bound on the facility of forming a connected water network among the three Glu residues in the presence of an excess proton.

Second, the fact that the putative contact ion pair well is approximately equidistant between the two Glu34 residues and appears to represent a global minimum, or something close to it, in the proton uptake process, indicates that when Complex I runs in reverse, the mechanism of PT may be qualitatively different from the forward mechanism. The reverse mechanism of Complex I is not nearly as well-studied as the forward mechanism, for the obvious reason that the forward process is far more biologically relevant; studies suggest that the reversal of PT in Complex I only occurs naturally in very high *pmf* environments. It would be a natural assumption, however, that the reverse process would be the opposite, mechanistically, of the forward process, such that uptake to $\mathrm{Glu34_{ND4L}}$ resulted in transfer to Asp66 and beyond, laterally down the ND1-ND4L chain. This result suggests that, in fact, lateral PT through ND2 (which would trace the same path as forward PT through that domain), is also an option. The idea that the reverse PT process may fork at the $\mathrm{Glu34_{ND2}}$-Glu70-$\mathrm{Glu34_{ND4L}}$ triplet warrants further investigation. This may represent an important step in demonstrating the predictivity of the FitRMD/MS-RMD methodology.

The broadness of the well in question merits a brief comment, because it is unusual in appearance compared to ordinary PT minima. While the typical global minimum is proton residence on a residue (or in the bulk, depending on the system) and such wells are usually narrow (as the well above -2 Å in **Figure 4.5**), the broad well does not seem to be due to a bonded interaction at all, but to a coulombic interaction. The three Glu residues are so close together (the maximum extent of the $\mathrm{Glu34_{ND4L}}$-$\mathrm{Glu34_{ND2}}$ distance being about 14 Å, and less when the sidechains are rotated toward each other) that the CEC can apparently form a contact ion pair with one or more of the Glu carboxylate groups by small movements.

4.4 Conclusions and Future Work

In this chapter, several important results have been presented. First, the models of Chapter 2 were successfully applied to a channel containing several titratable carboxylate residues resulting in a PMF that suggests that PT to other domains of Complex I even through the ordinarily dry Asp66-Glu34$_{\mathrm{ND4L}}$ region is feasible, when hydration is properly biased. A novel mechanism for the reverse process has also been suggested, involving the simultaneous interaction of three Glu residues with the CEC, based on simulations of proton uptake from the P-side, which suggests that the reverse mechanism of Complex I may be qualitatively different from the forward mechanism. Additionally, some areas for further investigation and model development have been identified. Further study of alternate hydration mechanisms in systems such as this is warranted, and the development or modification of order parameters that account for the potential slow degrees of freedom involved may be necessary to make such simulations tractable. Moreover, our results suggest that both further investigation of the lateral PT process between Asp66 and the other Complex I membrane domains and the reverse process beyond each of the Glu 34 residues will produce new insights into the function of this biologically essential protein.

CHAPTER 5

Concluding Remarks

Taken together, the work presented in this thesis is a multi-faceted effort to improve scientific understanding of proton transport systems in biomolecular systems not well-suited to simulation by either ordinary classical or quantum mechanical computational methods. In Chapters 2 and 3, methodological advances were presented that improve on prior parameterization techniques for multiscale reactive molecular dynamics models of amino acids. The results presented represent an application of physics-constrained machine learning, whereby fitting MS-RMD parameters to an appropriate choice of electronic data by minimizing the residual of the diabatic state vector produces a model that captures the correct underlying physics of the system, and is thus more transferrable among different environments than prior, system-specific models. New or improved models were presented for glutamic acid, aspartic acid, and histidine; the prospect was also raised of further refining the models and improving the iterative fitRMD procedure by following diabatic matching with force-matching or by employing a more versatile quantum mechanics procedure to the generation of the electronic structure data.

In Chapter 4, the application of the carboxylate models to a real system, mitochondrial respiratory complex I, was presented. The results of this work demonstrate that the models remain robust to changes in chemical environment, producing physically accurate results even when the frequency of de/protonation of a residue, the pKa, is many orders of magnitude different from its bulk value. This section also once again demonstrates the importance of explicit treatment of hydration when conducting studies of proton transport, by showing that transport through a crystallographically dry region of Complex I is relatively facile, and well within the physical limitations set by the redox potential available; indeed, the fact that the turnover rate of the entire enzyme is on the order of the rate of deprotonation of Asp66 suggests that the latter residue may be vitally important to the enzyme's overall PT mechanism. As a general matter, the results presented in this section show that even extremely large biological systems are tractable in the MS-RMD approach, and that quantitative accuracy can be achieved over large distances at reasonable cost, contingent upon appropriate models. Predictions of the importance of Asp66 and a trio of Glu residues have been made that are amenable to experimental validation, and can focus future studies of the enzyme on the aspects most important to PT.

