THE UNIVERSITY OF CHICAGO

MULTISCALE REACTIVE MOLECULAR DYNAMICS SIMULATION OF PROTON TRANSPORT IN PROTEINS

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

RUIBIN LIANG

CHICAGO, ILLINOIS

JUNE 2016

To my parents.

Table of Contents

List of Figures ..... **xi**

List of Tables ..... **xvii**

ACKNOWLEDGEMENTS ..... **xx**

ABSTRACT ..... **xxii**

Chapter 1 Background **1**

1.1 Reactive Simulation of Proton Transport in Proteins. 1

1.2 References 13

Chapter 2 Theory and Methodology **20**

2.1 Multiscale Reactive Molecular Dynamics 20

2.2 FitRMD parameterization scheme 23

2.3 The SCC-DFTB Method. 24

2.4 References 26

Chapter 3 Application of the SCC-DFTB Method to Hydroxide Water Clusters and Aqueous Hydroxide Solutions **28**

3.1 Introduction 28

3.2 Methods 30

3.2.1 The SCC-DFTB Method. 30

3.2.2 Cluster simulation setup. 32

3.2.3 Molecular dynamics simulation setup. 33

3.3 Results and Discussion 34

3.3.1 Potential Energy Surface of $H_3O_2^-$. 34

3.3.2 Hydroxide Water Clusters $OH^-(H_2O)_n$, n = 4–7 36

3.3.3 Hydroxide bulk simulations. 41

3.4 Conclusions 68

3.5 References 69

Chapter 4 Benchmark Study of the SCC-DFTB Approach for a Biomolecular Proton Channel 73

4.1 Introduction 73

4.2 Simulation details 76

4.2.1 SCC-DFTB method 76

4.2.2 LS2 QM/MM simulation 78

4.2.3 Carbon nanotube QM/MM simulation 83

4.3 Results 84

4.3.1 Water structure and dynamics in the LS2 channel 84

4.3.2 Proton solvation and transport in a wider region of the LS2 channel 92

4.3.3 Proton solvation and transport in a narrower region of the LS2 channel 105

4.3.4 Proton solvation and transport in (8,8) single-wall carbon nanotube 107

4.4 Conclusions 109

4.5 References 110

Chapter 5 Multiscale simulation reveals a multifaceted mechanism of proton permeation through the influenza A M2 proton channel 114

5.1 Introduction 114

5.2 Materials and Methods 118

5.2.1 Modeling of the pH-Dependent Conformational Ensemble and Water Dynamics. 118

5.2.2 Modeling of Proton Permeation. 118

5.3 Results and Discussion 120

5.3.1 pH-Dependent Conformational Ensemble and Water Dynamics. 120

5.3.2 Proton Permeation in WT AM2. 122

5.3.3 Proton Permeation in AM2 Mutants. 128

5.4 Conclusions 130

5.5 References 132

Chapter 6 Computationally Efficient Multiscale Reactive Molecular Dynamics to Describe Amino Acid Deprotonation in Proteins 138

6.1 Introduction 138

6.2 Methods 143

6.2.1 Multiscale reactive molecular dynamics (MS-RMD) 143

6.2.2 FitRMD parameterization scheme 146

6.2.3 Developing MS-RMD models using FitRMD for ClC-ec1 147

6.2.4 1D-PMF calculations of PT from E203 to E148 in ClC-ec1 151

6.2.5 1D-PMF calculation with QM/MM 153

6.2.6 Hydration dynamics in the central region of ClC-ec1 154

6.2.7 Developing MS-RMD models using FitRMD for CcO 155

6.2.8 2D-PMF for PT in the CcO hydrophobic cavity with MS-RMD 157

6.2.9 Comparing the computational efficiency of different methods 158

6.3 Results and Discussion 159

6.3.1 Comparing QM/MM and MS-RMD free energy profiles in the ClC-ec1 antiporter 159
6.3.2 Limitations of the QM/MM free energy profiles 164
6.3.3 MS-RMD can capture coupling between hydration and PT in CcO 168
6.4 Conclusions 172
6.5 References 175

Chapter 7 Multiscale simulations reveal key features of the proton pumping mechanism in cytochrome c oxidase. 181

7.1 Introduction 181
7.2 Material and Methods 184
7.3 Results and Discussion 185
7.3.1 Transport of the Pumped Proton and Hydration of the HC 185
7.3.2 Transport of the Chemical Proton 189
7.3.3 Rates of the Pumped and Chemical Proton Transport Events 190
7.3.4 Comparison to Proposed Mechanisms 193
7.3.5 Proton Transport through the D-Channel 195
7.4 Conclusions 198
7.5 References 201

Chapter 8 Multiscale simulations reveal acid activation mechanism of the influenza A M2 proton channel 205

8.1 Introduction 205
8.2 Method and Simulation details 207
8.2.1 Classical Molecular Dynamics (MD) Simulations 207

8.2.2 QM/MM simulation 209
8.2.3 MS-RMD Umbrella Sampling 212
8.2.4 Calculation of Proton Conductance 213
8.3 Results 214
8.3.1 M2 channel has closed C-terminal helices in high pH condition 214
8.3.2 M2 channel is inactivated in high pH condition 216
8.3.3 M2 channel is activated during +2 to +3 state transition 220
8.4 Conclusion 222
8.5 References 223
Appendix A Supporting Information for Benchmark Study of the SCC-DFTB Approach for a Biomolecular Proton Channel 229
A.1 Comparison to multi-state empirical valence bond simulations 229
A.1.1 System setup 229
A.1.2 Supplementary Tables 230
A.1.3 Supplementary Figures 231
A.2 Additional results for proton solvation and transport in a narrower region of the LS2 channel 238
A.2.1 Supplementary Tables 238
A.2.2 Supplementary Figures 240
A.3 B3LYP-D/GTH-DZVP improves description of forces and energies over BLYP-D/GTH-TZV2P 243
A.4 Supplementary References 244

Appendix B Supporting Information for Multiscale Simulation Reveals a Multifaceted Mechanism of Proton Permeation through the Influenza A M2 Proton Channel 246

B.1 Classical Molecular Dynamics (MD) Simulations..... 246

B.2 QM/MM Umbrella Sampling..... 248

B.3 MS-EVB Umbrella Sampling. 252

B.4 Multiscale bridging of MS-EVB and QM/MM PMFs. 253

B.5 Calculation of Proton Conductance..... 255

B.6 Limitations in Conductance Calculations. 256

B.7 Supplementary Figures..... 257

B.8 Supplementary References..... 258

Appendix C Supporting Information for Multiscale simulations reveal key features of the proton pumping mechanism in cytochrome c oxidase 262

C.1 System Setup and Structural Equilibration. 262

C.2 Multiscale reactive molecular dynamics (MS-RMD). 263

C.3 Finding the Proton Transport Pathways. 266

C.4 Developing MS-RMD Models for Protonatable Groups. 267

C.5 MS-RMD Free Energy Profiles in the HC and D-Channel..... 269

C.6 Definition of the PMF Collective Variables and Rate Calculation. 270

C.7 Justification for the choice of collective variables in the 2D PMF calculations. 272

C.8 Aspartic Acid Deprotonation in Bulk Water Modeled by the SCC-DFTB Method. 273

C.9 The Biprotonated E286 Mechanism..... 274

C.10 Supplementary Figures..... 277

C.11 Supplementary Tables 283

C.12 Supplementary References 287

List of Figures

Figure 3.1 Contour plots of $H_3O_2^-$ calculated with various electronic structure methods 35

Figure 3.2 Structures of the low-lying energy minima 38

Figure 3.3 Solvation of the hydroxide anion in bulk water. 42

Figure 3.4 Radial distribution functions of the hydroxide anion complex 43

Figure 3.5 Partial radial distribution function for the six nearest waters to the hydroxide 50

Figure 3.6 Radial distribution functions of the hydroxide anion complex 52

Figure 3.7 Potential of mean force for the hydroxide for proton transport in $H_3O_2^-$ 56

Figure 3.8 Time dependence of the identity of O_0 59

Figure 3.9 Forward proton hopping function 62

Figure 3.10 The snap shots of the neutral, hydroxide, and protonated bulk system 65

Figure 3.11 Radial distribution function for in hydrated hydroxide system 65

Figure 3.12 Results revealing hydroxide transport mechanism predicted by DFTB3-3OB method 66

Figure 4.1 LS2 channel system filled with a protonated water wire 75

Figure 4.2 Snapshots of protonated water wire in LS2 with excess proton CEC restrained at z = -6.5 Å from (A) BLYP-D simulation (B) SCC-DFTB simulation. 82

Figure 4.3 RDFs of (A) O_w-O_w and(B) O_w-H_w in the LS2 channel filled with water and no excess proton. 86

Figure 4.4 Distribution of the pore water oxygen atoms along the z axis of the unprotonated LS2 channel. 87

Figure 4.5 Semi-log plot of continuous hydrogen bond time correlation function for the (A) water-water and (B) water-serine hydrogen bonds. 90

Figure 4.6 RDF of O_w-O_s in the LS2 channel filled with water and no excess proton. ... 91

Figure 4.7 Radial distribution functions of O*-O_w and O_{1x}-O in protonated LS2 (A, B) and protonated water filled CNT (C, D). 94

Figure 4.8 Radial distribution functions of H*-O_w in (A) the protonated LS2, (B) the water filled protonated CNT. 96

Figure 4.9 Distribution of H*-O_w distances in protonated (A) LS2 channel and (B) CNT. 97

Figure 4.10 Distribution of O*-O distance, where O is the closest atom not hydrogen bonded to O* in the (A) protonated LS2 channel and (B) CNT. 99

Figure 4.11 Average number of different O_{1x} partners to O* as a function of time during non-transfer intervals for the (A) LS2 channel and (B) CNT. 100

Figure 4.12 Z coordinate of the excess proton CEC as a function of time for unconstrained simulations using the (A) BLYP-D, (B) B3LYP-D, and (C) SCC-DFTB methods 104

Figure 4.13 Radial distribution functions of (A) O*-O_s between the excess proton complex and serines in the protonated LS2 channel, and (B) O*-O, where O includes both O_w and O_s. 105

Figure 4.14 Radial distribution functions of (A) O*-O_w, (B) O*-O where O includes both O_w and O_s, and (C) H*-O_w, in a narrower region of protonated LS2 channel. 107

Figure 5.1 (A) Equilibrated structures of the AM2 channel transmembrane domain (see text for definition) and (B) The excess proton (purple) delocalized around entry water cluster and histidine residues. 115

Figure 5.2 Equilibrated AM2 channel transmembrane domain structures at various protonation states and aligned with either starting crystal structure 3LBW (a) or 3C9J (b). 124

Figure 5.3 Proton transport +3 free energy profiles (PMFs) (blue) compared with +2 (black). 128

Figure 5.4 Proton transport free energy profiles (PMFs) for mutants compared with the WT AM2 channel. The origin is the center of mass of 4 Cα of Gly34. The positions of Val27, His37 and Trp41 are labeled by the text boxes. The regions sampled by MS-EVB and QM/MM are labeled and separated by black dashed vertical lines. 130

Figure 6.1 (A) Overview of the structure of the ClC-ec1 antiporter and transport pathways for Cl^- (green dashed) and H^+ (red dashed) and (B) Representative configurations of the central region with (left) Cl^-_{cen} absent and (right) present for the "P" state in the PMFs in **Figure 6.2**. 149

Figure 6.2 PMFs of PT through central region of ClC-ec1 antiporter with (A) Cl^-_{cen} absent and (B) present, as calculated with MS-RMD (blue) and QM/MM (red). 161

Figure 6.3 In the system with Cl^-_{cen} absent, RDFs from E148 carboxyl carbon (CD in the PDB) to water oxygens are shown in A1 to A3, for the "R", "T", and "P" states (**Figure 6.2**), respectively. The RDFs from E203 carboxyl carbon to water oxygens are shown in B1 to B3 in the same order. The averaged positions of the four side chains are shown in

C1 to C3 in the same order, with the QM/MM structures in red and the MS-RMD structures in blue. 163

Figure 6.4 In the CLC system as in **Figure 6.3**, but with Cl^-_{cen} present. All notation is consistent with **Figure 6.3**. 164

Figure 6.5 A segment of PMF in **Figure 6.2** (A) in the range of $\zeta R = 0.18 - 0.26$ and additional structural analysis in the CV range. 166

Figure 6.6 Cytochrome c oxidase from *Rhodobacter sphaeroides*. 170

Figure 6.7 Two-dimensional free energy profiles (2D-PMFs) for PT from the E286 to the PLS in the P_M' state, as a function of the CEC coordinate through the hydrophobic cavity (HC) as the horizontal axis and the water hydration in the HC as the vertical axis. 171

Figure 6.8 Scaling plot for the MS-RMD, SCC-DFTB-based QM/MM and BLYP-based QM/MM simulations of C*c*O. 172

Figure 7.1 Illustration of the simulation setup for the full C*c*O from *Rhodobactor sphaeroides* in a membrane and surrounded by water. 184

Figure 7.2 Free energy profiles for PT from E286 to PLS in hydrophobic cavity and from D132 to E286 in D-channel. 189

Figure 7.3 Free energy diagram for the reaction sequence during the A→F transition. 200

Figure 8.1 Classical MD equilibration of the +0 to +2 state structures. 215

Figure 8.2 Free energy profiles (PMFs) for proton transport through influenza A M2 channel. 219

Figure 8.3 RMSD of the C_α atoms in the M2TM helices during the equilibration of NMR structure 2L0J (with amphipathic helices) in lipid bilayer. 222

Figure A.1 RDFs of (A) O*-O_w and (B) O_{1x}-O, where O includes both O* and O_w, in the protonated LS2 channel. 231

Figure A.2 RDFs of H*-O_w in the protonated (A) LS2 channel and (B) CNT channels, with coordination numbers indicated in the legend. 232

Figure A.3 Distribution of H*-O_w distance in the protonated (A) LS2 and (B) CNT channels..... 233

Figure A.4 Distribution of the O*-O distance, where O is the closest atom not hydrogen bonded to O*, for the protonated (A) LS2 and (B) CNT channels..... 234

Figure A.5 Average number of different O_{1x} partners to O* as a function of time during non-transfer intervals for the protonated (A) LS2 and (B) CNT channels. 235

Figure A.6 Z coordinate of the proton CEC as a function of time for unconstrained simulations using the (A) BLYP-D (B) B3LYP-D (C) SCC-DFTB and (D) MS-EVB3 methods. 236

Figure A.7 Proton hopping function for the free excess proton simulation initiated from the constrained case where the center of excess charge is trapped at a wide region. 237

Figure A.8 (A) RDF of O*-O_s in the protonated LS2 channel. 238

Figure A.9 RDFs of (A) O*-O_w, (B) O*-O , and (C) O_{1x}-O , where O includes both O_w and O_s, in a narrow region of the protonated LS2 channel. 240

Figure A.10 (A) RDF of H*-O_w in the narrow region of the protonated LS2 channel. (B) Distribution of H*-O distances, where O includes both O_w and O_s, in the narrow region of the protonated LS2 channel. 241

Figure A.11 Distribution of the O*-O distance, where O is the closest atom not hydrogen bonded to O*, for a narrow region of the protonated LS2 channel. 242

Figure A.12 Scatter plot of forces from B3LYP-D/6-31++G** vs B3LYP-D/GTH-DZVP (red) and BLYP-D/GTH-TZV2P (blue). The B3LYP-D/GTH-DZVP provides more accurate forces than BLYP-D/GTH-TZV2P. 244

Figure B.1 Proton transport +2 free energy profile (PMF). 257

Figure B.2 Comparison for the deprotonation barrier of a Q3 state His37 tetrad under different backbone conformations. 258

Figure B.3 (A) Pore radius profile and (B) water density for WT AM2/TM and mutants. 258

Figure C.1 Free energy profiles (PMFs) for deprotonation of aspartic acid in bulk water. 277

Figure C.2 Proton transfer free energy profile (PMF) in the D-channel for the biprotonated E286 mechanism in the P_M' state. 278

Figure C.3 Free energy profiles for PT through hydrophobic cavity. 279

Figure C.4 Free energy profiles for PT through D-channel. 280

Figure C.5 (A-E) Molecular structures during the PT from E286 to the PLS in the P_R state along the minimum free energy pathway (black line on the 2D-PMF). 281

Figure C.6 (A-D) Molecular structures during the PT from D132 to E286 in the P_R state along the minimum free energy pathway (black line on the 2D-PMF). 282

List of Tables

Table 3.1 Formation Energies (kcal/mol) of several isomers of $OH^-(H_2O)_n$, for n = 4–7 (Continued) 39

Table 3.2 Coordination number of hydroxide and first peak height of radial distribution function of O_0-O. 44

Table 3.3 Coordination number of O_2 and first peak height of radial distribution function of O_2-O. 48

Table 3.4 Coordination number of O_3 and first peak height of radial distribution function of O_3-O. 48

Table 3.5 Coordination number of hydrogen and second peak height of radial distribution function of O_0-H. 54

Table 4.1 Averaged number of hydrogen bonds per water molecule in the LS2 channel as a function of QM/MM methodology. 89

Table 4.2 Average water-water, water-serine hydrogen bond relaxation time, and rotatioanl correlation time in the LS2 channel as a function of QM/MM methodology. . 89

Table 4.3 Average hydrogen bond relaxation time for hydrogen bond around 4 Å of excess proton CEC in the LS2 channel and CNT as a function of QM/MM methodology. 101

Table 5.1 Calculated dominant H-bond lifetime (ps), between pore-lining residue Gly34 amide-I groups and their nearest water molecules. 122

Table 5.2 Conductance comparison between simulation and experiments calculated from the +2 process (defined in main text) 128

Table 5.3 Relative experimental (30) and calculated (present work) conductances for AM2 mutants normalized to the conductance of wild type (WT) 130

Table 6.1 The MS-RMD model parameters of E148 and E203 in ClC-ec1. 151

Table 6.2 The MS-RMD model parameters of E286, PRD$\underline{a}_3$ and PRAa_3 in C*c*O..... 157

Table 7.1 Calculated time constants (inverse of rate constants*) for PT of the pumped (E286→PLS), chemical (E286→BNC), and back leaked (PLS→E286) protons in the P_M' and P_R states 193

Table 7.2 Calculated time constants (inverse of the rate constants*) for PT in the D-channel from protonated D132 to deprotonated E286 in the P_R and F states 195

Table 8.1 Proton conductance comparison between simulations for the +0, +1 and +2 processes (defined in text) and experiments.(60) 220

Table A.1 Average hydrogen bond relaxation times for hydrogen bonds within 4 Å of the excess proton CEC in protonated LS2 channel and CNT..... 230

Table A.2 Average hydrogen bond relaxation time for hydrogen bonds within 4 Å of the excess proton CEC in the narrow region of the protonated LS2 channel..... 239

Table C.1 Description of the CcO redox and protonation states in the MS-RMD simulations. 283

Table C.2 The MS-RMD model parameters of E286, D132 and BNC in C*c*O. The redox states of the system and the PT processes for which the MS-RMD models are parameterized are labeled in the parenthesis under the protonatable groups..... 283

Table C.3 The MS-RMD model parameters of PRDa_3, PRAa_3 in C*c*O. The redox states of the system for which the MS-RMD models are parameterized are labeled in the parenthesis under the protonatable groups. 284

Table C.4 Charges for the Cu_B complex with water ligand. The total charge of the complex is +1 and the charges represent the F state. 285

Table C.5 Charges for the Cu_B complex with water ligand. The total charge of the complex is +2 and the charges represent the hypothetical state when water is formed at the BNC without electron transferred from heme *a* to the BNC. 286

Table C.6 Bond and angle parameters for the water ligand in the Cu_B complex. For the O-H bond Morse potential is used. For the H-O-H angle the harmonic potential is used. 287

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my advisor, Dr. Gregory A. Voth, for his excellent guidance and continuous support throughout my Ph.D. study. It is my great honor to work in his outstanding research team and have him as my advisor, who provides me precious opportunities, abundant resources and excellent atmosphere in my research experience. I could not be prouder to learn from him, who is a role model of scientist with vast knowledge, endless passions, dedicated work and rigorous attitude to science.

I would also like to thank the rest of my thesis committee: Dr. Ka Yee Lee and Dr. Suriyanarayanan Vaikuntanathan, for their encouragement and insightful suggestions and advices on the thesis.

I couldn't be more fortunate to have a co-advisor, Dr. Jessica M. J. Swanson, throughout my Ph.D. research. She is a professional sub-group leader in the Voth group and continuously provides invaluable and insightful advices to my research projects, and also helped me with my scientific writing for proposals, articles and this thesis.

I would also like to thank the team of talented scientists (past and present) in the Voth group including Dr. Yuxing Peng, Dr. Zhen Cao, Dr. Hui Li, Dr. Sangyun Lee, Dr. Gard Nelson, Dr. Isaiah Sumner, Dr. Mark C. Maupin, Dr. Christopher Knight, Dr. Christopher Arntsen, Dr. Ying-Lung Steve Tse, Dr. Heather Mayes, Dr. James Dama, Dr. Revati Kumar, Dr. Gerrick E. Lindberg, Dr. Rajib Biswas, Dr. Rui Sun, Dr. Chen Chen, Dr. Shule Liu.

Last but not least, I would like to thank my family: my father Kefa Liang and my mother Lin Guan, for their unconditional and endless love in my life.

ABSTRACT

The process of proton solvation and transport is involved in a wide array of phenomena. Although seemingly simple, it is a complex process that continues to engage research efforts. One unique aspect of the molecular mechanism of proton solvation and transport is that the hydrated proton is able to hop between neighboring water molecules by "Grotthuss shuttling", giving rise to anomalously high diffusion rate of protons compared to other simple cations. In addition, the hydrated proton is able to dynamically delocalize the excess charge defect over multiple water molecules within several solvation shells. Inside proteins, the structure and dynamics of the residues and water molecules that mediate proton transport are strongly influenced by the electrostatic and hydrophobic environment in the protein cavities. Hence, proton transport in proteins is very different from that in aqueous bulk solution and displays many interesting behaviors. Explicitly simulating proton transport in proteins is an inherently challenging problem. It requires both the explicit treatment of the excess proton, including its charge defect delocalization and Grotthuss shuttling through inhomogeneous moieties (water and amino acid residues), and extensive sampling of slow conformational changes of both the protein and water clusters inside protein cavities.

In recent years, the self-consistent charge density functional tight binding (SCC-DFTB) method has been increasingly applied to study proton transport (PT) in biological environments. However, recent studies revealing some significant limitations of SCC-

DFTB for proton and hydroxide solvation and transport in bulk aqueous systems call into question its accuracy for simulating PT in biological systems. The current work benchmarks the SCC-DFTB/MM method against more accurate DFT/MM by simulating PT in a synthetic leucine-serine channel (LS2), which emulates the structure and function of biomolecular proton channels. It is observed that SCC-DFTB/MM produces over-coordinated and less structured pore water, an over-coordinated excess proton, weak hydrogen bonds around the excess proton charge defect and qualitatively different PT dynamics. Similar issues are demonstrated for PT in a carbon nanotube, indicating that the inaccuracies found for SCC-DFTB are not due to the point charge based QM/MM electrostatic coupling scheme, but rather to the approximations of the semiempirical method itself. The results presented in this work highlight the limitations of the present form of the SCC-DFTB/MM approach for simulating PT processes in biological protein or channel-like environments, while providing benchmark results that may lead to an improvement of the underlying method.

An alternative approach that explicitly accounts for the reactive nature of the hydrated excess proton is multiscale reactive molecular dynamics (MS-RMD) method. In this approach, quantum mechanical forces from targeted quantum mechanics/molecular mechanics (QM/MM) calculations are bridged, in a multiscale fashion via a variational mathematical framework, into the reactive MD algorithm for the dynamics of system nuclei, thus allowing efficient and accurate description of Grotthuss shuttling and charge delocalization during PT. Herein, we have used a synthesis of the MS-RMD, QM/MM and classical MD simulations to study the PT mechanism in the influenza A virus M2 channel (AM2), which is crucial in the viral life cycle. Despite many previous

experimental and computational studies, the mechanism of the activating process in which proton permeation acidifies the virion to release the viral RNA and core proteins is not well understood. We report, to our knowledge, the first complete free-energy profiles for PT through the entire AM2 transmembrane domain at various pH values, including explicit treatment of excess proton charge delocalization and shuttling through the His37 tetrad. The free-energy profiles reveal that the excess proton must overcome a large free-energy barrier to diffuse to the His37 tetrad, where it is stabilized in a deep minimum reflecting the delocalization of the excess charge among the histidines and the cost of shuttling the proton past them. At lower pH values the His37 tetrad has a larger total charge that increases the channel width, hydration, and solvent dynamics, in agreement with recent 2D-IR spectroscopic studies. The PT barrier becomes smaller, despite the increased charge repulsion, due to backbone expansion and the more dynamic pore water molecules. The calculated conductances are in quantitative agreement with recent experimental measurements. In addition, the free-energy profiles and conductances for PT in several mutants provide insights for explaining our findings and those of previous experimental mutagenesis studies.

Another important challenge for simulation of PT in biomolecular systems is a quantitative description of the protonation and deprotonation process of amino acid residues. Despite the seeming simplicity of adding or removing a positively charged hydrogen nucleus, simulating the actual protonation/deprotonation process is inherently difficult. In a recent paper (J. Chem. Theory Comput. 2014, 10, 2729–2737), a multiscale approach was developed to map high-level quantum mechanics/molecular mechanics (QM/MM) data into a MS-RMD model in order to describe amino acid deprotonation in

bulk water. Herein, the fitting approach (called FitRMD) was extended to create MS-RMD models for ionizable amino acids within proteins. The resulting models are shown to faithfully reproduce the free energy profiles of the reference QM/MM Hamiltonian for PT inside an example protein, the ClC-ec1 H+/Cl– antiporter. Moreover, we show that the resulting MS-RMD models are computationally efficient enough to then characterize more complex 2-dimensional free energy surfaces due to slow degrees of freedom such as water hydration of internal protein cavities that can be inherently coupled to the excess proton charge translocation. The FitRMD method is thus shown to be an effective way to map ab initio level accuracy into a much more computationally efficient reactive MD method in order to explicitly simulate and quantitatively describe amino acid protonation/deprotonation in proteins.

The combination of FitRMD and our other multiscale methods is herein used to study proton transport in aa_3-type cytochrome c oxidase (C*c*O), which is the terminal enzyme in the respiratory electron transfer chain in the inner membrane of mitochondria and plasma membrane of bacteria. C*c*O catalyzes the reduction of O_2 to H_2O and couples the free energy of this exergonic reaction to the pumping of protons across the membrane, creating a transmembrane proton electrochemical gradient that drives, for example, ATP synthesis. We have used MS-RMD simulations to explicitly characterize (with free energy profiles and calculated rates) the internal PT events that enable pumping and chemistry during the $A \rightarrow P_R \rightarrow F$ transition in the aa_3-type C*c*O. Our results show that PT from amino acid residue E286 to both the pump loading site (PLS) and to the binuclear center (BNC) are thermodynamically driven by electron transfer from heme *a* to the BNC, but that the former (i.e., pumping) is kinetically favored while the latter (i.e.,

transfer of the chemical proton) is rate-limiting. The calculated rates are in quantitative agreement with experimental measurements. The back flow of the pumped proton from the PLS to E286 and from E286 to the inner side of membrane are prevented by the fast reprotonation of E286 through the D-channel and large free energy barriers for the back flow reactions. PT from E286 to the PLS through the hydrophobic cavity (HC) and from D132 to E286 through the D-channel are found to be strongly coupled to dynamical hydration changes in the corresponding pathways. This work presents a comprehensive description of the key steps in the proton pumping mechanism in C*c*O.

Lastly, we have used a synthesis of the MS-RMD, QM/MM and classical MD simulations to explicitly characterize (with free energy profiles and calculated conductances) the activation mechanism of influenza M2 channel from high to intermediate pH condition. Starting from a recently resolved crystal structure in high pH condition (4QKL), we found that the Trp41 side chains in the high pH condition block the channel, and the channel is gradually activated as the pH in the viral exterior is lowered, as a result of the increased positive charges on the His37 tetrad. The qualitative asymmetry of the PT free energy profile in high pH condition explains why the inward proton flux is allowed when the pH is low in viral exterior and high in viral interior, but outward proton flux is prohibited when the pH gradient is reversed. Our results also demonstrate that the amphipathic helices do not change the proton conduction mechanism in the AM2 transmembrane domain significantly.

Chapter 1

Background

1.1 Reactive Simulation of Proton Transport in Proteins.

The hydrated excess "proton" is in fact a unit of net positive charge (1) due to a missing electron that can be passed between and among biomolecules (2) in a seemingly simple dance of charge neutralization. However, the molecular nature of this dance is complicated, involving a dynamic charge delocalization between molecules and a constant restructuring of covalent and hydrogen bond topologies. Understanding and being able to characterize the migration of excess protons is important given the many roles that proton transport (PT) plays throughout biology. Virtually all biomolecules are sensitive to pH and many incorporate PT into their functional mechanisms, including transporters, proton pumps, proton channels, and enzymes. For example, cytochrome *c* oxidase (C*c*O), a proton pump in the respiratory chain of mitochondria and bacteria, reduces oxygen to water and utilizes the released free energy to pump protons across the membrane, contributing to the transmembrane electrochemical potential gradient that is necessary for ATP synthesis.(3-6) The influenza A M2 proton channel protein(7) transports the protons across the influenza virus membrane and triggers the dissociation of the viral matrix proteins, which is an essential step in the influenza virus replication cycle.(8) This is a short list of the many systems in which PT plays an important role in a functional mechanism.

Computational approaches can play an important role in the investigation of PT mechanisms in proteins, adding molecular level insight as well as increased temporal and spatial resolution to experimental data. However, it is very challenging to *explicitly* model the PT process, even in simple bulk water solution, because it involves charge defect delocalization, Grotthuss shuttling (proton hopping), and solvent reorganization. Moreover, the migration of an excess proton in proteins and other confined spaces can be non-trivially coupled with changes in the hydration along the PT pathway.(9-11) As will be discussed later, water molecules move in/out of internal protein cavities in response to the excess positive charge as it moves between water molecules and ionizable amino acids. Therefore, to more completely describe PT in biological systems, a computational method should (1) explicitly treat charge defect delocalization and Grotthuss shuttling of the excess proton(s) between water molecules and ionizable amino acids undergoing protonation/deprotonation; (2) allow exchange of water molecules between different protein internal cavities as well as between those cavities and bulk solution; and (3) be computationally efficient enough to achieve sufficient sampling of the charge translocation and protein and solvent configuration space for large-scale biomolecular systems, including protein, solvent, ions, and (where needed) membrane. Quantum mechanical approaches, such as *ab initio* molecular dynamics (AIMD) or quantum mechanics molecular mechanics (QM/MM) MD methods, can in principle explicitly treat the reactive nature of amino acid protonation/deprotonation and the Grotthuss hopping phenomenon. However, their computational expense limits the application of the former to systems much smaller than proteins, and restricts the free energy sampling of the latter when applied to large biomolecular systems. A lack of sufficient free energy sampling

generally leads to results with large systematic errors that depend on the initial state of system. In addition, QM/MM methods without adaptive partitioning(12, 13) prohibit exchange of MM and QM atoms across the QM/MM boundary, which is unphysical. Although adaptive QM/MM partitioning methods enable the exchange of MM and QM atoms across the boundary, they can suffer from both inaccuracy of forces on atoms near the boundary and a lack of sufficient energy conservation.(14) The QM/MM boundary may be especially problematic when the hydration along the PT pathway changes in response to the migration of the excess proton charge defect since this necessitates the exchange of water molecules between protein cavities and bulk solution, which often occurs beyond the nanosecond timescale.

Therefore, it is sometimes necessary to develop approximations to *ab initio* QM methods that offer a balance between accuracy and computational efficiency. The self-consistent charge density functional tight binding (SCC-DFTB) method(15) is a fast semi-empirical QM algorithm that has become popular in recent years for simulating chemical processes in biomolecular systems due to the high degree of interest in studying such systems.(16) The SCC-DFTB method is derived from density functional theory (DFT) by approximation and parameterization of multi-center electron integrals.(15) The computational speed gained by these approximations can be 2-3 orders of magnitude compared to more accurate and "first principles" standard DFT.(17) While the SCC-DFTB approach is understandably popular, the method has been shown to have substantial limitations for predicting structural, energetic and dynamic properties of proton transport (PT)(18) and hydroxide transport in bulk water. (19) Despite recent SCC-DFTB developments (17, 20, 21), these issues remain unresolved. The SCC-DFTB

method was also, e.g., recently shown to poorly describe noncovalent interactions involving sulfur atoms.(22) Given the increased use of quantum mechanics/molecular mechanics (QM/MM) with SCC-DFTB as the QM method (SCC-DFTB/MM) to study proton hydration and transport in biomolecular systems,(23-25) there is a need to benchmark its present level of accuracy and potential limitations in such environments. As a counter part of proton transport, hydroxide transport is an equally important reactive process in both biomolecular and material system, where SCC-DFTB method is widely utilized. Therefore, it is also necessary to benchmark the accuracy of SCC-DFTB to simulate hydroxide transport.

Chapter 3 (19) evaluates the accuracy of the current SCC-DFTB implementations, including the proposed O–H repulsive potential and 3rd order correction with both MIO and the 3OB parameter sets, focusing on both cluster and liquid water systems with the hydroxide ion (proton hole).(25) Chapter 4 (26) establishes a systematic benchmark of SCC-DFTB/MM method against arguably more accurate QM/MM methods that employ both generalized gradient approximation (GGA) level and hybrid-GGA level DFT theories for the QM calculation. The comparison is made in the context of condensed phase molecular dynamics (MD) simulations of PT in channel systems. All the methods under investigation have a comparable amount of sampling time, which has seldom been done to date. The synthetic leucine-serine channel (LS2, **Figure 4.1**)(27) is chosen as the simulation system. Although it is synthetic, LS2 possesses key features that are representative of proton channels in nature, such as high proton selectivity, a nonuniform pore radius along the channel axis, a hydrophilic pore due to pore-lining serine residues, and a macrodipole formed by parallel α helices.(27-30) In addition, it is small enough to

enable sufficient sampling for the convergence of statistical quantities extracted from MD simulations, which is both essential for condensed phase analysis and computationally demanding for the higher level QM/MM approaches.

An alternative approach that explicitly accounts for the reactive nature of the hydrated excess proton is multiscale reactive molecular dynamics (MS-RMD) method.(2, 31, 32) This approach describes the migration of an excess proton including explicit Grotthuss shuttling and charge defect delocalization by evolving the system on a reactive potential energy surface defined as a dynamical linear combination of diabatic basis states, as in the earlier multistate empirical valence bond (MS-EVB) method.(2, 33, 34) As will be shown in detail later, the newer MS-RMD approach differs from the older MS-EVB approach primarily because the underlying reactive force field of MS-RMD is largely derived from AIMD or QM/MM data via force-matching and other means, rendering it far less empirical than the original MS-EVB approach. MS-RMD is also "multiscale" in the sense that quantum information on the electronic states is variationally bridged upward in scale to describe the forces felt by the system nuclei in the RMD model. As shown later as well, MS-RMD is several orders of magnitude more computationally efficient than MD from QM/MM, while still describing the charge delocalization and reactive nature of the PT. The MS-RMD method is also not complicated by the QM/MM boundary issues because the molecules that participate in the proton charge defect charge delocalization are dynamically redefined at each time step in such a way that the forces on the atoms are more continuous. Thus, as water molecules or residues move away from the excess proton, the forces acting on the atoms gradually transition to those described by the classical force field. These advantages make the MS-RMD method (and MS-EVB

before it) a powerful tool for investigating PT in a variety of biological systems.

In the Chapter 5 (35), a powerful multiscale combination of MS-RMD, classical and QM/MM MD simulations is employed to systematically investigate the proton solvation and transport mechanism through the influenza A M2 protein transmembrane domain (AM2/TM). The influenza type A virus is a highly pathogenic RNA virus that causes flu in birds and mammals (36). The AM2 protein contains a homotetramer channel that transports protons across the viral membrane and acidifies the virion interior, enabling the dissociation of the viral matrix proteins, which is a crucial step in viral replication (8). The protein has been the target of antiviral drugs amantadine and rimantadine (37, 38). Much effort has been devoted to discovering the structure and proton transport (PT) mechanism of the AM2 channel, resulting in many crystal structures available in the protein data bank (39-47). Based on the crystal structures and electrophysiology experiments, several PT models have been suggested. (47-54).

Although many computational studies have investigated the features of AM2 that may influence its PT in recent years (45, 55-71), only a few have explicitly simulated any aspect of the explicit PT process (55, 56, 60, 68, 69). This is because it is challenging to accurately model the charge delocalization and Grotthuss shuttling of the hydrated excess proton in a computationally tractable way. Among these studies, only one so far has provided a free-energy profile (potential of mean force, or PMF) for PT across the entire AM2/TM channel (56), which is the essential property for understanding the full proton permeation mechanism. However, this particular free energy study was limited by the approximation that the His37 tetrad remains in a fixed protonation state during the proton conduction; therefore, it could not capture the more plausible shuttle mechanism (47, 50-

54). More recently, Carnevale et. al (60) used a quantum mechanics/molecular mechanics (QM/MM) approach to investigate PT in the specific region of the His37 tetrad of AM2. Although they could not achieve sufficient sampling to calculate a free energy profile, this work did allow the His37 tetrad to change protonation states as necessitated by the shuttle mechanism. Their work in fact helped to lay the foundation for the study of PT in the His37 tetrad using a QM/MM approach in that region of the channel.

In Chapter 5 (35), by employing a classical force field, the conformational ensemble of AM2/TM is characterized starting from two recent high-resolution structures in low and intermediate pH conditions (PDB code 3C9J (42) and 3LBW (45), respectively). The influence of pH on the protein and water dynamics is then investigated. It is found that as the pH is lowered, the channel adopts a more open conformation, the channel pore is more hydrated, and the pore water molecules are more mobile. Following this we utilize a synthesis of the reactive multi-state empirical valence bond (MS-EVB) (2) and QM/MM approaches to calculate complete proton permeation free energy profiles (PMFs) through the AM2/TM channel. Since the MS-EVB potential can be derived by force-matching of *ab initio* MD simulation data (72), and since the force-matching algorithm provides a PMF for the reference potential, this unique combination of approaches yields a consistent multiscale computational framework for obtaining an *ab intio* level quality PT free energy profile. This multiscale combination of methods is thus used to calculate the PMFs for excess protons permeating through the entire AM2/TM channel, accounting for the charge delocalization and shuttling of hydrated excess protons through both water and the His37 tetrad explicitly. By virtue of this

methodology, it is found that at lower pH values the free energy barrier for a proton to diffuse to His37 is decreased and the proton conductance increased, explaining the unique pH-dependent activation mechanism of AM2. The PMFs of several AM2 mutants, including the amantadine-resistant and transmissible mutants V27A and S31N are also reported. The calculated conductances of both the wild-type AM2 and its mutants are in close agreement with experimental results and provide microscopic explanations for their trends.

In Chapter 6 (73), we move on to addressing the primary challenge of properly parameterizing the MS-RMD method in order to faithfully simulate PT in proteins. In ref (74), an approach for parameterizing MS-RMD models for ionizable amino acids in bulk water was developed. In this fitting approach, forces from QM/MM calculations are bridged via an iterative variational framework into the reactive MS-RMD model. In particular, the MS-RMD model parameters are obtained by sampling configurations with a guess MS-RMD model and then fitting the model parameters to best reproduce forces from QM/MM calculations that are as accurate as possible. Our fitting procedure was partially motivated by the work of Wang and co-workers, in which force fields for liquid water were developed with an adaptive force matching method.(75, 76) It also bears similarity to the work of Zhou and Pu on reactive path force matching.(77) Due to the use of configurations from condensed phase MS-RMD trajectories, the resulting model takes into account the condensed phase environment, which is not captured in gas phase fitting. For a given MS-RMD potential energy functional form and with restrictions on the ranges of parameter values, this fitting approach was shown to work well for investigating glutamic acid (Glu) and aspartic acid (Asp) deprotonation in bulk water.(74)

However, one cannot simply use these models to study deprotonation of amino acids such as Asp and Glu within proteins. The electrostatic and hydration environment affecting ionizable amino acid protonation/deprotonation inside biomolecules can be very different from that in bulk solution. For example, deprotonated Asp and Glu residues often form salt bridges with positively charged residues in proteins, and the surrounding hydration structure is rarely bulk-like. As a consequence, the pK_a values of amino acids inside proteins are often largely shifted from those measured in bulk solution. These differences require reparameterization of the MS-RMD models to fit the protein environment. In Chapter 6 (73), we extended the FitRMD method to parameterize MS-RMD models for ionizable amino acid in proteins.

In Chapter 7 (78), we apply the FitRMD and MS-RMD methods to investigate the proton transport mechanism in a challenging protein system, the cytochrome c oxidase (CcO). C*c*O is the terminal enzyme in the respiratory electron transfer chain in the inner membrane of mitochondria and plasma membrane of bacteria. It catalyzes the reduction of O_2 to H_2O and couples the free energy of this exergonic reaction to the pumping of protons across the membrane, creating a transmembrane proton electrochemical gradient that drives, for example, ATP synthesis. During each reaction cycle eight protons are taken up from the negatively-charged inside (N-side) of the membrane and either react with oxygen (referred to as 'chemical' protons below) or are pumped to the positively-charged outside (P-side) of the membrane (referred to as 'pumped' protons below). In aa_3-type C*c*O, as found in mitochondria, the D-channel is responsible for uptake of all four pumped protons and at least one out of four chemical protons. Protons on the N-side are taken into D-channel via the amino acid residue D132 at the channel entrance, and

then transferred to residue E286 in the middle of the membrane. By Grotthuss shuttling through the water molecules in the hydrophobic cavity (HC) above E286, each proton is either transferred to react with oxygen in the binuclear center (BNC), consisting of heme a_3 and the Cu_B complex, or transferred to the pump loading site (PLS) and then further released to the P-side of the membrane (c.f. **Figure 7.1**). Despite decades of study, the C*c*O proton pumping mechanism, which entails the transport of two protons from the N-side of the membrane (one to be pumped and one for catalysis) coupled to a single electron transferred to the BNC, is still incompletely understood at the atomistic level. It has been unclear, for example, how electron transfer (ET) is coupled with the multiple proton transport (PT) events, in what order the charge transport processes happen, and how C*c*O prevents back flow of pumped protons during the transfer of chemical protons.