Bibliography

(1) Mitchell, P. Coupling of Phosphorylation to Electron and Hydrogen Transfer by a Chemi-Osmotic type of Mechanism. *Nature* **1961**, *191* (4784), 144-148. DOI: 10.1038/191144a0.
(2) Decoursey, T. E. Voltage-gated proton channels and other proton transfer pathways. *Physiol Rev* **2003**, *83* (2), 475-579. DOI: 10.1152/physrev.00028.2002 From NLM.
(3) 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. DOI: https://doi.org/10.1016/j.bbabio.2006.06.017.
(4) Gunner, M. R.; Amin, M.; Zhu, X.; Lu, J. Molecular mechanisms for generating transmembrane proton gradients. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2013**, *1827* (8), 892-913. DOI: https://doi.org/10.1016/j.bbabio.2013.03.001.
(5) Gunner, M. R.; Koder, R. The design features cells use to build their transmembrane proton gradient. *Physical Biology* **2017**, *14* (1), 013001. DOI: 10.1088/1478-3975/14/1/013001.
(6) Tsukihara, T.; Shimokata, K.; Katayama, Y.; Shimada, H.; Muramoto, K.; Aoyama, H.; Mochizuki, M.; Shinzawa-Itoh, K.; Yamashita, E.; Yao, M.; et al. The low-spin heme of cytochrome c oxidase as the driving element of the proton-pumping process. *Proc Natl Acad Sci U S A* **2003**, *100* (26), 15304-15309. DOI: 10.1073/pnas.2635097100 From NLM.
(7) Stanley, K. S.; Meyer, D. J.; Gatto, C.; Artigas, P. Intracellular Requirements for Passive Proton Transport through the Na(+),K(+)-ATPase. *Biophys J* **2016**, *111* (11), 2430-2439. DOI: 10.1016/j.bpj.2016.09.042 From NLM.
(8) Wright, J. J.; Biner, O.; Chung, I.; Burger, N.; Bridges, H. R.; Hirst, J. Reverse Electron Transfer by Respiratory Complex I Catalyzed in a Modular Proteoliposome System. *Journal of the American Chemical Society* **2022**, *144* (15), 6791-6801. DOI: 10.1021/jacs.2c00274.
(9) Agmon, N.; Bakker, H. J.; Campen, R. K.; Henchman, R. H.; Pohl, P.; Roke, S.; Thämer, M.; Hassanali, A. Protons and Hydroxide Ions in Aqueous Systems. *Chemical Reviews* **2016**, *116* (13), 7642-7672. DOI: 10.1021/acs.chemrev.5b00736.
(10) Kaur, D.; Khaniya, U.; Zhang, Y.; Gunner, M. R. Protein Motifs for Proton Transfers That Build the Transmembrane Proton Gradient. *Frontiers in Chemistry* **2021**, *9*, Review. DOI: 10.3389/fchem.2021.660954.
(11) Henderson, J. A.; Huang, Y.; Beckstein, O.; Shen, J. Alternative proton-binding site and long-distance coupling in <i>Escherichia coli</i> sodium–proton antiporter NhaA. *Proceedings of the National Academy of Sciences* **2020**, *117* (41), 25517-25522. DOI: doi:10.1073/pnas.2005467117.
(12) Maloney, P. C.; Kashket, E. R.; Wilson, T. H. A protonmotive force drives ATP synthesis in bacteria. *Proc Natl Acad Sci U S A* **1974**, *71* (10), 3896-3900. DOI: 10.1073/pnas.71.10.3896 From NLM.
(13) Accardi, A.; Picollo, A. CLC channels and transporters: Proteins with borderline personalities. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2010**, *1798* (8), 1457-1464. DOI: https://doi.org/10.1016/j.bbamem.2010.02.022.
(14) Aoi, W.; Marunaka, Y. Importance of pH homeostasis in metabolic health and diseases: crucial role of membrane proton transport. *Biomed Res Int* **2014**, *2014*, 598986. DOI: 10.1155/2014/598986 From NLM.
(15) Sachs, G. Proton pump inhibitors and acid-related diseases. *Pharmacotherapy* **1997**, *17* (1), 22-37. From NLM.
(16) Kasper, D.; Planells-Cases, R.; Fuhrmann, J. C.; Scheel, O.; Zeitz, O.; Ruether, K.; Schmitt, A.; Poët, M.; Steinfeld, R.; Schweizer, M.; et al. Loss of the chloride channel ClC-7 leads to lysosomal storage disease and neurodegeneration. *Embo j* **2005**, *24* (5), 1079-1091. DOI: 10.1038/sj.emboj.7600576 From NLM.
(17) Parekh, A. Recent developments of proton exchange membranes for PEMFC: A review. *Frontiers in Energy Research* **2022**, *10*, Mini Review. DOI: 10.3389/fenrg.2022.956132.
(18) Knight, C.; Voth, G. A. The Curious Case of the Hydrated Proton. *Accounts of Chemical Research* **2012**, *45* (1), 101-109. DOI: 10.1021/ar200140h.

(19) Clever, H. L. The hydrated hydronium ion. *Journal of Chemical Education* **1963**, *40* (12), 637. DOI: 10.1021/ed040p637.

(20) Zundel, G.; Metzger, H. Energiebänder der tunnelnden Überschuß-Protonen in flüssigen Säuren. Eine IR-spektroskopische Untersuchung der Natur der Gruppierungen H5O2+. *Zeitschrift für Physikalische Chemie* **1968**, *58* (5_6), 225-245. DOI: doi:10.1524/zpch.1968.58.5_6.225 (acccessed 2023-03-09).

(21) Eigen, M. Proton Transfer, Acid-Base Catalysis, and Enzymatic Hydrolysis. Part I: ELEMENTARY PROCESSES. *Angewandte Chemie International Edition in English* **1964**, *3* (1), 1-19. DOI: 10.1002/anie.196400011.

(22) 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. DOI: 10.1021/jacs.1c08552.

(23) Grotthuss, C. J. T. d. Sur la decomposition de l'eau et des corps qu'elle tient en dissolution a l'aide dl'electrique galvanique. *Ann. Chim.* **1806**, *58*, 54-73.

(24) Agmon, N. The Grotthuss mechanism. *Chemical Physics Letters* **1995**, *244* (5), 456-462. DOI: https://doi.org/10.1016/0009-2614(95)00905-J.

(25) Swanson, J. M.; Simons, J. Role of charge transfer in the structure and dynamics of the hydrated proton. *J Phys Chem B* **2009**, *113* (15), 5149-5161. DOI: 10.1021/jp810652v From NLM.

(26) Zhu, Z.; Paddison, S. J. Perspective: Morphology and ion transport in ion-containing polymers from multiscale modeling and simulations. *Frontiers in Chemistry* **2022**, *10*, Review. DOI: 10.3389/fchem.2022.981508.

(27) Day, T. J. F.; Soudackov, A. V.; Čuma, M.; Schmitt, U. W.; Voth, G. A. A second generation multistate empirical valence bond model for proton transport in aqueous systems. *The Journal of Chemical Physics* **2002**, *117* (12), 5839-5849. DOI: 10.1063/1.1497157.

(28) Bernal, J. D.; Fowler, R. H. A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions. *The Journal of Chemical Physics* **1933**, *1* (8), 515-548. DOI: 10.1063/1.1749327.

(29) Cukierman, S. Et tu, Grotthuss! and other unfinished stories. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2006**, *1757* (8), 876-885. DOI: https://doi.org/10.1016/j.bbabio.2005.12.001.

(30) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Physical Review Letters* **1985**, *55* (22), 2471-2474. DOI: 10.1103/PhysRevLett.55.2471.

(31) Pezeshki, S.; Davis, C.; Heyden, A.; Lin, H. Adaptive-Partitioning QM/MM Dynamics Simulations: 3. Solvent Molecules Entering and Leaving Protein Binding Sites. *Journal of Chemical Theory and Computation* **2014**, *10* (11), 4765-4776. DOI: 10.1021/ct500553x.

(32) Stuart, S.; Tutein, A.; Harrison, J. A reactive potential for hydrocarbons with intermolecular interactions. *The Journal of Chemical Physics* **2000**, *112*, 6472-6486. DOI: 10.1063/1.481208.

(33) Van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *The Journal of Physical Chemistry A* **2001**, *105* (41), 9396-9409. DOI: 10.1021/jp004368u.

(34) 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. DOI: 10.1021/ct3006437.

(35) Nelson, J. G.; Peng, Y.; Silverstein, D. W.; Swanson, J. M. J. Multiscale Reactive Molecular Dynamics for Absolute pKa Predictions and Amino Acid Deprotonation. *Journal of Chemical Theory and Computation* **2014**, *10* (7), 2729-2737. DOI: 10.1021/ct500250f.

(36) 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. DOI: 10.1021/acs.jctc.5b01109.

(37) Liang, R.; Li, H.; Swanson, J. M. J.; 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. DOI: 10.1073/pnas.1401997111 (acccessed 2023/03/12).