To address these questions, experimental results must be complemented with molecular-level insight from computer simulation. However, in large and complicated biomolecular systems it is challenging to simulate PT, which requires an explicit treatment of the positive charge defect associated with a hydrated excess proton, including its delocalization and Grotthuss shuttling (1). To overcome this challenge, a multiscale reactive molecular dynamics (MS-RMD) method has been extensively developed and applied in our group to study PT in aqueous and biological contexts [see, e.g., Refs. (2, 31, 32, 74, 79, 80)]. Here we have carried out extensive MS-RMD free energy simulations to study the proton pumping mechanism in C*c*O. In our MS-RMD approach, quantum mechanical forces from targeted quantum mechanics/molecular mechanics (QM/MM) calculations are bridged, in a multiscale fashion via a variational mathematical framework, into the reactive MD algorithm (MS-RMD) for the dynamics of

system nuclei, thus including chemical bond breaking and making. In this way, we explicitly simulate PT between proton binding sites, including Grotthuss shuttling of the excess proton(s) through residues and intervening water molecules.

We focus on the $A \rightarrow P_R \rightarrow F$ transition in C*c*O, which occurs during the oxidation of the fully reduced enzyme, because there is extensive experimental information to which we can compare our computational results. Based on our simulations, we provide a quantitative, comprehensive and molecular-level description of proton uptake, pumping, and chemical proton transfer during the $A \rightarrow P_R \rightarrow F$ transition. Our results show that both PT events (E286 to the PLS and E286 to the BNC) are driven by ET from heme *a* to the BNC. The transfer of the pumped proton is kinetically favored while that of the chemical proton is rate limiting. Our calculated rate for the chemical proton transfer is in quantitative agreement with experimental measurements. (81, 82) These results also explain how C*c*O prevents the decoupling of pumping from the chemical reaction with kinetic gating. The fast pumping process precedes transfer of the chemical proton to the BNC, and fast D-channel PT to E286 after pumping prevents proton back flow from the PLS. Given the computational accuracy and efficiency of the MS-RMD methodology (80), a critical component of our results is the explicit characterization of the coupling between PT and hydration changes in both the HC and D-channel, revealing a remarkable and dynamic coupling between the migration of the excess proton and hydration. Finally, we present results that argue against the possibility of E286 being biprotonated during the pumping process.

Chapter 8 is dedicated to the characterization of the acid activation mechanism of influenza M2 channel from high to intermediate pH condition, which was not

investigated in Chapter 5 (35). The study was motivated by a recent work by Thomaston et al (83), where several M2TM crystal structures were resolved in both high and low pH conditions in lipid cubic phase. In that work, it was found that the backbones of the AM2 channel adopt similar structures in both low and high pH conditions. In all structures, the C-terminal helices are wide open, allowing sufficient solvation of His37 and Trp41 tetrad. The similarity of the new set of crystal structures in both low and high pH conditions give rise to questions regarding how the channel conduction is activated upon lowering the pH value.

Building upon the previous simulations in Chapter 5 (35) and the recently published crystal structures (83), in Chapter 8 we further investigated the AM2 acid activation mechanism for the transition from high to intermediate pH condition. We performed multiscale simulations starting from the high pH room temperature crystal structure in ref (83) (PDB code 4QKL). Our results show that in membrane environment, the channel closes its C-terminal helices in high pH condition and the resulting large deprotonation barrier of His37 blocks the channel. Lowering the pH value increases the charge state of His37, and as a result the Trp41 gate is gradually opened. This increases the channel conductance and leads to activation of the AM2 channel in intermediate pH condition.

In the paper by Sharma et al (47), the NMR structure of AM2 conductance domain (residues 22-62) was published (PDB code 2L0J). The structure includes both M2TM (residues 26-46) and the amphipathic helices. However, Ma et al (84) concluded that the M2TM helices conserves the essential features of the proton transport in the full length AM2 channel, and it is for this reason that our previous simulations (35, 56) and many other computational and experimental works (45, 52, 53, 60, 64, 68, 71) used the M2TM

helices to investigate the proton conduction mechanism of AM2. Here in this work we also investigated the effect of including the amphipathic helices on the proton conduction mechanism of the channel by performing simulations starting from the 2L0J NMR structure. Our results show that the amphipathic helices do not significantly change the proton conduction mechanism of the AM2.

1.2 References

1. Knight C & Voth GA (2012) The Curious Case of the Hydrated Proton. *Accounts of Chemical Research* 45(1):101-109.

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

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

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

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

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

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

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

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

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

11. Peng Y, Swanson JMJ, Kang S-g, Zhou R, & Voth GA (2015) Hydrated Excess Protons Can Create Their Own Water Wires. *The Journal of Physical Chemistry B* 119(29):9212-9218.

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

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

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

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

16. Elstner M (2006) The SCC-DFTB method and its application to biological systems. *Theor Chem Acc* 116(1-3):316-325.

17. Gaus M, Cui QA, & Elstner M (2011) DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J Chem Theory Comput* 7(4):931-948.

18. Maupin CM, Aradi B, & Voth GA (2010) The Self-Consistent Charge Density Functional Tight Binding Method Applied to Liquid Water and the Hydrated Excess Proton: Benchmark Simulations. *J Phys Chem B* 114(20):6922-6931.

19. Choi TH, Liang R, Maupin CM, & Voth GA (2013) Application of the SCC-DFTB Method to Hydroxide Water Clusters and Aqueous Hydroxide Solutions. *J Phys Chem B* 117(17):5165-5179.

20. Goyal P, Elstner M, & Cui Q (2011) Application of the SCC-DFTB Method to Neutral and Protonated Water Clusters and Bulk Water. *J Phys Chem B* 115(20):6790-6805.

21. Gaus M, Goez A, & Elstner M (2013) Parametrization and Benchmark of DFTB3 for Organic Molecules. *J Chem Theory Comput* 9(1):338-354.

22. Petraglia R & Corminboeuf C (2013) A Caveat on SCC-DFTB and Noncovalent Interactions Involving Sulfur Atoms. *J. Chem. Theory Comput.* 9(7):3020-3025.

23. Goyal P*, et al.* (2011) Proton Storage Site in Bacteriorhodopsin: New Insights from Quantum Mechanics/Molecular Mechanics Simulations of Microscopic pK(a) and Infrared Spectra. *J Am Chem Soc* 133(38):14981-14997.

24. Riccardi D, Yang S, & Cui Q (2010) Proton transfer function of carbonic anhydrase: Insights from QM/MM simulations. *Bba-Proteins Proteom* 1804(2):342-351.

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

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

27. Lear JD, Wasserman ZR, & Degrado WF (1988) Synthetic Amphiphilic Peptide Models for Protein Ion Channels. *Science* 240(4856):1177-1181.

28. Wu YJ, Ilan B, & Voth GA (2007) Charge delocalization in proton channels, II: The synthetic LS2 channel and proton selectivity. *Biophys J* 92(1):61-69.

29. Voth GA (2003) The computer simulation of proton transport in biomolecular systems. *Front Biosci* 8:S1384-S1397.

30. Wu YJ & Voth GA (2003) A computer simulation study of the hydrated proton in a synthetic proton channel. *Biophys J* 85(2):864-875.

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

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

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

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

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

36. Cannell JJ, Zasloff M, Garland CF, Scragg R, & Giovannucci E (2008) On the epidemiology of influenza. *Virol J.* 5:29.

37. Hay AJ, Wolstenholme AJ, Skehel JJ, & Smith MH (1985) The Molecular-Basis of the Specific Anti-Influenza Action of Amantadine. *EMBO J.* 4(11):3021-3024.

38. Wang C, Takeuchi K, Holsinger LJ, Lamb RA, & Pinto LH (1993) Ion Channel Activity of the M2 Protein of Influenza a-Virus. *Biophys. J.* 64(2):A94-A94.

39. Wang JF, Kim S, Kovacs F, & Cross TA (2001) Structure of the transmembrane region of the M2 protein H+ channel. *Protein Sci.* 10(11):2241-2250.

40. Nishimura K, Kim S, Zhang L, & Cross TA (2002) The Closed State of a H^+ Channel Helical Bundle Combining Precise Orientational and Distance Restraints from Solid State NMR. *Biochemistry* 41(44):13170-13177.

41. Schnell JR & Chou JJ (2008) Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* 451(7178):591-595.

42. Stouffer AL*, et al.* (2008) Structural basis for the function and inhibition of an influenza virus proton channel. *Nature* 451(7178):596-U513.

43. Cady SD, Mishanina TV, & Hong M (2009) Structure of Amantadine-Bound M2 Transmembrane Peptide of Influenza A in Lipid Bilayers from Magic-Angle-Spinning Solid-State NMR: The Role of Ser31 in Amantadine Binding. *J. Mol. Biol.* 385(4):1127-1141.

44. Cady SD*, et al.* (2010) Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. *Nature* 463(7281):689-692.

45. Acharya R*, et al.* (2010) Structure and mechanism of proton transport through the transmembrane tetrameric M2 protein bundle of the influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* 107(34):15075-15080.

46. Pielak RM & Chou JJ (2010) Solution NMR structure of the V27A drug resistant mutant of influenza A M2 channel. *Biochem. Biophys. Res. Commun.* 401(1):58-63.

47. Sharma M*, et al.* (2010) Insight into the Mechanism of the Influenza A Proton Channel from a Structure in a Lipid Bilayer. *Science* 330(6003):509-512.

48. Sansom MSP, Kerr ID, Smith GR, & Son HS (1997) The Influenza A Virus M2 Channel: A Molecular Modeling and Simulation Study. *Virology* 233(1):163-173.

49. Okada A, Miura T, & Takeuchi H (2001) Protonation of Histidine and Histidine-Tryptophan Interaction in the Activation of the M2 Ion Channel from Influenza A Virus. *Biochemistry* 40(20):6053-6060.

50. Pinto LH*, et al.* (1997) A functionally defined model for the M2 proton channel of influenza A virus suggests a mechanism for its ion selectivity. *Proceedings of the National Academy of Sciences* 94(21):11301-11306.

51. Hong M, Fritzsching KJ, & Williams JK (2012) Hydrogen-Bonding Partner of the Proton-Conducting Histidine in the Influenza M2 Proton Channel Revealed From H-1 Chemical Shifts. *J. Am. Chem. Soc.* 134(36):14753-14755.

52. Hu FH, Schmidt-Rohr K, & Hong M (2012) NMR Detection of pH-Dependent Histidine-Water Proton Exchange Reveals the Conduction Mechanism of a Transmembrane Proton Channel. *J. Am. Chem. Soc.* 134(8):3703-3713.

53. Hu FH, Luo WB, & Hong M (2010) Mechanisms of Proton Conduction and Gating in Influenza M2 Proton Channels from Solid-State NMR. *Science* 330(6003):505-508.

54. Williams JK*, et al.* (2013) Drug-Induced Conformational and Dynamical Changes of the S31N Mutant of the Influenza M2 Proton Channel Investigated by Solid-State NMR. *J. Am. Chem. Soc.* 135(26):9885-9897.

55. Smondyrev AM & Voth GA (2002) Molecular dynamics simulation of proton transport through the influenza A virus M2 channel. *Biophys. J.* 83(4):1987-1996.

56. Chen HN, Wu YJ, & Voth GA (2007) Proton transport Behavior through the influenza a M2 channel: Insights from molecular simulation. *Biophys. J.* 93(10):3470-3479.

57. Yi M, Cross TA, & Zhou HX (2008) A secondary gate as a mechanism for inhibition of the M2 proton channel by amantadine. *J. Phys. Chem. B* 112(27):7977-7979.

58. Khurana E*, et al.* (2009) Molecular dynamics calculations suggest a conduction mechanism for the M2 proton channel from influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* 106(4):1069-1074.

59. Yi MG, Cross TA, & Zhou HX (2009) Conformational heterogeneity of the M2 proton channel and a structural model for channel activation. *Proc. Natl. Acad. Sci. U. S. A.* 106(32):13311-13316.

60. Carnevale V, Fiorin G, Levine BG, DeGrado WF, & Klein ML (2010) Multiple Proton Confinement in the M2 Channel from the Influenza A Virus. *J. Phys. Chem. C* 114(48):20856-20863.

61. Balannik V*, et al.* (2010) Functional Studies and Modeling of Pore-Lining Residue Mutants of the Influenza A Virus M2 Ion Channel. *Biochemistry* 49(4):696-708.

62. Phongphanphanee S, Rungrotmongkol T, Yoshida N, Hannongbua S, & Hirata F (2010) Proton Transport through the Influenza A M2 Channel: Three-Dimensional Reference Interaction Site Model Study. *J. Am. Chem. Soc.* 132(28):9782-9788.

63. Zhou HX (2010) Diffusion-Influenced Transport of Ions across a Transmembrane Channel with an Internal Binding Site. *J. Phys. Chem. Lett.* 1(13):1973-1976.

64. Wang J*, et al.* (2011) Molecular Dynamics Simulation Directed Rational Design of Inhibitors Targeting Drug-Resistant Mutants of Influenza A Virus M2. *J. Am. Chem. Soc.* 133(32):12834-12841.

65. Gu RX*, et al.* (2011) Free energy calculations on the two drug binding sites in the M2 proton channel. *J. Am. Chem. Soc.* 133(28):10817-10825.

66. Zhou HX (2011) A Theory for the Proton Transport of the Influenza Virus M2 Protein: Extensive Test against Conductance Data. *Biophys. J.* 100(4):912-921.

67. Gu RX, Liu LA, Wang YH, Xu Q, & Wei DQ (2013) Structural comparison of the wild-type and drug-resistant mutants of the influenza a m2 proton channel by molecular dynamics simulations. *J. Phys. Chem. B* 117(20):6042-6051.

68. Dong H, Fiorin G, DeGrado WF, & Klein ML (2013) Exploring Histidine Conformations in the M2 Channel Lumen of the Influenza A Virus at Neutral pH via Molecular Simulations. *J. Phys. Chem. Lett.* 4(18):3067-3071.

69. Bankura A, Klein ML, & Carnevale V (2013) Proton affinity of the histidine-tryptophan cluster motif from the influenza A virus from ab initio molecular dynamics. *Chem. Phys.* 422:156-164.

70. Wang J*, et al.* (2013) Structure and inhibition of the drug-resistant S31N mutant of the M2 ion channel of influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* 110(4):1315-1320.

71. Wei C & Pohorille A (2013) Activation and Proton Transport Mechanism in Influenza A M2 Channel. *Biophys. J.* 105(9):2036-2045.

72. Knight C, Maupin CM, Izvekov S, & Voth GA (2010) Defining Condensed Phase Reactive Force Fields from ab Initio Molecular Dynamics Simulations: The Case of the Hydrated Excess Proton. *J. Chem. Theory Comput.* 6(10):3223-3232.

73. Lee S, Liang R, Voth GA, & Swanson JMJ (2016) Computationally Efficient Multiscale Reactive Molecular Dynamics to Describe Amino Acid Deprotonation in Proteins. *J. Chem. Theory Comput.* 12(2):879-891.

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

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

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

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

78. Liang R, Swanson JMJ, Peng Y, Wikström M, & Voth GA (2015) Multiscale simulations reveal key features of the proton pumping mechanism in cytochrome c oxidase. *(submitted)*.

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

80. Lee S, Liang R, Voth GA, & Swanson JMJ (2016) Computationally Efficient Multiscale Reactive Molecular Dynamics to Describe Amino Acid Deprotonation in Proteins. *J. Chem. Theory Comput.* (In Press):DOI: 10.1021/acs.jctc.1025b01109.

81. Faxen K, Gilderson G, Adelroth P, & Brzezinski P (2005) A mechanistic principle for proton pumping by cytochrome c oxidase. *Nature* 437(7056):286-289.

82. Belevich I, Verkhovsky MI, & Wikstrom M (2006) Proton-coupled electron transfer drives the proton pump of cytochrome c oxidase. *Nature* 440(7085):829-832.

83. Thomaston JL*, et al.* (2015) High-resolution structures of the M2 channel from influenza A virus reveal dynamic pathways for proton stabilization and transduction. *Proc. Natl. Acad. Sci. U. S. A.* 112(46):14260-14265.

84. Ma CL*, et al.* (2009) Identification of the functional core of the influenza A virus A/M2 proton-selective ion channel. *Proc. Natl. Acad. Sci. U. S. A.* 106(30):12283-12288.

Chapter 2

Theory and Methodology

2.1 Multiscale Reactive Molecular Dynamics

The MS-RMD (and MS-EVB) method describes electronic delocalization of the excess proton as a linear combination of distinct states with different chemical bonding topologies. (See Figure S1 in the SI.) The Hamiltonian is defined as:

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

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

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

The forces are evaluated by the Hellmann-Feynman theorem and are used to propagate the system in an MD simulation. It is important to note that the method explicitly treats the excess proton charge defect delocalization, Grotthuss shuttling, and part of the polarization effect, and the resulting MD trajectory is continuous and deterministic to within numerical error.

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

the state with protonated glutamate (GLUH) is described as,

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

where the first three terms are the inter- and the intra-molecular potentials of protonated glutamate and all other surrounding molecules, such as waters, other protein residues, lipids, and ions in the system. They are computed with CHARMM22 force field,(1) with the exception of the O-H bond in the carboxyl (-COOH) group of glutamate. To properly represent its dissociation as a proton transfers from the carboxyl oxygen to the water oxygen, the harmonic bond stretch potential is replaced with a standard Morse potential, $U^{Morse}(r)$:

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

where r is the O-H bond length, and D, α, and r_0 are parameters, which are taken from our previous work.(2) Since the classical force fields between two protonated forms of glutamate and water do not share a common energy origin, V_{ii} is introduced to compensate the constant energy shift between the two states.

In order to correct overestimated electrostatic interaction between opposite charges on a hydronium and deprotonated glutamate at a short distance(3), two repulsive terms, $V_{OO_k}^{rep}$ and $V_{H_jO_k}^{rep}$, are introduced in h_{ii} corresponding to the state with deprotonated glutamate:

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

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

where R_{OO_ϵ} is the distance between the hydronium oxygen, O, and the carboxyl oxygen of glutamate, O_ϵ (OE1 and OE2 in the PDB), and $R_{H_jO_\epsilon}$ is the distance between each of

three hydronium hydrogen atoms, H_j, and the carboxyl oxygen of glutamate. The functional forms for the repulsive terms are the same as those used in the MS-EVB3.1 model. (2) B, b, b', C, and c are fitted parameters, and d_{OO}^0 and d_{OH}^0 are fixed the same value used in MS-EVB3.1, which are 2.4 and 1.0 Å, respectively.

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

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

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

The MS-RMD (and MS-EVB) formalism also provides a convenient and physically intuitive description of the excess proton center of the excess charge (CEC), defined as:(4)

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

where the N_{EVB} is the total number of EVB states, $c_i^2(\mathbf{r})$ is the population of state i contributing to the MS-RMD ground state, and $\mathbf{r}_i^{\mathrm{COC}}$ is the geometric center-of-charge for state i. This CEC definition allows the use of a continuous reaction coordinate (further discussed below) for the PMF calculation of PT in biological systems.[2] The protonated moiety in the state with the largest coefficient, c_1 (Equation 8), possesses the majority of

the excess positive charge. When in bulk water this state is the most hydronium-like species(5-8) (or the so-called “pivot” hydronium, a technical term used below).

2.2 FitRMD parameterization scheme

One possible fitting procedure of an MS-RMD potential energy function to QM/MM data was first described in ref (2). In such an approach, configurations along the PT reaction coordinate are sampled by MS-RMD umbrella sampling simulations with initial guess MS-RMD model parameter set. In the present paper, the initial guess amino acid models were taken from the previous work done in bulk water,(2) except that the constant energy shift between protonated/deprotonated states in the model (V_{ii}) was determined by the difference in the Coulomb energy of the RMD (EVB) complex between the most favorable hydronium state and the protonated state. The range of the PT reaction coordinate was set to sample configurations for both protonated/deprotonated states (further defined below). Next, for each configuration a QM/MM calculation was performed to collect the reference forces on each atom in the MS-RMD reactive complex. Then the MS-RMD model parameter set was optimized by minimizing the variational residual:

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

where N_C and N_A are the number of configurations and the number atoms in each configuration, respectively. The weight of each atom $w(\mathbf{r}_{ij})$ is set to unity here, but it should be noted other choices are possible. The atomic force $\mathbf{F}_{ij}$ is the one obtained from

the current MS-RMD model parameter set and $\mathbf{F}_{ij}^{\mathrm{QM/MM}}$ is the reference force from the QM/MM calculations. The whole set of the model parameters were then divided into two groups: 1) the diagonal terms V_{ii}, and 2) the off-diagonal and repulsive terms. The model optimization was done individually for each group. First, the off-diagonal and repulsive terms were fit with the value of V_{ii} fixed, and then V_{ii} was refit with new values of the off-diagonal and repulsive terms held fixed.

2.3 The SCC-DFTB Method.

The SCC-DFTB method is based on the Taylor expansion of the Kohn-Sham total energy in terms of the charge density fluctuations, with the Hamiltonian matrix elements being evaluated using a minimal basis set of pseudoatomic orbitals, along with a two-center approximation. The energy function for the SCC-DFTB method is expressed as

$$E_{total}^{SCC-DFTB} = \sum_{i\mu v} c_{\mu}^{i} c_{\mu}^{i} H_{\mu v}^{0} + E_{rep} + \frac{1}{2}\sum_{\alpha\beta} \gamma_{\alpha\beta} \Delta q_{\alpha} \Delta q_{\beta} \ , \qquad (1)$$

where $H_{\mu v}^{0}$ are the Hamiltonian matrix elements and c_{μ}^{i} and c_{μ}^{i} are the wave function expansion coefficients, E_{rep} accounts for short-range repulsive term, and Δq_{α} and Δq_{β} are the charge fluctuation terms associated with atoms α and β, respectively. The $\gamma_{\alpha\beta}$ function is determined by:

$$\gamma_{\alpha\beta} = \frac{1}{R_{\alpha\beta}} - S_{\alpha\beta} \times f_{\alpha\beta},$$

$$f_{\alpha\beta} = \begin{cases} \exp\left[-\left(\dfrac{U_{\alpha} + U_{\beta}}{2}\right)^{\zeta} R_{\alpha\beta}^{2}\right], & \text{if } (\alpha \text{ or } \beta = \mathrm{H}) \\ \text{or} & \\ 1, & \text{else} \end{cases} \qquad (2)$$

where $S_{\alpha\beta}$ is an exponentially decaying short-range function. The damping function $f_{\alpha\beta}$ is introduced to correctly describe hydrogen-bonding interactions, where U_α and U_β are the atomic Hubbard parameters related to the chemical hardness of atoms, α and β, respectively. The parameter ζ is determined by fitting to hydrogen bonding energies for select clusters from high level QM calculations.

The third order correction term was initially introduced with only the on-site third order term, and the complete third-order term has been recently implemented in gas phase calculations.(9) In the present study, we considered two types of 3rd order DFTB corrections, the first being the on-site 3rd order correction and the second being full 3rd order correction. The two corrections are described by:

$$E_{on-site}^{3rd} = \frac{1}{6}\sum_{\alpha} U_\alpha^d \Delta q_\alpha^3 \tag{3}$$

$$E_{Full}^{3rd} = \frac{1}{3}\sum_{\alpha\beta} \Gamma_{\alpha\beta} \Delta q_\alpha^2 \Delta\, q_\beta \tag{4}$$

respectively, where U_α^d is the charge derivative of the atomic Hubbard parameter U_α, and $\Gamma_{\alpha\beta}$ describe derivatives of γ function with respect to atomic charges. The full correction is an extension of the on-site correction and was shown to improve hydrogen bonding energies and proton affinities over the latter.(9) Furthermore, the O–H repulsive potential has been modified for better description of proton transport (PT) barrier, and represented as follows:

$$U_{OH}^{Rep}(R) = U_{OH}^{Rep,std}(R) + a_1 \exp\left[-\frac{(R-a_2)^2}{a_3}\right] S(R) \tag{5}$$

where $S(R)$ is the switching function and the parameters, a_1, a_2, and a_3 were optimized using a genetic algorithm and the SCC-DFTB potential including the hydrogen-bonding correction. A more detailed description can be found in refs (10) and (11).

In ref (12) a new parameter set named 3OB was designed such that both the electronic and repulsive potential parameters (appearing in the first and second terms in eq 1, respectively) are optimized for full 3rd order SCC-DFTB calculations. In gas phase calculations this parameter set was shown to outperform the old MIO parameter set due to an improved geometry of nonbonded interactions and reduction of overbinding errors.(12) In Chapter 3 and 4, we benchmark the newest, full 3rd order SCC-DFTB method using the 3OB parameter set. The SCC-DFTB related parameters are taken from ref (12). In Chapter 3, we also investigate the effect of replacing MIO with the 3OB parameter set in the 3rd order SCC-DFTB calculations.

2.4 References

1. MacKerell AD*, et al.* (1998) All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B* 102(18):3586-3616.

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

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

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

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

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

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

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

9. Gaus M, Cui QA, & Elstner M (2011) DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J Chem Theory Comput* 7(4):931-948.

10. Riccardi D*, et al.* (2006) Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *J. Phys. Chem. B* 110(13):6458-6469.

11. Goyal P, Elstner M, & Cui Q (2011) Application of the SCC-DFTB Method to Neutral and Protonated Water Clusters and Bulk Water. *J Phys Chem B* 115(20):6790-6805.

12. Gaus M, Goez A, & Elstner M (2013) Parametrization and Benchmark of DFTB3 for Organic Molecules. *J Chem Theory Comput* 9(1):338-354.

Chapter 3

Application of the SCC-DFTB Method to Hydroxide Water Clusters and Aqueous Hydroxide Solutions

3.1 Introduction

The energetic and dynamical nature of aqueous systems that contain excess protons and hydroxide (proton hole) has been widely studied due to its fundamental role in acid-base chemistry and biological processes.(1-5) Although the use of high level quantum mechanics (QM) can accurately describe the static potential energy surface of cluster systems and the reactive process of the transfer media in aqueous solution, the high computational cost of QM calculations limits its use in modeling relevant sized systems and/or long time dynamical process.(6) The alternative to full QM calculations is to develop models or employ approximate QM methods such as semiempirical electronic structure methods. These approaches are parameterized using higher level *ab initio* calculations and their development is still ongoing based on benchmark studies. (7-9)

One of the most widely used approximate QM methods is the self-consistent charge density functional tight binding (SCC-DFTB) method as an approximation to density functional theory (DFT).(9-26) Compared to other semi-empirical electronic structure methods such as AM1, PM3, and NDDO,(27-32) the relatively more accurate description of geometries, energies, and vibrational frequencies that SCC-DFTB yields has made it widely utilized for the study of organic, biomolecular systems, and material systems, (23-26) although sometimes without a good degree of accuracy. (33) Moreover, in our previous simulations based on the released parameter sets at the time for several

prototype protonated water clusters including the magic cluster, $H^+(H_2O)_{21}$, and bulk water in the presence of an excess proton, the SCC-DFTB method gives generally inaccurate results for the protonated clusters, bulk water, and the excess proton in water. (18, 19) These studies led to a re-examination by Goyal et al., who suggested that the inaccurate standard γ_{OH} repulsive potential of the SCC-DFTB method resulted in a reduced proton-transfer barrier, and stabilization of the Zundel form of the hydrated proton over the Eigen form in gas and condensed phase. These authors addressed these shortcomings and some of the properties were marginally improved through the implementation of 3rd order correction and an empirically modified O–H repulsive potential.(20) In the very recent work by Gaus *et al*., a new parameter set, 3OB, was introduced and showed promising results for benchmark tests in the gas phase.(34) 3OB stands for design for DFTB3 of organic and biological applications while the previous parameter set, MIO, was designed for material and biological system.

In the present work, we evaluate the accuracy of the current SCC-DFTB implementations, including the proposed O–H repulsive potential and 3rd order correction with both MIO and the 3OB parameter sets, focusing on both cluster and liquid water systems with the hydroxide ion (proton hole).(35)

For the detailed analysis of the hydroxide cluster, basin hopping Monte Carlo simulations(36-39) are utilized to investigate the low-lying energy minimum structures of the small OH^- water cluster ion, $OH^-(H_2O)_n$, for n = 4–7. For the condensed phase, molecular dynamics (MD) simulations of the hydroxide ion in an aqueous environment with Ewald summation long-range electrostatics calculated with the supercell method was utilized.

3.2 Methods

3.2.1 The SCC-DFTB Method.

The SCC-DFTB method is based on the Taylor expansion of the Kohn-Sham total energy in terms of the charge density fluctuations, with the Hamiltonian matrix elements being evaluated using a minimal basis set of pseudoatomic orbitals, along with a two-center approximation. The energy function for the SCC-DFTB method is expressed as

$$E_{total}^{SCC-DFTB} = \sum_{i\mu v} c_{\mu}^{i} c_{\mu}^{i} H_{\mu v}^{0} + E_{rep} + \frac{1}{2}\sum_{\alpha\beta} \gamma_{\alpha\beta} \Delta q_{\alpha} \Delta q_{\beta} \ , \tag{1}$$

where $H_{\mu v}^{0}$ are the Hamiltonian matrix elements and c_{μ}^{i} and c_{μ}^{i} are the wave function expansion coefficients, E_{rep} accounts for short-range repulsive term, and Δq_{α} and Δq_{β} are the charge fluctuation terms associated with atoms α and β, respectively. The $\gamma_{\alpha\beta}$ function is determined by:

$$\gamma_{\alpha\beta} = \frac{1}{R_{\alpha\beta}} - S_{\alpha\beta} \times f_{\alpha\beta},$$

$$f_{\alpha\beta} = \begin{cases} \exp\left[-\left(\frac{U_{\alpha} + U_{\beta}}{2}\right)^{\zeta} R_{\alpha\beta}^{2}\right], & \text{if } (\alpha \text{ or } \beta = \text{H}) \\ \text{or} & \\ 1 \quad , & \text{else} \end{cases} \tag{2}$$

where $S_{\alpha\beta}$ is an exponentially decaying short-range function. The damping function $f_{\alpha\beta}$ is introduced to correctly describe hydrogen-bonding interactions, where U_{α} and U_{β} are the atomic Hubbard parameters related to the chemical hardness of atoms, α and β, respectively. The parameter ζ is determined by fitting to hydrogen bonding energies for select clusters from high level QM calculations.

The third order correction term was initially introduced with only the on-site third order term, and the complete third-order term has been recently implemented in gas phase calculations.(21) In the present study, we considered two types of 3rd order DFTB corrections, the first being the on-site 3rd order correction and the second being full 3rd order correction. The two corrections are described by:

$$E_{on-site}^{3rd} = \frac{1}{6}\sum_{\alpha} U_{\alpha}^{d}\Delta q_{\alpha}^{3} \tag{3}$$

$$E_{Full}^{3rd} = \frac{1}{3}\sum_{\alpha\beta} \Gamma_{\alpha\beta}\Delta q_{\alpha}^{2}\Delta\, q_{\beta} \tag{4}$$

respectively, where U_{α}^{d} is the charge derivative of the atomic Hubbard parameter U_{α}, and $\Gamma_{\alpha\beta}$ describe derivatives of γ function with respect to atomic charges. The full correction is an extension of the on-site correction and was shown to improve hydrogen bonding energies and proton affinities over the latter.(21) Furthermore, the O–H repulsive potential has been modified for better description of proton transport (PT) barrier, and represented as follows:

$$U_{OH}^{Rep}(R) = U_{OH}^{Rep,std}(R) + a_1 \exp\left[-\frac{(R-a_2)^2}{a_3}\right] S(R) \tag{5}$$

where $S(R)$ is the switching function and the parameters, a_1, a_2, and a_3 were optimized using a genetic algorithm and the SCC-DFTB potential including the hydrogen-bonding correction. A more detailed description can be found in refs (16) and (20).

In ref (34), a new parameter set named 3OB was designed, where both electronic parameters and repulsive potential parameters (appearing in the first and second term in eq 1, respectively) are optimized for full 3rd order SCC-DFTB calculations. In gas phase calculations, it was shown to outperform the old MIO parameter set due to an improved

geometry of nonbonded interactions and reduction of the overbinding errors. The effect of replacing MIO with the 3OB parameter set in the 3rd order SCC-DFTB calculations will also be investigated in this work.

In the following sections, we will use the same notations from ref (20) for denoting the different variants of the SCC-DFTB method. The standard second-order SCC-DFTB as expressed in eq 1 will be denoted as "DFTB2", the hydrogen-bonding correction is indicated by adding "-γ^{h}". The 3rd order DFTB methods with the on-site third order correction (eq 3) and with full third order correction (eq 4) are expressed as "DFTB3-diag" and "DFTB3-Full", respectively. Those DFTB methods using a modified O–H repulsive potential (eq 5) are indicated by the additional "gaus" keyword. The above notation applies to the SCC-DFTB methods using the MIO parameter set while the 3rd order DFTB method using the new 3OB parameter set will be denoted as "DFTB3-3OB". All 3rd order methods used in this study include the hydrogen-bonding correction. In the case of the cluster simulations, the dispersion interaction term is added because it plays a major role in determining the low-energy minima based on our benchmark calculations with neutral and protonated water clusters.(18)

3.2.2 Cluster simulation setup.

The basin-hopping Monte Carlo method uses Monte Carlo walks combined with gradient-based optimizations to locate local potential energy minima. At each move, the structure is optimized to the local minimum in its "basin", and acceptance of moves is based on the energies of the minima. As a result, barriers are effectively bypassed, making this an efficient approach for searching for low-energy minima. In the present

study, the Monte Carlo moves involve both translational and rotational moves at a reduced temperature of 2.0 kJ/mol, with a target acceptance ratio of 0.4, which has been shown to work reasonably well.(18) At each Monte Carlo step, the number of molecules to be moved was chosen at random. Randomly selected initial structures of $OH^-(H_2O)_n$, with $n = 4 - 7$, were then selected and subsequently used to carry out eight independent sets of the basin hopping Monte Carlo runs of 100,000 steps each.

3.2.3 Molecular dynamics simulation setup.

An initial bulk water system was created by simulating 128 modified TIP3P classical water molecules for 500 ps of molecular dynamics (MD) in the constant NVE ensemble followed by 500 ps in the constant NPT ensemble. The NPT pre-equilibration simulations resulted in box dimensions of $15.8 \times 15.8 \times 15.8$ $Å^3$ and a corresponding density of $\rho = 0.97$ g/mL. A randomly selected water molecule from the previous simulations was replaced with a hydroxide anion. This newly created hydroxide system was then duplicated and two 60 ps equilibration runs in the constant NVT ensemble were simulated in each of the various DFTB+ codes.(40) The latter simulations were conducted at a temperature of 300 K using an Anderson thermostat, an integration time step of 1 fs, and an SCC convergence criteria of 10^{-6} atomic units. After equilibration, the systems were subjected to production runs of 50~100 ps in the constant NVE ensemble. The electrostatics for all of the DFTB+ simulations were treated by the supercell sampling using 2×2×2 Monkhorst-Pack scheme.(41) In the SCC-DFTB simulations, ζ in the modified function in eq. 2 was set to 4.5 for the DFTB2-γ^h method and 4.95 for

DFTB3-diag and DFTB3-diag+gaus methods, as indicated in previous work.(20) For DFTB3-Full+gaus method, the values for ζ and Hubbard derivatives U^d were taken from Table 2 "DFTB3 fit" entry in ref (21). For the DFTB3-3OB calculations the parameters were taken from Table 1 in ref (34).

3.3 Results and Discussion

3.3.1 Potential Energy Surface of $H_3O_2^-$.

Figure 3.1 shows the potential energy contour diagrams for $H_3O_2^-$ system using several different electronic structure calculations. In generating the potential energy surfaces, the geometry for the $H_3O_2^-$ cluster was optimized at the MP2/6-31G++(2d,2p) level of theory. The potential energy surface was then created by scanning the H between OH and H_2O at fixed OO distances. This scanning calculation was repeated by increasing the OO distance by 0.05 Å from 2.0 to 3.5 Å. Other bond lengths and angles were fixed with the optimized geometry during scanning energies. In **Figure 3.1**, the y axis represents the OO distance between two oxygen atoms, and the x axis depicts the distance ratio, R_{OH}/R_{OO}, between the OH and OO distance, which indicates the "Zundel" or "Eigen" like hydroxide structures effectively.(42, 43)

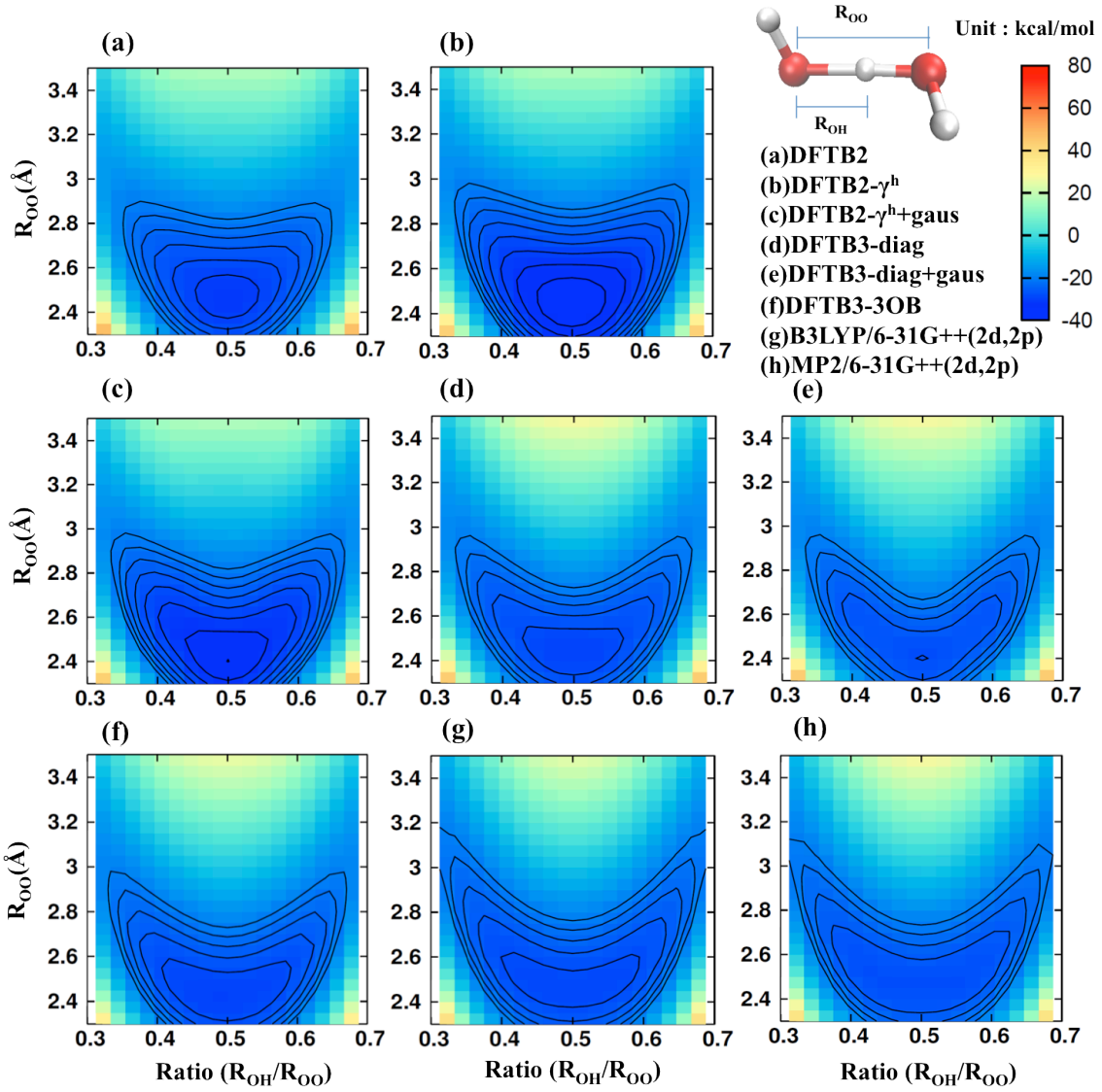


Figure 3.1 Contour plots of $H_3O_2^-$ calculated with various electronic structure methods: (a) DFTB2, (b) DFTB2-γ^h, (c) DFTB2-γ^h+gaus, (d) DFTB3-diag, and (e) DFTB3-diag+gaus, (f) DFTB3-3OB (g) B3LYP/6-31+G(2d,2p), and (h) MP2/6-31+G(2d,2p) levels of theory. The difference between two contour lines is 2 kcal/mol.

The shapes of the potential energy surfaces of the DFTB2 (a) and DFTB2-γ^h (b) methods are very similar, but the potential energy depth in DFTB2-γ^h is about 4 kcal/mol deeper than that in DFTB2. (Eight contour lines in DFTB2-γ^h (b), but only six lines in DFTB2 method (a)) These potential shapes are significantly different from those obtained from the B3LYP (f) or MP2 methods (g), which possess a much broader valley near the

minima located at about 0.45 and 0.55 Å regions. When the on-site third order correction is included (d and e), the potential shapes become more similar to those of high level electronic structure calculations. The potential energies including the O–H repulsive correction (c and e) indicate a change in the overall surface especially when the hydrogen is located between two oxygen atoms because this O–H repulsive potential was designed to avoid perturbing hydrogen-bonding energies in the barrier region. This correction provides a better description of the proton transfer barrier, but the total potential surface appears distorted when compared to high level calculations. The potential energy surface calculated with the 3OB parameter set (f) is similar to the DFTB3-diag potential, but has a broader curvature in the minimum potential area. Overall, the potential energy surface of the DFTB3- 3OB method shows the best contour shape of the $H_3O_2^-$ cluster among all the SCC-DFTB potentials compared to that of the MP2 calculation.