(38) 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-E6964. DOI: 10.1073/pnas.1615471113 (acccessed 2023/03/12).
(39) Lee, S.; Swanson, J. M.; Voth, G. A. Multiscale Simulations Reveal Key Aspects of the Proton Transport Mechanism in the ClC-ec1 Antiporter. *Biophys J* **2016**, *110* (6), 1334-1345. DOI: 10.1016/j.bpj.2016.02.014 From NLM.
(40) 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. DOI: 10.1021/jacs.6b06683.
(41) 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), 5924-5929. DOI: 10.1073/pnas.1703654114 (acccessed 2023/03/12).
(42) 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.; et al. Proton movement and coupling in the POT family of peptide transporters. *Proceedings of the National Academy of Sciences* **2017**, *114* (50), 13182-13187. DOI: 10.1073/pnas.1710727114 (acccessed 2023/03/12).
(43) 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. DOI: 10.1021/jacs.8b07614.
(44) 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. DOI: 10.1021/jacs.7b11463.
(45) Watkins, L. C.; Liang, R.; Swanson, J. M. J.; DeGrado, W. F.; Voth, G. A. Proton-Induced Conformational and Hydration Dynamics in the Influenza A M2 Channel. *Journal of the American Chemical Society* **2019**, *141* (29), 11667-11676. DOI: 10.1021/jacs.9b05136.
(46) Watkins, L. C.; DeGrado, W. F.; Voth, G. A. Influenza A M2 Inhibitor Binding Understood through Mechanisms of Excess Proton Stabilization and Channel Dynamics. *Journal of the American Chemical Society* **2020**, *142* (41), 17425-17433. DOI: 10.1021/jacs.0c06419.
(47) 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. DOI: https://doi.org/10.1016/j.bpj.2020.07.027.
(48) 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. DOI: 10.1073/pnas.2101932118 (acccessed 2023/03/12).
(49) Watkins, L. C.; Degrado, W. F.; Voth, G. A. Multiscale Simulation of an Influenza A M2 Channel Mutant Reveals Key Features of Its Markedly Different Proton Transport Behavior. *Journal of the American Chemical Society* **2022**, *144* (2), 769-776. DOI: 10.1021/jacs.1c09281.
(50) Li, C.; Yue, Z.; Newstead, S.; Voth, G. A. Proton coupling and the multiscale kinetic mechanism of a peptide transporter. *Biophys J* **2022**, *121* (12), 2266-2278. DOI: 10.1016/j.bpj.2022.05.029 From NLM.
(51) Kaila, V. R. I. Long-range proton-coupled electron transfer in biological energy conversion: towards mechanistic understanding of respiratory complex I. *Journal of The Royal Society Interface* **2018**, *15* (141), 20170916. DOI: 10.1098/rsif.2017.0916 (acccessed 2023/03/12).
(52) Hirst, J. Mitochondrial Complex I. *Annual Review of Biochemistry* **2013**, *82* (1), 551-575. DOI: 10.1146/annurev-biochem-070511-103700 (acccessed 2023/03/12).
(53) Brandt, U. Energy converting NADH:quinone oxidoreductase (complex I). *Annu Rev Biochem* **2006**, *75*, 69-92. DOI: 10.1146/annurev.biochem.75.103004.142539 From NLM.
(54) Sazanov, L. A. A giant molecular proton pump: structure and mechanism of respiratory complex I. *Nat Rev Mol Cell Biol* **2015**, *16* (6), 375-388. DOI: 10.1038/nrm3997 From NLM.

(55) Verkhovskaya, M. L.; Belevich, N.; Euro, L.; Wikström, M.; Verkhovsky, M. I. Real-time electron transfer in respiratory complex I. *Proceedings of the National Academy of Sciences* **2008**, *105* (10), 3763-3767. DOI: 10.1073/pnas.0711249105 (acccessed 2023/03/12).

(56) de Vries, S.; Dörner, K.; Strampraad, M. J. F.; Friedrich, T. Electron Tunneling Rates in Respiratory Complex I Are Tuned for Efficient Energy Conversion. *Angewandte Chemie International Edition* **2015**, *54* (9), 2844-2848, https://doi.org/10.1002/anie.201410967. DOI: https://doi.org/10.1002/anie.201410967 (acccessed 2023/03/12).

(57) Sazanov, L. A.; Hinchliffe, P. Structure of the hydrophilic domain of respiratory complex I from Thermus thermophilus. *Science* **2006**, *311* (5766), 1430-1436. DOI: 10.1126/science.1123809 From NLM.

(58) Yoshida, M.; Muneyuki, E.; Hisabori, T. ATP synthase--a marvellous rotary engine of the cell. *Nat Rev Mol Cell Biol* **2001**, *2* (9), 669-677. DOI: 10.1038/35089509 From NLM.

(59) Röpke, M.; Saura, P.; Riepl, D.; Pöverlein, M. C.; Kaila, V. R. I. Functional Water Wires Catalyze Long-Range Proton Pumping in the Mammalian Respiratory Complex I. *Journal of the American Chemical Society* **2020**, *142* (52), 21758-21766. DOI: 10.1021/jacs.0c09209.

(60) Lambert, A. J.; Brand, M. D. Superoxide production by NADH:ubiquinone oxidoreductase (complex I) depends on the pH gradient across the mitochondrial inner membrane. *Biochem J* **2004**, *382* (Pt 2), 511-517. DOI: 10.1042/bj20040485 From NLM.

(61) Murphy, M. P. How mitochondria produce reactive oxygen species. *Biochem J* **2009**, *417* (1), 1-13. DOI: 10.1042/bj20081386 From NLM.

(62) Pryde, K. R.; Hirst, J. Superoxide is produced by the reduced flavin in mitochondrial complex I: a single, unified mechanism that applies during both forward and reverse electron transfer. *J Biol Chem* **2011**, *286* (20), 18056-18065. DOI: 10.1074/jbc.M110.186841 From NLM.

(63) Kampjut, D.; Sazanov, L. A. The coupling mechanism of mammalian respiratory complex I. *Science* **2020**, *370* (6516), eabc4209. DOI: 10.1126/science.abc4209 (acccessed 2023/03/12).