3.3.2 Hydroxide Water Clusters $OH^-(H_2O)_n$, $n = 4-7$.

The low-lying energy minima of $OH^-(H_2O)_n$, $n = 4-7$, was evaluated using basin hopping Monte Carlo simulations. Several low-lying energy minima were selected, then re-optimized with the MP2/6-31G++(2d,2p) level of theory. These structures are displayed in **Figure 3.2** for $OH^-(H_2O)_n$, n = 4–7, and ***Table 3.1*** shows the formation energies of these small hydroxide water clusters using several different electronic structure calculations. Contrasted to the protonated water clusters, most low-lying minima of the small hydroxide water clusters do not have Zundel-like structures. There are some Zundel-like structures in $OH^-(H_2O)_4$ in **Figure 3.2**, calculated with the DFTB2-

γ^h potential, but these structures (**Figure 3.2** 4C and 4D) collapsed back to Eigen-like structures. ***Table 3.1*** also indicates the lowest energy structure in each electronic structure method (i.e., bold numbers). All the lowest energy structures (4A, 5A, 6A, 7A) are the same in the modified SCC-DFTB methods (DFTB2-γ^h, DFTB2-γ^h +gaus, DFTB3-diag+gaus, and DFTB3-diag+gaus and DFTB3-3OB). However the original SCC-DFTB method (DFTB2) shows different lowest energy structures, 4E and 7C, for the $OH^-(H_2O)_4$ and $OH^-(H_2O)_7$ clusters, respectively. This observation indicates the hydrogen-bonding correction ($-\gamma^h$) is of significant magnitude to alter the energetics of low-lying energy structures for the small hydroxide water clusters.

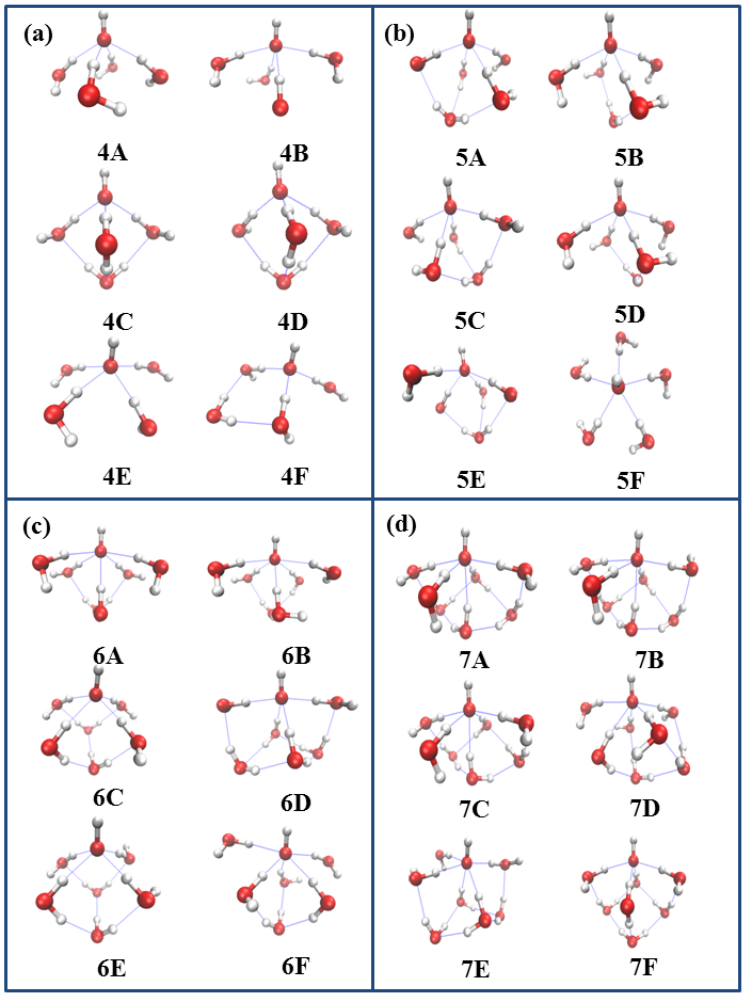


Figure 3.2 Structures of the low-lying energy minima of (a) $OH^-(H_2O)_4$, (b) $OH^-(H_2O)_5$, (c) $OH^-(H_2O)_6$, and (d) $OH^-(H_2O)_7$. Geometries are optimized with the MP2/6-31+G(2d,2p) level of theory.

Table 3.1 Formation Energies (kcal/mol) of several isomers of $OH^-(H_2O)_n$, for n = 4–7

$OH^-(H_2O)_4$	DFTB2	DFTB2-γ^h	DFTB2-γ^h +gaus	DFTB3-diag	DFTB3-diag+gaus	DFTB3-3OB	B3LYP/6-31G++(2d,2p)	MP2/6-31G++(2d,2p)
4A	-81.78	[b]**-95.22**	**-95.22**	**-95.14**	**-95.14**	**-91.13**	-84.42	-87.55
4B	-80.82	-94.78	-94.78	-94.41	-94.41	-90.09	-83.22	-85.16
4C	-80.58	-94.51	-94.40	-93.30	-93.30	-88.83	-85.12	**-87.85**
4D	-80.39	-94.27	-92.52	-91.86	-89.64	-88.32	-84.66	-87.30
4E	**-81.82**	-94.16	-94.14	-94.41	-94.41	-90.09	-83.22	-85.53
4F	-80.83	-93.54	-93.53	-89.08	-89.08	-86.62	**-85.16**	-86.48
rmsd[a]	1.28	1.05	1.43	2.68	3.20	2.16	0.98	0

$OH^-(H_2O)_5$	DFTB2	DFTB2-γ^h	DFTB2-γ^h +gaus	DFTB3-diag	DFTB3-diag+gaus	DFTB3-3OB	B3LYP/6-31G++(2d,2p)	MP2/6-31G++(2d,2p)
5A	**-94.93**	**-111.90**	**-111.90**	**-112.46**	**-112.46**	**-105.56**	**-99.01**	**-103.68**
5B	-94.55	-111.21	-111.21	-111.52	-111.52	-104.08	-97.60	-102.08
5C	-94.22	-110.90	-110.90	-110.98	-110.98	-104.68	-98.41	-102.98
5D	-94.56	-110.97	-110.97	-111.17	-111.17	-104.64	-98.36	-102.88
5E	-94.27	-110.43	-110.43	-110.48	-110.48	-103.01	-97.23	-101.45
5F	-94.76	-109.81	-109.81	-109.36	-109.36	-103.63	-94.97	-99.69
rmsd[a]	1.64	0.92	0.92	0.60	0.60	0.86	0.22	0

$OH^-(H_2O)_6$	DFTB2	DFTB2-γ^h	DFTB2-γ^h +gaus	DFTB3-diag	DFTB3-diag+gaus	DFTB3-3OB	B3LYP/6-31G++(2d,2p)	MP2/6-31G++(2d,2p)
6A	**-107.77**	**-127.50**	**-127.50**	**-129.41**	**-129.41**	**-119.97**	-110.34	-116.85
6B	-107.07	-126.51	-126.51	-128.37	-128.37	-119.26	-111.02	-116.08
6C	-107.19	-125.91	-125.88	-126.74	-126.74	-117.98	**-113.55**	**-119.52**

Table 3.1 Formation Energies (kcal/mol) of several isomers of $OH^-(H_2O)_n$, for n = 4–7 (Continued)

6D	-106.64	-125.94	-125.94	-127.51	-127.51	-117.78	-112.13	-118.01
6E	-106.66	-125.56	-125.54	-125.62	-125.62	-117.65	-113.18	-119.04
6F	-106.45	-125.32	-125.32	-126.53	-126.53	-117.37	-113.55	-114.69
rmsd[a]	1.98	2.67	2.68	3.52	3.52	2.98	2.31	0
$OH^-(H_2O)_7$	DFTB2	DFTB2-γ^h	DFTB2-γ^h +gaus	DFTB3-diag	DFTB3-diag+gaus	DFTB3-3OB	B3LYP/6-31G++(2d,2p)	MP2/6-31G++(2d,2p)
7A	-119.33	**-141.18**	**-141.18**	**-143.73**	**-143.73**	**-132.22**	-123.96	-131.67
7B	-119.12	-141.12	-141.12	-143.97	-143.97	-131.74	-123.16	-130.79
7C	**-119.36**	-141.11	-141.11	-143.73	-143.73	-132.14	-123.79	-131.57
7D	-118.93	-140.74	-140.74	-143.38	-143.38	-131.76	-125.37	-131.63
7E	-118.72	-140.36	-140.36	-143.06	-143.06	-131.72	-126.56	-130.51
7F	-118.17	-139.76	-138.03	-140.33	-138.20	-130.25	**-127.05**	**-134.13**
rmsd[a]	1.41	1.63	2.32	2.45	3.30	1.85	1.67	0

[a]rmsd given relative to the MP2 energies

[b]The bold number indicates the lowest energy structure in each method

The low-lying energy structures calculated with the DFTB methods were re-optimized with the B3LYP and MP2 level of theory. It was found that the minimum energy structures are changed for the $OH^-(H_2O)_n$ clusters, n = 4, 6, and 7. Only the $OH^-(H_2O)_5$ cluster shows the same minimum structure (5A) for all electronic structure calculations. ***Table 3.1*** also presents the root mean square deviations (rmsd) relative to the MP2 energies for all the clusters. Thus, for the relative energies of these small clusters, the use

of third order correction with either MIO or 3OB parameter set does not result in the best accuracy for the $OH^-(H_2O)_n$ cluster, n = 4, 6, and 7, compared to high level QM calculations. In fact, the incorporation of the third order term results in a less accurate method than the DFTB2 and DFTB2-γ^h methods. It is also found that the O–H repulsive correction (+gaus) typically does not change the energy or the structures in these particular cluster calculations because the modified O–H repulsive correction usually affects the structures in the barriers, not in the minima.

While the lowing-lying energy minimum structures of $OH^-(H_2O)_4$ accept three or four hydrogen-bonds (HBs), those of the $OH^-(H_2O)_n$, n = 5–7, clusters coordinate four or five HBs. Typically tetra-coordinated structures around OH^- are observed in the lowest-energy minima of the cluster ($n \geq 4$) when calculated with B3LYP and MP2 levels of theory. The SCC-DFTB method therefore tends to over coordinate the HBs around the OH^- anion when compared to high level electronic structure calculations.

3.3.3 Hydroxide bulk simulations.

3.3.3.1 Structural Properties

For the following discussions pertaining to the structural and dynamical properties of the hydrated hydroxide ion it is beneficial to describe the contributing molecules as outlined in **Figure 3.3**, which is a typical snapshot from a DFTB2-γ^h simulation. In **Figure 3.3** the oxygen and hydrogen of the central hydroxide are denoted as O_0 and H′, respectively. Due to the dynamical nature of the bonding topology and solvent environment, the first solvation shell has been included in the description of the central

hydroxide. The oxygen atoms on the surrounding waters are denoted in general as O_w. The nearest solvating waters are denoted as O_1, O_2, O_3 , …, O_n, as determined by the distance between O_0 and the surrounding water oxygens. In this section, the structural properties predicted by the various DFTB methods are compared with the structural properties predicted by the CPMD results (4, 5), which are consistent with known experimental data in terms of both structural and dynamical properties.

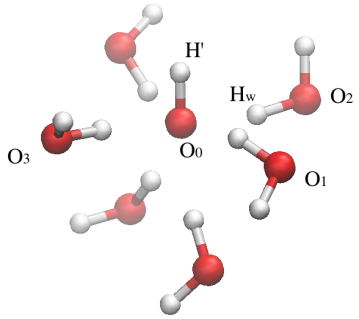


Figure 3.3 Solvation of the hydroxide anion in bulk water. The hydroxide oxygen and hydrogen are labeled as O_0, H′ respectively. The nearest three surrounding water oxygens are labeled as O_1, O_2, O_3. The water hydrogen is labeled as H_w.

The radial distribution function (RDF) between the hydroxide's oxygen (O_0) and the surrounding water oxygens (O) can be found in **Figure 3.4** (A). Evaluation of the first peak's location indicates that the various SCC-DFTB methods predict the first solvation shell to reside at $R_{O_0\text{-}O} = 2.68 \pm 0.01$ Å, larger than the CPMD result, which predicts $R_{O_0\text{-}O} = 2.6$ Å. This suggests that in bulk water, the SCC-DFTB methods overestimate the distance between hydroxide ion and its first solvation shell by nearly 0.1 Å. In addition, the first peak predicted by SCC-DFTB is much narrower than that in CPMD, which indicates that SCC-DFTB methods predict a more ordered first solvation shell around the hydroxide ion in bulk water as compared to CPMD.

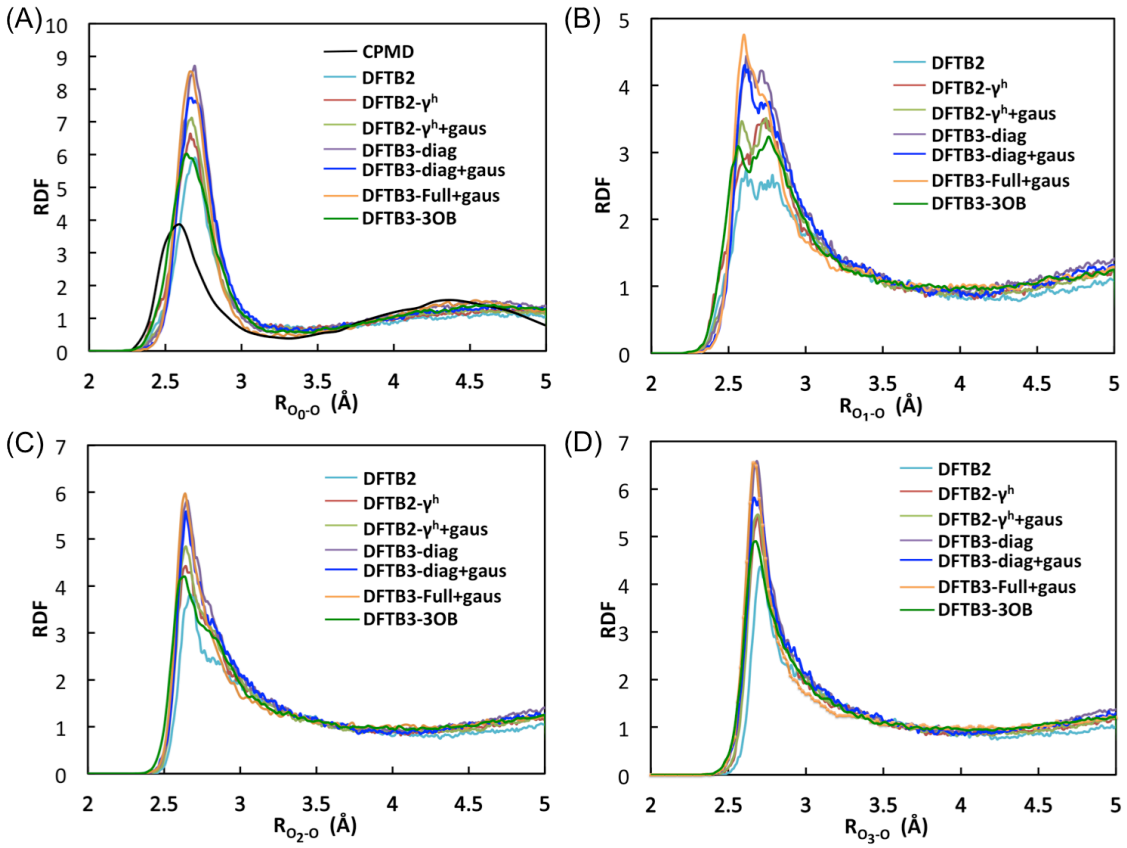


Figure 3.4 Radial distribution functions of the hydroxide anion complex for (A) $g_{O_0\text{-}O}(r)$, (B) $g_{O_1\text{-}O}(r)$, (C) $g_{O_2\text{-}O}(r)$ and (D) $g_{O_3\text{-}O}(r)$ (see **Figure 3.3** for definitions). Each panel has data for DFTB2 (light blue), DFTB2-γ^h (red), DFTB2-γ^h+gaus (light green), DFTB3-diag (purple), DFTB3-diag+gaus (deep blue), DFTB3-Full+gaus (orange) and DFTB3-3OB (deep green) simulations. For $g_{O_0\text{-}O}(r)$, CPMD simulation data is shown in black.

The oxygen coordination number of the central hydroxide, which indicates the number of solvating waters, can be determined by integrating the first peak of RDF of O_0-O for the SCC-DFTB and CPMD data:

$$n = 4\pi\rho \int_{r_1}^{r_2} r^2 g_{O_0\text{-}O}(r)dr \tag{6}$$

where n is the oxygen coordination number, ρ is the bulk density of water oxygen, r_1 and r_2 are the starting and ending radial distance of the first peak, respectively. The results of the first peak integration and first peak height are found in **Table 3.2**. From

Table 3.2, it is obvious that the coordination numbers predicted by the various SCC-DFTB methods are much larger than the CPMD result. This indicates that the SCC-DFTB methods have a tendency to oversolvate the hydroxide ion. The oversolvation of the hydrated hydroxide ion is similar to what the cluster calculations predict and have also been seen in both the evaluation of the hydrated excess proton and the bulk water characteristics as described by the SCC-DFTB method.(19)

Table 3.2 Coordination number of hydroxide and first peak height of radial distribution function of O_0-O.

Method	Coordination number	First peak height
CPMD[4,5]	4.8	4.5
DFTB2	6.62 ±0.03	5.7±0.2
DFTB2-γ^h	6.87 ±0.04	6.5±0.2
DFTB2-γ^h +gaus	6.94 ±0.05	6.9±0.3
DFTB3-diag	7.87 ±0.04	8.68±0.03
DFTB3-diag+gaus	7.82 ±0.07	7.78±0.07
DFTB3-full+gaus	7.36±0.03	8.6±0.1
DFTB3-3OB	7.0±0.1	6.07±0.05

The structural information in **Figure 3.4** and **Table 3.2** indicates that from DFTB2 to DFTB3-diag, the progressive addition of correction term to the SCC-DFTB description of the hydroxide ion has the major effect of increasing the height of the first peak as well as the coordination number in the first solvation shell for the hydroxide ion. It is found that the DFTB2-γ^h causes slight oversolvation of hydroxide with respect to the DFTB2, which

has the best agreement with CPMD data. The increase in the first peak height and coordination number caused by the γ^h function is in line with the increased depth of in the potential energy surface for $H_3O_2^-$ cluster, as already mentioned in section 3.3.1 (**Figure 3.1** (a) and (b)). The enhanced OH^--H_2O dimer interaction energy predicted by DFTB2-γ^h is responsible for a more structured first solvation shell compared to DFTB2. The addition of the O–H repulsive potential to DFTB2-γ^h causes a further slight increase in first peak height and coordination number. The introduction of third order correction terms while using the MIO parameter set has a major adverse impact on the solvation structure as compared to the CPMD simulations. For the three methods DFTB3-diag, DFTB3-diag+gaus and DFTB3-Full+gaus, the first peak height is greatly enhanced and the hydroxide accommodates an additional water molecule in the 1st solvation shell when compared to the DFTB2 method prediction. Additional changes in the RDF include a narrowing of the first peak and a more pronounced second peak, indicating more ordering in both the 1st and 2nd solvation shells. The increased over solvation and structuring in the 1st solvation shell are in line with the increase in the interaction energy of hydroxide clusters for the DFTB3 and DFTB3-diag+gaus methods as compared to the DFTB2 methods (***Table 3.1***). It is worth noting that the addition of the repulsive potential to DFTB3-diag method and the use of the full 3rd order correction do not result in significant improvement, but the use of the 3OB parameter set in the DFTB3-3OB method does generates a better result than the MIO parameter set. The use of the 3OB parameter set reduces the hydroxide oversolvation and makes the solvation shells less ordered, which is in agreement with the decrease in the interaction energy of hydroxide clusters for DFTB3-3OB compared to DFTB3-diag (***Table 3.1***). The reason for the

improvement may be that the 3OB parameter set is optimized for SCC-DFTB with full 3rd order expansion, but the MIO parameter set was developed for the original DFTB2. However, the DFTB3-3OB method still predicts results qualitatively different from CPMD. Interestingly, the newest state-of-art DFTB3-3OB method provides results with an accuracy only similar to the original DFTB2 methods.

Further analysis of the structural properties requires the inspection of the solvation environment for the first solvation shell around the hydroxide ion. The RDF for O_1-O (**Figure 3.4** (B)) indicates the presence of a special pair, like that seen for the hydrated excess proton, (19) although to a much smaller extent. In **Figure 3.4** (B) it is clear that the first peak is more pronounced in the third order methods with MIO parameter set than in the second order method, indicating that with a more ordered solvation environment the presence of a special pair becomes more obvious and localized. The first subpeak in the first main peak is more dominant in DFTB3 with MIO parameter set than the DFTB2 methods, which indicates that the O_1 is more localized (i.e. held tighter to O_0). Similar to the previously reported trends, the DFTB3-3OB method exhibits an improvement over the DFTB3 methods using the MIO parameter set, providing results similar to the DFTB2 methods.

In the RDF for O_2-O and O_3-O (**Figure 3.4** (C) and (D)), the first peaks are located at 2.65±0.01 Å and 2.69±0.02Å respectively, indicating that there is degeneracy in the distance and preference for a particular water in this solvation shell. The first peak's height and coordination number are listed in **Table 3.3** and **Table 3.4** for O_2-O and O_3-O respectively, indicating an oversolvation behavior. The peak position and starting point of the first peak is observed to shift slightly to the left as one changes the method from

DFTB2 to DFTB3-full+gaus. The peak's height and coordination number reflect the same trend in the RDF of O_0-O and O_1-O, revealing an oversolvation and ordering of the waters. The DFTB3 methods using MIO parameter set greatly enhanced the height of the first peak as well as the coordination number compared to DFTB2 methods. These structural changes are due to the overall increased ordering in the solvation structure and number of solvating waters. The DFTB3-3OB reduces these problems and leads to results comparable to DFTB2 methods. It is also obvious that all the SCC-DFTB methods predict a very broad density depletion region between the 1st and 2nd solvation shell and the depletion is weak in magnitude. This has the impact of causing abnormally large coordination numbers (**Table 3.4**) due to the fuzzy cutoff region between solvation shells. The same structural feature was observed in our previous paper for hydrated excess proton system.(19)

Table 3.3 Coordination number of O_2 and first peak height of radial distribution function of O_2-O.

Method	Coordination number	First peak height
DFTB2	9.36±0.08	3.7±0.1
DFTB2-γ^h	10.15±0.01	4.48±0.08
DFTB2-γ^h +gaus	10.14±0.02	4.8±0.1
DFTB3-diag	11.12±0.03	5.83±0.02
DFTB3-diag+gaus	10.9±0.1	5.5±0.1
DFTB3-full+gaus	10.5±0.1	5.99±0.02
DFTB3-3OB	10.31±0.06	4.19±0.05

Table 3.4 Coordination number of O_3 and first peak height of radial distribution function of O_3-O.

Method	Coordination number	First peak height
DFTB2	9.40±0.08	4.3±0.1
DFTB2-γ^h	10.11±0.06	5.418±0.002
DFTB2-γ^h +gaus	10.14±0.06	5.3±0.2
DFTB3-diag	11.10±0.08	6.57±0.01
DFTB3-diag+gaus	10.94±0.08	5.9±0.1
DFTB3-full+gaus	10.5±0.1	6.6±0.1
DFTB3-3OB	10.26±0.05	4.90±0.01

To gain more insight into the detailed structure of the first solvation shell of the hydroxide ion, the partial RDF's with reference to O_0 for various DFTB methods are plotted in **Figure 3.5**. It is observed that the DFTB3 methods with the MIO parameter set

lead to the narrowing of the partial RDF, which is in agreement with the narrowing of the first peak in O_0-O RDF. Looking at the relative shapes of the subpeaks one can see that as the peaks move farther away from the hydroxide they increase in width. This is indicative of waters that associate less favorable than closer waters (i.e., reduced interaction energy). However, this trend is not pronounced for the DFTB3 methods with the MIO parameter set, again indicating the stronger interaction between the hydroxide ion and the surrounding waters. Closer inspection of **Figure 3.5** reveals that those DFTB3 method's subpeaks are packed closer to each other than DFTB2 methods, indicating a tighter and closer solvation shell with reduced movements and fluctuations. These observations are in agreement with our earlier conclusions that the DFTB3 methods with the MIO parameter set have a more ordered structure in the first solvation shell. The DFTB3-3OB method leads to a less structured first solvation shell and provides similar results to the DFTB2 methods.

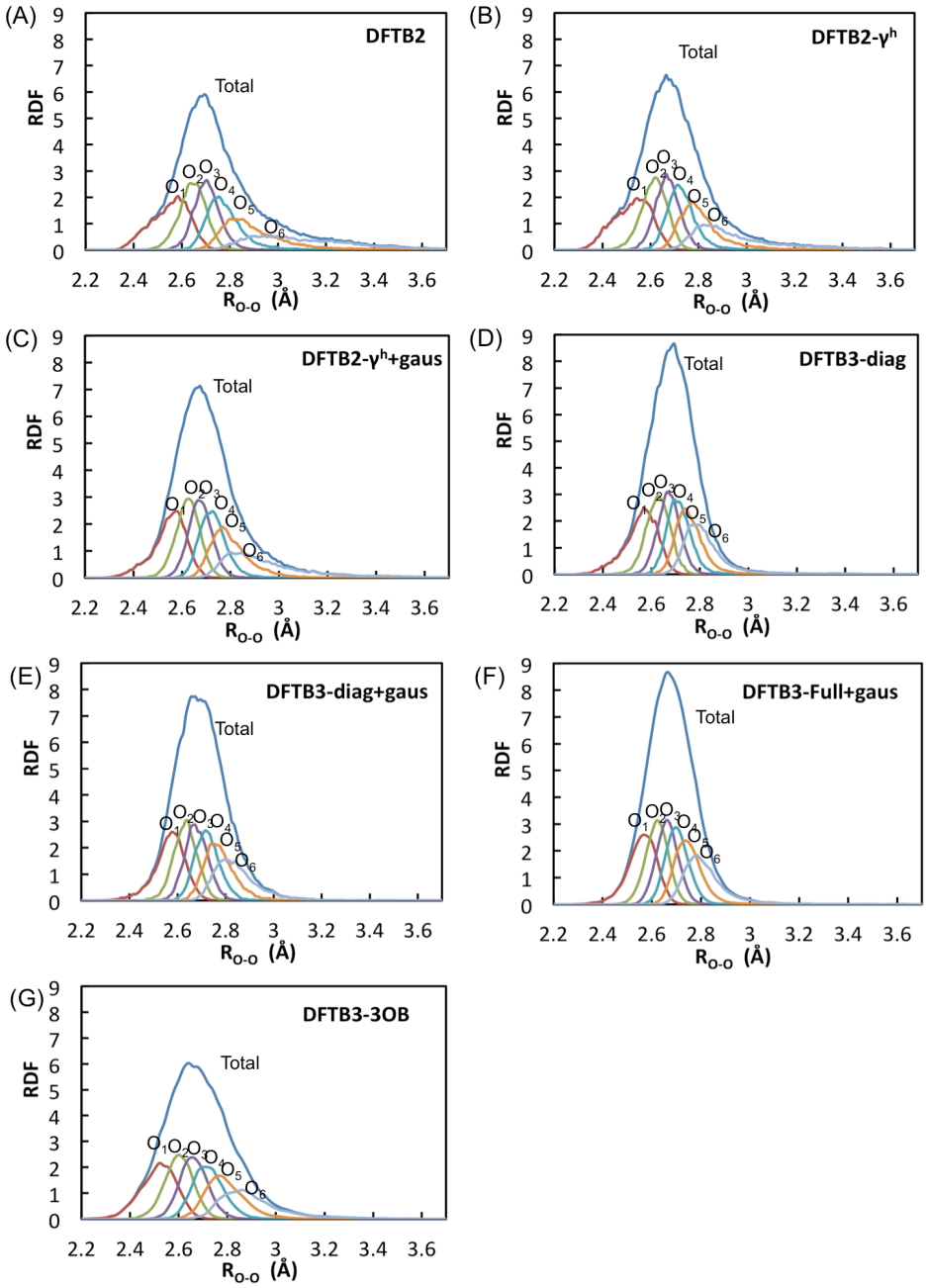


Figure 3.5 Partial radial distribution function for the six nearest waters to the hydroxide for (A) DFTB2, (B) DFTB2-γ^h, (C) DFTB2-γ^h+gaus, (D) DFTB3-diag, (E) DFTB3-diag+gaus, (F) DFTB3-Full+gaus and (G) DFTB3-3OB simulations. The notation for "O_1", "O_2", "O_3" … etc stands for partial RDF of O_0-O_1, O_0-O_2, O_0-O_3 … etc.

Overall, the variety of hydroxide oxygen-oxygen RDF's indicate that the DFTB3 methods with the MIO parameter set make the solvation structure more ordered and densely packed in regions near the hydroxide ion, which may exacerbate the observed

"water voids" issue (see section 3.3.3.5). It is found that the DFTB3 methods with the MIO parameter set generate results more at odds with those of the CPMD calculation than the results generated with the DFTB2 methods, while the DFTB3-3OB method is found to be a modest improvement over the MIO parameter set.

In addition to analyzing the oxygen-oxygen structures it is also beneficial to inspect the hydrogen-oxygen structures and the resulting hydrogen bonding network. The RDF of H′-O_w are shown in **Figure 3.6** (A). The δ in **Figure 3.6** (A) is a PT reaction coordinate defined(5) as: $|\delta| = \min|R_{O_0H} - R_{O_aH}|$, where O_a is any possible oxygen atom that is involved in a hydrogen bond with the hydroxide, and R_{O_0H} and R_{O_aH} are the distances between the shared hydrogen and the two oxygens. A small δ correspond to the transition state of the PT reaction, while a large δ correspond to the system residing in the reactant or product well (i.e., resting state). The first peak's height from the DFTB3 methods is near 2 and is more pronounced when compared to those of the CPMD methods, which is below 0.8 for both the resting and transition states. Integration over the first peak in the RDF of H′-O_w for DFTB3-diag, DFTB3-diag+gaus, DFTB3-full-gaus and DFTB3-3OB gives hydrogen coordination number of 0.93±0.02, 0.86±0.01 and 0.93±0.03, 0.83±0.02, respectively, larger than the value of 0.72 predicted by CPMD simulation.(4) It is also observed that the positions of the RDF's first peaks for the DFTB3 methods are shifted to smaller separation distances as compared to the DFTB2 methods. This indicates that the DFTB3 methods strengthen the hydrogen bond donated by the hydroxide to the nearby water oxygen, and for the majority of the simulation time the hydroxide hydrogen is hydrogen bonded to a water molecule's oxygen, which is in sharp contrast with CPMD conclusions.(4, 5) This hydrogen accepting water molecule

accounts for the additional water molecule solvating the hydroxide as compared to the DFTB2 methods. The DFTB3-3OB method slightly reduces the coordination number and the peak height, but is still qualitatively different from DFTB2 and CPMD results, suggesting that the 3OB parameter set does not significantly improve the hydrogen bonding interaction between water and the hydroxide ion.

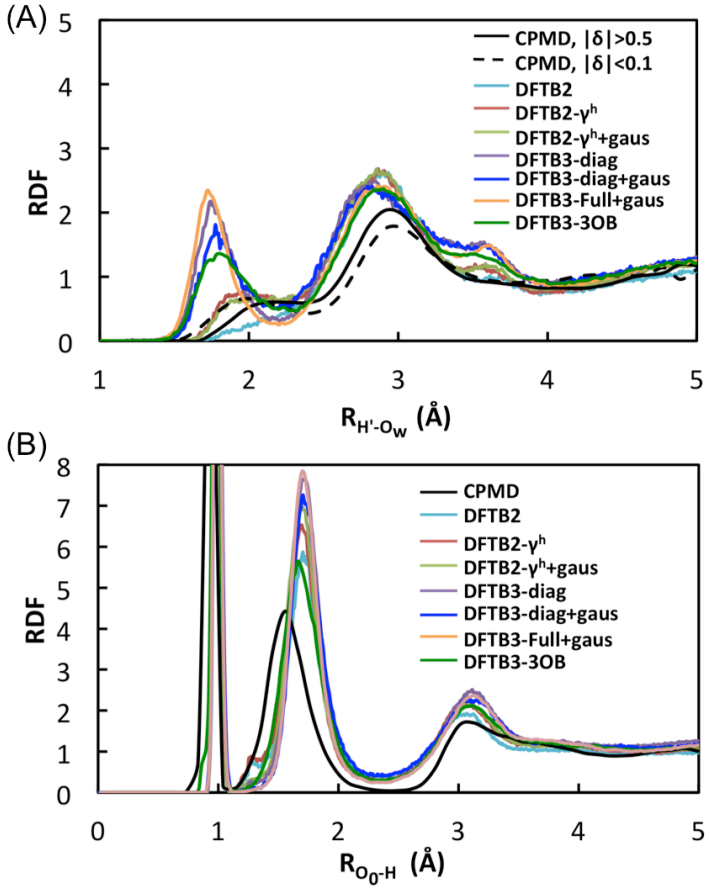


Figure 3.6 Radial distribution functions of the hydroxide anion complex for (A) $g_{H'-O_w}(r)$ and (B) $g_{O_0-H}(r)$. Each panel has data for DFTB2 (light blue), DFTB2-γh (red), DFTB2-γh+gaus (light green), DFTB3-diag (purple), DFTB3-diag+gaus (deep blue), DFTB3-Full+gaus (orange) and DFTB3-3OB (deep green) simulations. CPMD simulation data is shown in black.

The RDF for the central oxygen, O_0-H, is shown in **Figure 3.6** (B), and the integration of the second peak is the hydrogen coordination number for the hydroxide oxygen as depicted in **Table 3.5**. The second peaks for the various SCC-DFTB methods reside at 1.70 ± 0.02 Å, whereas the CPMD simulations predict 1.6 Å, indicating a slight elongation of the oxygen hydrogen distance for the SCC-DFTB methods. This suggests that the SCC-DFTB methods predict a larger donating hydrogen bond length, and is in line with the observation from the RDF of O_0-O_w that reveals the distance between the hydroxide ion and its first solvation shell to be about 0.1 Å larger in SCC-DFTB than in CPMD. The use of the 3OB parameter set does not result in an appreciable change in the location of the 2nd peak, which again suggests that it fails to significantly improve the description of hydrogen bonds between water and hydroxide. Moreover, the second peak width in RDF of O_0-H in SCC-DFTB methods is narrower than that in CPMD simulation, again suggesting a more ordered solvation shell. In addition, the SCC-DFTB methods appear to have an overall tendency to donate more hydrogen bonds to the central hydroxide oxygen, and as observed from the hydrogen coordination number, the hydroxide does not accept more hydrogen bonds from the first solvation shell waters in the DFTB3 methods as compared to DFTB2 methods. This is in agreement with the observation from the RDF of H′-O_w, which indicates that the additional solvating water in the DFTB3 methods is accepting hydrogen bonds from the hydroxide rather than donating hydrogen bonds. Although the hydrogen coordination numbers are similar for the various SCC-DFTB methods, it is observed that the second peak's height increases significantly in going from DFTB2 methods to DFTB3 methods with MIO parameter set, again indicating an increased ordering and localization of atoms around the hydroxide ion.

The DFTB3-3OB again reduces the degree of over structuring, providing similar results to the DFTB2 methods.

Table 3.5 Coordination number of hydrogen and second peak height of radial distribution function of O_0-H.

Method	Coordination number	Second peak height
CPMD[4, 5]	4.0	4.5
DFTB2	5.02±0.06	5.7±0.2
DFTB2-γ^h	5.35±0.07	6.4±0.1
DFTB2-γ^h +gaus	5.47±0.08	6.8±0.3
DFTB3-diag	5.98±0.05	7.7±0.1
DFTB3-diag+gaus	5.8±0.1	7.1±0.2
DFTB3-full+gaus	5.60±0.01	7.9±0.1
DFTB3-3OB	5.26±0.04	5.6±0.1

In addition, from the comparison between DFTB2-γ^h vs. DFTB2-γ^h+gaus, and DFTB3-diag vs. DFTB3-diag+gaus, we can see the shoulders in the second peak are reduced when the O–H repulsive potential is added. This indicates that the O–H repulsive potential creates a density depletion area that is unfavorable for the surrounding waters hydrogen atoms to enter. The absence of hydrogen density in the shoulder area is in line with the effect of the O–H repulsive potential in the region from 1~1.5 Å. It is important to note that the absence of the shoulder region in the second peak is also seen in CPMD simulations.(4, 5)

A closer examination of **Figure 3.3**, a typical snapshot from the DFTB2-γ^h method, provides useful clues as to the origin of the oversolvation of the hydroxide ion. In the dynamical hypercoordination scenario of the CPMD simulations,(4, 5) the hydroxide accepts four hydrogen bonds from four surrounding waters in a nearly square planar arrangement (Figure 1e in ref (4)). However, in the DFTB2-γ^h method, in addition to the 4 hydrogen-donating waters arranged in a roughly planar structure, there are two more waters donating hydrogen bonds from below the square planar structural plane. These two waters account for the oversolvation trend of the hydroxide ion by the SCC-DFTB methods. The increased number of hydrogen bonds in the DFTB methods compared to CPMD method is also in agreement with the $OH^-(H_2O)_n$ cluster simulation result mentioned in section 3.3.2 and the conclusions from the RDF of O_0-H. As will be discussed in the next sections, this oversolvation causes incorrect energetic and dynamical properties of the hydroxide hopping.

3.3.3.2 Energetic Properties

In **Figure 3.7**, the potential of mean force (PMF) for the transfer of a proton along the reaction coordinate δ, described above, in various DFTB methods are depicted. The PMF is calculated as

$$F = -k_B T \ln P(\delta) \tag{7}$$

where $P(\delta)$ is the probability that a certain value of δ is observed during the MD run.

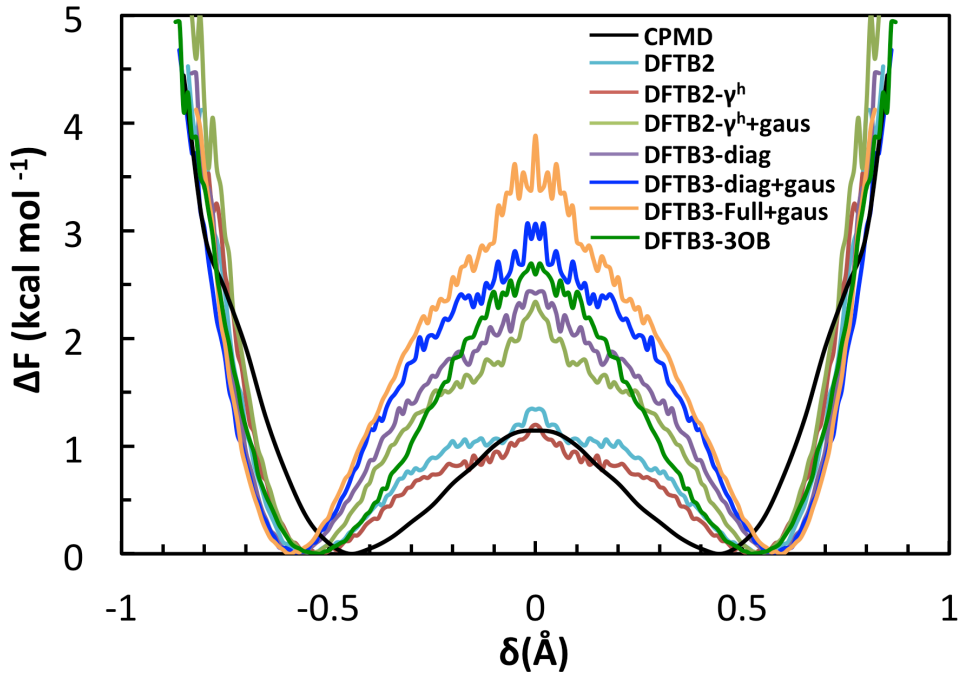


Figure 3.7 Potential of mean force for the hydroxide for proton transport in $H_3O_2^-$ for CPMD (black), DFTB2 (light blue), DFTB2-γ^h (red), DFTB2-γ^h+gaus (light green), DFTB3-diag (purple), DFTB3-diag+gaus (deep blue), DFTB3-Full+gaus (orange) and DFTB3-3OB (deep green) simulations. The potential of mean force is represented as a function of the δ coordinate described in the text.

From **Figure 3.7**, the following observations are made:

1. All the SCC-DFTB methods predict a larger minima δ value for the PMF. The larger minima δ value predicted by SCC-DFTB can be explained by the larger separation between hydroxide ion and its surrounding water molecules in SCC-DFTB compared to CPMD simulations. This result is in line with the conclusions drawn from the RDFs for O_0-O_w and O_0-H.

2. The DFTB2 and DFTB2-γ^h methods give similar PT energy barriers compared to the CPMD value, $\Delta E \approx 1.15$kcal/mol .(4, 5) However, the addition of the on-site 3rd

order correction raises hydroxide hopping barrier by at least 1 kcal/mol, and the addition of the full 3rd order correction raises the barrier even more. This behavior indicates that the PT will be less pronounced in the DFTB3 results than in the DFTB2 results. The reduction in the hydroxide transport is caused by the extreme oversolvation of the hydroxide and the large energy required to reorganize the hydrogen bond network topology during PT. The DFTB3-3OB method improves the results over the MIO parameter set and reduces the barrier to 2.5 kcal/mol, but it is still higher than the DFTB2 methods, which are in better agreement with the CPMD results.