(64) Mühlbauer, M. E.; Saura, P.; Nuber, F.; Di Luca, A.; Friedrich, T.; Kaila, V. R. I. Water-Gated Proton Transfer Dynamics in Respiratory Complex I. *Journal of the American Chemical Society* **2020**, *142* (32), 13718-13728. DOI: 10.1021/jacs.0c02789.

(65) Parey, K.; Haapanen, O.; Sharma, V.; Köfeler, H.; Züllig, T.; Prinz, S.; Siegmund, K.; Wittig, I.; Mills, D. J.; Vonck, J.; et al. High-resolution cryo-EM structures of respiratory complex I: Mechanism, assembly, and disease. *Science Advances 5* (12), eaax9484. DOI: 10.1126/sciadv.aax9484 (acccessed 2023/03/12).

(66) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99* (6), 1129-1143. DOI: 10.1016/j.neuron.2018.08.011.

(67) Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*; Cambridge University Press, 2009. DOI: DOI: 10.1017/CBO9780511609633.

(68) Groenhof, G. Introduction to QM/MM Simulations. Humana Press, 2013; pp 43-66.

(69) Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y.; Shin, Y. K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M. J.; Aktulga, H. M.; et al. The ReaxFF reactive force-field: development, applications and future directions. *npj Computational Materials* **2016**, *2* (1), 15011. DOI: 10.1038/npjcompumats.2015.11.

(70) Knight, C.; Lindberg, G. E.; Voth, G. A. Multiscale reactive molecular dynamics. *The Journal of Chemical Physics* **2012**, *137* (22), 22A525. DOI: 10.1063/1.4743958.

(71) Voth, G. A. Computer Simulation of Proton Solvation and Transport in Aqueous and Biomolecular Systems. *Accounts of Chemical Research* **2006**, *39* (2), 143-150. DOI: 10.1021/ar0402098.

(72) 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. DOI: 10.1021/acs.jpcb.1c05992.

(73) Reijenga, J.; van Hoof, A.; van Loon, A.; Teunissen, B. Development of Methods for the Determination of pKa Values. *Anal Chem Insights* **2013**, *8*, 53-71. DOI: 10.4137/ACI.S12304 PubMed.

(74) Li, C.; Voth Gregory, 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. DOI: 10.1073/pnas.2113141118 (acccessed 2022/07/07).

(75) Maupin, C. M.; Wong, K. F.; Soudackov, A. V.; Kim, S.; Voth, G. A. A multistate empirical valence bond description of protonatable amino acids. *J Phys Chem A* **2006**, *110* (2), 631-639. DOI: 10.1021/jp053596r From NLM.
(76) Wu, Y.; Chen, H.; Wang, F.; Paesani, F.; Voth, G. A. An Improved Multistate Empirical Valence Bond Model for Aqueous Proton Solvation and Transport. *The Journal of Physical Chemistry B* **2008**, *112* (2), 467-482. DOI: 10.1021/jp076658h.
(77) Swanson, J. M. J.; Maupin, C. M.; Chen, H.; Petersen, M. K.; Xu, J.; Wu, Y.; Voth, G. A. Proton Solvation and Transport in Aqueous and Biomolecular Systems: Insights from Computer Simulations. *The Journal of Physical Chemistry B* **2007**, *111* (17), 4300-4314. DOI: 10.1021/jp070104x.
(78) Kaduk, B.; Kowalczyk, T.; Van Voorhis, T. Constrained Density Functional Theory. *Chemical Reviews* **2012**, *112* (1), 321-370. DOI: 10.1021/cr200148b.
(79) Becke, A. D. A multicenter numerical integration scheme for polyatomic molecules. *The Journal of Chemical Physics* **1988**, *88* (4), 2547-2553. DOI: 10.1063/1.454033 (acccessed 2022/05/26).
(80) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; et al. 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. DOI: 10.1063/5.0007045 (acccessed 2022/05/25).
(81) Chai, J.-D.; Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *The Journal of Chemical Physics* **2008**, *128* (8), 084106. DOI: 10.1063/1.2834918 (acccessed 2022/05/25).
(82) Gao, F.; Han, L. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Computational Optimization and Applications* **2012**, *51* (1), 259-277. DOI: 10.1007/s10589-010-9329-3.
(83) Nelder, J. A.; Mead, R. A Simplex Method for Function Minimization. *The Computer Journal* **1965**, *7* (4), 308-313. DOI: 10.1093/comjnl/7.4.308 (acccessed 5/26/2022).
(84) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; et al. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications* **2022**, *271*, 108171. DOI: https://doi.org/10.1016/j.cpc.2021.108171.
(85) 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. DOI: https://doi.org/10.1016/j.cpc.2013.09.018.
(86) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications* **2009**, *180* (10), 1961-1972. DOI: https://doi.org/10.1016/j.cpc.2009.05.011.
(87) 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 χ₁ and χ₂ Dihedral Angles. *Journal of Chemical Theory and Computation* **2012**, *8* (9), 3257-3273. DOI: 10.1021/ct300400x.
(88) Wu, Y.; Tepper, H. L.; Voth, G. A. Flexible simple point-charge water model with improved liquid-state properties. *The Journal of Chemical Physics* **2006**, *124* (2), 024503. DOI: 10.1063/1.2136877 (acccessed 2022/05/26).
(89) 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. DOI: 10.1021/acs.jpcb.5b09466.
(90) 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. DOI: https://doi.org/10.1016/j.softx.2015.06.001.
(91) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *The Journal of Chemical Physics* **1992**, *97* (4), 2635-2643. DOI: 10.1063/1.463940.