3. The addition of the O–H repulsive potential raises the barrier height for hydroxide transfer thereby favoring the resting state (i.e., product and or reactant). The hydroxide hopping barrier is raised by about 1 kcal/mol from DFTB2-γ^h to DFTB2-γ^h+gaus and by about 0.5 kcal/mol from DFTB3-diag to DFTB3-diag+gaus. These observations indicate that the transition state for hydroxide hopping is disfavored in the DFTB2-γ^h+gaus and DFTB3-diag+gaus results. This is due to the fact that when the donating proton reaches the hopping region where $|\delta|$ approaches zero, the proton is interacting with the Gaussian repulsive potential in an unfavorable fashion. Therefore, the "reverse" Grotthuss hopping is energetically disfavored. The increased hopping barrier caused by the O–H repulsive potential is also in accordance with the recent report from Goyal et al. (ref (20)) for the hydrated proton in water. The incorrect free energy profiles predicted by SCC-DFTB methods result in incorrect hydroxide hopping dynamical properties, which we discuss below.

3.3.3.3 Dynamical Properties

For each SCC-DFTB version, the hydroxide oxygen's identity was followed during the MD simulation to gain insight into the mechanism for hydroxide transport. The results of the hydroxide identity are plotted in **Figure 3.8**. The DFTB2 and DFTB2-γ^h methods give similar frequency for identity change of O_0 as that predicted in the dynamical hypercoordination scheme in previous CPMD simulations. (4, 5) The thick line region for the DFTB2 methods is slightly more pronounced than that found for the CPMD simulations, indicating a longer oscillatory shuttling, where the proton is transferred back and forth between the same two oxygen atoms. It is clear from **Figure 3.8** (C) and (E) that the O–H repulsive potential has a deleterious effect on the hopping rate. As mentioned in the free energy analysis, this is because the repulsive potential tends to disfavor the hydrogen located at the midpoint of two sharing oxygens. The overall result is a significant reduction in the number of successful hopping events and a resulting model that favors vehicular (standard diffusive) transport over the Grotthus type bond rearrangement-process.

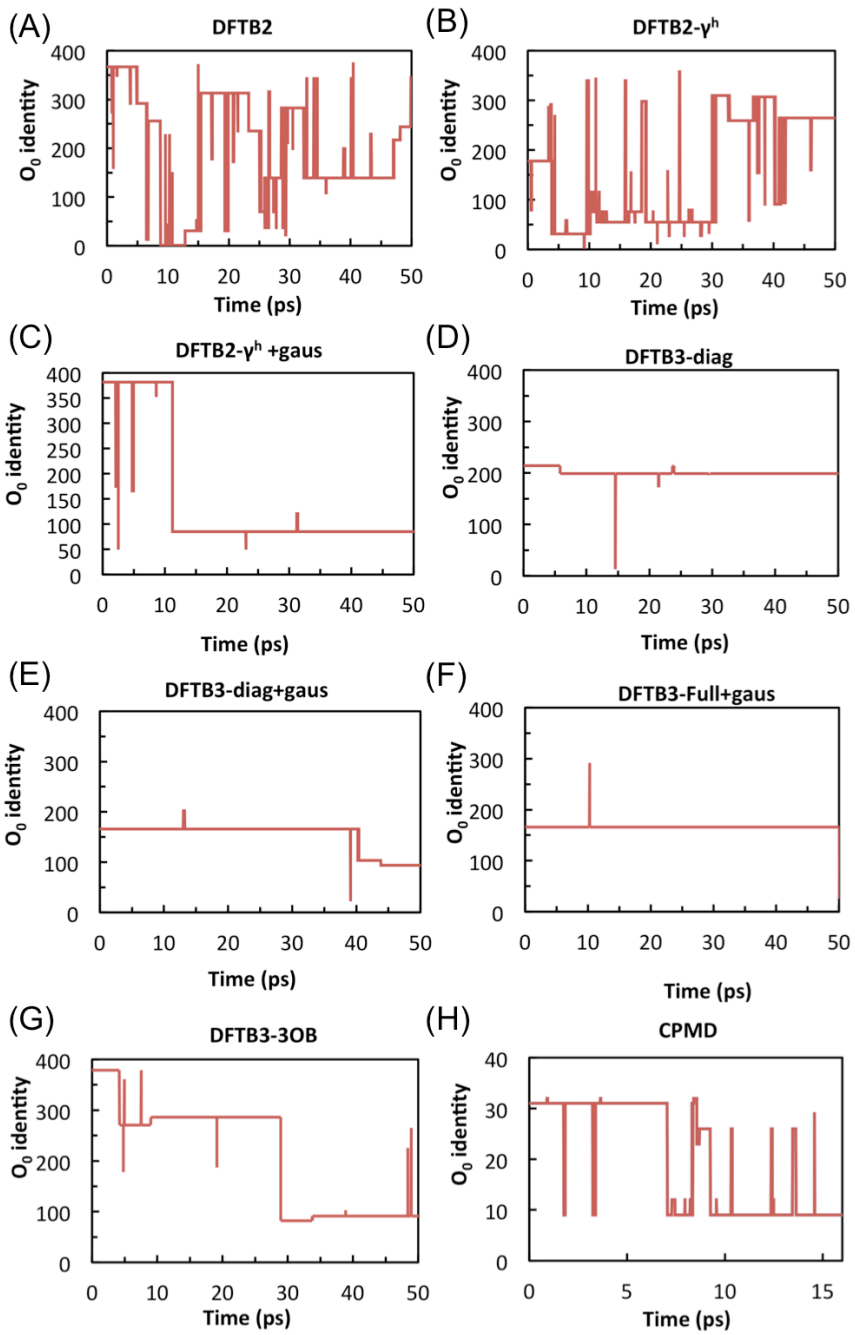


Figure 3.8 Time dependence of the identity of O_0 for the (A) DFTB2, (B) DFTB2-γ^h, (C) DFTB2-γ^h+gaus, (D) DFTB3-diag, (E) DFTB3-diag+gaus, (F) DFTB3-Full+gaus, (G) DFTB3-3OB and (H) CPMD simulations. The data depicted here is for one of the two independent trajectories. The reader should note the differences in scale of the horizontal axis between the various panels of the figure.

Evaluation of the DFTB3 methods with the MIO parameter set clearly reveals a greatly suppressed Grotthuss hopping mechanism, almost to the extent of eliminating Grotthuss hopping and utilizing only vehicular transport. The favoring of the vehicular transport is due to the oversolvated hydroxide and the densely packed first solvation shell. In order for a successful hopping to occur, the hydrogen bonding topology of the proton donating water and proton accepting hydroxide must undergo large structural changes, which are unfavorable. It is also clear that PT is further hindered by the O–H repulsive potential in the third order method with the MIO parameter set, which effectively eliminates all Grotthuss PT mechanisms. Thus in DFTB3-diag+gaus and DFTB3-full+gaus simulations, there are essentially only vehicular hydroxide transport.

The Grotthuss shuttling rate is calculated by the accumulation functions:

$$h(\Delta t) = h(\Delta t - 1) + \Delta h(\Delta t) \qquad (7)$$
$$h(0) = 0$$

where Δt is the time step and $\Delta h(\Delta t)$ is evaluated as follows:

$$\Delta h(\Delta t) = \qquad (8)$$
$$\begin{cases} 0, \text{ if no proton hop } (O_0 \text{ identity remains unchanged}) \\ 1, \text{ if the proton hops to a new donor} \\ -1, \text{ if the proton hops to the last donor} \end{cases}$$

In **Figure 3.9**, the Grotthuss shuttling rate for various DFTB methods is depicted. (The difference in the horizontal scale between the various figure panels should be noted.) The DFTB2 and DFTB2-γ^h give similar result for proton hopping as the CPMD simulation. It is observed that there are long intervals of oscillatory shuttling. This makes the $h(t)$ fluctuate around a value rather than increase successively as time passes. The O–H

repulsive potential makes the hopping less frequent and extends the period of oscillatory shuttling. The third order method greatly hinders the hopping mechanism as expected, further suppressing hopping in the DFTB3-diag+gaus and DFTB3-full+gaus methods and resulting in essentially no successful Grotthuss hopping for the 50 ps trajectory. The DFTB3-3OB method improves the results over the MIO parameter set and predicts more frequent Grotthuss hopping events. This is in line with the lowered free energy barrier and may be attributed to the reduction in the coordination number of the OH^- ion, but the hopping frequency is still much lower than that determined from the CPMD method.

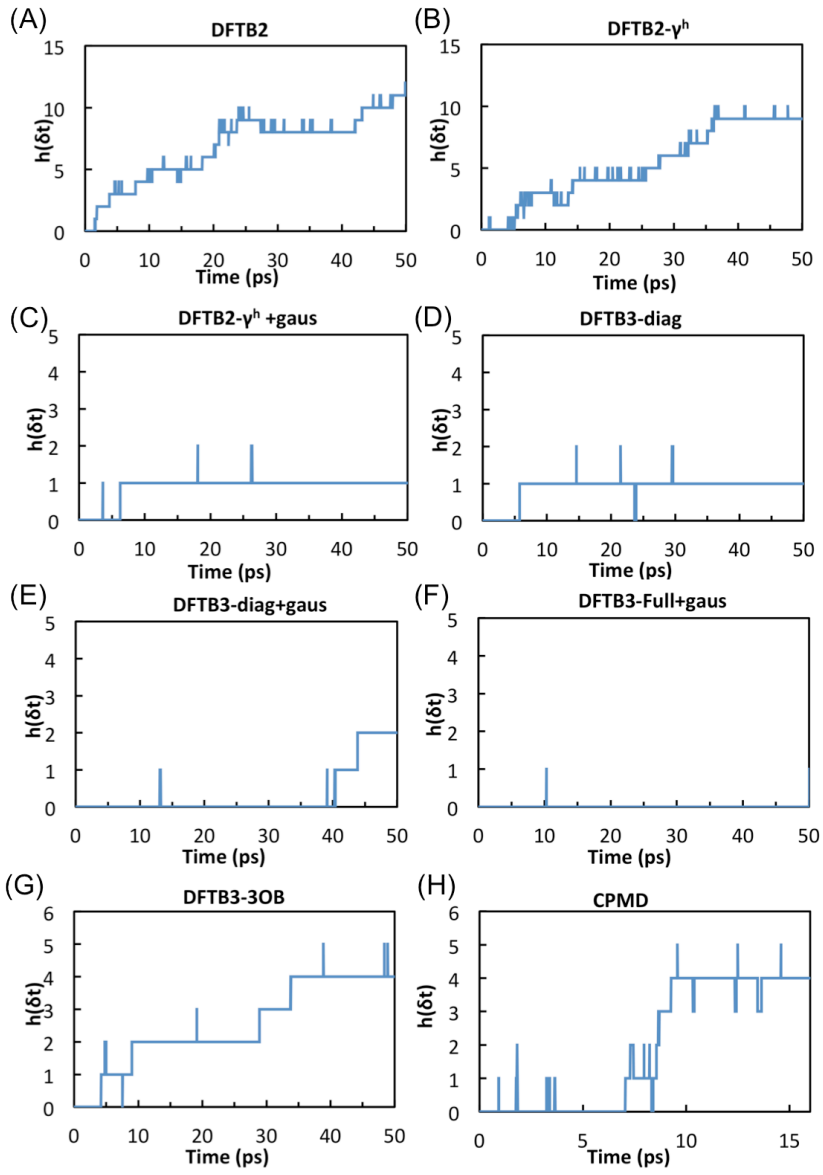


Figure 3.9 Forward proton hopping function calculated from (A) DFTB2, (B) DFTB2-γ^h, (C) DFTB2-γ^h+gaus, (D) DFTB3-diag, (E) DFTB3-diag+gaus, (F) DFTB3-Full+gaus, (G) DFTB3-3OB and (H) CPMD simulations. The data depicted here is for one of the two independent trajectories. The reader should note the differences in scale of the horizontal axis between the various panels of the figure.

3.3.3.4 DFTB mechanism for hydroxide ion transport

A more detailed analysis on the DFTB hydroxide transport mechanism indicates that the DFTB methods captures some features of the dynamical coordination mechanism, (4, 5) but fail to reproduce the entire picture. Here, the DFTB-3OB is taken as an example to

illustrate the similarities and differences. **Figure 3.12** (A) shows the probability distribution for the number of HBs accepted by O_0, noted as P(n*). P(n*) is largest at 5 and 6, with a small value at 4, which is in agreement with the hydroxide ion oversolvation issue we discussed in the Section 3.3.3.1 . The change in n* near the PT event is analyzed in more detail in **Figure 3.12** (B), by plotting n* as a function of the PT reaction coordinate $|\delta|$, denoted as $n^*(|\delta|)$. It is observed that $n^*(|\delta|)$ is at its maximum value far from the PT event (large $|\delta|$ value), and decreases near the PT event (small $|\delta|$ value). This means the hydroxide ion becomes less solvated near the transition state, and this could be a driving force for the hydroxide transport in the DFTB mechanism. The PMFs for hydroxide ions with 4, 5 and 6-coordinated waters is shown in **Figure 3.12** (C) further confirm the hypothesis that a decrease in coordination number drives the PT event. It is shown that the 4-coordinated hydroxide ions have a lower PT energy barrier than the 5-coordinated systems, which have a lower barrier compared to the 6-coordinated systems. For the 5 and 6-coordinated hydroxide ions, the energy barriers are larger than 2.5 kcal/mol and are unlikely to undergo PT events from thermal fluctuations. In fact, even if PT events do occur, they will not lead to a new stabilized $OH^-[H_2O]_5$ or $OH^-[H_2O]_6$ complex, and the proton will be most likely return to the original water oxygen in a short amount of time. In contrast, the 4-coordinated hydroxide ion has sufficiently low barrier for the PT.

However, one important feature of the DFTB mechanism that is qualitatively different from the dynamical hypercoordination mechanism is shown in the RDF of H'-Ow, **Figure 3.12** (D). The H'-Ow RDF reveals that for configurations both far and close to the PT event ($|\delta| > 0.5$ and $|\delta| < 0.1$, respectively), that the first peak is always pronounced.

This indicates that the HB donated by the hydroxide ion is always present, and is qualitatively different from the dynamical hypercoordination mechanism, which predicts that the hydroxide ion only donates a HB close to the PT event (small $|\delta|$). It is interesting to see that this aspect of the DFTB mechanism is similar to that of the "traditional mirror image" mechanism, which also predicts a constantly donated HB by the hydroxide ion. However, "traditional mirror image" also differs from the DFTB mechanism in that it does not predict a reduction in the O_0 coordination number near the PT event.(4, 5) Therefore, the DFTB mechanism cannot be interpreted as either the "dynamical hypercoordination" or the "traditional mirror image" mechanism. It combines elements from these two qualitatively different mechanisms and forms a hydroxide transport mechanism of its own kind.

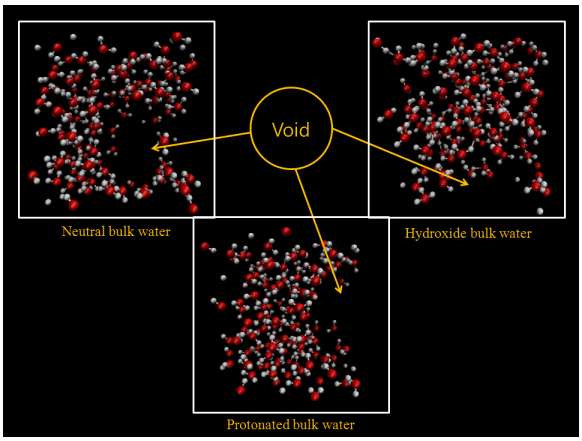


Figure 3.10 The snap shots of the neutral, hydroxide, and protonated bulk system after 50 ps of the MD time evolution using the DFTB3-diag+gaus method. The voids are detected in all simulations.

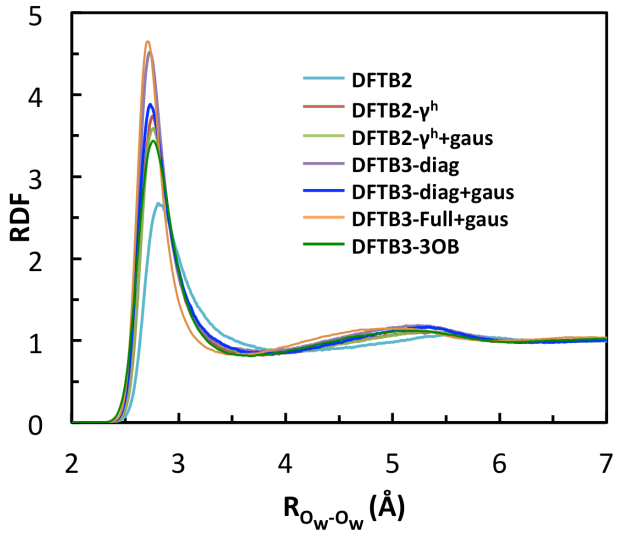


Figure 3.11 Radial distribution function for $g_{O_w\text{-}O_w}(r)$ in hydrated hydroxide system. Data shown here includes DFTB2 (light blue), DFTB2-γ^h (red), DFTB2-γ^h+gaus (light green), DFTB3-diag (purple), DFTB3-diag+gaus (deep blue), DFTB3-Full+gaus (orange) and DFTB3-3OB (deep green) simulations.

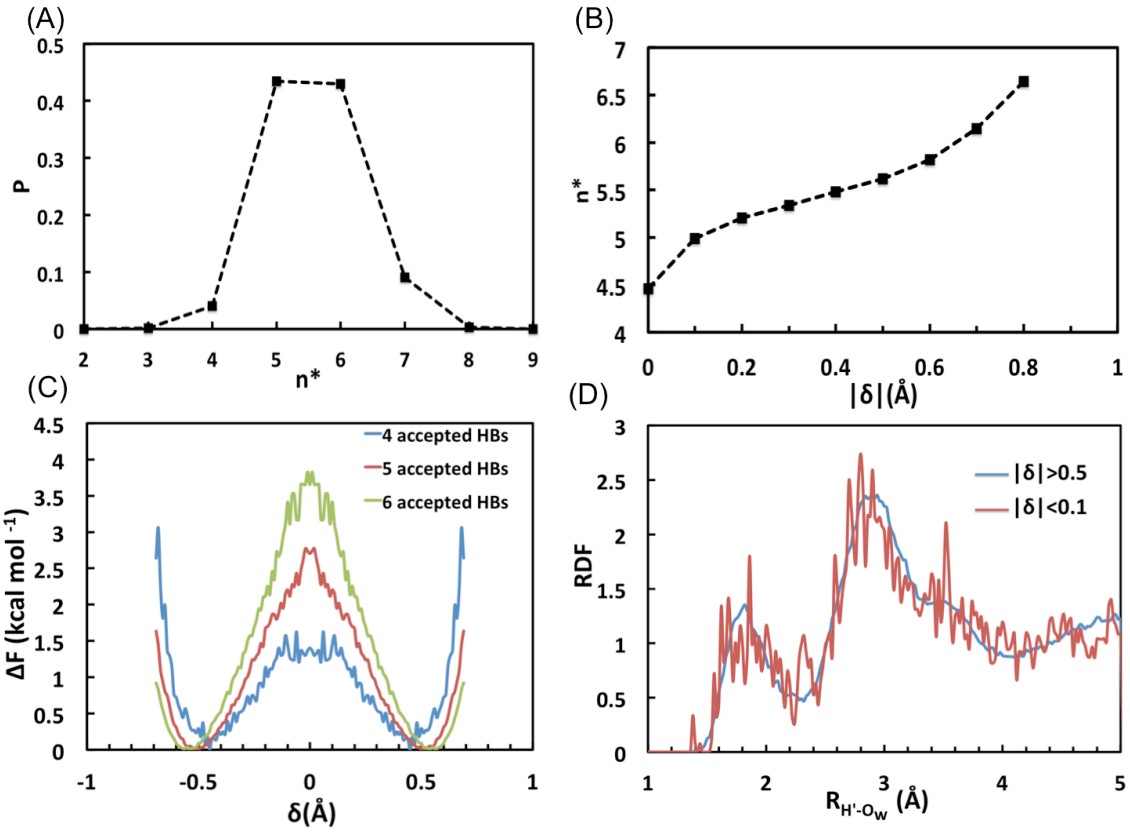


Figure 3.12 Results revealing hydroxide transport mechanism predicted by DFTB3-3OB method. (A) Probability distribution of number of hydrogen bonds accepted by O_0 (B) Number of hydrogen bonds accepted O_0 as a function of |δ| (C) Proton transport reaction free energy profiles for different number of hydrogen bonds accepted by O_0 (D) Radial distribution function for $g_{H'\text{-}O_w}(r)$ at both large and small |δ|.

3.3.3.5 Formation of water voids in the simulation.

In all the SCC-DFTB methods, pockets of voids with no or very little water density were observed in the MD simulations. The formation of the voids was seen to occur in simulations of the hydrated excess proton, the hydrated hydroxide, and bulk water simulations with no ions. **Figure 3.10** shows the snapshots of the bulk water, hydrated hydroxide, and hydrated excess proton system after 50 ps of NVT molecular dynamics simulations at a density of 0.97g/ml using the DFTB3-diag+gaus method. It is noted that these voids also form in simulations in the constant NVE ensemble. The presences of

significant voids were detected in all of the simulated systems, and these voids are found to fluctuate in size and position. This phenomenon is observed to occur primarily due to the over-coordinated water molecules. Due to the presence of the voids in the simulation cell it is clear that not all portions of the system are completely solvated or solvated in a homogeneous fashion.

For all the SCC-DFTB methods, the water coordination number calculated from the O_w-O_w RDF (**Figure 3.11**) is larger by at least 1 compared to that calculated from the CPMD simulations using the HCTH120 functional,(19) which is in line with the results from Table 6 in ref (20) and results in ref (19). This indicates an increased water density in the first solvation shell that causes a density heterogeneity in the simulation box that was pre-equilibrated using the TIP3P water model at ambient pressure and temperature. Comparing DFTB2-γ^h with DFTB2, it is observed that in the DFTB2-γ^h method, the first and second peaks are more pronounced and their locations are shifted to the left (i.e., smaller separation distance). Also, the first peak is narrowed, indicating an increased localization. This observation is in agreement with ref (19) and implies that the water molecules are closer to each other and become more ordered when the hydrogen-bonding correction is added. Combining the effect of the overcoordination of water molecules with the shortened first peak distance in the O_w-O_w RDF introduced by the γ^h function, the density is not likely to be physical and water void formation is almost unavoidable during constant volume simulation. It therefore seems likely that a SCC-DFTB simulation at constant NPT would reduce the simulation box volume and result in a water density significantly higher than reality.

3.4 Conclusions

Presented here is a comparison of various SCC-DFTB methods and their predicted structural and dynamical properties for the hydrated hydroxide ion. The SCC-DFTB methods have been applied to small sized hydroxide water clusters, $OH^-(H_2O)_n$, n = 4–7, and to simulations of the hydroxide ion in bulk water. Special attention has been given to newer SCC-DFTB versions, including the on-site and full third order term in the energy expansions based on the reference density, the modified O–H repulsive potential and the newest 3OB parameter set. These versions of the methodology are implemented because they have been shown to modestly improve the description of the hydrated excess proton in both gas-phase clusters and bulk water.(20) Interestingly, the results indicate that the original DFTB2 method and its hydrogen-bonding corrected variant more closely reproduce CPMD results while the incorporation of the O–H repulsive potential and the use of the 3rd order correction with the original MIO parameter set effectively eliminate Grotthuss hopping and yield an oversolvated (hyper coordinated) OH^- ion that diffuses mainly through vehicular transport. The DFTB3-3OB parameter set outperforms the MIO parameter set for some of the structural, energetic and dynamic properties studied in this work, but it still does not outperform the accuracy of the DFTB2 methods. Importantly, the need for redesigning the parameters in the first and second terms of eq 1 to improve the accuracy of DFTB3 may imply that the physically motivated 3rd order term alone does not necessarily improve the accuracy of the SCC-DFTB method, and for system such as hydroxide ions solvated in bulk water it may even produce worse result than the

DFTB2 method. Therefore, it is possible that the accuracy of the SCC-DFTB method is more limited by the approximations and empiricism in the first and second term of eq 1 than by the early truncation at the 2nd order expansion.(44-46) In addition, the over-coordination issue of the underlying SCC-DFTB water model causes increased local water density and water voids in constant volume simulations of the hydrated excess proton, the hydrated hydroxide, and bulk water. The present results highlight the extreme difficulties in formulating an accurate, yet efficient, approximate QM approach for the simulation of aqueous and related condensed phase systems. However, our results can serve as important benchmarks for future improvement of the SCC-DFTB methodology.

3.5 References

1. Knight C & Voth GA (2012) The Curious Case of the Hydrated Proton. *Accounts of Chemical Research* 45(1):101-109.

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

3. Ludwig R (2003) New insight into the transport mechanism of hydrated hydroxide ions in water. *Angew Chem Int Edit* 42(3):258-260.

4. Tuckerman ME, Chandra A, & Marx D (2006) Structure and dynamics of OH-(aq). *Accounts of Chemical Research* 39(2):151-158.

5. Marx D, Chandra A, & Tuckerman ME (2010) Aqueous Basic Solutions: Hydroxide Solvation, Structural Diffusion, and Comparison to the Hydrated Proton. *Chem Rev* 110(4):2174-2216.

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

7. Xie WS & Gao JL (2007) Design of a next generation force field: The X-POL potential. *J. Chem. Theory Comput.* 3(6):1890-1900.

8. Song LC, Han JB, Lin YL, Xie WS, & Gao JL (2009) Explicit Polarization (X-Pol) Potential Using ab Initio Molecular Orbital Theory and Density Functional Theory. *J Phys Chem A* 113(43):11656-11664.

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

10. Elstner M, Hobza P, Frauenheim T, Suhai S, & Kaxiras E (2001) Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment. *J. Chem. Phys.* 114(12):5149-5155.

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

12. Elstner M*, et al.* (2003) Modeling zinc in biomolecules with the self consistent charge-density functional tight binding (SCC-DFTB) method: Applications to structural and energetic analysis. *J Comput Chem* 24(5):565-581.

13. Lin CS*, et al.* (2005) Simulation of water cluster assembly on a graphite surface. *J Phys Chem B* 109(29):14183-14188.

14. Kruger T, Elstner M, Schiffels P, & Frauenheim T (2005) Validation of the density-functional based tight-binding approximation method for the calculation of reaction energies and other data. *J. Chem. Phys.* 122(11).

15. Zhechkov L, Heine T, Patchkovskii S, Seifert G, & Duarte HA (2005) An efficient a Posteriori treatment for dispersion interaction in density-functional-based tight binding. *J. Chem. Theory Comput.* 1(5):841-847.

16. Riccardi D*, et al.* (2006) Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *J. Phys. Chem. B* 110(13):6458-6469.

17. Elstner M (2006) The SCC-DFTB method and its application to biological systems. *Theor Chem Acc* 116(1-3):316-325.

18. Choi TH & Jordan KD (2010) Application of the SCC-DFTB Method to H+(H2O)(6), H+(H2O)(21), and H+(H2O)(22). *J Phys Chem B* 114(20):6932-6936.

19. Maupin CM, Aradi B, & Voth GA (2010) The Self-Consistent Charge Density Functional Tight Binding Method Applied to Liquid Water and the Hydrated Excess Proton: Benchmark Simulations. *J Phys Chem B* 114(20):6922-6931.

20. Goyal P, Elstner M, & Cui Q (2011) Application of the SCC-DFTB Method to Neutral and Protonated Water Clusters and Bulk Water. *J Phys Chem B* 115(20):6790-6805.

21. Gaus M, Cui QA, & Elstner M (2011) DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J Chem Theory Comput* 7(4):931-948.

22. Choi TH (2012) Simulation of the (H2O)(8) cluster with the SCC-DFTB electronic structure method. *Chem Phys Lett* 543:45-49.

23. Ohta Y, Okamoto Y, Irle S, & Morokuma K (2008) Rapid growth of a single-walled carbon nanotube on an iron cluster: Density-functional tight-binding molecular dynamics simulations. *Acs Nano* 2(7):1437-1444.

24. Qian HJ, van Duin ACT, Morokuma K, & Irle S (2011) Reactive Molecular Dynamics Simulation of Fullerene Combustion Synthesis: ReaxFF vs DFTB Potentials. *J Chem Theory Comput* 7(7):2040-2048.

25. Yang Y, Yu HB, York D, Elstner M, & Cui Q (2008) Description of Phosphate Hydrolysis Reactions with the Self-Consistent-Charge Density-Functional-Tight-Binding (SCC-DFTB) Theory. 1. Parameterization. *J Chem Theory Comput* 4(12):2067-2084.

26. Xu DG, Guo H, & Cui Q (2007) Antibiotic binding to dizinc beta-lactamase L1 from Stenotrophomonas maltophilia: SCC-DFTB/CHARMM and DFT Studies. *J. Phys. Chem. A* 111(26):5630-5636.

27. Dewar MJS, Zoebisch EG, Healy EF, & Stewart JJP (1985) The Development and Use of Quantum-Mechanical Molecular-Models .76. Am1 - a New General-Purpose Quantum-Mechanical Molecular-Model. *J Am Chem Soc* 107(13):3902-3909.

28. Stewart JJP (1989) Optimization of Parameters for Semiempirical Methods .1. Method. *J Comput Chem* 10(2):209-220.

29. Ventura ON, Coitino EL, Lledos A, & Bertran J (1989) Am1 Study of Hydrogen-Bonded Complexes of Water. *Theochem-J Mol Struc* 56:55-68.

30. Csonka GI (1993) Analysis of the Core-Repulsion Functions Used in Am1 and Pm3 Semiempirical Calculations - Conformational-Analysis of Ring-Systems. *J Comput Chem* 14(8):895-898.

31. Csonka GI & Angyan JG (1997) The origin of the problems with the PM3 core repulsion function. *J Mol Struc-Theochem* 393:31-38.

32. Sattelmeyer KW, Tirado-Rives J, & Jorgensen WL (2006) Comparison of SCC-DFTB and NDDO-based semiempirical molecular orbital methods for organic molecules. *J Phys Chem A* 110(50):13551-13559.

33. Islam SM & Roy PN (2012) Performance of the SCC-DFTB Model for Description of Five-Membered Ring Carbohydrate Conformations: Comparison

to Force Fields, High-Level Electronic Structure Methods, and Experiment. *J. Chem. Theory Comput.* 8(7):2412-2423.

34. Gaus M, Goez A, & Elstner M (2013) Parametrization and Benchmark of DFTB3 for Organic Molecules. *J Chem Theory Comput* 9(1):338-354.

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

36. Wales DJ & Scheraga HA (1999) Review: Chemistry - Global optimization of clusters, crystals, and biomolecules. *Science* 285(5432):1368-1372.

37. Doye JPK & Wales DJ (1997) Structural consequences of the range of the interatomic potential - A menagerie of clusters. *J Chem Soc Faraday T* 93(24):4233-4243.

38. Doye JPK & Wales DJ (1999) Structural transitions and global minima of sodium chloride clusters. *Phys Rev B* 59(3):2292-2300.

39. Doye JPK, Wales DJ, & Miller MA (1998) Thermodynamics and the global optimization of Lennard-Jones clusters. *J Chem Phys* 109(19):8143-8153.

40. Aradi B, Hourahine B, & Frauenheim T (2007) DFTB+, a sparse matrix-based implementation of the DFTB method. *J Phys Chem A* 111(26):5678-5684.

41. Monkhorst HJ & Pack JD (1976) Special Points for Brillouin-Zone Integrations. *Phys. Rev. B* 13(12):5188-5192.

42. Roberts ST*, et al.* (2009) Observation of a Zundel-like transition state during proton transfer in aqueous hydroxide solutions. *Proc. Natl. Acad. Sci. U. S. A.* 106(36):15154-15159.

43. Ufimtsev IS, Kalinichev AG, Martinez TJ, & Kirkpatrick RJ (2009) A multistate empirical valence bond model for solvation and transport simulations of OH- in aqueous solutions. *Phys. Chem. Chem. Phys.* 11(41):9420-9430.

44. Giese TJ & York DM (2010) Density-functional expansion methods: Evaluation of LDA, GGA, and meta-GGA functionals and different integral approximations. *J. Chem. Phys.* 133(24).

45. Giese TJ & York DM (2011) Density-functional expansion methods: Generalization of the auxiliary basis. *J. Chem. Phys.* 134(19).

46. Giese TJ & York DM (2012) Density-functional expansion methods: grand challenges. *Theor. Chem. Acc.* 131(3).

Chapter 4

Benchmark Study of the SCC-DFTB Approach for a Biomolecular Proton Channel

4.1 Introduction

Calculations using *ab initio* quantum mechanical (QM) methods can in principle be quite accurate for describing various kinds of systems at the atomic level. QM methods are also well suited for describing chemical reactions in which bond breaking and formation occur. However, the cost of the QM calculations limits not only the system size but also the sampling that is required to calculate statistically meaningful quantities for condensed phase systems, such as free energies of binding or a potential of mean force (PMF) for solute or ion transport. Therefore, it is sometimes necessary to develop approximations to *ab initio* QM methods that offer a balance between accuracy and computational efficiency. The self-consistent charge density functional tight binding (SCC-DFTB) method(1) is a fast semi-empirical QM algorithm that has become popular in recent years for simulating chemical processes in biomolecular systems due to the high degree of interest in studying such systems.(2) The SCC-DFTB method is derived from density functional theory (DFT) by approximation and parameterization of multi-center electron integrals.(1) The computational speed gained by these approximations can be 2-3 orders of magnitude compared to more accurate and "first principles" standard DFT.(3) While the SCC-DFTB approach is understandably popular, the method has been shown to have substantial limitations for predicting structural, energetic and dynamic properties of proton transport (PT)(4) and hydroxide transport in bulk water. (5) Despite recent SCC-DFTB developments (3, 6, 7), these issues remain unresolved. The SCC-DFTB method

was also, e.g., recently shown to poorly describe noncovalent interactions involving sulfur atoms.(8) Given the increased use of quantum mechanics/molecular mechanics (QM/MM) with SCC-DFTB as the QM method (SCC-DFTB/MM) to study proton hydration and transport in biomolecular systems,(9-11) there is a need to benchmark its present level of accuracy and potential limitations in such environments.

The current work establishes a systematic benchmark of SCC-DFTB/MM method against arguably more accurate QM/MM methods that employ both generalized gradient approximation (GGA) level and hybrid-GGA level DFT theories for the QM calculation. The comparison is made in the context of condensed phase molecular dynamics (MD) simulations of PT in channel systems. All the methods under investigation have a comparable amount of sampling time, which has seldom been done to date. The synthetic leucine-serine channel (LS2, **Figure 4.1**)(12) is chosen as the simulation system. Although it is synthetic, LS2 possesses key features that are representative of proton channels in nature, such as high proton selectivity, a nonuniform pore radius along the channel axis, a hydrophilic pore due to pore-lining serine residues, and a macrodipole formed by parallel α helices.(12-15) In addition, it is small enough to enable sufficient sampling for the convergence of statistical quantities extracted from MD simulations, which is both essential for condensed phase analysis and computationally demanding for the higher level QM/MM approaches.

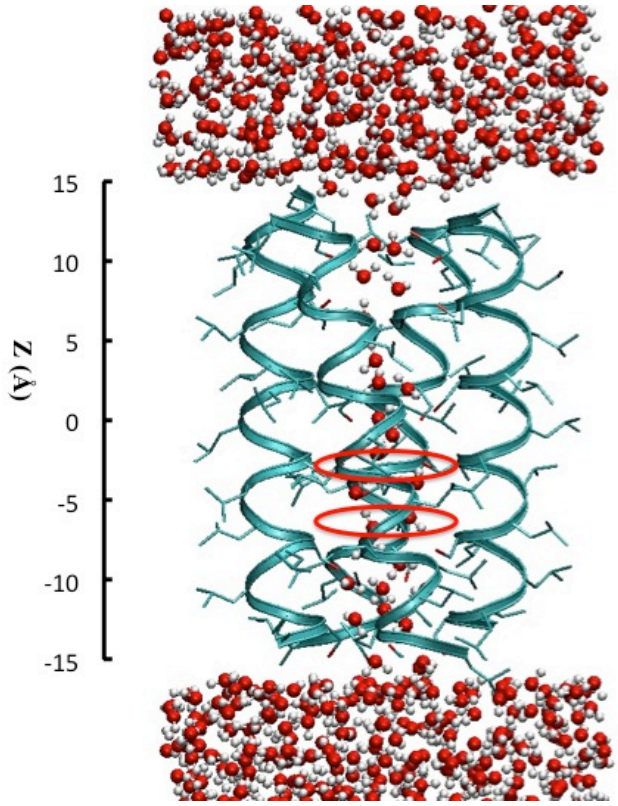


Figure 4.1 LS2 channel system filled with a protonated water wire. The two red circles denote the regions where the excess proton CEC is restrained in the QM/MM simulations. The protein backbones and side chains are depicted as the ribbons and sticks, respectively. On the left is the scale of the channel along the z direction.

It has been argued (6) that the conclusions drawn from previous SCC-DFTB benchmark studies of PT in bulk water (4) may not be relevant to PT in the biological systems for two reasons: first, the charge delocalization of the excess protonic defect may be smaller in biological systems; and second, the pKa values of biological titratable groups may dominate PT energetics. However, in this work it is observed that the present 'state-of-the-art' SCC-DFTB method fails to reproduce important aspects of the more accurate DFT results in a channel environment. Furthermore, by also running simulations in a hydrophobic channel (a nanotube) it is shown that the inaccuracies of SCC-DFTB/MM are more related to the approximations of the semiempirical QM method itself, rather

than to the approximations introduced by a simple point charge based QM/MM electrostatic coupling scheme.(16) These combined results provide an important benchmark of the current SCC-DFTB/MM method that can help researchers to gauge the reliability of this approach for studying PT in realistic biomolecular systems. These results can also help underpin possible improvement of the method if in fact such an improvement is possible without resorting to an excessive degree of empiricism that would undermine the purpose of a QM approach in the first place.

4.2 Simulation details

4.2.1 SCC-DFTB method

The SCC-DFTB method is derived by Taylor series expansion of the DFT Kohn-Sham energy in terms of the charge density fluctuations, and the Hamiltonian matrix elements are evaluated using a minimal basis set of pseudoatomic orbitals with a two-center approximation. The total energy for the full 3rd order SCC-DFTB method is expressed as:(3)

$$E_{total}^{SCC-DFTB} = \sum_{i\mu v} c_\mu^i c_\mu^i H_{\mu v}^0 + E_{rep} + \frac{1}{2}\sum_{\alpha\beta} \gamma_{\alpha\beta}\Delta q_\alpha \Delta q_\beta + \frac{1}{3}\sum_{\alpha\beta} \Gamma_{\alpha\beta}\, \Delta q_\alpha^2 \Delta q_\beta \qquad \textbf{Eq 1}$$

where $H_{\mu v}^0$ are the Hamiltonian matrix elements and c_μ^i and c_μ^i are the wave function expansion coefficients, E_{rep} accounts for short-range repulsive term, and Δq_α and Δq_β are the charge fluctuation terms associated with atoms α and β, respectively. The $\gamma_{\alpha\beta}$ function is determined by

$$\gamma_{\alpha\beta} = \frac{1}{R_{\alpha\beta}} - S_{\alpha\beta} \times f_{\alpha\beta}, \qquad \textbf{Eq 2}$$

$$f_{\alpha\beta} = \begin{cases} \exp\left[-\left(\frac{U_\alpha + U_\beta}{2}\right)^{\zeta} R_{\alpha\beta}^2\right], & \text{if } (\alpha \text{ or } \beta = \text{H}) \\ \text{or} & \\ 1 \quad , & \text{otherwise} \end{cases}$$

where $S_{\alpha\beta}$ is an exponentially decaying short-range function. The damping function $f_{\alpha\beta}$ is introduced to help describe hydrogen-bonding interactions, where U_α and U_β are the atomic Hubbard parameters related to the chemical hardness of atoms α and β, respectively. The parameter ζ is determined by fitting to hydrogen bonding energies for select clusters from higher-level QM calculations.(17) The last term in eq 1 is a third order correction that was initially introduced with only the on-site third order term,(17) but was recently implemented with the complete third-order term.(3) The $\Gamma_{\alpha\beta}$ describes derivatives of the γ function with respect to atomic charges. The introduction of this 3rd order correction is suggested to improve the description of charged systems.(3, 17)

In ref (7) a new parameter set named 3OB was designed such that both the electronic and repulsive potential parameters (appearing in the first and second terms in eq 1, respectively) are optimized for full 3rd order SCC-DFTB calculations. In gas phase calculations this parameter set was shown to outperform the old MIO parameter set due to an improved geometry of nonbonded interactions and reduction of overbinding errors.(7) In the present work, we benchmark the newest, full 3rd order SCC-DFTB method using the 3OB parameter set. The SCC-DFTB related parameters are taken from ref (7).

4.2.2 LS2 QM/MM simulation

The LS2 system (shown in **Figure 4.1**) was constructed using the same protocol as described in detail in ref (15). A reduced system was composed of the LS2 channel with pore waters inside and a 10 Å–thick layer of equilibrated water molecules on either side of the channel, but with the lipid bilayer surrounding LS2 removed. Two layers of virtual atoms were added at the position of lipid head groups to prevent water molecules from entering the membrane region (see **Figure 4.1**). The virtual atoms had zero charge, Lennard-Jones (LJ) parameters of $\sigma = 1.781797$Å and $\varepsilon = 0.11$ kcal/mol, and were hexagonally close-packed with spacing of 1.95 Å. The parameters of the virtual atoms were chosen to minimize their effect on the pore water, while preventing water molecules from leaking into the lipid bilayer region. To maintain the integrity of the LS2 channel structure, harmonic position restraints with force constants of 1000 kJ/mol/Å^2 were applied to the C_α atoms. These restraints were previously shown to maintain the channel geometry and pore water diffusion constants of the realistic composite system.(15) The protein was described by the CHARMM22 force field(18, 19) and the classical water molecules by the TIP3P(20) model modified for the CHARMM force field.