(92) Deserno, M.; Holm, C. How to mesh up Ewald sums. II. An accurate error estimate for the particle–particle–particle-mesh algorithm. *The Journal of Chemical Physics* **1998**, *109* (18), 7694-7701. DOI: 10.1063/1.477415.

(93) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters* **2008**, *100* (2), 020603. DOI: 10.1103/PhysRevLett.100.020603.

(94) Denisov, V. P.; Schlessman, J. L.; García-Moreno E, B.; Halle, B. Stabilization of Internal Charges in a Protein: Water Penetration or Conformational Change? *Biophysical Journal* **2004**, *87* (6), 3982-3994. DOI: 10.1529/biophysj.104.048454 (acccessed 2022/05/26).

(95) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981**, *52* (12), 7182-7190. DOI: 10.1063/1.328693 (acccessed 2022/05/26).

(96) 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, https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H. DOI: https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H (acccessed 2022/05/26).

(97) Ghosh, N.; Cui, Q. p<i>K</i>_a of Residue 66 in <i>Staphylococal nuclease</i>. I. Insights from QM/MM Simulations with Conventional Sampling. *The Journal of Physical Chemistry B* **2008**, *112* (28), 8387-8397. DOI: 10.1021/jp800168z.

(98) Karp, D. A.; Gittis, A. G.; Stahley, M. R.; Fitch, C. A.; Stites, W. E.; García-Moreno E, B. High Apparent Dielectric Constant Inside a Protein Reflects Structural Reorganization Coupled to the Ionization of an Internal Asp. *Biophysical Journal* **2007**, *92* (6), 2041-2053. DOI: 10.1529/biophysj.106.090266.

(99) 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, https://doi.org/10.1002/jcc.540130812. DOI: https://doi.org/10.1002/jcc.540130812 (acccessed 2022/06/15).

(100) Lide, D. R. *CRC Handbook of Chemistry and Physics*; CRC Press, 2005.

(101) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry* **1994**, *98* (45), 11623-11627. DOI: 10.1021/j100096a001.

(102) Gillan, M. J.; Alfè, D.; Michaelides, A. Perspective: How good is DFT for water? *The Journal of Chemical Physics* **2016**, *144* (13), 130901. DOI: 10.1063/1.4944633.

(103) Karp, D. A.; Stahley, M. R.; García-Moreno E, B. Conformational Consequences of Ionization of Lys, Asp, and Glu Buried at Position 66 in Staphylococcal Nuclease. *Biochemistry* **2010**, *49* (19), 4138-4146. DOI: 10.1021/bi902114m.

(104) Wu, X.; Brooks, B. R. Hydronium Ions Accompanying Buried Acidic Residues Lead to High Apparent Dielectric Constants in the Interior of Proteins. *The Journal of Physical Chemistry B* **2018**, *122* (23), 6215-6223. DOI: 10.1021/acs.jpcb.8b04584.

(105) 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. DOI: 10.1021/ct100578z.

(106) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *Journal of Chemical Theory and Computation* **2011**, *7* (7), 2284-2295. DOI: 10.1021/ct200133y.

(107) Najmanovich, R.; Kuttner, J.; Sobolev, V.; Edelman, M. Side-chain flexibility in proteins upon ligand binding. *Proteins: Structure, Function, and Genetics* **2000**, *39* (3), 261-268. DOI: 10.1002/(sici)1097-0134(20000515)39:3<261::aid-prot90>3.0.co;2-4.

(108) Zheng, L.; Chen, M.; Yang, W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proceedings of the National Academy of Sciences* **2008**, *105* (51), 20227-20232. DOI: 10.1073/pnas.0810631106.

(109) Fisk, J. D.; Powell, D. R.; Gellman, S. H. Control of Hairpin Formation via Proline Configuration in Parallel β-Sheet Model Systems. *Journal of the American Chemical Society* **2000**, *122* (23), 5443-5447. DOI: 10.1021/ja9929483.

(110) Langosch, D.; Heringa, J. Interaction of transmembrane helices by a knobs-into-holes packing characteristic of soluble coiled coils. *Proteins: Structure, Function, and Bioinformatics* **1998**, *31* (2), 150-159, https://doi.org/10.1002/(SICI)1097-0134(19980501)31:2<150::AID-PROT5>3.0.CO;2-Q. DOI: https://doi.org/10.1002/(SICI)1097-0134(19980501)31:2<150::AID-PROT5>3.0.CO;2-Q (acccessed 2023/01/26).

(111) Wiedemann, C.; Kumar, A.; Lang, A.; Ohlenschläger, O. Cysteines and Disulfide Bonds as Structure-Forming Units: Insights From Different Domains of Life and the Potential for Characterization by NMR. *Frontiers in Chemistry* **2020**, *8*, Mini Review. DOI: 10.3389/fchem.2020.00280.

(112) Dokainish, H. M.; Kitao, A. Computational Assignment of the Histidine Protonation State in (6-4) Photolyase Enzyme and Its Effect on the Protonation Step. *ACS Catalysis* **2016**, *6* (8), 5500-5507. DOI: 10.1021/acscatal.6b01094.

(113) Li, S.; Hong, M. Protonation, tautomerization, and rotameric structure of histidine: a comprehensive study by magic-angle-spinning solid-state NMR. *J Am Chem Soc* **2011**, *133* (5), 1534-1544. DOI: 10.1021/ja108943n From NLM.

(114) Morgan, J. E.; Verkhovsky, M. I.; Wikström, M. The histidine cycle: a new model for proton translocation in the respiratory heme-copper oxidases. *J Bioenerg Biomembr* **1994**, *26* (6), 599-608. DOI: 10.1007/bf00831534 From NLM.

(115) He, J.; Xu, S.; Mixson, A. J. The Multifaceted Histidine-Based Carriers for Nucleic Acid Delivery: Advances and Challenges. In *Pharmaceutics*, 2020; Vol. 12.

(116) Ren, X.; Tu, C.; Laipis, P. J.; Silverman, D. N. Proton Transfer by Histidine 67 in Site-Directed Mutants of Human Carbonic Anhydrase III. *Biochemistry* **1995**, *34* (26), 8492-8498. DOI: 10.1021/bi00026a033.

(117) Tu, C. K.; Silverman, D. N.; Forsman, C.; Jonsson, B. H.; Lindskog, S. Role of histidine 64 in the catalytic mechanism of human carbonic anhydrase II studied with a site-specific mutant. *Biochemistry* **1989**, *28* (19), 7913-7918. DOI: 10.1021/bi00445a054 From NLM.

(118) Duster, A. W.; Lin, H. Tracking Proton Transfer through Titratable Amino Acid Side Chains in Adaptive QM/MM Simulations. *Journal of Chemical Theory and Computation* **2019**, *15* (11), 5794-5809. DOI: 10.1021/acs.jctc.9b00649.

(119) Warshel, A.; Weiss, R. M. An empirical valence bond approach for comparing reactions in solutions and in enzymes. *Journal of the American Chemical Society* **1980**, *102* (20), 6218-6226. DOI: 10.1021/ja00540a008.

(120) and, U. W. S.; Voth, G. A. Multistate Empirical Valence Bond Model for Proton Transport in Water. *Journal of Physical Chemistry B* **1998**, *102*, 5547-5551.

(121) Walker, J. E. The ATP synthase: the understood, the uncertain and the unknown. *Biochem Soc Trans* **2013**, *41* (1), 1-16. DOI: 10.1042/bst20110773 From NLM.

(122) Di Luca, A.; Gamiz-Hernandez, A. P.; Kaila, V. R. I. Symmetry-related proton transfer pathways in respiratory complex I. *Proceedings of the National Academy of Sciences* **2017**, *114* (31), E6314-E6321. DOI: 10.1073/pnas.1706278114 (acccessed 2023/03/12).

(123) Letts, J. A.; Sazanov, L. A. Clarifying the supercomplex: the higher-order organization of the mitochondrial electron transport chain. *Nat Struct Mol Biol* **2017**, *24* (10), 800-808. DOI: 10.1038/nsmb.3460 From NLM.

(124) Zhao, R. Z.; Jiang, S.; Zhang, L.; Yu, Z. B. Mitochondrial electron transport chain, ROS generation and uncoupling (Review). *Int J Mol Med* **2019**, *44* (1), 3-15. DOI: 10.3892/ijmm.2019.4188 From NLM.

(125) Bridges, H. R.; Fedor, J. G.; Blaza, J. N.; Di Luca, A.; Jussupow, A.; Jarman, O. D.; Wright, J. J.; Agip, A.-N. A.; Gamiz-Hernandez, A. P.; Roessler, M. M.; et al. Structure of inhibitor-bound mammalian complex I. *Nature Communications* **2020**, *11* (1), 5261. DOI: 10.1038/s41467-020-18950-3.

(126) Clason, T.; Ruiz, T.; Schägger, H.; Peng, G.; Zickermann, V.; Brandt, U.; Michel, H.; Radermacher, M. The structure of eukaryotic and prokaryotic complex I. *J Struct Biol* **2010**, *169* (1), 81-88. DOI: 10.1016/j.jsb.2009.08.017 From NLM.

(127) Galkin, A.; Dröse, S.; Brandt, U. The proton pumping stoichiometry of purified mitochondrial complex I reconstituted into proteoliposomes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2006**, *1757* (12), 1575-1581. DOI: https://doi.org/10.1016/j.bbabio.2006.10.001.