To equilibrate the system, initial velocities were assigned consistent with a temperature of 300 K that was then maintained with a Nosé-Hoover thermostat using a relaxation time constant of 0.1 ps for 10ns of constant NVT simulation. The integration time step was 1 fs. The cutoff radius for LJ interactions was 12 Å employing a switching function starting at 10 Å. The cutoff radius for real space electrostatic interactions was 12 Å and the long-

range electrostatics were treated by Particle-Particle Particle-Mesh (PPPM) method (21)with an accuracy threshold of 10^{-4}.

To set up a protonated water wire inside the LS2 channel, an excess proton was placed near one of the channel waters, forming a classical hydronium cation parameterized consistently with the MS-EVB3 model.(22) A chloride counter ion was added to the bulk water region to maintain charge neutrality. A harmonic potential with a 4 kcal/mol/Å^2 force constant was applied to the hydronium oxygen atom to restrain its z coordinate around z = -6.5 Å, coinciding with a minimum in the previously calculated free energy profile for PT through the channel.(13) The protonated system was equilibrated for 1ns at 300 K using the same settings as described above. All equilibration simulations were performed with the LAMMPS package (http://lammps.sandia.gov).(23)

The last snapshot of the MS-EVB3 equilibration was used as the initial structure for the QM/MM simulation. For the unprotonated LS2 system, the QM region included all of the channel water molecules (22-25 water molecules). For the protonated system, the QM region included all channel water molecules plus the excess proton. For the DFT/MM setup, the QM box size was 12 Å × 12 Å × 32 Å, with the z direction aligned with the channel axis. A quadratic confining wall potential was used to restrain the QM atoms within the QM box, with a wall skin thickness of 1 Å × 1 Å × 2.5 Å. The QM box size was chosen such that the quadratic wall potential did not affect the QM atoms for the majority of the simulation. The QM region was treated with several different density functionals under the Gaussian plane wave (GPW) scheme.(24) Two generalized gradient

approximation (GGA) level exchange-correlation functionals were employed: the BLYP functional(25, 26) with a empirical dispersion corrections(27) and the HCTH/120 functional(28) (hereafter denoted as BLYP-D and HCTH, respectively). In addition, the hybrid GGA level DFT method B3LYP(29) with a dispersion correction(27) (B3LYP-D) was employed to improve accuracy. For BLYP-D and HCTH, the Goedecker-Teter-Hutter (GTH) pseudopotentials(30) were used and the Kohn-Sham orbitals were expanded in the Gaussian TZV2P basis set. The electron density was expanded by auxiliary plane wave basis set up to 300 Ry. For B3LYP-D, GTH pseudopotentials for BLYP were used, and the wavefunction was expanded in the Gaussian DZVP basis set with an energy cutoff of 300 Ry for the plane wave basis set. The Gaussian Expansion of the Electrostatic Potential (GEEP) scheme was used to treat the QM/MM electrostatic coupling with periodic boundary conditions (PBCs),(31, 32) and the spurious QM/QM periodic image interactions were decoupled as described in ref (33). The QM/MM van der Waals (vdW) interactions were described by the CHARMM22 force field Lennard-Jones (LJ) parameters. The motion of nuclei was integrated using a time step of 0.5 fs, and the wavefunction was optimized to the Born-Oppenheimer surface by an orbital transformation method(34) with a convergence criterion of 10^{-6}. All DFT/MM simulations were run using the CP2K package.(35)

For the SCC-DFTB/MM setup, the QM atoms were chosen in the same way as DFT/MM and planar restraining potentials were used to prevent the boundary QM atoms from diffusing out of the channel. Again, the QM atoms within the boundary defined by the planes were not affected. The point charge based Ewald summation was used to treat

QM/MM electrostatic coupling under PBCs.(16) Just as for the DFT/MM simulations, the SCC-DFTB/MM vdW interactions were described by the CHARMM22 force field LJ parameters. The time step was set as 0.5fs, and the SCC-DFTB convergence criterion of 10^{-8} was used. The SCC-DFTB/MM simulations were carried out using the CHARMM simulation package.(36)

In previous studies the PT free energy profile through the LS2 channel features several local minima separated by energy barriers that are several times that of the thermal energy, thus impeding the free translocation of the excess proton.(13) In order to make a statistically meaningful comparison between SCC-DFTB and DFT methods for the PT properties within a limited simulation time (~100 ps), we chose to enhance the sampling of the excess proton in two regions inside the LS2 channel (one wider and one narrower region). In particular, the z position was restrained with a harmonic potential around z = -6.5 Å (**Figure 4.2**), which corresponds to a wide region in the channel where the complete first solvation shell of the excess proton can be accommodated and a minimum in the PT free energy profile was previously reported (Figure 2 in ref (13)), and z = - 3.5 Å, corresponding to a narrow channel region and a barrier in the PT free energy profile.

The charge defect associated with the excess proton was tracked via the center of excess charge (CEC) defined in this work as:(37)

$$\vec{\xi} = \sum_{i=1}^{N_H} \vec{r}^{H_i} - \sum_{j=1}^{No} w^{O_j} \vec{r}^{O_j} - \sum_{i=1}^{N_H} \sum_{j=1}^{No} f_{sw}\left(d_{O_j H_i}\right)(\vec{r}^{H_i} - \vec{r}^{O_j}), \qquad \textbf{Eq 3}$$

where the "O" and "H" denotes oxygen atoms and hydrogen atoms in the QM region, the weighting factor w^{O_j}'s are set to 2, $d_{O_j H_i}$ is the distance between atom O_j and atom H_i,

and $f_{sw}\left(d_{O_jH_i}\right) = 1/(1 + \exp\left[(d_{O_jH_i} - r_{sw})/d_{sw}\right])$ is the switching function describing the coordination number of H_i to O_j, with the parameters chosen as $d_{sw} = 0.03$ Å, $d_{sw} = 0.13$ Å.(37) This excess proton CEC definition has been shown to adequately locate the position of the excess proton charge defect in previous QM/MM simulations of biological PT channels.(11, 37) This quantity will be referred to often in the following discussion so the readers should take note of this acronym and its definition.

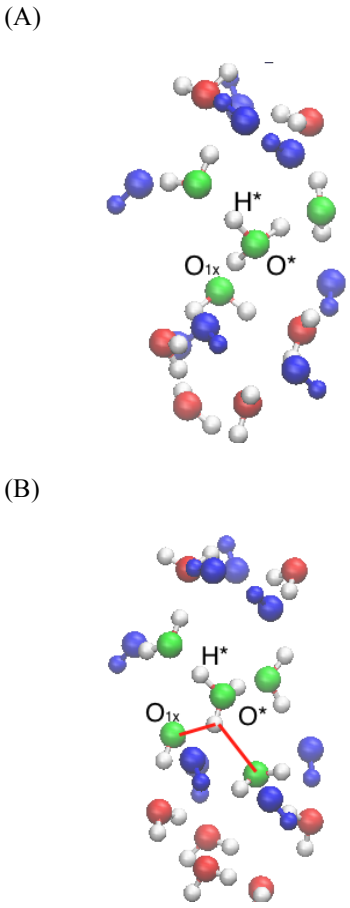


Figure 4.2 Snapshots of protonated water wire in LS2 with excess proton CEC restrained at z = -6.5 Å from (A) BLYP-D simulation (B) SCC-DFTB simulation. The serine residues are colored in blue, the most hydronium-like oxygen O* and the water oxygens within 3 Å of O* are colored in green, and the rest of the water oxygens are colored in red. One of the hydrogens closest to the O* is shown by H* and the next closest oxygen in the special pair, O_{1x}, is also shown in panel A. Notice the oversolvation of SCC-DFTB O* and the bifurcated hydrogen bond indicated by the red lines in panel B.

Following 200 steps of steepest descent geometry minimization, the systems were heated up and the temperature then maintained at 300 K using a Nosé-Hoover thermostat with a relaxation time constant of 0.1 ps for all of the methods. For the all of the methods, both restrained and unrestrained simulations of the protonated LS2 channel were run, with the latter starting from the last snapshot of the corresponding method's restrained simulation. All the QM/MM simulations were equilibrated for at least 30 ps and run for another 70-100 ps in the constant NVT ensemble.

4.2.3 Carbon nanotube QM/MM simulation

A (8,8) single-wall carbon nanotube (CNT) of length 29.47 Å was first solvated in a large box of equilibrated TIP3P water. The LJ parameters for carbon atoms were $\sigma = 3.40$ Å and $\varepsilon = 0.086$kcal/mol, corresponding to the sp^2 carbon atoms in the AMBER force field.(38) All the carbon atoms were fixed. The cutoffs for LJ and real space electrostatic interactions were 10 Å, employing switching function starting at 8 Å for the LJ interactions. To equilibrate the system, initial velocities were assigned to 300 K followed by 20 ps of constant NVT and then 2ns of constant NPT equilibration. On average 38 water molecules resided in the CNT during the NPT equilibration. One snapshot was taken from the last 10% of the equilibration run, in which 38 water molecules were inside the CNT, and water molecules outside the CNT were removed. Then the simulation box was set up with PBC of 20 Å × 20 Å × 29.47 Å, and the longest dimension of simulation box was aligned with tube axis to create an infinitely long CNT. The system was equilibrated at 300 K using Nosé-Hoover thermostat for 1ns. One excess proton was then added near one water molecule, and the equilibration was continued for 1ns using the

same classical hydronium model as used in the LS2 equilibration. All the equilibration simulations were again carried out with LAMMPS.

The last snapshot of equilibration was used to construct the QM/MM simulation. The QM atoms were chosen to be all the water molecules and the excess proton. Following 40 ps of constant NVT simulation at 300 K using a Nosé-Hoover thermostat, the production constant NVE runs were continued for at least 80 ps. The DFT/MM simulations were run with the BLYP-D and HCTH functionals. The QM/MM settings were the same as those used in the LS2 channel, except that the energy cutoff for plane wave basis set was increased to 360 Ry and the convergence criterion was decreased to 10^{-7} to better conserve total energy. For the SCC-DFTB/MM simulations, the convergence criterion was 10^{-8}. For all QM/MM simulations, the motion of nuclei was integrated using a time step of 0.5 fs.

4.3 Results

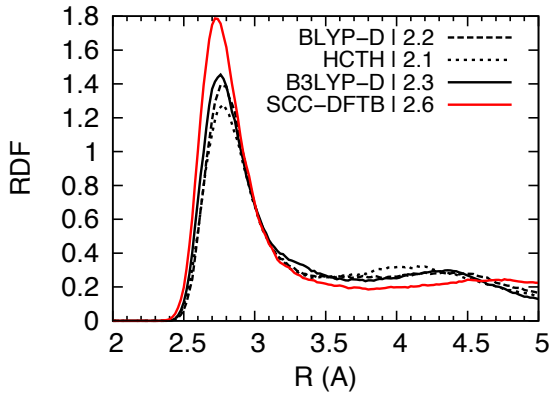
4.3.1 Water structure and dynamics in the LS2 channel

The water molecules filling the LS2 channel make hydrogen bonds with one another and with pore lining serine side chains, resulting in a continuous hydrogen-bonded water wire across the entire channel.(15)-(39) It is beneficial to first investigate the properties of pure water in the LS2 channel. The BLYP functional with dispersion corrections (BLYP-D) and HCTH/120 functional are chosen as the benchmark QM methods because both methods were shown to reduce the excessive structuring and sluggish dynamics of bulk water predicted by the BLYP functional.(40-44) These functionals were in turn benchmarked by simulations using the arguably more accurate B3LYP functional with

dispersion corrections (B3LYP-D). To save computational cost, B3LYP-D was combined with GTH pseudopotentials for BLYP and the DZVP basis set. We note that B3LYP should generally be used with a large basis set augmented by diffuse functions in order to guarantee a highly accurate hydrogen bond description.(45) Although this was not computationally feasible, the B3LYP-D employed in our QM/MM simulations still provide more accurate forces and energies than GGA level DFT methods. (See Supporting Information)

In the following, the water oxygen, water hydrogen and serine oxygen atoms will be denoted as O_w, H_w and O_s, respectively. To probe the structural properties of the water within the channel, radial distribution functions (RDFs) were calculated and normalized to bulk water density. From the O_w-O_w RDF (**Figure 4.3** (A)), it is apparent that SCC-DFTB predicts a more pronounced first peak. The coordination number, obtained by integrating over the first peak, is larger for SCC-DFTB due to a more compact and dense first solvation shell. In addition, the second peak in the O_w-O_w RDF and the density depletion region between second and third peaks in the O_w-H_w RDF (**Figure 4.3** (B)) are less pronounced for SCC-DFTB than for the DFT methods, which corresponds to a second solvation shell water encroaching on the first solvation shell. This suggests that the water wire predicted by SCC-DFTB is less ordered inside the LS2 channel. In line with this observation, SCC-DFTB predicts less distinct peaks and valleys in the pore water density profile along the z axis (**Figure 4.4**) compared to both DFT methods, indicating that the water molecules deep in the pore region (-10 Å to 10 Å) are more mobile.

(A)



(B)

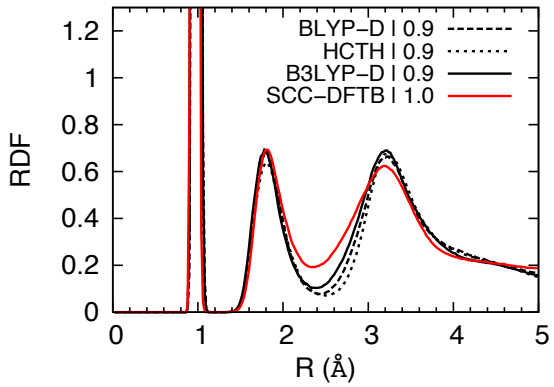


Figure 4.3 RDFs of (A) O_w-O_w and(B) O_w-H_w in the LS2 channel filled with water and no excess proton. The coordination numbers, indicated in the legend, are larger for the SCC-DFTB method due to less structured 2nd shell waters encroaching on the 1st solvation shell.

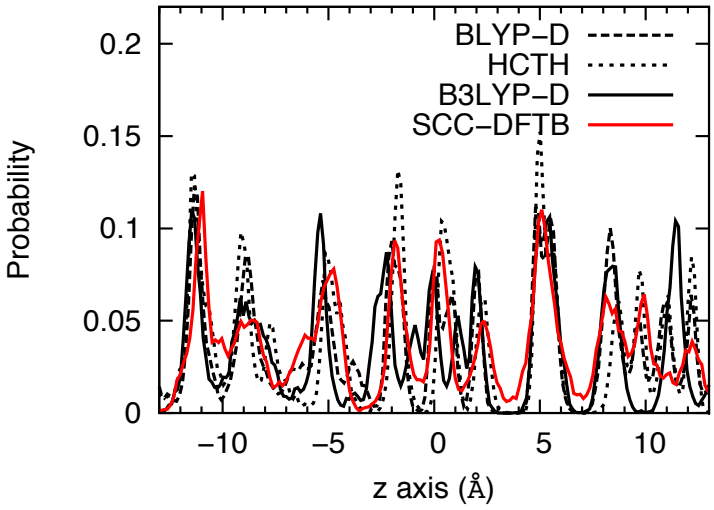


Figure 4.4 Distribution of the pore water oxygen atoms along the z axis of the unprotonated LS2 channel. The decreased peaks and valleys for the SCC-DFTB results indicate more mobile water within the channel.

The decreased structure in the LS2 channel water could be caused by weaker water-water hydrogen bonds in SCC-DFTB compared to those in the DFT methods. To quantitatively investigate the hydrogen bond properties, we define the water-water hydrogen bond as when the O_w---O_w separation is less than 3.5Å (corresponding to the location of the first minimum in the O_w-O_w RDF) and the H_w-O_w---O_w angle is less than 30°.(46, 47) The number of water-water hydrogen bonds each water molecule makes is averaged over the entire 90 ps production trajectory and listed in **Table 4.1**. It is observed that the channel water described by SCC-DFTB forms fewer hydrogen bonds than those described by the DFT methods, which are in close agreement with each other. To compare the longevity of the hydrogen bonds, we calculated the continuous hydrogen bond time correlation function $C_c(t) = \langle h(0) \cdot H(t) \rangle / \langle h \rangle$, where $h(t)$ is unity when a tagged pair of atoms are hydrogen bonded at time t, and zero otherwise. $H(t)$ is defined as unity when the tagged pair of atoms is continuously hydrogen bonded from time 0 to time t, and zero otherwise.

A semi-log plot of $C_c(t)$ is shown in **Figure 4.5** (A). The continuous hydrogen bond time correlation functions are then fit to the sum of 3 exponential functions, and the average relaxation time is estimated by the amplitude-averaged relaxation time,(48)(49) as summarized in **Table 4.2**. It is observed that SCC-DFTB produces a shorter hydrogen bond relaxation time compared to the DFT methods, again confirming shorter (and likely weaker) water-water hydrogen bond interactions.

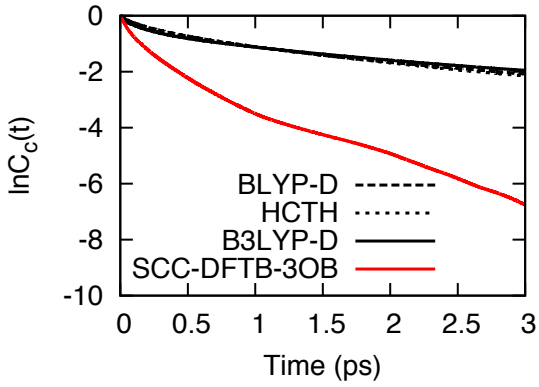
Table 4.1 Averaged number of hydrogen bonds per water molecule in the LS2 channel as a function of QM/MM methodology.

Methods	water-water hydrogen bond	serine-water hydrogen bond
BLYP-D	1.64 ± 0.03	1.14 ± 0.01
HCTH	1.66 ± 0.02	1.17 ± 0.02
B3LYP-D	1.60± 0.02	1.17 ± 0.02
SCC-DFTB	1.51 ± 0.01	1.06 ± 0.02

Table 4.2 Average water-water, water-serine hydrogen bond relaxation time, and rotatioanl correlation time in the LS2 channel as a function of QM/MM methodology.

Method	water-water hydrogen bond relaxation time (ps)	water-serine hydrogen bond relaxation time (ps)	water rotaiotnal correlation time (ps)
BLYP-D	1.26	0.86	24
HCTH	1.20	1.03	28
B3LYP-D	1.38	0.93	23
SCC-DFTB	0.21	0.42	7

(A)



(B)

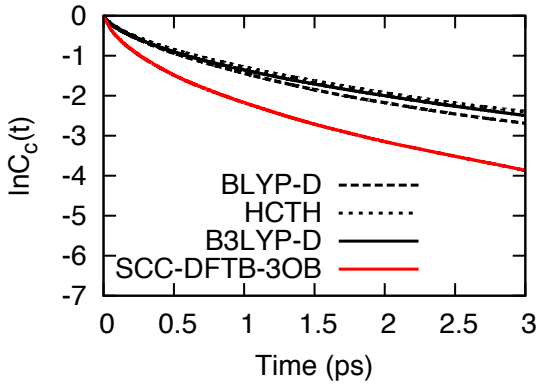


Figure 4.5 Semi-log plot of continuous hydrogen bond time correlation function for the (A) water-water and (B) water-serine hydrogen bonds. The faster decay of time-correlation functions reflects weaker hydrogen bonds described by SCC-DFTB compared to the DFT methods.

The interaction between water and the serine side chains also plays an important role in shaping the structure and slowing down the dynamics of water inside the LS2 channel.(39) The O_w-O_s RDF (**Figure 4.6**) reveals that SCC-DFTB fails to reproduce the DFT density depletion region after the first peak and the presence of a second peak. Similar to the situation with water-water interactions, the less ordered structure revealed in the SCC-DFTB RDF reflects weaker interactions between water and serine side chains. With the water-serine hydrogen bond defined using the same distance and angle

cutoff as the water-water hydrogen bond (described above), we calculated the average number of water-serine hydrogen bonds each water makes (listed in **Table 4.1**) and the continuous hydrogen bond correlation function together with the hydrogen bond relaxation time (**Figure 4.5** (B) and **Table 4.2**). The fewer number of water-serine hydrogen bonds with each water molecule and the faster relaxation of those hydrogen bonds for SCC-DFTB suggest that the weaker water-serine interaction results from fewer and weaker hydrogen bonds.

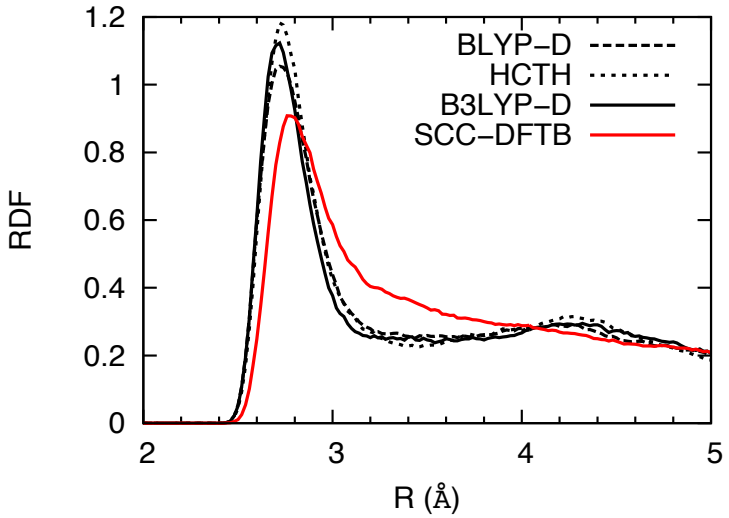


Figure 4.6 RDF of O_w-O_s in the LS2 channel filled with water and no excess proton. The less structured RDF produced by SCC-DFTB indicates weaker water-serine interactions due to weaker and fewer hydrogen bonds.

The consequence of weaker water-water and water-serine interactions is a less structured water wire inside the LS2 channel with accelerated orientational dynamics in the SCC-DFTB simulations. To probe the orientational dynamics, the second rank orientation time-correlation function $C_2(t) = \langle \mathrm{P}_2(\mathbf{u}(0) \cdot \mathbf{u}(t)) \rangle / \langle \mathrm{P}_2(\mathbf{u}(0) \cdot \mathbf{u}(0)) \rangle$ was calculated, where $\mathbf{u}$ is the unit vector in the direction of one of the O-H bonds in water and P_2 is second order Legendre polynomial. The last 30% of $C_2(t)$ is fit to a single exponential

function to get the rotational relaxation times, which are summarized in **Table 4.2**. It is found that the SCC-DFTB method predicts water reorientation much faster than the two DFT methods, again indicating more dynamic and disordered water in the LS2 channel.

Overall, SCC-DFTB predicts an overcoordinated and more disordered water structure featuring faster dynamics of water molecules compared to the DFT methods. These differences in the structural and dynamical properties of unprotonated water are in turn linked to differences in the structure and dynamics of PT, as will be discussed below.

4.3.2 Proton solvation and transport in a wider region of the LS2 channel

In the following the hydronium oxygen, the 3 closest hydrogen atoms to that oxygen, and the next closest oxygen atom will be denoted as O*, H* and O_{1x}, respectively (see **Figure 4.2** (A)). The O_{1x} is the so-called "special pair" oxygen that is forming an especially short strong hydrogen bond with the hydronium at that instant in time. Proton transfer often, but not always, occurs along this special pair coordinate.(50) From the RDFs of O*-O_w and O_{1x}-O (O here denotes O* and O_w) (**Figure 4.7** (A) and (B)), it is apparent that SCC-DFTB predicts an overcoordinated hydronium oxygen, with a coordination number of 3.56 compared to those predicted by the DFT methods (all ~3). Also, the solvation shell around the O* is less structured in the SCC-DFTB simulations than the DFT counterparts. Where the DFT methods predict a distinct density depletion after the first peak of the O*-O_w RDF (at separation distance of 2.8 Å corresponding to the boundary of the first solvation shell), SCC-DFTB predicts a density decay region that extends beyond 3.4 Å. The overcoordination of O* and lack of distinct density depletion beyond the first solvation shell suggests that second solvation shell water molecules penetrate into the

first, making the solvation structure less ordered compared to DFT simulations. In fact, it is observed that in DFT simulations the hydronium is consistently coordinated by 3 first solvation shell water molecules, forming a stable $H_9O_4^+$ Eigen complex, which is further stabilized by hydrogen bond interactions with nearby serine side chains and second solvation shell water molecules (**Figure 4.2** (A)). In contrast, the SCC-DFTB method predicts an overcoordinated hydronium due to a second solvation shell water molecule frequently entering the 3 Å range of O*, leading to a *bifurcated hydrogen bond* (one hydrogen coordinated to two water oxygens) and disrupting the Eigen complex solvation structure (**Figure 4.2** (B)). As will be discussed below, the SCC-DFTB hydronium is also directly coordinated to serine side chains much of the time, further increasing its overcoordination. In addition, the SCC-DFTB first solvation shell waters are also overcoordinated as indicated by the large coordination number shown in the RDF's of O_{1x}-O (**Figure 4.7** (B)), which is in line with this method's tendency to overcoordinate water as discussed in section 4.3.1 .

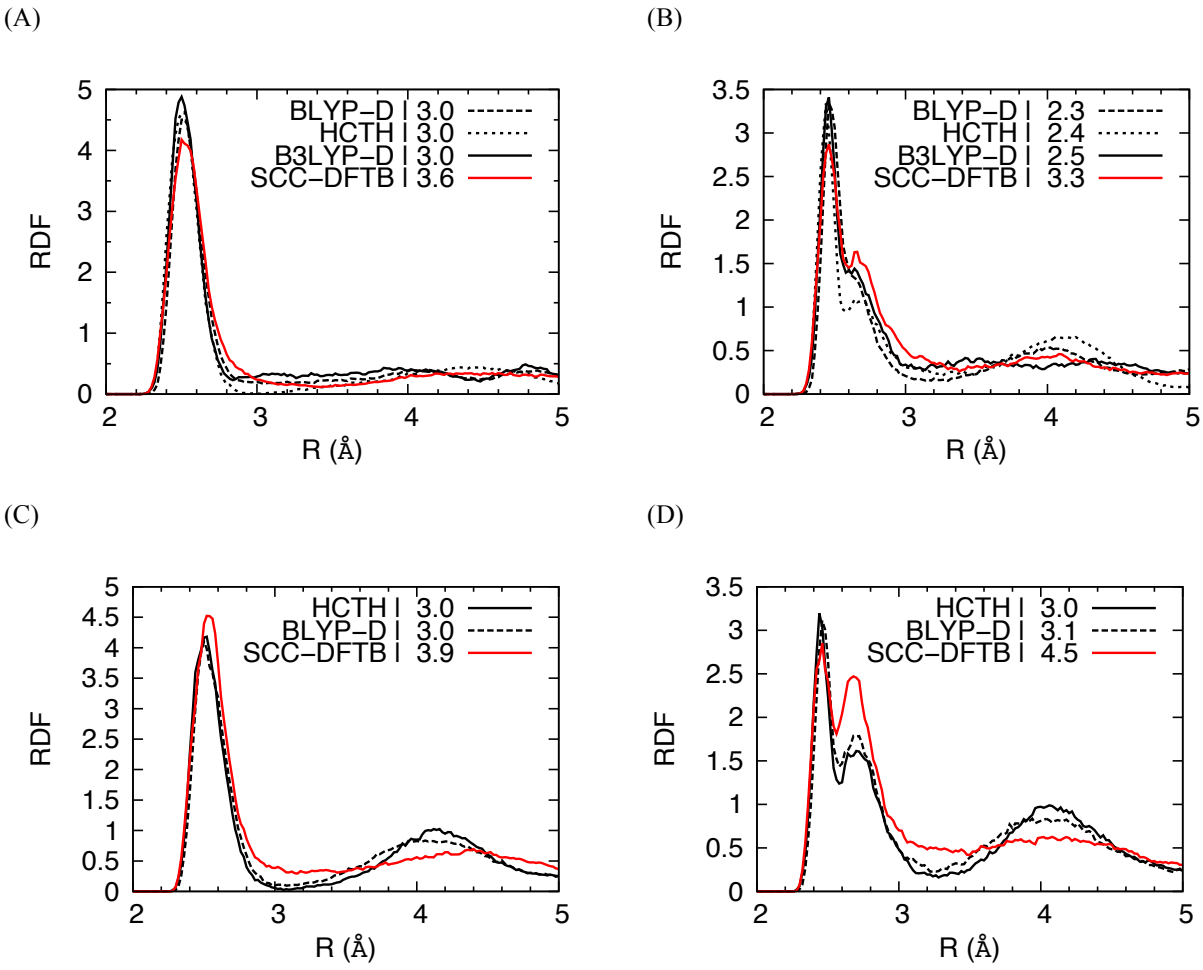
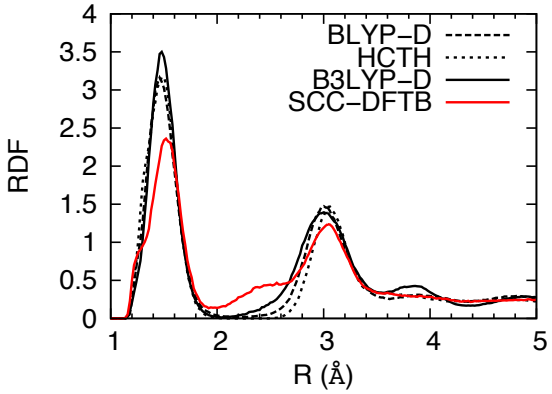


Figure 4.7 Radial distribution functions of O*-O_w and O_{1x}-O in protonated LS2 (A, B) and protonated water filled CNT (C, D). O includes both O* and O_w. Coordination numbers are indicated in the legend. SCC-DFTB predicts an overcoordinated excess proton and less structured solvation complex compared to the DFT methods.

The overcoordinated excess proton structure described by SCC-DFTB leads to an abnormal hydrogen bond network around the excess proton. To illustrate this, the RDF of H*-O_w, where H* is any of the three hydrogens bound to O*, is plotted for LS2 (**Figure 4.8** (A)). The absence of density depletion in the region 2-2.5 Å in the SCC-DFTB results reflects the encroaching second solvation shell water and bifurcated hydrogen bond donated by the hydronium. This observation is perhaps even more clear in **Figure 4.9**

(A), which shows the distribution of 1st and 2nd closest water oxygens to each H*. Clearly, there is an overlap region between the two distributions for SCC-DFTB simulations, which is absent in the DFT results. The overlap is a consequence of the SCC-DFTB method not localizing the first solvation shell water molecules, making it more likely to form a bifurcated hydrogen bond with two waters in its first solvation shell (see **Figure 4.2** (B)).

(A)



(B)

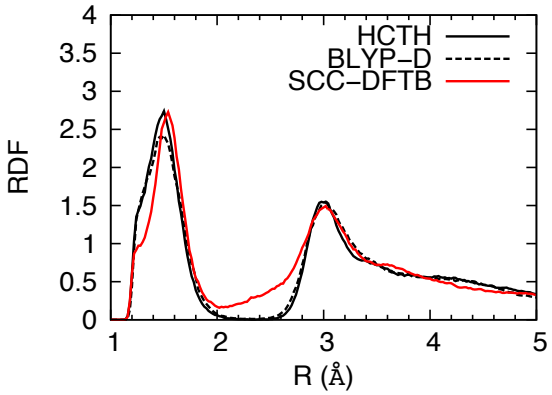
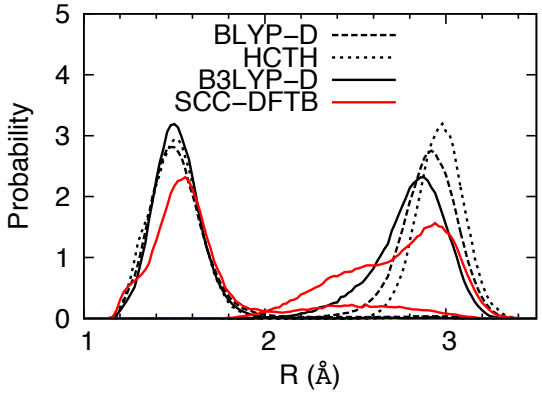


Figure 4.8 Radial distribution functions of H*-O_w in (A) the protonated LS2, (B) the water filled protonated CNT. The absence of density depletion in the distance region 2-2.5 Å in the SCC-DFTB results reflects the encroaching 2nd solvation shell water and bifurcated hydrogen bond donated by the hydronium.

(A)



(B)

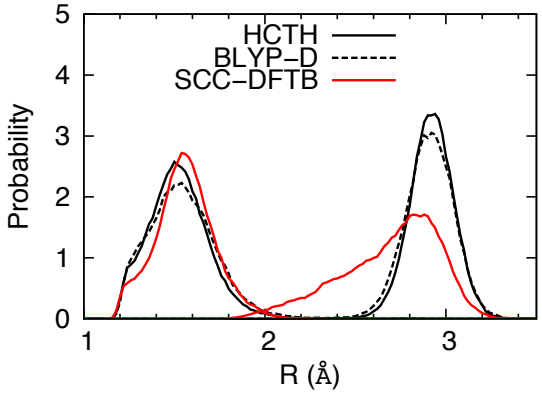
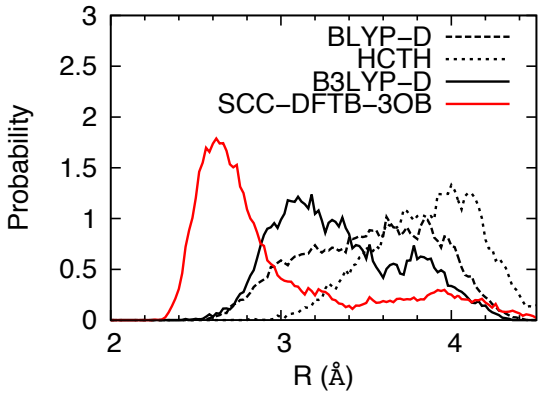


Figure 4.9 Distribution of H*-O_w distances in protonated (A) LS2 channel and (B) CNT. The first and second peaks correspond to the closest and second closest O_w to H*. The overlap of the two peaks in the SCC-DFTB simulation suggests that hydrogen bonding around the excess proton is too weak to localize the position of its first solvation shell water molecules, making it more likely to form a bifurcated hydrogen bond with two waters in its first solvation shell.

Although SCC-DFTB results in an increased probability of forming a bifurcated hydrogen bond, the water penetrating the first solvation shell is not hydrogen bonded to the O*-H* at all for a significant amount of time. **Figure 4.10** (A) shows the distribution of the distance between O* and the nearest water oxygen atoms that are not hydrogen bonded to it. In the SCC-DFTB simulations, for both the LS2 and CNT channels, an

unphysical configuration is frequently observed where a water molecule within the range of first solvation shell (2.8 Å of O*) is not hydrogen bonded to it. More surprisingly, the O_w not hydrogen bonded to or forming a bifurcated hydrogen bond with O* can even belong to the closest water molecule for a significant portion of the simulation time. It has been shown that in a bulk aqueous system during the long intervals between PT events (*aka* non-PT intervals), the closest oxygen to O* (O_{1x}) switches between the 3 surrounding water oxygens in the Eigen complex, resulting in a "special pair dance" mechanism.(50) The existence of a penetrating water molecule that is closest to O* but not hydrogen bonded to it will alter PT dynamics and limit SCC-DFTB's ability to correctly capture the special pair dance dynamics in the LS2 channel. To probe this, we calculated the average number of distinct O_w atoms that have been O_{1x} during non-PT intervals. As shown in **Figure 4.11** the DFT simulations show a mechanism in which O_{1x} rotates among the 3 hydrogen bond accepting waters around O*.(50) Correspondingly, the number of different water oxygens that have once been O_{1x} levels off at 3 in 0.4 ps. However, the SCC-DFTB method predicts more than 3 O_{1x} water molecules exhibiting this behavior. Moreover, the curve shows no sign of leveling off near 0.5 ps. In the SCC-DFTB simulations, the O_w chosen as special partner, O_{1x}, sometimes belongs to a water molecule accepting a hydrogen bond from O*, while at other times it belongs to the penetrating water molecule not hydrogen bonded to *or* forming one of the bifurcated hydrogen bonds with O*. These situations interchange frequently resulting in a qualitatively different dynamical picture than that predicted by DFT.

(A)



(B)

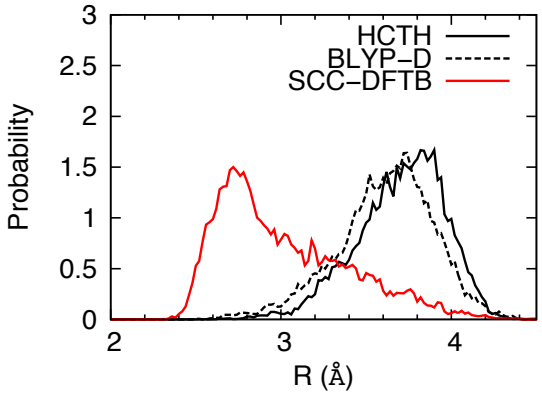
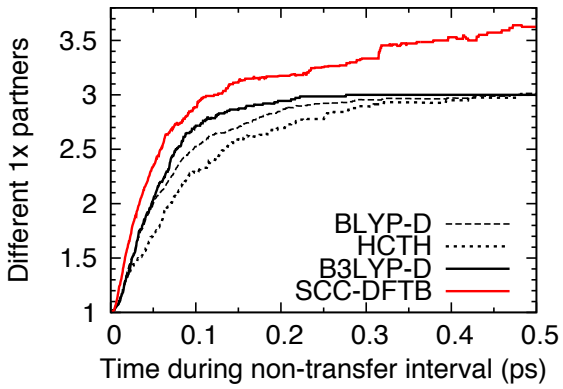


Figure 4.10 Distribution of O*-O distance, where O is the closest atom not hydrogen bonded to O* in the (A) protonated LS2 channel and (B) CNT. The noticeable peak before 2.8 Å produced by SCC-DFTB indicates that on average some first solvation shell water molecules are not hydrogen bonded to the hydronium, which perturbs the "special pair dance" dynamics in the resting state of PT (see main text).

(A)



(B)

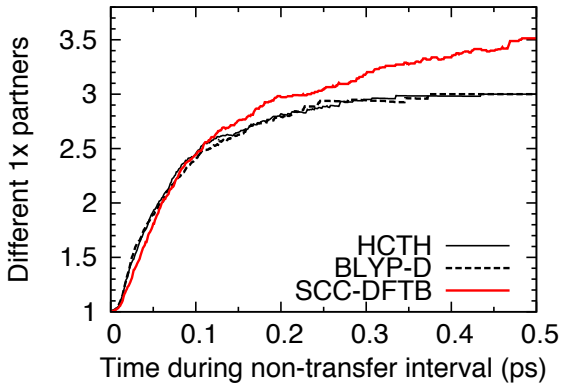


Figure 4.11 Average number of different O_{1x} partners to O* as a function of time during non-transfer intervals for the (A) LS2 channel and (B) CNT. The SCC-DFTB curve fails to level off and indicates that it does not reproduce the "special pair dance" behavior seen from the DFT results.

To further illustrate the weak hydrogen bond around O* in SCC-DFTB from a dynamical point of view, we applied the same analysis of average hydrogen bond relaxation times as was shown for the unprotonated water in LS2. **Table 4.3** compares the relaxation times of hydrogen bonds within 4 Å of the excess proton CEC (defined in section 3.2.2) for different methods in both the LS2 and CNT systems. Not only is the SCC-DFTB hydrogen bond relaxation time much shorter than those predicted by the DFT methods,

but the differences between SCC-DFTB and the DFT methods are much larger than those found for the unprotonated situation. This suggests that the effect of charge defect delocalization in enhancing hydrogen bond strength around the excess proton CEC is relatively weaker for SCC-DFTB than for the DFT methods. Thus, improving the description of hydrogen bond around the hydrated excess proton needs to be addressed in future SCC-DFTB developments.

Table 4.3 Average hydrogen bond relaxation time for hydrogen bond around 4 Å of excess proton CEC in the LS2 channel and CNT as a function of QM/MM methodology.

Method	LS2 hydrogen bond relaxation time (ps)	CNT hydrogen bond relaxation time (ps)
BLYP-D	17.3	13.7
HCTH	19.8	10.0
B3LYP-D	11.2	NA
SCC-DFTB	1.8	1.4

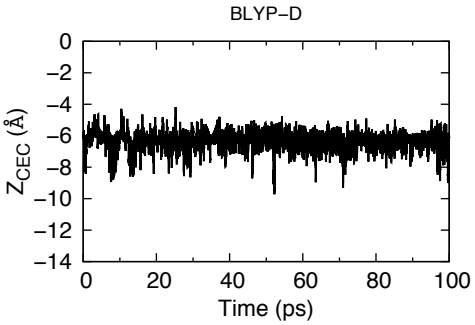
Overall, compared to the arguably more accurate DFT approaches, the SCC-DFTB method predicts the incorrect solvation structure of water and the excess proton, weak water-water and water-serine hydrogen bond interactions, a bifurcated hydrogen bond donated by the hydronium-like oxygen O*, and an unphysical special pair dynamics pattern related to the weak hydrogen bonds around the excess proton CEC. These all lead to qualitatively different PT dynamics in the LS2 and CNT channels. The PT dynamics can also be described by the time evolution of the excess proton CEC z coordinate after the quadratic restraints on it are released, as shown in **Figure 4.12**. For the DFT methods, the excess charge is stabilized in the form of an Eigen complex, which is further

stabilized by the surrounding water and serine residues. During more than 70 ps of simulation the excess proton CEC z coordinate fluctuates around z = -6 Å, which is close to the reference point where the restraint was added. This suggests that the chosen region in the LS2 channel corresponds to a distinct free energy minimum in the free energy profile described by the DFT methods, in agreement with the previous MS-EVB result.(13) In sharp contrast, the excess proton CEC z coordinate in the SCC-DFTB simulations deviates as far as 6 Å from the previous reference point, which indicates that the excess proton is not stabilized in an Eigen complex. It hops around the overcoordinated, less ordered and loosely bound solvation shell in an unpredictable manner, and gradually hops away from the region where it was originally restrained. This can be attributed to the combined effect of weaker water-water and water-serine interactions, the overcoordinated hydronium O*, the weak and sometimes bifurcated hydrogen bond around the excess proton CEC, as well as the unphysical O_{1x} switching events.

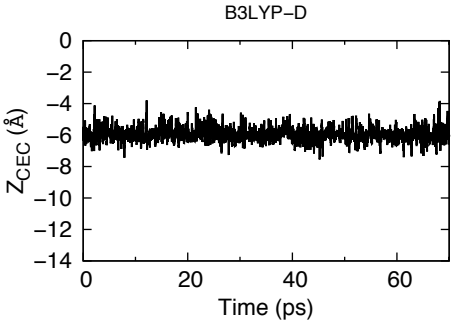
The tendency for SCC-DFTB to yield irregular excess proton solvation and transport remains even when the excess proton CEC motion is more trapped in the constrained simulations. In the constrained simulation SCC-DFTB predicts an O*-O_s RDF featuring a spurious first peak within 3 Å (**Figure 4.13** (A)). This is caused by O* being directly hydrogen bonded to serine side chain hydroxyl groups for significant amounts of the simulation, something that is not observed in the DFT simulations. It is important to note that this close interaction between serine side chain oxygen atoms and the hydronium further exacerbates the overcoordination issue already caused by excessive water

molecules around the SCC-DFTB hydronium in the LS2 channel. To illustrate this, the RDF of O*-O, where O denotes both O_w and O_s, is plotted in **Figure 4.13** (B). Comparing the coordination numbers in **Figure 4.7** (A) and **Figure 4.13** (B) (3.6 versus 3.9) demonstrates that the overcoordination around the excess proton is more pronounced when the serine oxygen atoms are included. **Figure 4.13** (B) shows that SCC-DFTB predicts an O* coordinated by four oxygen atoms, in contrast to three as predicted by the DFT methods. This observation is revealing because it shows that the SCC-DFTB hydronium O* favors an oxygen coordination number of 4. When its water coordination number is less than 4 in a confined environment (as shown in **Figure 4.7** (A)), the serine side chain oxygen atoms tend to further coordinate O* to saturate its solvation shell if allowed by local geometry.

(A)



(B)



(C)

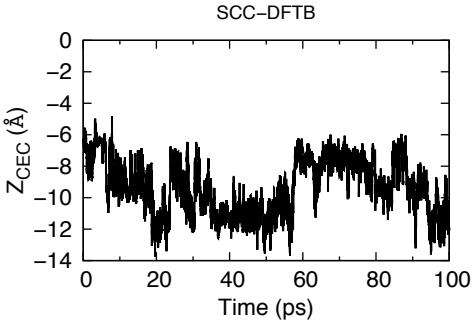
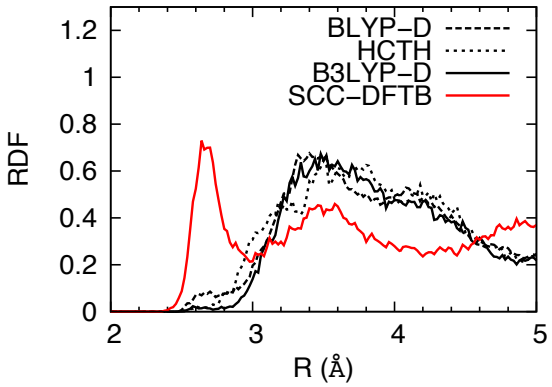


Figure 4.12 *Z* coordinate of the excess proton CEC as a function of time for unconstrained simulations using the (A) BLYP-D, (B) B3LYP-D, and (C) SCC-DFTB methods. The SCC-DFTB simulations show deviations as far as 6 Å from the previous reference point of the restraint, which indicates that the excess proton is not stabilized in an Eigen complex.

(A)



(B)

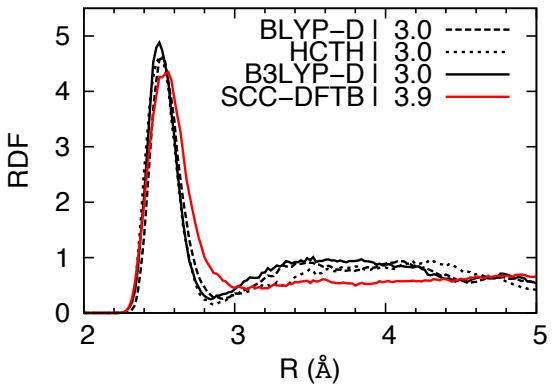


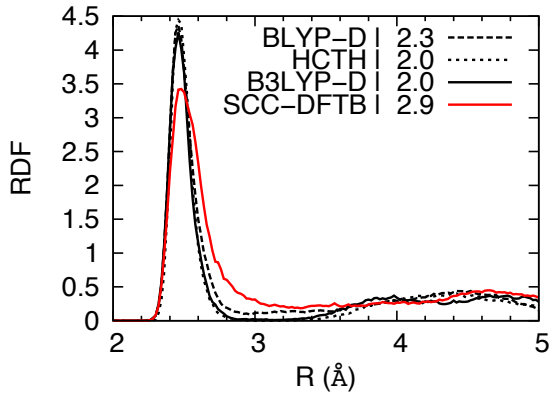
Figure 4.13 Radial distribution functions of (A) O*-O_s between the excess proton complex and serines in the protonated LS2 channel, and (B) O*-O, where O includes both O_w and O_s. The first peak in (A) predicted by SCC-DFTB method reveals that the hydronium is directly hydrogen bonded to serine side chain hydroxyl groups for significant amounts of the simulation. Compared with **Figure 4.9** (A), the coordination numbers in (B) indicate that the serine side chains exacerbates the overcoordination issue of O* in SCC-DFTB simulations.

4.3.3 Proton solvation and transport in a narrower region of the LS2 channel

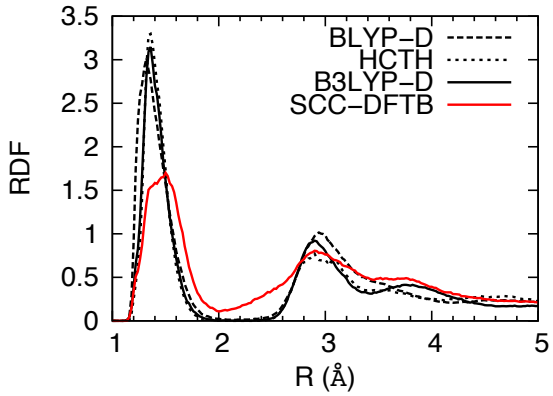
Clearly, narrow channel regions also play an important role in PT processes, as previously demonstrated for the LS2 channel.(13, 15) It is, therefore, both important and instructive to also study the proton solvation structure and transport mechanism in a

narrower region (z = - 3.5 Å) close to the top of a barrier in a previously reported LS2 PMF.(13) As shown in **Figure 4.14** (A) and (B), the DFT methods predict a $H_7O_3^+$ solvation structure of the excess proton, i.e., the hydronium solvated by two nearby water molecules and one serine hydroxyl group. In contrast, the SCC-DFTB method predicts an overcoordinated excess proton, just as it did in the wider region of the LS2 channel, with an additional water and serine hydroxyl in the first solvation shell. Also similar to the wider region, the absence of density depletion (around 2.0-2.5 Å in **Figure 4.14** (C)) in the SCC-DFTB results reflects a bifurcated hydrogen bond being donated by the hydronium. Therefore, the SCC-DFTB method displays similar defects in a more confined local environment. More results for the narrower region of the protonated LS2 channel can be found in the Supporting Information.

(A)



(C)



(B)

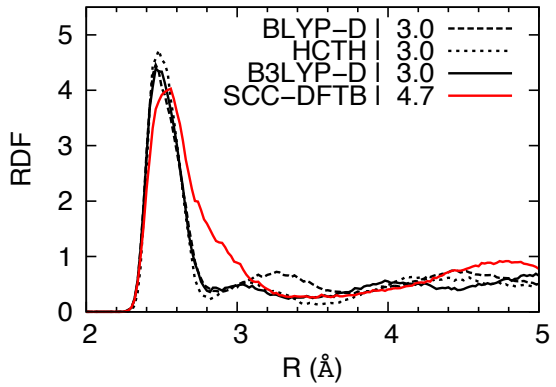


Figure 4.14 Radial distribution functions of (A) O*-O_w, (B) O*-O where O includes both O_w and O_s, and (C) H*-O_w, in a narrower region of protonated LS2 channel. The coordination numbers in panels A and B indicate that SCC-DFTB's overcoordination of O* remains in the narrow region of the channel. The absence of density depletion in panel C (around 2-2.5 Å) in the SCC-DFTB results reflects the encroaching 2nd solvation shell water and bifurcated hydrogen bond donated by the hydronium ion are very similar to the case of the wider region (compare to **Figure 4.8** (A)).

4.3.4 Proton solvation and transport in (8,8) single-wall carbon nanotube

The QM/MM electrostatic coupling scheme employed in the current work follows that of ref (16). The electrostatic interaction between SCC-DFTB atoms and MM atoms are

calculated based on point charges. The point charges on QM atoms are the Mulliken charges Δq_α defined in the SCC-DFTB energy expression (Eq 1). In the recent work of Hou et al.(51) a new QM/MM coupling scheme was implemented for SCC-DFTB/MM, which takes into account the finite size of QM and MM charge distributions by using a Klopman-Ohno functional form for the QM/MM electrostatic coupling. The new SCC-DFTB/MM coupling scheme improves the short range QM/MM interactions when the QM atoms are highly charged. In the LS2 channel some of the QM atoms (the excess proton and solvating water molecules) are positively charged and are close to the serine side chains described by MM. In order to test whether the inaccuracies of the SCC-DFTB/MM LS2 simulations described above are a result of the original point charge based electrostatic coupling scheme,(16) a hydrophobic proton channel system was created from a (8,8) single-wall carbon nanotube (CNT) filled with water molecules and one excess proton. In this system the QM and MM atoms only interact via the same LJ potential for all the methods under comparison (see section 3.2.3). The effective pore radius of the CNT (~ 4 Å) is comparable to the wide region of the LS2 channel, accommodating one complete solvation shell around the hydronium O* (i.e., the $H_9O_4^+$ Eigen complex). Similar properties were calculated for the CNT as were for the LS2 channel, and are plotted in **Figure 4.7** (C), (D), and **Figure 4.8**-**Figure 4.11** (B). These results reveal that the SCC-DFTB inaccuracies discussed above for LS2 remain in the CNT, including the overcoordinated excess proton, weak hydrogen bond around the excess proton CEC, and PT dynamics at odds with the DFT results. Therefore, the inaccurate description of PT in the LS2 channel by SCC-DFTB/MM should not be

attributed to using an approximate electrostatic coupling scheme, but rather to inherent limitations of SCC-DFTB method itself.

4.4 Conclusions

This benchmark study has shown that the SCC-DFTB method provides qualitatively different behavior for the solvation and transport of an excess proton in the LS2 channel. These differences, when compared to more exact DFT methods which the SCC-DFTB approach is meant to approximate, include a less ordered water structure, an over-coordinated excess proton, weak hydrogen bonding around the excess proton CEC, and altered PT dynamics. These differences are also seen for the hydrated excess proton in a carbon nanotube involving only van der Waals interactions between QM and MM atoms, verifying that the SCC-DFTB inaccuracies relative to DFT in the LS2 channel result from the semiempirical QM method itself, rather than the QM/MM electrostatic coupling scheme. In addition, SCC-DFTB is shown to predict higher local water density in channel environments, similar to that which has been shown in a previous study to cause clearly unphysical large pockets of vacuum in SCC-DFTB simulations of bulk aqueous systems.(4, 5) The less accurate descriptions of water and the hydrated excess proton results in qualitatively and quantitatively different structural and dynamical properties for PT in confined channel environments, calling into question the ability of the current SCC-DFTB method to accurately simulate excess proton transport in realistic biomolecular systems. At the same time, the presented results provide an important set of benchmarks by which the method can potentially be improved in the future to bring it more in line with more accurate, but more computationally costly, DFT-based QM/MM methods.

4.5 References

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

2. Elstner M (2006) The SCC-DFTB method and its application to biological systems. *Theor Chem Acc* 116(1-3):316-325.

3. Gaus M, Cui QA, & Elstner M (2011) DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J Chem Theory Comput* 7(4):931-948.

4. Maupin CM, Aradi B, & Voth GA (2010) The Self-Consistent Charge Density Functional Tight Binding Method Applied to Liquid Water and the Hydrated Excess Proton: Benchmark Simulations. *J Phys Chem B* 114(20):6922-6931.

5. Choi TH, Liang R, Maupin CM, & Voth GA (2013) Application of the SCC-DFTB Method to Hydroxide Water Clusters and Aqueous Hydroxide Solutions. *J Phys Chem B* 117(17):5165-5179.

6. Goyal P, Elstner M, & Cui Q (2011) Application of the SCC-DFTB Method to Neutral and Protonated Water Clusters and Bulk Water. *J Phys Chem B* 115(20):6790-6805.

7. Gaus M, Goez A, & Elstner M (2013) Parametrization and Benchmark of DFTB3 for Organic Molecules. *J Chem Theory Comput* 9(1):338-354.

8. Petraglia R & Corminboeuf C (2013) A Caveat on SCC-DFTB and Noncovalent Interactions Involving Sulfur Atoms. *J. Chem. Theory Comput.* 9(7):3020-3025.

9. Goyal P*, et al.* (2011) Proton Storage Site in Bacteriorhodopsin: New Insights from Quantum Mechanics/Molecular Mechanics Simulations of Microscopic pK(a) and Infrared Spectra. *J Am Chem Soc* 133(38):14981-14997.

10. Riccardi D, Yang S, & Cui Q (2010) Proton transfer function of carbonic anhydrase: Insights from QM/MM simulations. *Bba-Proteins Proteom* 1804(2):342-351.

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

12. Lear JD, Wasserman ZR, & Degrado WF (1988) Synthetic Amphiphilic Peptide Models for Protein Ion Channels. *Science* 240(4856):1177-1181.

13. Wu YJ, Ilan B, & Voth GA (2007) Charge delocalization in proton channels, II: The synthetic LS2 channel and proton selectivity. *Biophys J* 92(1):61-69.

14. Voth GA (2003) The computer simulation of proton transport in biomolecular systems. *Front Biosci* 8:S1384-S1397.

15. Wu YJ & Voth GA (2003) A computer simulation study of the hydrated proton in a synthetic proton channel. *Biophys J* 85(2):864-875.

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

17. Yang Y, Yu HB, York D, Cui Q, & Elstner M (2007) Extension of the self-consistent-charge density-functional tight-binding method: Third-order expansion of the density functional theory total energy and introduction of a modified effective coulomb interaction. *J Phys Chem A* 111(42):10861-10873.

18. MacKerell AD*, et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102(18):3586-3616.

19. Mackerell AD, Feig M, & Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25(11):1400-1415.

20. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein ML (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 79(2):926-935.

21. Hockney RW & Eastwood JW (1988) *Computer Simulation Using Particles* (Taylor & Francis).

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

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

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

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

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

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

28. Boese AD, Doltsinis NL, Handy NC, & Sprik M (2000) New generalized gradient approximation functionals. *J. Chem. Phys.* 112(4):1670-1678.

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

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

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

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

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

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

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

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

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

38. Cornell WD*, et al.* (1995) A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* 117(19):5179-5197.

39. Randa HS, Forrest LR, Voth GA, & Sansom MSP (1999) Molecular dynamics of synthetic leucine-serine ion channels in a phospholipid membrane. *Biophys J* 77(5):2400-2410.

40. Lin IC, Seitsonen AP, Coutinho-Neto MD, Tavernelli I, & Rothlisberger U (2009) Importance of van der Waals Interactions in Liquid Water. *J Phys Chem B* 113(4):1127-1131.

41. Izvekov S & Voth GA (2005) Ab initio molecular-dynamics simulation of aqueous proton solvation and transport revisited. *J Chem Phys* 123(4).

42. Izvekov S & Swanson JMJ (2011) Using force-matching to reveal essential differences between density functionals in ab initio molecular dynamics simulations. *J Chem Phys* 134(19).

43. VandeVondele J*, et al.* (2005) The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water. *J Chem Phys* 122(1).

44. Schmidt J*, et al.* (2009) Isobaric-Isothermal Molecular Dynamics Simulations Utilizing Density Functional Theory: An Assessment of the Structure and Density of Water at Near-Ambient Conditions. *J Phys Chem B* 113(35):11959-11964.

45. Rablen PR, Lockman JW, & Jorgensen WL (1998) Ab initio study of hydrogen-bonded complexes of small organic molecules with water. *J. Phys. Chem. A* 102(21):3782-3797.

46. Shepherd LMS & Morrison CA (2010) Simulating Proton Transport through a Simplified Model for Trans-Membrane Proteins. *J Phys Chem B* 114(20):7047-7055.

47. Luzar A & Chandler D (1996) Hydrogen-bond kinetics in liquid water. *Nature* 379(6560):55-57.

48. Bandyopadhyay S, Chakraborty S, & Bagchi B (2005) Secondary structure sensitivity of hydrogen bond lifetime dynamics in the protein hydration layer. *J. Am. Chem. Soc.* 127(47):16660-16667.

49. Sinha SK & Bandyopadhyay S (2012) Local heterogeneous dynamics of water around lysozyme: a computer simulation study. *Phys Chem Chem Phys* 14(2):899-913.

50. Markovitch O*, et al.* (2008) Special pair dance and partner selection: Elementary steps in proton transport in liquid water. *J Phys Chem B* 112(31):9456-9466.

51. Hou GH, Zhu X, Elstner M, & Cui Q (2012) A Modified QM/MM Hamiltonian with the Self-Consistent-Charge Density-Functional-Tight-Binding Theory for Highly Charged QM Regions. *J. Chem. Theory Comput.* 8(11):4293-4304.

Chapter 5

Multiscale simulation reveals a multifaceted mechanism of proton permeation through the influenza A M2 proton channel

5.1 Introduction

The influenza type A virus is a highly pathogenic RNA virus that causes flu in birds and mammals (1). The influenza A M2 (AM2) protein (2) contains a homotetramer channel that transports protons across the viral membrane and acidifies the virion interior, enabling the dissociation of the viral matrix proteins, which is a crucial step in viral replication (3). The protein has been the target of antiviral drugs amantadine and rimantadine (4, 5). Much effort has been devoted to discovering the structure and proton transport (PT) mechanism of the AM2 channel, resulting in many crystal structures available in the protein data bank (6-14). Based on the crystal structures and electrophysiology experiments, several PT models have been suggested. These mechanisms can be divided into two main categories, delineated by the role of the four histidine residues (aka the His37 tetrad), that reside in the middle of the AM2 transmembrane domain (AM2/TM, see **Figure 5.1** (A)), which has been experimentally shown (15) to account for the proton permeation behavior of the full AM2 protein. The 'shutter' mechanism (16, 17) suggests that the His37 tetrad works as a gate. At low pH the gate opens due to the electrostatic repulsion between the bi-protonated, positively-charged histidine residues. The excess proton is then transferred through continuous water wire via the Grotthuss mechanism, without changing the protonation state of His37.

In contrast, the 'shuttle' mechanism (14, 18-22) suggests that at acidic pH values the His37 tetrad changes protonation states (see **Figure 5.1** (B)) as it shuttles the proton through the His37-Trp41 quartet region (Trp41 lies just below His37 and is also thought to play an important role in the PT mechanism).

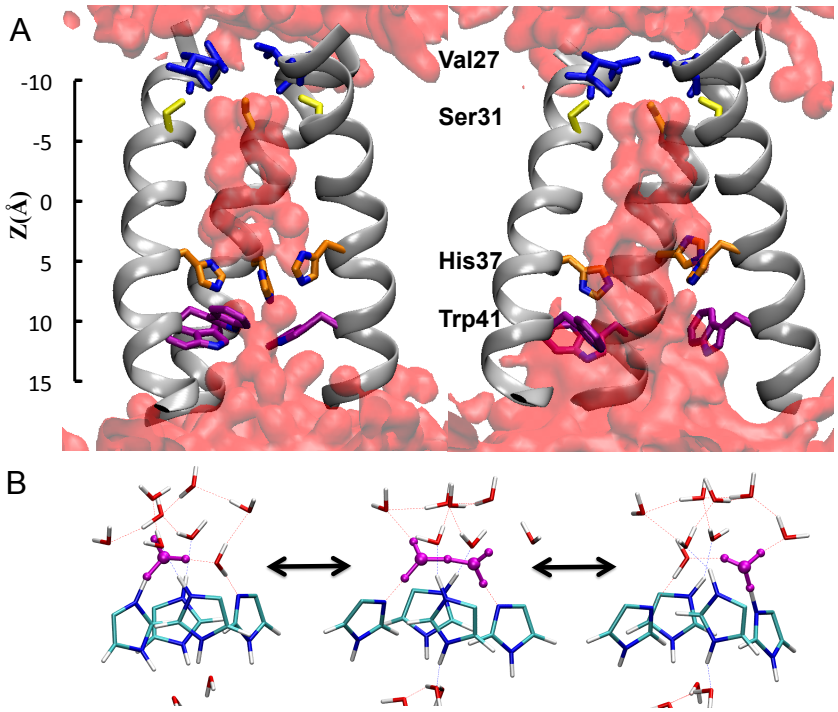


Figure 5.1 (A) Equilibrated structures of the AM2 channel transmembrane domain (see text for definition) and (B) The excess proton (purple) delocalized around entry water cluster and histidine residues.

The structure and dynamics of water in the AM2 protein has also gained attention, since it is the essential medium for proton permeation. The high-resolution crystal structure (PDB code 3LBW) crystallized at pH 6.5 revealed layers of well-ordered water clusters above the His37 tetrad (12). However, the water dynamics in the AM2 protein probed using two-dimensional infrared (2D-IR) spectroscopy revealed that the well-ordered "ice-like" pore water dynamics at pH = 8.0 change to more mobile and "liquid-like" dynamics (on the timescale of a few picoseconds) at pH = 3.2 (23). This result

suggests an interesting pH-dependent behavior of the AM2 protein that is highly relevant for understanding its PT mechanism.

Although many computational studies have investigated the features of AM2 that may influence its PT in recent years (12, 24-40), only a few have explicitly simulated any aspect of the explicit PT process (24, 25, 29, 37, 38). This is because it is challenging to accurately model the charge delocalization and Grotthuss shuttling of the hydrated excess proton in a computationally tractable way. Among these studies, only one so far has provided a free-energy profile (potential of mean force, or PMF) for PT across the entire AM2/TM channel (25), which is the essential property for understanding the full proton permeation mechanism. However, this particular free energy study was limited by the approximation that the His37 tetrad remains in a fixed protonation state during the proton conduction; therefore, it could not capture the more plausible shuttle mechanism (14, 18-22). More recently, Carnevale et. al (29) used a quantum mechanics/molecular mechanics (QM/MM) approach to investigate PT in the specific region of the His37 tetrad of AM2. Although they could not achieve sufficient sampling to calculate a free energy profile, this work did allow the His37 tetrad to change protonation states as necessitated by the shuttle mechanism. Their work in fact helped to lay the foundation for the study of PT in the His37 tetrad using a QM/MM approach in that region of the channel.

In the present study, a powerful multiscale combination of classical, reactive, and *ab initio* (QM/MM) molecular dynamics (MD) simulations is employed to systematically investigate the proton solvation and transport mechanism through the full AM2 protein. By employing a classical force field, the conformational ensemble of AM2/TM is

characterized starting from two recent high-resolution structures in low and intermediate pH conditions (PDB code 3C9J (9) and 3LBW (12), respectively). The influence of pH on the protein and water dynamics is then investigated. It is found that as the pH is lowered, the channel adopts a more open conformation, the channel pore is more hydrated, and the pore water molecules are more mobile. Following this we utilize a synthesis of the reactive multi-state empirical valence bond (MS-EVB) (41) and QM/MM approaches to calculate complete proton permeation free energy profiles (PMFs) through the AM2/TM channel. Since the MS-EVB potential can be derived by force-matching of *ab initio* MD simulation data (42), and since the force-matching algorithm provides a PMF for the reference potential, this unique combination of approaches yields a consistent multiscale computational framework for obtaining an *ab intio* level quality PT free energy profile. This multiscale combination of methods is thus used to calculate the PMFs for excess protons permeating through the entire AM2/TM channel, accounting for the charge delocalization and shuttling of hydrated excess protons through both water and the His37 tetrad explicitly. By virtue of this methodology, it is found that at lower pH values the free energy barrier for a proton to diffuse to His37 is decreased and the proton conductance increased, explaining the unique pH-dependent activation mechanism of AM2. We also report the PMFs of several AM2 mutants, including the amantadine-resistant and transmissible mutants V27A and S31N. The calculated conductances of both the wild-type AM2 and its mutants are in close agreement with experimental results and provide microscopic explanations for their trends.

5.2 Materials and Methods

5.2.1 Modeling of the pH-Dependent Conformational Ensemble and Water Dynamics.

Two high-resolution structures for the G34A mutant (3BLW and 3C9J - crystallized at pH=6.5 and pH=5.3, respectively) (9, 12), were used as the initial structures for MD simulations. To provide a clear picture of the channel conformation change upon acidification, the His37 tetrad was modeled at each possible protonation state. For the Q1, Q2, and Q3 states the neutral histidine residues are assigned to be in the ε tautomer form (12). Further details for system equilibration using classical force field and the simulation results for the different starting conformations are provided in the Appendix B

.

5.2.2 Modeling of Proton Permeation.

The solvation and transport of an excess proton requires special treatment due to its delocalized nature and the dynamic rearrangement of covalent and hydrogen bonding via the Grotthuss mechanism. To characterize free energy profiles of PT in the AM2/TM, umbrella sampling was used. The collective variable was defined by the difference between the z coordinate of the center of mass of C_α atoms of the four Gly34 residues and that of the excess proton center of excess charge. The collective variable was restrained by a harmonic potential in the direction of the channel's principal axis (z direction). The simulations of the windows centered outside the His37 tetrad ($z < -4$ Å or $z > 12$ Å for the intermediate pH PMF and $z < 1$ Å or $z > 10$ Å for the low pH PMF) were performed

with the multi-state empirical valence bond version 3 (MS-EVB3) model (43). Details of the MS-EVB methodology are provided in previous studies and reviews (41, 42, 44-48). The MS-EVB simulations were performed using the RAPTOR software (49) interfaced with the LAMMPS MD package (http://lammps.sandia.gov) (50). For the simulations of windows centered near and within the His37 tetrad (-4 Å < z < 12 Å for the intermediate pH PMF and 1 Å < z < 10 Å for the low pH PMF), a QM/MM MD approach was used to describe the excess charge delocalized around the ionizable His37 tetrad and nearby water molecules. The QM region, including the His37 tetrad and surrounding water molecules, was treated with density functional theory using the Becke exchange (51) and Lee–Yang–Parr correlation (52) functionals, corrected by semiempirical dispersion terms (53). The QM/MM simulations were performed with the CP2K software package (54). Further details on the MS-EVB and QM/MM simulations are provided in Appendix B .

The final PMFs for PT across the entire channel were generated in multiscale fashion by combining MS-EVB and QM/MM windows obtained using WHAM (55-57). The statistical errors were estimated by a block-average analysis. For the mutants, the trajectories for the wild type QM/MM windows were used in the WHAM calculations. (See Appendix B for more details.)

The single channel maximum proton conductances were estimated by using the one-dimensional approximation of the Poisson-Nernst-Planck (PNP) theory, as described in Appendix B . The PT PMFs through the channel region, and diffusion constants, are used to estimate the proton conductances. This approach to calculate conductance is justified by ref (58).

5.3 Results and Discussion

5.3.1 pH-Dependent Conformational Ensemble and Water Dynamics.

As described in recent measurements of the four pKa values for the His37 tetrad (20, 59), lowering the pH corresponds to increasing the number of protonated His37 residues from one to four. To illustrate the dependence of AM2/TM conformation on pH value, eight equilibrated AM2/TM structures, representing each possible protonation state and initially aligned with either PDB code 3LBW or 3C9J, are shown in **Figure 5.2**. The notation Q1, Q2, Q3, and Q4 is used to denote the His37 tetrad in its +1, +2, +3 and +4 protonation states, respectively. The two starting crystal structures 3C9J and 3LBW (denoted by prefixes S and D, respectively) were crystalized at different pH values and thought to potentially represent different protonation states of the His37 tetrad. After extensive simulation in the high pH environment (namely the Q1 and Q2 states), the Trp41 residues at the C-terminus are tightly packed in the C4 symmetry, similar to the 3LBW structure and regardless of the starting structure. See Appendix B for more details. As shown in **Figure 5.2**, lowering the pH value (i.e., increasing the number of positively charged His37 residues to the Q3 and Q4 states) resulted in a more open packing of the α-helices at C-terminus, allowing more water molecules to enter the channel (27). To probe the change of pore water dynamics upon lowering pH value in an experimentally verifiable way, we calculated the lifetimes of the Gly34 amide-I group hydrogen bonds (HBs) for various protonation states (**Table 5.1**). The HB donor was defined as water oxygen atoms, and the HB acceptor was the Gly34 carbonyl oxygen atoms. The HB autocorrelation function $C(\tau) = \langle s_i(t)s_i(t+\tau)\rangle$ was calculated for the HB between Gly34 and water (60), where $s_i(t)$ is unity if the HB exists at time t, and 0

otherwise. The autocorrelation function was then integrated to obtain the average HB lifetime. The high-pH simulations (Q1, Q2) feature long HB relaxation times (6.1-9.4 ps), suggesting the water molecules around each Gly34 carbonyl group remain immobilized. In contrast, the low-pH simulations indicate faster water reorientation, with the Q4 state HB lifetime being 0.9-2.4 ps and that of the Q3 state being 2.7-3.2 ps. In a recent study, Ghosh et al. probed the hydration of Gly34 carbonyl group by 2D-IR spectroscopy and reported pH-dependent water dynamics in the AM2/TM (23). At pH 6.2, the spectral dynamics has a relaxation time of ~1.3 ps, whereas at pH=8.0 the relaxation time is ~10 ps (23). Our simulations quantitatively capture the experimentally reported acceleration of channel water HB dynamics upon lowering the pH value. An important consequence of this is that PT in the water cluster would be hindered by the slow water dynamics at high pH values, whereas it would be facilitated at low pH values. This will be discussed further in the next section.

Table 5.1 Calculated dominant H-bond lifetime (ps), between pore-lining residue Gly34 amide-I groups and their nearest water molecules.The simulations are started from the 3LBW crystal structure. The Q2, Q3, Q4 notation denotes +2, +3 and +4 protonation states of the His37 tetrad in the classical MD simulation, respectively. Experimental values are taken from Ref. (23).

Helix	Q2	Q3	Q4
A	9.4	2.7	1.4
B	6.0	2.9	1.0
C	9.1	2.7	1.4
D	6.1	2.8	2.2
Experimental	~10	~1.3	

5.3.2 Proton Permeation in WT AM2.

The quantity of interest and of most value in understanding proton permeation through AM2 is the free energy profile (PMF) for the PT. Although finding the reaction path of proton permeation through a simple channel is not a significant challenge, combining a simulation methodology that captures the essential physics with a sampling protocol that enables statistical convergence is. These features are especially challenging for a system like AM2 where multiple reactive sites (e.g., four protonatable His37 residues) are in close proximity such that the reactive pathway through them is not clear and their coupled participation in the reaction is likely. To capture PT through the His37 tetrad with the MS-EVB method alone would require developing titratable MS-EVB models for the four His37 moieties that simultaneously allowed for proton transfer between each other and with the surrounding solvent (44). This approach would further require the treatment of multiple shuttling excess protons in the His37 tetrad region. Although

possible, these MS-EVB models can be hard to develop and benchmark in highly complex environments. In contrast, QM analysis of the His37 tetrad is more likely to accurately treat the requisite charge delocalization and multiple proton shuttling events between the histidine residues. Given the size of the full system, QM/MM simulation is necessary to capture the surrounding environment. On the other hand, the cost of QM/MM analysis on a system this large also limits the sampling to tens to hundreds ps, and is not feasible for simulating PT over the entire channel length. This limitation is especially true when slower conformational restructuring of the protein is important. Indeed, as described below, both solvation and protein structural changes are coupled to PT in the regions above and below the His37 tetrad. Hence much longer sampling times are required to converge the full channel PT free energy profile than can be obtained with QM/MM MD simulations alone.

In the present work, the challenges described above are overcome by combining the advantages of the MS-EVB and QM/MM methods to calculate the full free energy profile for PT across the entire AM2/TM channel. The PT through the regions above and below the His37-Trp41 quartet are described with the MS-EVB method, which has been extensively validated for water mediated PT in many systems (41, 45-48), while the PT through the His37-Trp41 quartet is described with the QM/MM MD method to allow for the explicit protonation and deprotonation of the His37 tetrad. Two complete potentials of mean force (PMFs) corresponding to the proton permeation process at different pH conditions are then investigated by bridging the MS-EVB and QM/MM PMFs in a multiscale fashion (as explained in Materials and Methods and Appendix B). The first PMF (denoted the +2 PMF) describes the PT process at an intermediate pH value,

wherein the channel begins in the Q2 state (2 bi-protonated His37 residues) and a third excess proton passes through the channel. The second PMF (denoted the +3 PMF) describes the process at lower pH value with the channel in the Q3 state and a fourth proton passing through. The charge states for the two processes are chosen based on the experimental finding that AM2 conduction occurs when three or more His37 residues are protonated (20, 59) (which includes in our case the translocating proton).

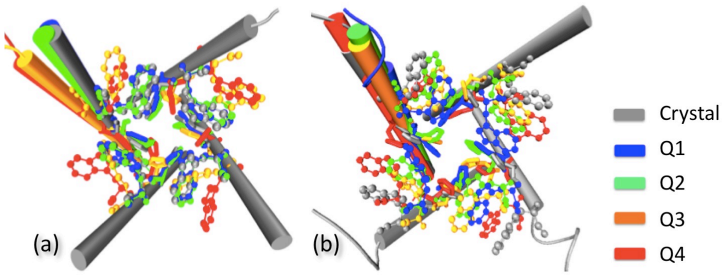


Figure 5.2 Equilibrated AM2 channel transmembrane domain structures at various protonation states and aligned with either starting crystal structure 3LBW (a) or 3C9J (b). The notation Q1, Q2, Q3, and Q4 is used to denote the His37 tetrad in its +1, +2, +3 and +4 protonation states, respectively. His37 are shown with the licorice model, and Trp41 are illustrated with the ball-stick model. Helices B, C and D in the equilibrated structures are removed, for the purpose of clarity.

The +2 PMF (**Figure 5.3**) reveals several important events along the proton conduction pathway. First, there is a small free energy minimum as the excess proton reaches the interface between bulk water and the hydrophobic Val27 residues ($z \sim -14$ Å). This minimum is caused by the anisotropic solvation structure of the hydrated excess proton, which makes it somewhat amphiphilic in nature. Such interfacial enrichment of the hydrated excess proton has been predicted by MS-EVB simulations (61), and subsequently confirmed in experiments (62, 63). Second, the free energy around Val27 (-13 Å$< z <-10$ Å) increases due to proton desolvation and hydrophobic interactions with

the Val27 tetrad. Unlike previous classical MD simulations (26) that were lacking a shuttling explicit excess proton in the system, our results can now confirm an important role of Val27 for the PT through AM2. Third, and possibly most important, there is a large free energy barrier for the proton to diffuse from Val27 to the His37 tetrad. Several factors contribute to this free energy barrier. As discussed above and in Ref (23), the water is more ice-like inside the channel in the Q2 state and exhibits slower orientational dynamics. Since a successful proton transfer event from one water molecule to the next requires collective rearrangement of the surrounding HB network (64, 65), the slow dynamics of channel water inhibits PT. Furthermore, the positively charged His37 tetrad (+2 in the Q2 state) both repels the excess proton and creates an electric field that orients the water dipoles in a way further increases this free energy barrier.

When the excess proton reaches the so called 'entry' water cluster (12) (two solvation shells above the His37 tetrad, -4 Å< z < 0 Å), there is a plateau in the free energy. The flat free energy profile for PT in the entry cluster agrees well with the more limited QM/MM sampling result of Carnevale et. al. (29), which suggested that the excess proton diffuses across this region in a nearly barrierless fashion. The plateau is likely caused by a balancing of the excess proton's charge interactions – repulsive from the positively charged His37 residues but attractive to the unprotonated N_δ of the neutral histidines to which it forms water-mediated hydrogen bonds. As the excess proton then protonates the 3rd histidine, the system transitions from the Q2 to the Q3 state and there is a free energy decrease in the PMF culminating in a deep minimum ($z \approx 6$ Å). This result verifies that the shuttle mechanism, involving the explicit protonation of His37, dominates and that the excess proton is substantially stabilized by charge delocalization among the His37

tetrad and nearby water molecules. The stability of the triply charged His37 tetrad is also in line with the result of Carnevale et. al. (29), who did not calculate a PMF but did observe in their QM/MM simulations that an unrestrained excess proton starting from the entry cluster prefers to protonate the uncharged His37 residue, while no proton in the charged His37 residue was transferred to the entry cluster. Following this, the deprotonation of histidine gives rise to a large free energy barrier of ~10 kcal/mol. Such a large deprotonation barrier would make histidine deprotonation the rate-limiting step in the overall proton permeation process, as suggested in previous experimental work (14, 66). Finally, after the proton passes the Trp41 tetrad there is large free energy decrease, consistent with the fact that it is energetically unfavorable for the excess proton to reach the His37 tetrad from virus interior. The energetic cost of this reverse process may be ascribed to proton desolvation, steric hindrance from the Trp41 tetrad, and repulsion from the positively charged histidines.

Unlike other studies to date, the simulations reported herein were not limited to the protonation/deprotonation of one presumed histidine residue. Since our collective variable, which defines the location of the excess proton, can be fully delocalized among the entire histidine tetrad and surrounding water molecules (see Appendix B), all possible protonation/deprotonation processes were accounted for. It was found that more than one histidine residue participates in proton shuttling. In fact, once the Q3 state is reached, the three positively charged histidines are nearly indistinguishable due to charge delocalization. It is actually the solvation structure below the histidines that determines which residue gets deprotonated. The most likely candidate is the residue that has the most accessible water molecule below it, not the His37 that was protonated by the

entering excess proton as one might assume. As shown in **Table 5.2**, it is encouraging that the calculated conductance (see Appendix B for details) agrees well with recent experimental values (67).

The free energy profile for PT at low pH (the +3 PMF in **Figure 5.3**) is similar in shape to the +2 PMF, but with lower barriers. In this process, the channel is more solvated such that continuous water wires form around the histidine residues. Since our collective variable definition does not impose an *a priori* bias towards either the 'shutter' mechanism or 'shuttle' mechanism (see Appendix B), at lower pH one might expect to see the proton diffuse more along the water wire via the Grotthuss mechanism (i.e., the 'shutter' mechanism) rather than protonation of a His37 residue. However, this is not the case. The histidine protonation/deprotonation (i.e., the 'shuttle' mechanism) process still occurs. The calculated conductance from the low pH PMF is higher (14 fs), which is in reasonable agreement with the experimental conductance (20) and supports the unique acid-activation feature of this channel. The increased conductance can be ascribed to several factors. First, the histidine deprotonation barrier is lower due to the opening of the Trp gate and more water molecules accessible to His37 residues. Second, the water molecules between Val27 and His37 are more dynamic as described earlier, facilitating hydrogen bond network rearrangement and thus PT. Finally, the pore radius is larger and the presence of more water molecules increases the screening of the electrostatic repulsion between the His37 tetrad and the incoming proton, which further lowers the free energy barrier for PT from the Val27 to His37.

As proton passes through the histidine tetrad, it is again observed that more than one histidine is involved in protonation/deprotonation. In the Q4 state with four identical

positively charged histidine residues, the deprotonation process is controlled by the availability of water molecules.

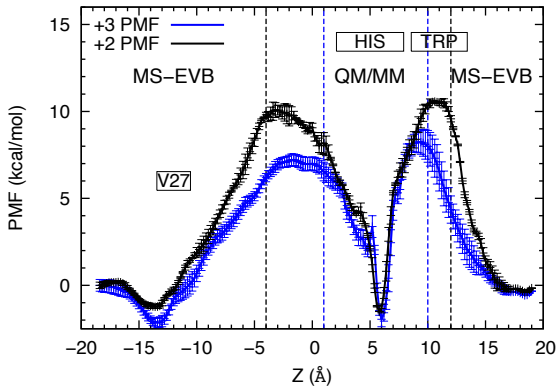


Figure 5.3 Proton transport +3 free energy profiles (PMFs) (blue) compared with +2 (black).The origin is the center of mass of 4 Cα of Gly34. The positions of Val27, His37 and Trp41 are labeled with text boxes. The regions sampled by MS-EVB and QM/MM are labeled and separated by black dashed vertical lines for +2 PMF, and blue dashed lines for +3 PMF.

Table 5.2 Conductance comparison between simulation and experiments calculated from the +2 process (defined in main text)

	Conductance (fS)
Experimental (56)*	1-4 fS
Simulation	1.2 fS

* For the experimental value, the conductance is measured at T=291K, pH_{out}=6 and extrapolated to T=310K (56).