(128) Jones, A. J.; Blaza, J. N.; Varghese, F.; Hirst, J. Respiratory Complex I in Bos taurus and Paracoccus denitrificans Pumps Four Protons across the Membrane for Every NADH Oxidized. *J Biol Chem* **2017**, *292* (12), 4987-4995. DOI: 10.1074/jbc.M116.771899 From NLM.

(129) Wikström, M.; Hummer, G. Stoichiometry of proton translocation by respiratory complex I and its mechanistic implications. *Proceedings of the National Academy of Sciences* **2012**, *109* (12), 4431-4436. DOI: doi:10.1073/pnas.1120949109.

(130) Mimaki, M.; Wang, X.; McKenzie, M.; Thorburn, D. R.; Ryan, M. T. Understanding mitochondrial complex I assembly in health and disease. *Biochim Biophys Acta* **2012**, *1817* (6), 851-862. DOI: 10.1016/j.bbabio.2011.08.010 From NLM.

(131) Fiedorczuk, K.; Letts, J. A.; Degliesposti, G.; Kaszuba, K.; Skehel, M.; Sazanov, L. A. Atomic structure of the entire mammalian mitochondrial complex I. *Nature* **2016**, *538* (7625), 406-410. DOI: 10.1038/nature19794.

(132) Zhu, J.; Vinothkumar, K. R.; Hirst, J. Structure of mammalian respiratory complex I. *Nature* **2016**, *536* (7616), 354-358. DOI: 10.1038/nature19095.

(133) Letts, J. A.; Fiedorczuk, K.; Degliesposti, G.; Skehel, M.; Sazanov, L. A. Structures of Respiratory Supercomplex I+III2 Reveal Functional and Conformational Crosstalk. *Molecular Cell* **2019**, *75* (6), 1131-1146.e1136. DOI: https://doi.org/10.1016/j.molcel.2019.07.022.

(134) Röpke, M.; Riepl, D.; Saura, P.; Di Luca, A.; Mühlbauer, M. E.; Jussupow, A.; Gamiz-Hernandez, A. P.; Kaila, V. R. I. Deactivation blocks proton pathways in the mitochondrial complex I. *Proceedings of the National Academy of Sciences* **2021**, *118* (29), e2019498118. DOI: 10.1073/pnas.2019498118 (acccessed 2023/03/12).

(135) Gamiz-Hernandez, A. P.; Jussupow, A.; Johansson, M. P.; Kaila, V. R. I. Terminal Electron-Proton Transfer Dynamics in the Quinone Reduction of Respiratory Complex I. *J Am Chem Soc* **2017**, *139* (45), 16282-16288. DOI: 10.1021/jacs.7b08486 From NLM.

(136) 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. DOI: doi:10.1073/pnas.2113141118.

(137) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *J Chem Phys* **2007**, *126* (5), 054103. DOI: 10.1063/1.2432340 From NLM.

(138) Díaz Leines, G.; Ensing, B. Path Finding on High-Dimensional Free Energy Landscapes. *Physical Review Letters* **2012**, *109* (2), 020601. DOI: 10.1103/PhysRevLett.109.020601.

(139) 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. DOI: 10.1021/jp5095118.

(140) Pan, A. C.; Sezer, D.; Roux, B. Finding transition pathways using the string method with swarms of trajectories. *J Phys Chem B* **2008**, *112* (11), 3432-3440. DOI: 10.1021/jp0777059 From NLM.

(141) Chen, H.; Ogden, D.; Pant, S.; Cai, W.; Tajkhorshid, E.; Moradi, M.; Roux, B.; Chipot, C. A Companion Guide to the String Method with Swarms of Trajectories: Characterization, Performance, and Pitfalls. *Journal of Chemical Theory and Computation* **2022**, *18* (3), 1406-1422. DOI: 10.1021/acs.jctc.1c01049.

(142) Roux, B. String Method with Swarms-of-Trajectories, Mean Drifts, Lag Time, and Committor. *The Journal of Physical Chemistry A* **2021**, *125* (34), 7558-7571. DOI: 10.1021/acs.jpca.1c04110.

(143) *WHAM: the weighted histogram anlaysis method*; 2022. http://membrane.urmc.rochester.edu/wordpress/?page_id=126 (accessed.

(144) Kratochvil, H., Watkins, LC, Mravic, M, Thomaston, JL, Nicoludis, JM, Liu, L, Voth, GA, DeGrado, WF. Transient Water Wires Mediate Selective Proton Conduction in Designed Channel Proteins. *Nature Chemistry* **2023**, *In Press*.
(145) Haapanen, O.; Sharma, V. Role of water and protein dynamics in proton pumping by respiratory complex I. *Scientific Reports* **2017**, *7* (1), 7747. DOI: 10.1038/s41598-017-07930-1.
(146) Eyring, H. The Activated Complex in Chemical Reactions. *The Journal of Chemical Physics* **1935**, *3* (2), 107-115. DOI: 10.1063/1.1749604.