5.3.3 Proton Permeation in AM2 Mutants.

In addition to wild type (WT) AM2, the PT PMF for several important mutants (30) were studied. The PMFs are plotted in **Figure 5.4** and the conductances relative to WT AM2

are listed in **Table 5.3**. For S31N, the overall free energy profile along the proton conduction pathway is similar to WT; thus, the conductance is similar to wild type. This is in line with the fact that a virus with the S31N mutation is just as transmissible as those with a WT AM2 channel (30). For V27A, the channel entrance radius is larger; thus, the entrance barrier height is reduced resulting in increased conductance. This behavior once again confirms the role of Val27 for secondary modulation of the proton permeation. In addition, the WT free energy minimum in this region disappears in this mutant, because the Ala27 residues are not bulky enough to form a hydrophobic interface with the bulk water. For V27R, there exists a free energy barrier before the proton enters the channel due to the Arg27 at the channel entrance electrostatically repelling the proton. The overall conductance is thus lowered compared to WT. Finally, it is encouraging that our simulation results reproduce the experimental conductance trends (30) (in nearly quantitative agreement), which further supports the reliability of the multiscale simulation approach taken herein. One approximation in the mutant PMF calculations is the use of WT QM/MM PMF for the His37-Trp41 region. This approximation was necessary due to the great computational cost of the QM/MM simulations, and was considered to be reasonable since the positions of the mutated residues are far from the His37-Trp41 region. It was further supported by calculating the pore radius and water density in the His37-Trp41 region, which remains nearly identical to WT (see **Figure B.3**).

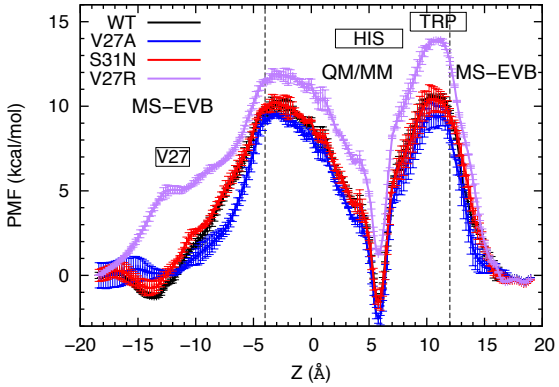


Figure 5.4 Proton transport free energy profiles (PMFs) for mutants compared with the WT AM2 channel. The origin is the center of mass of 4 Cα of Gly34. The positions of Val27, His37 and Trp41 are labeled by the text boxes. The regions sampled by MS-EVB and QM/MM are labeled and separated by black dashed vertical lines.

Table 5.3 Relative experimental (30) and calculated (present work) conductances for AM2 mutants normalized to the conductance of wild type (WT)

	Experiment	Simulation
WT	1.0	1.0
S31N	1.3	1.4
V27A	1.5	2.2
V27R	0.2	0.3

5.4 Conclusions

We have presented the complete free energy profile and conductance calculations for proton permeation through the influenza A M2 channel, as well as several key experimental mutants. The results were obtained via a combination of classical, reactive, and *ab initio* MD simulations. Classical MD simulations first revealed that as the pH is

lowered, the channel is expanded to accommodate more water molecules with increased mobility, in agreement with experimental 2D-IR results. Free energy profiles (PMFs) were then calculated using a multiscale simulation approach for both the intermediate and low pH ranges. These provide explicit confirmation of current proposals for the AM2 conduction mechanism, such as the importance of V27 for proton channel entry, the role of the His37 tetrad in excess proton charge delocalization (i.e., the 'shuttle' mechanism), the barrier for His37 deprotonation, and a role for the Trp41 conformation. Importantly, the complete free energy profiles also provide new insights. First, the presence of a large barrier (~ 10 kcal/mol) for an excess proton to shuttle to the His37 tetrad has not been previously discussed. This barrier is caused by a combination of proton desolvation, charge repulsion from the positively charged His37 tetrad, and immobile water in this region. Interestingly, this barrier decreases at lower pH value despite the increased charge of the His37 tetrad. The decreased barrier is induced by channel broadening, increased solvation, and more mobile hydrogen bond dynamics. For both pH conditions, the deprotonation of the His37 tetrad is rate limiting, and largely influenced by the water structure and dynamics below the histidines and through the Trp41 tetrad. The histidine shuttling can involve more than one residue, and at low pH condition the 'shuttle' mechanism is still present. The mutant studies, which are in good agreement with experimental conductance results, not only help to validate our computational methodology for studying the AM2 channel, but point to various physical explanations for the altered mutant conductance rates. The insight presented in this work, particularly the presence of multiple targetable portions of the channel and their relationship to the proton conductance, may help guide future efforts to alter or block the AM2 channel.

5.5 References

1. Cannell JJ, Zasloff M, Garland CF, Scragg R, & Giovannucci E (2008) On the epidemiology of influenza. *Virol J.* 5:29.

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

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

4. Hay AJ, Wolstenholme AJ, Skehel JJ, & Smith MH (1985) The Molecular-Basis of the Specific Anti-Influenza Action of Amantadine. *EMBO J.* 4(11):3021-3024.

5. Wang C, Takeuchi K, Holsinger LJ, Lamb RA, & Pinto LH (1993) Ion Channel Activity of the M2 Protein of Influenza a-Virus. *Biophys. J.* 64(2):A94-A94.

6. Wang JF, Kim S, Kovacs F, & Cross TA (2001) Structure of the transmembrane region of the M2 protein H+ channel. *Protein Sci.* 10(11):2241-2250.

7. Nishimura K, Kim S, Zhang L, & Cross TA (2002) The Closed State of a H^+ Channel Helical Bundle Combining Precise Orientational and Distance Restraints from Solid State NMR. *Biochemistry* 41(44):13170-13177.

8. Schnell JR & Chou JJ (2008) Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* 451(7178):591-595.

9. Stouffer AL*, et al.* (2008) Structural basis for the function and inhibition of an influenza virus proton channel. *Nature* 451(7178):596-U513.

10. Cady SD, Mishanina TV, & Hong M (2009) Structure of Amantadine-Bound M2 Transmembrane Peptide of Influenza A in Lipid Bilayers from Magic-Angle-Spinning Solid-State NMR: The Role of Ser31 in Amantadine Binding. *J. Mol. Biol.* 385(4):1127-1141.

11. Cady SD*, et al.* (2010) Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. *Nature* 463(7281):689-692.

12. Acharya R*, et al.* (2010) Structure and mechanism of proton transport through the transmembrane tetrameric M2 protein bundle of the influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* 107(34):15075-15080.

13. Pielak RM & Chou JJ (2010) Solution NMR structure of the V27A drug resistant mutant of influenza A M2 channel. *Biochem. Biophys. Res. Commun.* 401(1):58-63.

14. Sharma M, *et al.* (2010) Insight into the Mechanism of the Influenza A Proton Channel from a Structure in a Lipid Bilayer. *Science* 330(6003):509-512.

15. Ma CL, *et al.* (2009) Identification of the functional core of the influenza A virus A/M2 proton-selective ion channel. *Proc. Natl. Acad. Sci. U. S. A.* 106(30):12283-12288.

16. Sansom MSP, Kerr ID, Smith GR, & Son HS (1997) The Influenza A Virus M2 Channel: A Molecular Modeling and Simulation Study. *Virology* 233(1):163-173.

17. Okada A, Miura T, & Takeuchi H (2001) Protonation of Histidine and Histidine-Tryptophan Interaction in the Activation of the M2 Ion Channel from Influenza A Virus. *Biochemistry* 40(20):6053-6060.

18. Pinto LH, *et al.* (1997) A functionally defined model for the M2 proton channel of influenza A virus suggests a mechanism for its ion selectivity. *Proceedings of the National Academy of Sciences* 94(21):11301-11306.

19. Hong M, Fritzsching KJ, & Williams JK (2012) Hydrogen-Bonding Partner of the Proton-Conducting Histidine in the Influenza M2 Proton Channel Revealed From H-1 Chemical Shifts. *J. Am. Chem. Soc.* 134(36):14753-14755.

20. Hu FH, Schmidt-Rohr K, & Hong M (2012) NMR Detection of pH-Dependent Histidine-Water Proton Exchange Reveals the Conduction Mechanism of a Transmembrane Proton Channel. *J. Am. Chem. Soc.* 134(8):3703-3713.

21. Hu FH, Luo WB, & Hong M (2010) Mechanisms of Proton Conduction and Gating in Influenza M2 Proton Channels from Solid-State NMR. *Science* 330(6003):505-508.

22. Williams JK, *et al.* (2013) Drug-Induced Conformational and Dynamical Changes of the S31N Mutant of the Influenza M2 Proton Channel Investigated by Solid-State NMR. *J. Am. Chem. Soc.* 135(26):9885-9897.

23. Ghosh A, Qiu J, DeGrado WF, & Hochstrasser RM (2011) Tidal surge in the M2 proton channel, sensed by 2D IR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* 108(15):6115-6120.

24. Smondyrev AM & Voth GA (2002) Molecular dynamics simulation of proton transport through the influenza A virus M2 channel. *Biophys. J.* 83(4):1987-1996.

25. Chen HN, Wu YJ, & Voth GA (2007) Proton transport Behavior through the influenza a M2 channel: Insights from molecular simulation. *Biophys. J.* 93(10):3470-3479.

26. Yi M, Cross TA, & Zhou HX (2008) A secondary gate as a mechanism for inhibition of the M2 proton channel by amantadine. *J. Phys. Chem. B* 112(27):7977-7979.

27. Khurana E*, et al.* (2009) Molecular dynamics calculations suggest a conduction mechanism for the M2 proton channel from influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* 106(4):1069-1074.

28. Yi MG, Cross TA, & Zhou HX (2009) Conformational heterogeneity of the M2 proton channel and a structural model for channel activation. *Proc. Natl. Acad. Sci. U. S. A.* 106(32):13311-13316.

29. Carnevale V, Fiorin G, Levine BG, DeGrado WF, & Klein ML (2010) Multiple Proton Confinement in the M2 Channel from the Influenza A Virus. *J. Phys. Chem. C* 114(48):20856-20863.

30. Balannik V*, et al.* (2010) Functional Studies and Modeling of Pore-Lining Residue Mutants of the Influenza A Virus M2 Ion Channel. *Biochemistry* 49(4):696-708.

31. Phongphanphanee S, Rungrotmongkol T, Yoshida N, Hannongbua S, & Hirata F (2010) Proton Transport through the Influenza A M2 Channel: Three-Dimensional Reference Interaction Site Model Study. *J. Am. Chem. Soc.* 132(28):9782-9788.

32. Zhou HX (2010) Diffusion-Influenced Transport of Ions across a Transmembrane Channel with an Internal Binding Site. *J. Phys. Chem. Lett.* 1(13):1973-1976.

33. Wang J*, et al.* (2011) Molecular Dynamics Simulation Directed Rational Design of Inhibitors Targeting Drug-Resistant Mutants of Influenza A Virus M2. *J. Am. Chem. Soc.* 133(32):12834-12841.

34. Gu RX*, et al.* (2011) Free energy calculations on the two drug binding sites in the M2 proton channel. *J. Am. Chem. Soc.* 133(28):10817-10825.

35. Zhou HX (2011) A Theory for the Proton Transport of the Influenza Virus M2 Protein: Extensive Test against Conductance Data. *Biophys. J.* 100(4):912-921.

36. Gu RX, Liu LA, Wang YH, Xu Q, & Wei DQ (2013) Structural comparison of the wild-type and drug-resistant mutants of the influenza a m2 proton channel by molecular dynamics simulations. *J. Phys. Chem. B* 117(20):6042-6051.

37. Dong H, Fiorin G, DeGrado WF, & Klein ML (2013) Exploring Histidine Conformations in the M2 Channel Lumen of the Influenza A Virus at Neutral pH via Molecular Simulations. *J. Phys. Chem. Lett.* 4(18):3067-3071.

38. Bankura A, Klein ML, & Carnevale V (2013) Proton affinity of the histidine-tryptophan cluster motif from the influenza A virus from ab initio molecular dynamics. *Chem. Phys.* 422:156-164.

39. Wang J, *et al.* (2013) Structure and inhibition of the drug-resistant S31N mutant of the M2 ion channel of influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* 110(4):1315-1320.

40. Wei C & Pohorille A (2013) Activation and Proton Transport Mechanism in Influenza A M2 Channel. *Biophys. J.* 105(9):2036-2045.

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

42. Knight C, Maupin CM, Izvekov S, & Voth GA (2010) Defining Condensed Phase Reactive Force Fields from ab Initio Molecular Dynamics Simulations: The Case of the Hydrated Excess Proton. *J. Chem. Theory Comput.* 6(10):3223-3232.

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

44. Maupin CM, Wong KF, Soudackov AV, Kim S, & Voth GA (2006) A multistate empirical valence bond description of protonatable amino acids. *J Phys Chem A* 110(2):631-639.

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

46. Maupin CM, *et al.* (2011) Chemical Rescue of Enzymes: Proton Transfer in Mutants of Human Carbonic Anhydrase II. *J. Am. Chem. Soc.* 133(16):6223-6234.

47. Yamashita T & Voth GA (2011) Insights into the Mechanism of Proton Transport in Cytochrome c Oxidase. *J. Am. Chem. Soc.* 134(2):1147-1152.

48. Li H, *et al.* (2011) Enhancement of proton conductance by mutations of the selectivity filter of aquaporin-1. *J. Mol. Biol.* 407(4):607-620.

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

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

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

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

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

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

55. Roux B (1995) The Calculation of the Potential of Mean Force Using Computer-Simulations. *Comput. Phys. Commun.* 91(1-3):275-282.

56. Allen MP & Tildesley DJ (1990) *Computer Simulation of Liquids* (Oxford University Press New York, NY).

57. Kumar S, Bouzida D, Swendsen RH, Kollman PA, & Rosenberg JM (1992) The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J. Comput. Chem.* 13(8):1011-1021.

58. Allen TW, Andersen OS, & Roux B (2004) Energetics of ion conduction through the gramicidin channel. *Proc. Natl. Acad. Sci. U. S. A.* 101(1):117-122.

59. Hu J*, et al.* (2006) Histidines, heart of the hydrogen ion channel from influenza A virus: Toward an understanding of conductance and proton selectivity. *Proc. Natl. Acad. Sci. U. S. A.* 103(18):6865-6870.

60. Rapaport DC (1983) Hydrogen-Bonds in Water Network Organization and Lifetimes. *Mol. Phys.* 50(5):1151-1162.

61. Petersen MK, Iyengar SS, Day TJF, & Voth GA (2004) The hydrated proton at the water liquid/vapor interface. *J Phys Chem B* 108(39):14804-14806.

62. Petersen PB & Saykally RJ (2005) Evidence for an enhanced hydronium concentration at the liquid water surface. *J. Phys. Chem. B* 109(16):7976-7980.

63. Yamaguchi S, Kundu A, Sen P, & Tahara T (2012) Communication: Quantitative estimate of the water surface pH using heterodyne-detected electronic sum frequency generation. *J. Chem. Phys.* 137(15).

64. Lapid H, Agmon N, Petersen MK, & Voth GA (2005) A bond-order analysis of the mechanism for hydrated proton mobility in liquid water. *J Chem Phys* 122(1).

65. Berkelbach TC, Lee HS, & Tuckerman ME (2009) Concerted Hydrogen-Bond Dynamics in the Transport Mechanism of the Hydrated Proton: A First-Principles Molecular Dynamics Study. *Phys. Rev. Lett.* 103(23).

66. Williams JK, Zhang Y, Schmidt-Rohr K, & Hong M (2013) pH-Dependent Conformation, Dynamics, and Aromatic Interaction of the Gating Tryptophan Residue of the Influenza M2 Proton Channel from Solid-State NMR. *Biophys. J.* 104(8):1698-1708.

67. Leiding T, Wang J, Martinsson J, DeGrado WF, & Arskold SP (2010) Proton and cation transport activity of the M2 proton channel from influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* 107(35):15409-15414.

Chapter 6

Computationally Efficient Multiscale Reactive Molecular Dynamics to Describe Amino Acid Deprotonation in Proteins.

6.1 Introduction

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

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

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

the excess proton(s) between water molecules and ionizable amino acids undergoing protonation/deprotonation; (2) allow exchange of water molecules between different protein internal cavities as well as between those cavities and bulk solution; and (3) be computationally efficient enough to achieve sufficient sampling of the charge translocation and protein and solvent configuration space for large-scale biomolecular systems, including protein, solvent, ions, and (where needed) membrane. Quantum mechanical approaches, such as *ab initio* molecular dynamics (AIMD) or quantum mechanics molecular mechanics (QM/MM) MD methods, can in principle explicitly treat the reactive nature of amino acid protonation/deprotonation and the Grotthuss hopping phenomenon. However, their computational expense limits the application of the former to small systems much smaller than proteins, and restricts the free energy sampling of the latter when applied to large biomolecular systems. A lack of sufficient free energy sampling generally leads to results with large systematic errors that depend on the initial state of system. In addition, QM/MM methods without adaptive partitioning(24, 25) prohibit exchange of MM and QM atoms across the QM/MM boundary, which is unphysical. Although adaptive QM/MM partitioning methods enable the exchange of MM and QM atoms across the boundary, they can suffer from both inaccuracy of forces on atoms near the boundary and a lack of sufficient energy conservation.(26) The QM/MM boundary may be especially problematic when the hydration along the PT pathway changes in response to the migration of the excess proton charge defect since this necessitates the exchange of water molecules between protein cavities and bulk solution, which often occurs beyond the nanosecond timescale.

An alternative approach that explicitly accounts for the reactive nature of the hydrated

excess proton is multiscale reactive molecular dynamics (MS-RMD) method.(2, 27, 28) This approach describes the migration of an excess proton including explicit Grotthuss shuttling and charge defect delocalization by evolving the system on a reactive potential energy surface defined as a dynamical linear combination of diabatic basis states, as in the earlier multistate empirical valence bond (MS-EVB) method.(2, 29, 30) As will be shown in detail later, the newer MS-RMD approach differs from the older MS-EVB approach primarily because the underlying reactive force field of MS-RMD is largely derived from AIMD or QM/MM data via force-matching and other means, rendering it far less empirical than the original MS-EVB approach. MS-RMD is also "multiscale" in the sense that quantum information on the electronic states is variationally bridged upward in scale to describe the forces felt by the system nuclei in the RMD model. As shown later as well, MS-RMD is several orders of magnitude more computationally efficient than MD from QM/MM, while still describing the charge delocalization and reactive nature of the PT. The MS-RMD method is also not complicated by the QM/MM boundary issues because the molecules that participate in the proton charge defect charge delocalization are dynamically redefined at each time step in such a way that the forces on the atoms are more continuous. Thus, as water molecules or residues move away from the excess proton, the forces acting on the atoms gradually transition to those described by the classical force field. These advantages make the MS-RMD method (and MS-EVB before it) a powerful tool for investigating PT in a variety of biological systems.

On the other hand, the primary challenge of the MS-RMD method is that it needs to be properly parameterized in order to faithfully simulate PT. The parameterization of MS-RMD models is the focus of this work. In ref (31), an approach for parameterizing

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

In this work, we demonstrated the use of FitRMD parameterization to generate MS-

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

6.2 Methods

6.2.1 Multiscale reactive molecular dynamics (MS-RMD)

The MS-RMD (and MS-EVB) method describes electronic delocalization of the excess proton as a linear combination of distinct states with different chemical bonding topologies. (See Figure S1 in the SI.) The Hamiltonian is defined as:

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

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

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

The forces are evaluated by the Hellmann-Feynman theorem and are used to propagate the system in an MD simulation. It is important to note that the method explicitly treats the excess proton charge defect delocalization, Grotthuss shuttling, and part of the polarization effect, and the resulting MD trajectory is continuous and deterministic to within numerical error.

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

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

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

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

where r is the O-H bond length, and D, α, and r_0 are parameters, which are taken from our previous work.(31) Since the classical force fields between two protonated forms of

glutamate and water do not share a common energy origin, V_{ii} is introduced to compensate the constant energy shift between the two states.

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

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

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

where R_{OO_ϵ} is the distance between the hydronium oxygen, O, and the carboxyl oxygen of glutamate, O_ϵ (OE1 and OE2 in the PDB), and $R_{H_jO_\epsilon}$ is the distance between each of three hydronium hydrogen atoms, H_j, and the carboxyl oxygen of glutamate. The functional forms for the repulsive terms are the same as those used in the MS-EVB3.1 model. (31) B, b, b', C, and c are fitted parameters, and d_{OO}^0 and d_{OH}^0 are fixed the same value used in MS-EVB3.1, which are 2.4 and 1.0 Å, respectively.

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

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

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

The MS-RMD (and MS-EVB) formalism also provides a convenient and physically

intuitive description of the excess proton center of the excess charge (CEC), defined as:(2)

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

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

6.2.2 FitRMD parameterization scheme

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

performed to collect the reference forces on each atom in the MS-RMD reactive complex. Then the MS-RMD model parameter set was optimized by minimizing the variational residual:

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

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

6.2.3 Developing MS-RMD models using FitRMD for ClC-ec1

MS-RMD simulations of the PT process in the bacterial ClC homolog, ClC-ec1, are extensively described in ref (39). The FitRMD method was used to parameterize the MS-RMD models from QM/MM data for the two glutamate residues in the central region, namely the E148 (Glu_{ex}) and the E203 (Glu_{in}), which have been identified as intermediate sites for proton binding along the transport pathway (see **Figure 6.1**). Depending on the presence of Cl^- at the central site (Cl^-_{cen}), two different systems were setup. For each

system setup, two sets of umbrella sampling simulations were performed (deprotonating either E148 or E203) to sample configurations for generation of the QM/MM force data. In order to obtain the initial configurations for the umbrella sampling, 50 ns of unconstrained classical MD was run both with Cl^-_{cen} present and absent. In this time water penetrated the central region from the intracellular side of the membrane, forming a continuous hydrogen bonded network between the carboxyl oxygens of E148 and E203. A harmonic potential with a force constant of 20 $\text{kcal} \cdot \text{mol}^{-1} \cdot \text{Å}^{-2}$ was applied to a collective variable (CV), defined as $|\mathbf{r}_{CEC} - \mathbf{r}_X|$, where "*X*" is the carboxyl center of mass for either E148 or E203. The centers of the windows ranged from 2.0 to 4.0 Å and were separated by 0.25 Å. At CV = 2.0 Å all configurations have the amino acid fully protonated, while at CV=4.0 Å all were deprotonated with protonated state contributing less than 0.01% ($c_i^2(\mathbf{r}) \leq 10^{-4}$).

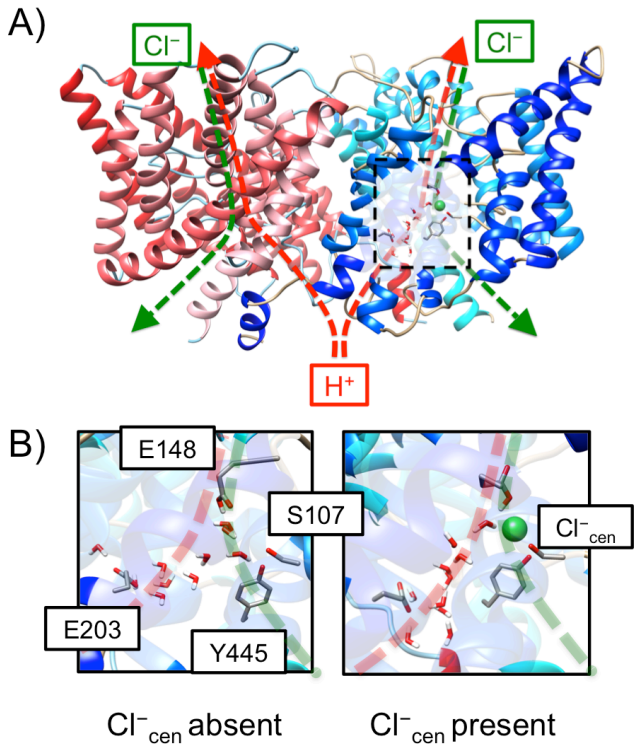


Figure 6.1 (A) Overview of the structure of the ClC-ec1 antiporter and transport pathways for Cl^- (green dashed) and H^+ (red dashed) and (B) Representative configurations of the central region with (left) Cl^-_{cen} absent and (right) present for the "P" state in the PMFs in **Figure 6.2**.

The configurations were selected with a 2 ps interval from each window for ~500 configurations for each protonation site. For each configuration, a single point QM/MM calculation was performed to evaluate the forces acting on the QM atoms. As shown in **Figure 6.1** (B), the QM region included Cl^-_{cen}, if present, the water molecules in the central region, and the side chain of the pore-lining residues, including E113, E148, E202, E203, Y445, and S107. The waters in the central region within the third solvation shell from E148 and E203 were also included in the QM region. The QM box size was set to be 20 – 30 Å in each dimension to ensure it was 6-8 Å larger than the size of the QM atoms in each dimension. The Gaussian Expansion of the Electrostatic Potential

(GEEP) scheme was used to treat the QM/MM electrostatic coupling with periodic boundary conditions (PBCs),(40, 41) and the spurious QM/QM periodic image interactions were decoupled as described in ref (42). The $C_\alpha - C_\beta$ chemical bonds that crossed the QM/MM boundary were capped with hydrogen atoms, the forces on which were calculated following the IMOMM scheme(43) with a scaling factor of 1.50. The QM atoms were treated with density function theory (DFT) using the BLYP functional(44, 45) with empirical dispersion corrections,(46) under the Gaussian and plane wave (GPW) scheme.(47) Goedecker-Teter-Hutter (GTH) pseudopotentials(48) were used and the Kohn-Sham orbitals were expanded in the Gaussian TZV2P basis set.

After the QM/MM forces were obtained, the MS-RMD model parameters were optimized by minimizing the residual given in eq 9 using a genetic algorithm.(49) All atoms in the reactive complex, as defined by the MS-RMD state-selection algorithm,(30) were included in the fitting (i.e., the glutamic acid side chains and solvent atoms in the central region). The V_{ii} and the off-diagonal terms were iteratively optimized for 3 – 6 rounds, depending on the system, until they changed less than 1 %. The models for E203 and E148 were developed independently because they are separated by 10-14 Å, depending on the presence of Cl^-_{cen}, and therefore never participate in coupled delocalization of the excess charge. The MS-RMD simulations were performed with the RAPTOR software(28) interfaced with the LAMMPS MD package (http://lammps.sandia.gov),(50) and the umbrella potentials were controlled by the PLUMED package.(51) All single point QM/MM calculations are performed by CP2K package.(52) The FitRMD calculation was performed by an in-house code.(31) Parameters for E148 and E203 for ClC-ec1 with Cl^-_{cen} either absent or present are given in **Table 6.1**.

Table 6.1 The MS-RMD model parameters of E148 and E203 in ClC-ec1.

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

6.2.4 1D-PMF calculations of PT from E203 to E148 in ClC-ec1

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

$$\zeta_R = \frac{|\mathbf{r}_{\mathrm{CEC}} - \mathbf{r}_{203}|}{|\mathbf{r}_{\mathrm{CEC}} - \mathbf{r}_{203}| + |\mathbf{r}_{148} - \mathbf{r}_{\mathrm{CEC}}|} \quad \textbf{(10)}$$

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

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

where the $\mathbf{r}^{X_j}$ is the position of the *j*th heavy atom in the QM region, which is either the water oxygen atoms or the carboxyl oxygen atoms in the E203 or the E148, and the $\mathbf{r}^{H_i}$

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

The CV ζ_R varies from 0 to 1 as the excess proton CEC moves from E203 to E148. The centers of the harmonic umbrella potentials were separated by 0.1 – 0.2 Å between adjacent windows. The umbrella sampling force constant for ζ_R was chosen to be in the range of 3000 – 7000 $kcal \cdot mol^{-1}$, depending on the sampling overlap between umbrella windows. Given that the denominator of ζ_R, $|\mathbf{r}_{\mathrm{CEC}} - \mathbf{r}_{203}| + |\mathbf{r}_{\mathrm{CEC}} - \mathbf{r}_{148}|$, is ~ 15 Å, the effective force constant acting on $|\mathbf{r}_{\mathrm{CEC}} - \mathbf{r}_{203}|$ was 15 – 30 kcal $\cdot$ mol^{-1} $\cdot$ Å^{-2}. Because the central region is accessible by the water molecules from the intracellular bulk, the CEC can locate outside the central region at the windows near E203 when a restraint is only placed on ζ_R. To avoid sampling irrelevant positions (leading to the intracellular bulk phase), an additional restraint defining the upper limit for $|\mathbf{r}_{\mathrm{CEC}} - \mathbf{r}_{148}|$ was applied such that the CEC was always situated between E148 and E203. The upper limit was chosen to be 2 σ above the average of $|\mathbf{r}_{\mathrm{CEC}} - \mathbf{r}_{148}|$ in each window.

The details of the preparation of the initial configuration and the classical MD equilibration for the ClC-ec1 antiporter are described in ref (39). The initial

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

6.2.5 1D-PMF calculation with QM/MM

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

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

Comparative structural analysis in ClC-ec1

Structural elements of the ClC-ec1 protein are shown in **Figure 6.1**-**Figure 6.5**. To obtain these structures the production run simulation conformations were averaged in each umbrella window following three steps. First, configurations were aligned based on alpha carbon atoms at least 15 Å away from the central region. Second, the positions of the side chains of the four residues in the central region (S107, E148, E203, and Y445) and Cl^-_{cen}, if present, were averaged and full moieties representative of these averages were depicted. Third, the positions of the water molecules and the excess proton were taken from the last MD frame.

6.2.6 Hydration dynamics in the central region of ClC-ec1

To calculate the timescale for the transition from a low to high hydration state in the central region of ClC-ec1, the low hydration state first needed to be created in the MS-RMD structure. To do this, a high hydration configuration was taken from the last MS-RMD snapshot at umbrella window $\zeta_R = 0.24$ (**Figure 6.5** (D)). An additional harmonic

potential was applied to the water density CV (N_w) in the pre-defined box (see ref (23) and (39) and for definition). The force constant of the harmonic potential was set to be 10 $kcal \cdot mol^{-1}$, and the center of the potential was at $N_w = 0$. The center of the box was chosen to be the midpoint between the center of mass of each carboxyl group of E148 and E203, and the box size was set to be 2 x 2 x 2 $Å^3$, which covered the position of the second water. The second water present in the initial configuration was expelled from the central region by the harmonic potential centered at $N_w = 0$. The MS-RMD simulation was run for 200 ps, until the system was equilibrated in the absence of the second water.

Ten independent trajectories were initiated from the last MD frame of the simulation above, after the velocities of all atoms were randomized and the harmonic potential centered at $N_w = 0$ was released. The coordination number of the pivot hydronium was defined as the number of oxygen atoms (from water or the E203 carboxyl group) within 3 Å from the pivot hydronium oxygen. The time for the second water entering the central region was estimated by averaging the ten simulations time to when there was a productive (not transient) transition of the coordination number from 2 to 3.

6.2.7 Developing MS-RMD models using FitRMD for CcO

The FitRMD method was also used to parameterize the MS-RMD models from QM/MM data for three protonatable sites in the hydrophobic cavity of C*c*O, including the E286, PRDa_3, PRAa_3 (**Figure 6.6**). Umbrella sampling simulations were first carried out along the PT pathway identified by the MS-RMD metadynamics (MTD) simulation.(56, 57) Then for each protonatable site ~100 configurations were selected from the trajectories from windows within 4 Å of the transition state of proton dissociation

(defined as the windows with ~50% of the configurations having the largest amplitude on the amino acid and the other 50% on the first water molecule) for each protonatable site. Single point QM/MM calculations were then performed for each configuration using the B3LYP level density functional theory.(58) The MM models were the CHARMM22 and CHARMM36(59) force fields for the protein and lipids, respectively. The QM region included the side chain of each protonatable amino acid, the hydrated excess proton, and water molecules within 3 solvation shells of the carboxyl group (**Figure 6.6** (B)). In all calculations, the QM box size was chosen to be 6-8 Å larger than the actual size of the QM atoms in each dimension. The GEEP scheme was used to treat the QM/MM electrostatic coupling with periodic boundary conditions (PBCs), and the spurious QM/QM periodic image interactions were decoupled as described in ref (42). The $C_\alpha - C_\beta$ chemical bonds that cross the QM/MM boundary were capped with hydrogen atoms, the forces on which were calculated following the IMOMM scheme with a scaling factor of 1.50. The forces generated by the QM/MM calculations were used to parameterize the MS-RMD parameters for the protonatable sites using FitRMD approach. The MS-RMD simulations were performed with the RAPTOR software interfaced with the LAMMPS MD package, as described earlier. The QM/MM calculation was performed with the CP2K package, and FitRMD was carried out with in-house software, again as described earlier. Parameters for E286, PRDa_3 and PRAa_3 are given in **Table 6.2**.

Table 6.2 The MS-RMD model parameters of E286, PRDa_3 and PRAa_3 in C*c*O.

	E286	PRDa_3	PRAa_3
B	0.037588	0.995945	0.994509
b	0.208388	1.999090	1.993038
b'	0.533589	0.000009	0.006308
d_{OO}^0	2.4	2.4	2.4
C	4.925621	0.988369	0.322227
c	1.975947	1.990330	1.981954
d_{OH}^0	1.0	1.0	1.0
r_s^l	3.5	3.5	3.5
r_s^h	4.0	4.0	4.0
V_{ii}	-135.809617	-149.99445	-146.174
c_1	-21.659933	-31.931186	-38.677822
c_2	2.785857	2.689357	1.903222
c_3	1.299987	1.147862	1.245548
D	143.003	143.003	143.003
α	1.8	1.8	1.8
r_0	0.975	0.975	0.975

6.2.8 2D-PMF for PT in the CcO hydrophobic cavity with MS-RMD

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

time step was 1 fs. The CV ζ_R was collected every ten time steps (10 fs). The 2D PMF was constructed by the WHAM.

6.2.9 Comparing the computational efficiency of different methods

The computational speed for the MS-RMD, self-consistent density functional tight binding (SCC-DFTB)-based QM/MM(61, 62) and BLYP-based QM/MM methods were compared for the MD simulation of the C*c*O system. (The QM/MM with B3LYP is much too slow for viable QM/MM MD in these systems.) For MS-RMD, the setup was the same as described in the above section. For the BLYP QM/MM method, the QM/MM setup was the same as described in the above section for FitRMD, except that the QM atoms were treated by the BLYP functional with empirical dispersion corrections. More details on the BLYP-based QM/MM MD simulation setup are presented in ref (60). For the SCC-DFTB-based QM/MM setup, the QM atoms were the same as in the BLYP-based QM/MM setup. The point charge based Ewald summation was used to treat QM/MM electrostatic coupling under PBCs.(63) More details on the SCC-DFTB-based QM/MM setup are described in ref (64). The MS-RMD simulation and BLYP-based QM/MM simulations were performed as described previously. The SCC-DFTB-based QM/MM simulation was performed with CHARMM package.(65)

6.3 Results and Discussion

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

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

PT through the central region of ClC antiporters is one of the essential intermediate steps in Cl^-/H^+ exchange. The migration of an excess proton from the internal E203 to the external E148 through water molecules in the central region is coupled with migration of approximately 2 Cl^- ions in the opposite direction (**Figure 6.1**). Therefore, the PMFs for

PT in the central region of the ClC-ec1 antiporter were calculated both in the presence and absence of Cl^- at the central binding site, Cl^-_{cen}.

The MS-RMD and QM/MM PMFs (**Figure 6.2** (A) and (B)) show excellent agreement for most parts of the reaction coordinate and for both states of the system (with and without Cl^-_{cen}), although there are some differences (discussed below). The free energy barrier of the PT process is significantly decreased when Cl^-_{cen} is present, mainly due to the electrostatic interaction between the excess proton and the Cl^-_{cen} ion. These results demonstrate that the FitRMD approach is capable of generating MS-RMD models that reproduce the free energy profile of the reference QM/MM Hamiltonian, using only forces on the atomic nuclei from a relatively small set of configurations as input for the fitting. Moreover, the FitRMD approach is robust enough to quantitatively capture the effect of the Cl^-_{cen} ion on the free energy surface of the reference QM/MM Hamiltonian. This suggests that the FitRMD approach is also capable of describing the shift in proton affinity in different protein environments, which has significant value for simulating PT in different protein systems with the MS-RMD method.

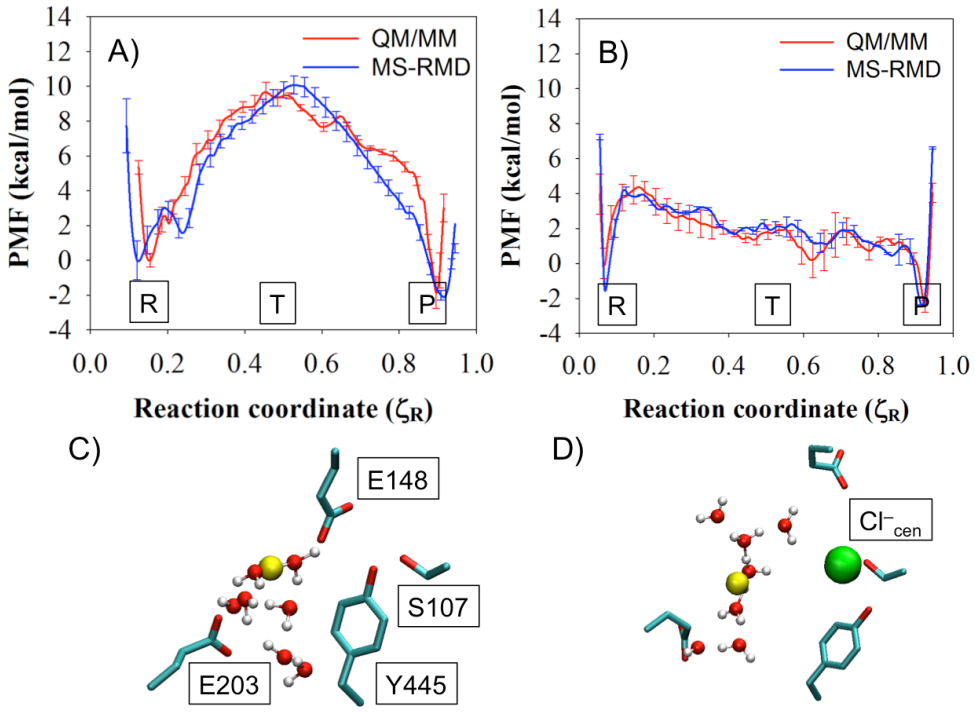


Figure 6.2 PMFs of PT through central region of ClC-ec1 antiporter with (A) Cl^-_{cen} absent and (B) present, as calculated with MS-RMD (blue) and QM/MM (red). Labels "R", "T", and "P" indicate the reactant state, the center of the CV range, and the product state, respectively. The CV ζ_R varies from 0 to 1 as the excess proton CEC moves from E203 to E148. The average side chain positions of four important residues are shown for state "T" in the MS-RMD systems with (C) Cl^-_{cen} absent or (D) present. The solvent positions and proton center of the excess charge (CEC, yellow sphere) are taken from the last snapshot.

The hydration structure surrounding the reactive protein residues was also compared. The radial distribution functions (RDFs) from the carboxyl carbon of either E148 or E203 to the water and hydronium oxygens were calculated with the excess proton CEC restrained in the reactant, transition, and product states of the PMF (the positions are labeled as "R", "T", and "P" in **Figure 6.2** (A) and (B) respectively). The RDFs (**Figure 6.3** and **Figure 6.4** with and without Cl^-_{cen}, respectively) demonstrate that the solvation structure around E203 and E148 is quite similar between the QM/MM and MS-RMD methods. This result provides additional evidence that the model generated by the

FitRMD approach faithfully reproduces the underlying free energy landscape of the reference QM/MM Hamiltonian, even though only QM/MM forces on the atom nuclei were used as input for the fitting. The discrepancies in the RDFs shown in **Figure 6.3** and **Figure 6.4** are likely due to the different water behavior of QM versus MM water as well as short QM/MM sampling where the water dynamics are slow in the confined space of the central region, which will be discussed below. We note, however, that this level of agreement may not be expected for more bulk-like water environments since the QM and MM water will have such different structural properties.

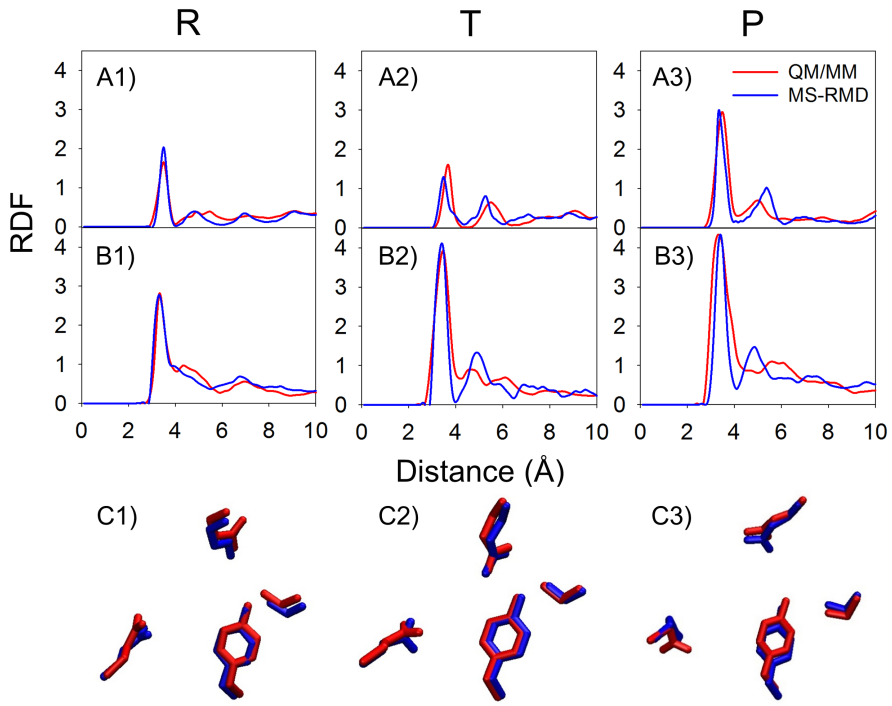


Figure 6.3 In the system with Cl^-_{cen} absent, RDFs from E148 carboxyl carbon (CD in the PDB) to water oxygens are shown in A1 to A3, for the "R", "T", and "P" states (**Figure 6.2**), respectively. The RDFs from E203 carboxyl carbon to water oxygens are shown in B1 to B3 in the same order. The averaged positions of the four side chains are shown in C1 to C3 in the same order, with the QM/MM structures in red and the MS-RMD structures in blue.

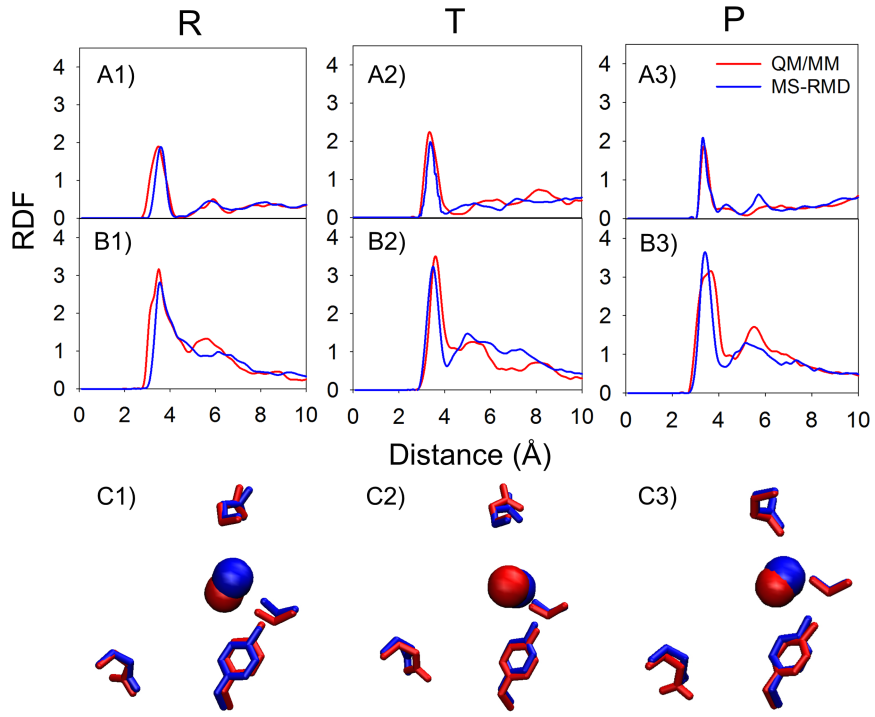
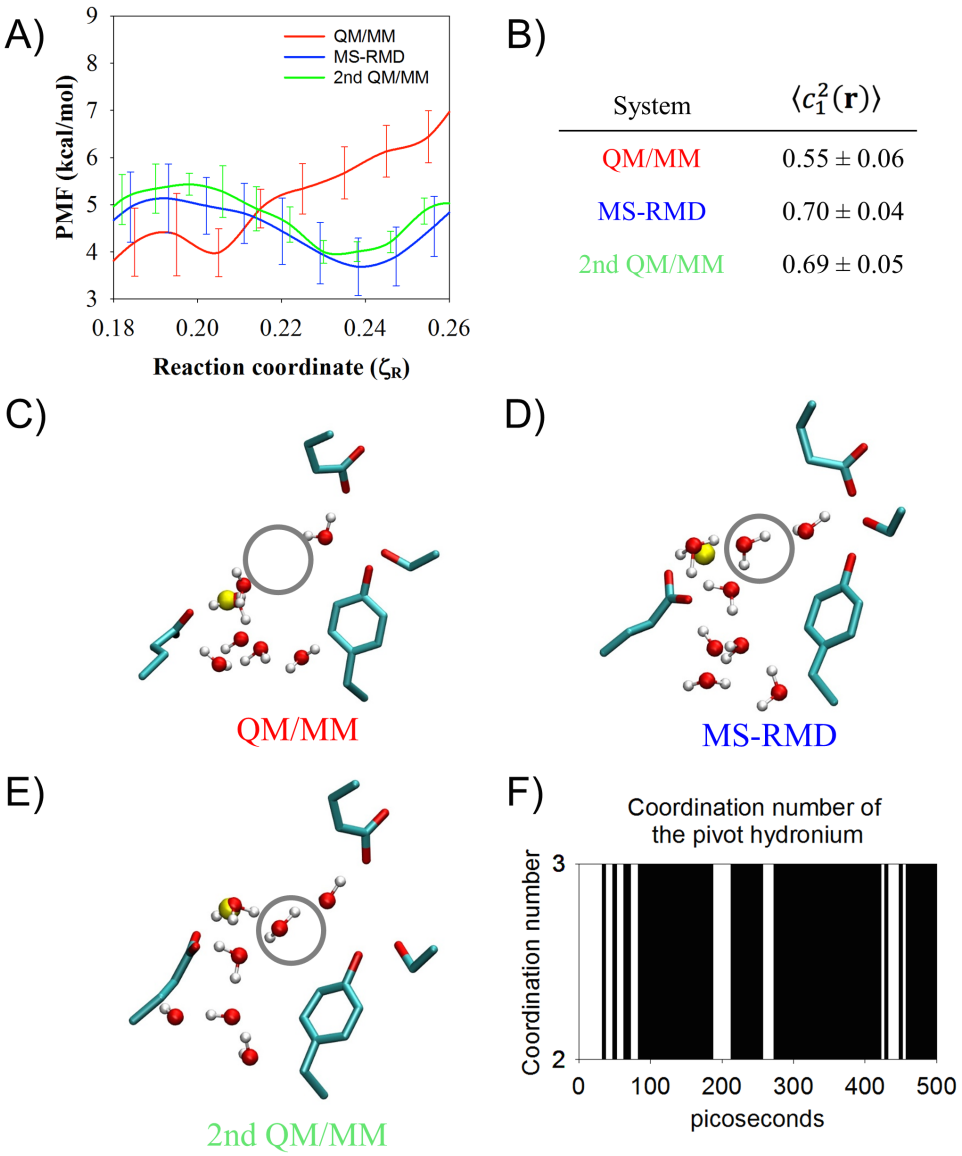


Figure 6.4 In the CLC system as in **Figure 6.3**, but with Cl^-_{cen} present. All notation is consistent with **Figure 6.3**.

6.3.2 Limitations of the QM/MM free energy profiles

Although the PMFs in **Figure 6.2** (A) and (B) show that the MS-RMD and QM/MM PMFs generally agree well, some discrepancies appear. Focusing on the most significant, **Figure 6.5** (A) highlights the region $0.18 < \zeta_R < 0.26$ from **Figure 6.2** (A), where the MS-RMD PMF (blue) dips to a modest metastable minimum but the QM/MM PMF (red) shows an uphill rise with only a slight dip at $\zeta_R \cong 0.21$. This discrepancy between the two PMFs is caused by the change of the local environment for the excess proton. The structures of the central region captured from the $\zeta_R = 0.24$ umbrella windows, where the two PMFs disagree most, are shown in **Figure 6.5** (C) (QM/MM) and (D) (MS-RMD). The MS-RMD structure shows one additional water (grey circle) compared to the

QM/MM structure. The center of excess charge (yellow sphere) is close to the first water coordinated to E203 in this window for both the QM/MM and MS-RMD configurations, meaning that the first water is the most hydronium-like species (the "pivot" hydronium, which has the largest MS-RMD state coefficient, c_1, in Equation 8). To better understand the difference in solvation structures and charge delocalization, we calculated the average value of c_1^2 for all configurations in the $\zeta_R = 0.24$ umbrella windows from the QM/MM and MS-RMD simulations (**Figure 6.5** (B)). It should be kept in mind that the c_1^2 value for an Eigen cation ($H_9O_4^+$) is ~ 0.65, which indicates that the pivot hydronium holds ~ 65% of the positive charge (hydronium-like) defect, while that for a Zundel cation ($H_5O_2^+$) is ~ 0.5 since in that case excess charge is more equally shared between two water molecules. The MS-RMD simulations yield larger $\langle c_1^2(\mathbf{r})\rangle$ values (0.7) because the pivot hydronium is coordinated to two waters in the central region, in addition to the carboxyl group, leading to a stable Eigen-like ($CO\text{-}H_7O_3^+$) complex. However, one of these waters is missing in the QM/MM simulations (**Figure 6.5** (C) - gray circle), shifting $\langle c_1^2(\mathbf{r})\rangle$ to a lower value (0.55) and the delocalization to more of a Zundel-like complex between the carboxyl group and first water molecule ($CO\text{-}H_3O^+$). The MS-RMD Eigen-like complex forms a contact ion pair (CIP) with the carboxyl group of deprotonated E203 and the stabilizing electrostatic attraction between oppositely charged ions causes a small energy well (~1 kcal/mol) in the MS-RMD PMF. However, in the QM/MM configurations, no hydrogen bond acceptor is found near one of hydrogen atoms of the pivot hydronium, causing the PMF to continue its uphill climb.



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

Figure 6.5 A segment of PMF in **Figure 6.2** (A) in the range of $\zeta_R = 0.18 - 0.26$ and additional structural analysis in the CV range.Panel (A): The PMF from **Figure 6.2** (A) for $\zeta_R = 0.18 - 0.26$. The second QM/MM PMF (green) was initiated from the final MS-RMD configuration in each window. (B) The average and the standard error of $c_1^2(\mathbf{r})$, reflecting the delocalization of the hydrated excess proton (see text), from the original QM/MM, the MS-RMD, and the second QM/MM umbrella sampling at window $\zeta_R = 0.24$. Panels (C-E): The averaged structure of the four residues and the water structure in the last snapshot from the (C) the QM/MM, (D) the MS-RMD, and (E) the second QM/MM umbrella sampling at the window $\zeta_R = 0.24$. The region circled by a gray line represents the position of the second water. (See further description in the main text.) Panel (F): The coordination number of the pivot hydronium to oxygen atoms of surrounding water molecules or the carboxyl group of E203 in the MS-RMD simulation, initiated from the configuration at low hydration level, which mimics the configuration of the QM/MM in (C).

To see if the PMF discrepancy was caused by sampling different conformational phase space, another set of QM/MM umbrella sampling simulations was run for 5 ps (as described earlier), initiated from the last snapshots from the MS-RMD umbrella sampling simulations in the CV range $0.18 < \zeta_R < 0.26$. The resulting QM/MM PMF (green in **Figure 6.5** (A)) shows much better agreement with the MS-RMD PMF. **Figure 6.5** (B) and (E) show that for this second QM/MM simulation the value of $\langle c_1^2(\mathbf{r}) \rangle$ is close to the MS-RMD value and the missing second water is present. Thus, the PMF discrepancy is indeed caused by sampling different conformational phase space with that sampled by MS-RMD having one additional water molecule to stabilize the excess proton.

To estimate the time scale required for the additional water molecule (missing in the original QM/MM simulations) to enter the central region, an MS-RMD configuration was prepared (see Methods) at $\zeta_R = 0.24$ in which this water was removed (mimicking the QM/MM solvation structure shown in **Figure 6.5** (C)). The system was equilibrated in this low solvation state, and then ten independent MS-RMD simulations were run for 500 ps. **Figure 6.5** (F) shows the coordination number of the pivot hydronium to the oxygen atoms of the surrounding water molecules or the carboxyl group of E203 in one of the trajectories. The coordination number at $t = 0$ ps is 2, when the pivot hydronium is coordinated to one water oxygen and E203's carboxyl oxygen. Although there are some transient transitions to a coordination of 3 around $t \simeq$ 30-70 ps, due to an additional water oxygen, the second water is not stable for another 100 ps. The second water then enters the central region from the intracellular bulk. The coordination number remains 3 for most of the last 400 ps of the trajectory (~ 89 % of the time). This trajectory is

representative of the other nine trajectories, in which the second water did not enter the central region until t = 60 - 185 ps ($\langle t \rangle$ = 99 ps), after which the coordination number remained 3 for 88 $\pm$ 5 % of the rest of the trajectories. These results suggest that the entrance of the second water is energetically favored in this umbrella window, but that the time scale of this event exceeds the sampling time possible in the QM/MM simulations.

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

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

C*c*O offers another example of PT through an interior protein region being coupled to hydration changes. During the reaction cycle of *aa*$_3$-type C*c*O, as found in mitochondria, protons from the intracellular side of the membrane are transported through the so-called D-channel to the glutamic acid E286 in the middle of the membrane. The protons are then transported through water molecules in a hydrophobic cavity (the HC) to either the pump loading site (PLS) to be further released to the

periplasmic side of the membrane, or to the binuclear center (BNC) to react with oxygen and form water. A particularly interesting aspect of the proton pumping mechanism in C*c*O is the role of water molecules in the HC during PT from amino acid E286 to the PLS or BNC (**Figure 6.6** (B)). The number of water molecules in the HC and their role in PT has been the focus of much debate.(21, 22, 66-70) To further complicate this issue, the water molecules can move in and out of the HC in C*c*O in response to the migration of the excess proton, and the two processes can be intrinsically coupled to each other. Capturing this type of cooperativity often requires computationally demanding enhanced sampling of multiple degrees of freedom. Here we show that the computationally efficient MS-RMD method parameterized by FitRMD approach allowed us to address this challenge.

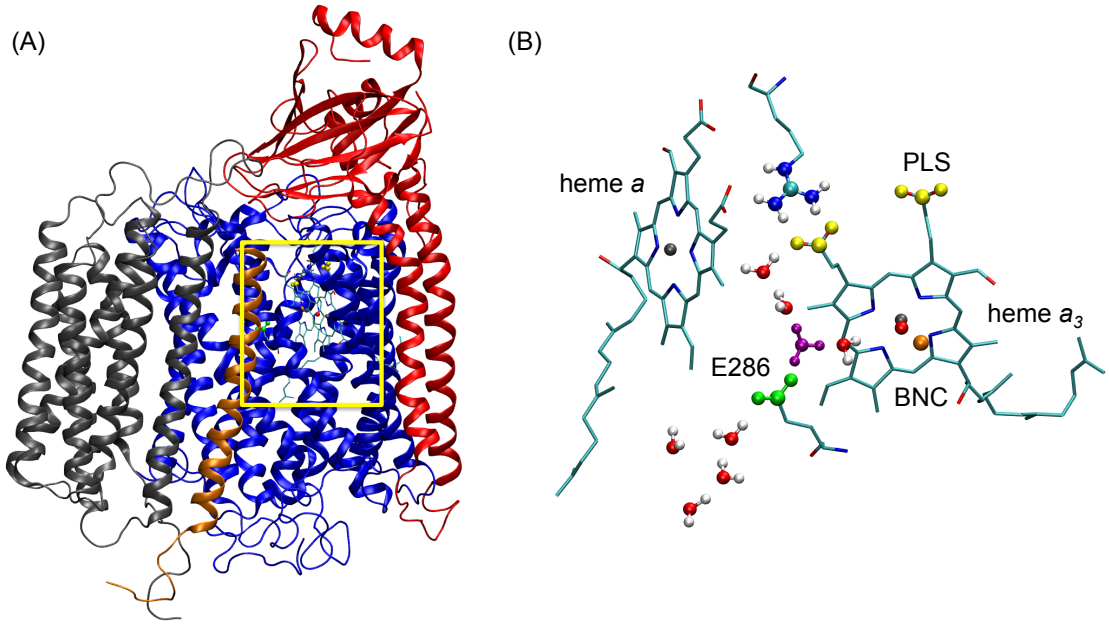


Figure 6.6 Cytochrome c oxidase from *Rhodobacter sphaeroides*. (A) The subunit I, II, III and IV are colored in blue, red, gray and orange, respectively. The yellow rectangular box highlights the E286 residue and the heme-copper groups that are essential for proton pumping and reaction. (B) Enlargement of the region where the E286, heme *a*, heme a_3 and BNC are located. The E286 is shown in green, the propionate groups of the PLS in yellow, the hydrated excess proton in purple, the iron atom of the heme groups in gray and the copper atom in the BNC in orange. The heme groups are shown as sticks. In the QM/MM calculation, the QM atoms include the E286 side chain, the hydronium and the 3 solvation shells of water molecules around the E286 residue (shown in VDW representation).

We calculated 2D PMFs for the PT from E286 to the PLS in different redox states during the A→ F transition of C*c*O. The collective variables used to define these 2D PMFs are (1) the progress of the excess proton CEC through the HC (horizontal axis) and (2) the degree of hydration of the HC (vertical axis) (**Figure 6.7**, see Methods for more discussion). The 2D PMF shown in **Figure 6.7** clearly reveals the cooperativity between the PT and dynamic hydration in the HC. The minimum free energy pathway (black line) along the 2D PMF (**Figure 6.7**) verifies that as the excess proton migrates from E286 to

the PRDa_3, the hydration level of HC gradually increases. The non-horizontal feature of the minimum free energy pathway indicates that the HC hydration is intrinsically coupled to the proton charge defect translocation in this activated process.

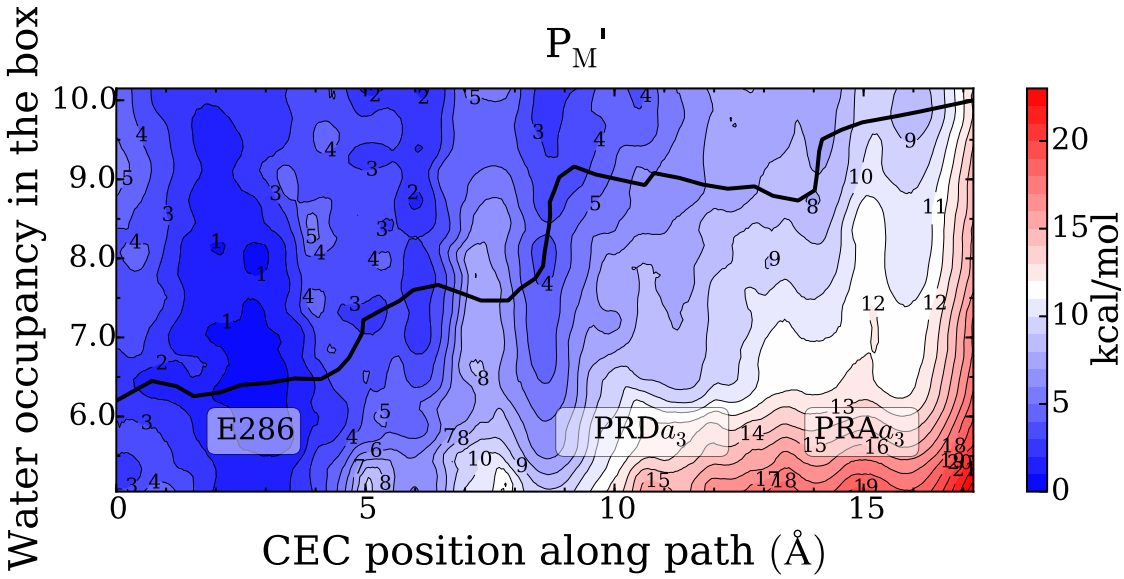


Figure 6.7 Two-dimensional free energy profiles (2D-PMFs) for PT from the E286 to the PLS in the P_M' state, as a function of the CEC coordinate through the hydrophobic cavity (HC) as the horizontal axis and the water hydration in the HC as the vertical axis. The minimum free energy pathways (black lines) are diagonal in nature, indicating the two processes are coupled. The statistic errors of the 2D-PMFs is ~ 0-3 kcal/mol.

It is also important to emphasize how the computational efficiency of the MS-RMD method makes it feasible to calculate such 2D PMFs. **Figure 6.8** shows a relative speed and scaling over processors plot of the MS-RMD, SCC-DFTB-based QM/MM and BLYP-based QM/MM MD methods for C*c*O. It is seen that MS-RMD is at least three orders of magnitude faster than the BLYP-based QM/MM simulation, and two orders of magnitude faster than the SCC-DFTB-based QM/MM simulation. MS-RMD also scales better over processors than the other two methods (the SCC-DFTB CHARMM code is not scalable at all.)

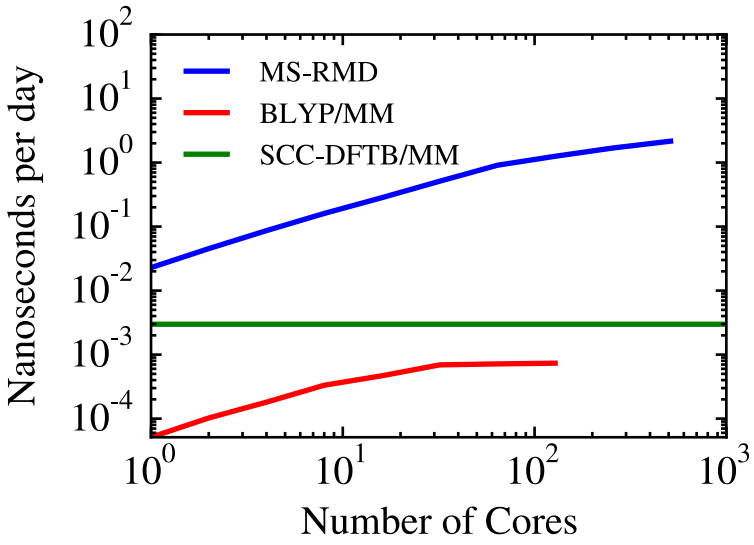


Figure 6.8 Scaling plot for the MS-RMD, SCC-DFTB-based QM/MM and BLYP-based QM/MM simulations of C*c*O.The blue curve is for MS-RMD, the green is for SCC-DFTB, and the red curve is for BLYP-based QM/MM. The MS-RMD method is at least 3 orders of magnitude faster than BLYP-based QM/MM method and 2 orders of magnitude faster than SCC-DFTB-based QM/MM, while more favorable scaling properties over processors (the SCC-DFTB CHARMM code is not scalable).

6.4 Conclusions

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

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

As presented here, the FitRMD approach variationally maps quantum data (DFT-level QM/MM forces in this case) onto the MS-RMD nonlinear reactive force field. We have demonstrated how FitRMD can be used to parameterize MS-RMD models for amino acids in proteins with two example systems, ClC-ec1 and C*c*O. In ClC-ec1 the MS-RMD models were shown to faithfully reproduce the PT PMFs of the reference QM/MM MD, both in the presence and absence of a Cl^- ion in the central region. Thus, the FitRMD parameterization was robust enough to capture quite different free energy profiles due to the presence or absence of a single ion. Moreover the local structure of the protein and confined water was shown to be quite similar between the MS-RMD and QM/MM simulations. In C*c*O the developed MS-RMD models were shown to be efficient enough to capture the coupling between PT and hydration changes in the HC region. Two-dimensional free energy surfaces in which both PT and hydration levels were explicitly sampled are required for this analysis. Given that individual umbrella

windows often required over 500ps to converge, these calculations would not be possible with a QM/MM MD approach. Moreover QM/MM boundary issues, such as the lack of exchange of water molecules, can lead to large systematic errors in such PT PMFs. Even for the ClC-ec1 system where exchange of water across the QM/MM boundary is minimal due to protein confinement, lack of water exchange was argued to introduce errors in the calculated QM/MM free energy profile.

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

focus on improving the FitRMD protocol to make it more robust to choice of the trial MS-RMD model, as free as possible from co-evolution of parameters, and insensitive to discrepancies between the QM and MM descriptions of non-reactive atoms in the system.

6.5 References

1. Knight C & Voth GA (2012) The Curious Case of the Hydrated Proton. *Accounts of Chemical Research* 45(1):101-109.

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

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

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

5. Chen TY (2005) Structure and function of ClC channels. *Annual review of physiology* 67(1):809-839.

6. Estevez R & Jentsch TJ (2002) CLC chloride channels: correlating structure with function. *Current opinion in structural biology* 12(4):531-539.

7. Jentsch TJ (2008) CLC chloride channels and transporters: from genes to protein structure, pathology and physiology. *Critical reviews in biochemistry and molecular biology* 43(1):3-36.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Chapter 7

Multiscale simulations reveal key features of the proton pumping mechanism in cytochrome c oxidase.

7.1 Introduction

Cytochrome *c* oxidase (C*c*O, **Figure 7.1**) is the terminal enzyme in the respiratory electron transfer chain in the inner membrane of mitochondria and plasma membrane of bacteria. It catalyzes the reduction of O_2 to H_2O and couples the free energy of this exergonic reaction to the pumping of protons across the membrane, creating a transmembrane proton electrochemical gradient that drives, for example, ATP synthesis. During each reaction cycle eight protons are taken up from the negatively-charged inside (N-side) of the membrane and either react with oxygen (referred to as 'chemical' protons below) or are pumped to the positively-charged outside (P-side) of the membrane (referred to as 'pumped' protons below). In aa_3-type C*c*O, as found in mitochondria, the D-channel is responsible for uptake of all four pumped protons and at least one out of four chemical protons. Protons on the N-side are taken into D-channel via the amino acid residue D132 at the channel entrance, and then transferred to residue E286 in the middle of the membrane. By Grotthuss shuttling through the water molecules in the hydrophobic cavity (HC) above E286, each proton is either transferred to react with oxygen in the binuclear center (BNC), consisting of heme a_3 and the Cu_B complex, or transferred to the pump loading site (PLS) and then further released to the P-side of the membrane (c.f. **Figure 7.1**). Despite decades of study, the C*c*O proton pumping mechanism, which

entails the transport of two protons from the N-side of the membrane (one to be pumped and one for catalysis) coupled to a single electron transferred to the BNC, is still incompletely understood at the atomistic level. It has been unclear, for example, how electron transfer (ET) is coupled with the multiple proton transport (PT) events, in what order the charge transport processes happen, and how C*c*O prevents back flow of pumped protons during the transfer of chemical protons.

To address these questions, experimental results must be complemented with molecular-level insight from computer simulation. However, in large and complicated biomolecular systems it is challenging to simulate PT, which requires an explicit treatment of the positive charge defect associated with a hydrated excess proton, including its delocalization and Grotthuss shuttling (1). To overcome this challenge, a multiscale reactive molecular dynamics (MS-RMD) method has been extensively developed and applied in our group to study PT in aqueous and biological contexts [see, e.g., Refs. (2-7)]. Here we have carried out extensive MS-RMD free energy simulations to study the proton pumping mechanism in C*c*O. In our MS-RMD approach, quantum mechanical forces from targeted quantum mechanics/molecular mechanics (QM/MM) calculations are bridged, in a multiscale fashion via a variational mathematical framework, into the reactive MD algorithm (MS-RMD) for the dynamics of system nuclei, thus including chemical bond breaking and making. In this way, we explicitly simulate PT between proton binding sites, including Grotthuss shuttling of the excess proton(s) through residues and intervening water molecules.

We focus in this work on the $A \rightarrow P_R \rightarrow F$ transition in C*c*O, which occurs during the oxidation of the fully reduced enzyme, because there is extensive experimental information to which we can compare our computational results. Based on our simulations, we provide a quantitative, comprehensive and molecular-level description of proton uptake, pumping, and chemical proton transfer during the $A \rightarrow P_R \rightarrow F$ transition. Our results show that both PT events (E286 to the PLS and E286 to the BNC) are driven by ET from heme *a* to the BNC. The transfer of the pumped proton is kinetically favored while that of the chemical proton is rate limiting. Our calculated rate for the chemical proton transfer is in quantitative agreement with experimental measurements. (8, 9) These results also explain how C*c*O prevents the decoupling of pumping from the chemical reaction with kinetic gating. The fast pumping process precedes transfer of the chemical proton to the BNC, and fast D-channel PT to E286 after pumping prevents proton back flow from the PLS. Given the computational accuracy and efficiency of the MS-RMD methodology (7), a critical component of our results is the explicit characterization of the coupling between PT and hydration changes in both the HC and D-channel, revealing a remarkable and dynamic coupling between the migration of the excess proton and hydration. Finally, we present results that argue against the possibility of E286 being biprotonated during the pumping process.

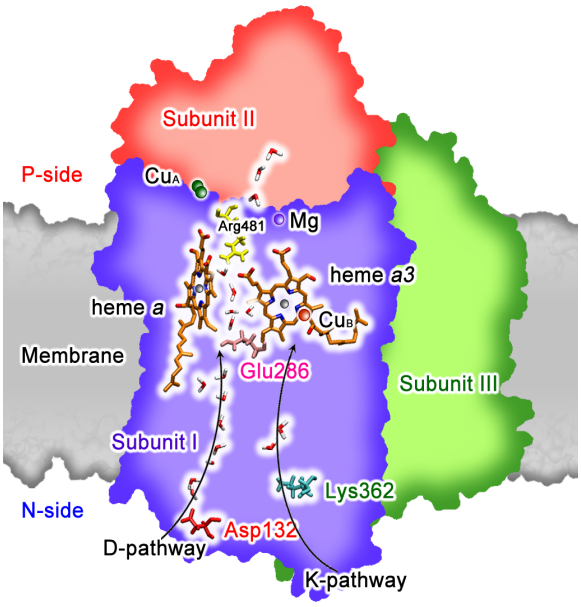


Figure 7.1 Illustration of the simulation setup for the full C*c*O from *Rhodobactor sphaeroides* in a membrane and surrounded by water. The D- and K-channels, as well as metal centers, key residues, and internal water molecules are depicted.

7.2 Material and Methods

The full structure of CcO from *Rhodobactor sphaeroides* (PDB code 1M56 (10)) was embedded in a dimyristoylphosphatidylcholine (DMPC) lipid bilayer and solvated by water molecules on each side of the membrane. MS-RMD simulations using metadynamics (MTD) (11) were performed to identify the PT pathways (12) in both the D-channel and HC. The FitRMD method (6, 7) was used to parameterize the MS-RMD models from QM/MM data for protonatable sites in CcO. The MS-RMD umbrella sampling calculation in the HC and D-channel were carried out by restraining 1) the excess proton CEC position (see *Appendix C* for definition) along the PT pathway defined from the MTD procedure, and 2) the water density in a predefined box (see ref

(13) for definition). Further details on simulation methods are provided in the Appendix C .

7.3 Results and Discussion

In the current work, three intermediate redox states during the A→P_R→F transition were simulated: P_M', P_R and F (**Table C.1**). The first intermediate state is P_M', where both Cu_A and heme *a* are reduced, the BNC is oxidized, and Cu_B has a hydroxide ligand. Electron transfer from heme *a* to the tyrosine radical in the BNC converts P_M' into the P_R state. Following this, PT from E286 to the PLS and a second PT to the Cu_B bound hydroxide, forming a water molecule, converts P_R into F. Partial electron transfer from Cu_A to heme *a* and proton release to the P-side of the membrane, which we did not simulate in this work, completes the A→P_R→F transition.

7.3.1 Transport of the Pumped Proton and Hydration of the HC

Previous results have suggested that internal PT from E286 to the PLS and BNC are coupled to the redox states of heme *a* and the BNC (9, 14-24). However, controversy remains regarding how they are coupled, in what order the PT and ET events occur, and what features enable pumping while preventing the back flow of protons from the P-side to the N-side (8, 9, 17-19, 23, 25). Particularly controversial has been the role of water in the HC during the proton pumping process with some authors arguing for a low hydration state (3-5 waters) (26-30), while others have suggested a high hydration state (> 6 waters) (31). Recent computational work proposed a stepwise pumping mechanism in which an

excess proton is first transported from E286 to PRDa_3 (the putative PLS) through a poorly hydrated HC, followed by an increase in the HC hydration (31), which was used to describe the transfer of the chemical proton (24). This issue is challenging to resolve because, as recently reported in ref (13), the migration of a hydrated excess proton can be strongly coupled with a dynamically changing solvation environment, to the extent that protons can even create their own "water wires" in otherwise dry hydrophobic spaces. In other words, as shown herein for C*c*O the two processes (PT and dynamic hydration) happen cooperatively, with waters entering and leaving the HC during the charge migration processes, and with the water hydration being intrinsically coupled to the proton charge defect translocation. Capturing this type of cooperativity often requires computationally demanding enhanced sampling of multiple degrees of freedom. The MS-RMD approach has allowed us to overcome this challenge for C*c*O.

To address the above mentioned controversies, we have simulated PT to the PLS and the BNC focusing on the coupled hydration changes. Starting with PT from E286 to the PLS we calculated *two-dimensional* free energy profiles, or 2D potentials of mean force (2D PMFs), in the P_M' (before ET) and P_R (after ET) states during the A→ F transition (**Table C.1**). The collective variables used to define these 2D PMFs are (1) the progress of the excess proton center of excess charge (CEC) through the HC (horizontal axis) and (2) the degree of hydration of the HC (vertical axis). (see Appendix C for definitions and more discussion). The 2D PMFs (**Figure 7.2** A and B) and minimum free energy pathways (black lines) verify that as the proton moves from E286 to the D-propionate on heme a_3 (PRDa_3) during the corresponding activated rate processes described by these pathways,

the HC becomes more hydrated (increasing from six to ten waters in our chosen hydration "box", which also includes approximately two waters outside of the traditionally defined inter-heme region). The curvy and non-horizontal nature of the minimum free energy pathways indicates that the two processes are indeed cooperative and coupled. When E286 is protonated, the HC favors a low hydration state (~ four waters in the inter-heme region) (**Figure C.5** A). As the excess proton moves to the water above E286 (**Figure C.5** B) the PRDa_3 side chain rotates down to interact with the positive charge (CEC). This weakens the interactions between PRDa_3 and nearby W172 and R481, allowing more water molecules to enter the HC. The hydration level in HC reaches its maximum just before (P_R) or as (P_M') PRDa_3 is protonated (**Figure C.5** C). The PRDa_3 then rotates from downward to upward orientation through the transition state (**Figure C.5** D). Subsequently, the proton moves to the A-propionate on heme a_3 (PRAa_3, the final PLS identified in this work as discussed below), and the high hydration state remains stable as long as E286 remains deprotonated (**Figure C.5** E). (This behavior was further confirmed by 30 ns of classical simulations of the P_M' and P_R states with both E286 and PRDa_3 deprotonated and PRAa_3 protonated.)

The 1D PMFs traced out along the minimum free energy pathways (**Figure 7.2** C), which should describe the dominant activated reactive energetics for the water-mediated transport of the pumped protons in the P_M' and P_R states, also reveal several important findings. First, the free energy minimum at the PRAa_3 in the pumping PMF of the P_R state suggests that PRAa_3 is the major PLS, in agreement with the conclusions of ref (32, 33). It is interesting to note that in ba_3-type C*c*O, PRAa_3 is also suggested to be the PLS

(34). However, PRDa_3 is also actively involved in the pumping process since it shuttles the proton from E286 to the PLS (breaking its salt bridge to R481 and hydrogen bond with W172), as suggested by Wikstrom et al (35) based on non-reactive classical MD simulations. The participation of PRDa_3 in the PT from E286 to PLS was also discussed by Yamashita et al. (21) in a reduced model for the C*c*O system. Second, PT from E286 to the PLS is thermodynamically unfavorable by ~ 8 kcal/mol in P_M' state, but favorable by 3 kcal/mol in P_R state (**Figure 7.2** C). Thus, ET to the BNC provides a thermodynamic driving force for PT to the PLS. Combined with the experimental results showing that ET is not complete without PT to the PLS in the E286Q mutant (36), one might conclude that ET to the BNC and PT to the PLS are coupled, each driving the other to its thermodynamically favored state.

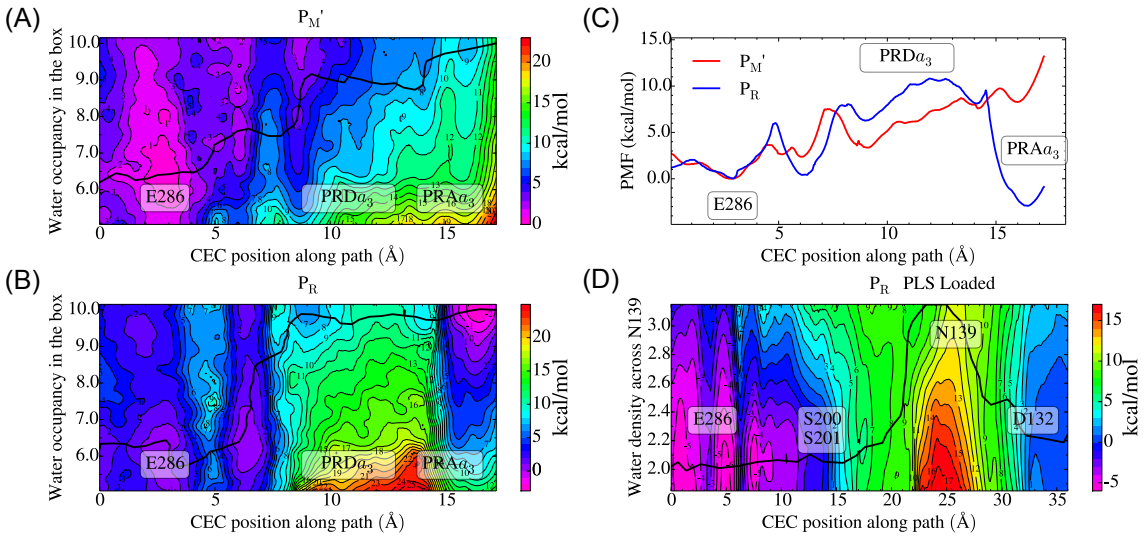


Figure 7.2 Free energy profiles for PT from E286 to PLS in hydrophobic cavity and from D132 to E286 in D-channel. (A) and (B): Two-dimensional free energy profiles (2D-PMFs) for PT from singly protonated E286 to the deprotonated PLS in the P_M' and P_R states, respectively, as a function of the excess proton center of excess charge (CEC) coordinate through the hydrophobic cavity (HC) as the horizontal axis and the water hydration in the HC as the vertical axis. The minimum free energy pathways (black lines) are diagonal in nature, indicating the two processes are coupled. (C) 1D free energy profiles (PMFs) for PT in the HC along the minimum free energy pathway for the P_M' (red) and P_R (blue) states. (D) 2D-PMF for PT from singly protonated D132 to the deprotonated E286 through the D-channel in the P_R state, with proton preloaded at PLS. The 2D-PMF is a function of the CEC coordinate through the D-channel as the horizontal axis and the water hydration in the asparagine gate region as the vertical axis. The minimum free energy pathway is depicted by a black line. The strongly coupled behavior of the PT CEC and the water hydration in the asparagine gate region along the 1D minimum free energy path (black line) is clearly evident. The statistical errors of the 2D-PMFs in (A), (B) and (D) are in the range ~ 0.1-3 kcal/mol. The statistical errors of the 1D-PMFs in (C) are in the range ~ 0.1-1 kcal/mol. In all the plots, the positions of E286, PRDa_3, PRAa_3, D132, N139, S200 and S201 are labeled with text boxes.

7.3.2 Transport of the Chemical Proton

Focusing next on the transfer of the chemical proton, we start from the low hydration state of the HC and calculate 1D PMFs for PT from E286 to the BNC in P_M' and P_R

states, both with and without the PLS (PRAa_3) protonated. The low hydration state is chosen for several reasons. First, the PT pathway is roughly horizontal to the membrane such that PRDa_3 remains deprotonated throughout the PT process, and never rotates down to increase the hydration level in the HC. Second, classical MD simulations starting from the high hydration state and PRAa_3 protonated relax to low hydration state within 30 ns once E286 is also protonated (i.e., following E286 reprotonation from the D channel, which is fast as further described below). The PMFs (**Figure C.3**) show that PT from E286 to the Cu_B bound hydroxide in the BNC is thermodynamically unfavorable by more than 10 kcal/mol in the P_M' state, but favorable by ~ 5 kcal/mol in the P_R state. This suggests that ET from heme *a* to BNC also provides a thermodynamic driving force for the PT from E286 to the BNC (i.e., the formation of the F state) that leads to the chemical reaction. Interestingly, this conclusion is independent of whether the PLS is loaded (protonated) or unloaded (deprotonated) (**Figure C.3** E), and also is not sensitive to the hydration level in the HC (**Figure C.3** A-D).

7.3.3 Rates of the Pumped and Chemical Proton Transport Events

Analysis of the rates (**Table 7.1**; reported as time constants, i.e., the inverse of rate constants) for PT to the PLS and BNC in the P_M' and P_R states provides deeper mechanistic insight into the proton pumping mechanism during the A→ P_R→ F transition. In the P_M' state (before ET), PT from E286 to the PLS is thermodynamically unfavorable, but still faster than the experimental A→P_R transition rate (50 μs for *Rhodobactor sphaeroides* (8) and ~25 μs for *Paracoccus denitrificans* (37)). However, PT from E286 to the Cu_B bound hydroxide in the BNC is significantly slower. Thus,

before ET it is kinetically prohibitive to transfer the chemical proton from E286 to the BNC, which would short-circuit pumping if it were to happen before PT to the PLS (29). Although forward PT from E286 to PLS is kinetically possible in the P_M' state, it is energetically unfavorable. Moreover, the reverse PT (PLS to E286) is even faster and outcompetes reprotonation of E286 through D-channel (**Table 7.2**). Thus, PT to the PLS is minimal before the ET. In contrast, after ET (in the P_R state) proton back flow from the PLS to deprotonated E286 is slower than the reprotonation of E286 through the D-channel. This allows timely reprotonation of E286 and prevents the proton loaded at PLS from leaking back to E286 and subsequently being consumed at the BNC. Thus, full loading of the PLS is achieved only after or concurrently with ET. In line with this, we note that the PT to PLS is still faster than PT to the BNC after the ET (**Table 7.2**, P_R state). This again prevents the above-mentioned short-circuiting and ensures that for the entire PT/ET process during A→F transition, regardless of the redox states of heme *a* and the BNC, PT to the PLS is not short-circuited by PT to the BNC.

Transfer of the chemical proton from E286 to the BNC is much faster after ET occurs because the electrostatic repulsion from the BNC is smaller in the P_R state than it is in the P_M' state (**Table 7.1**), reducing the PT free energy barrier (**Figure C.3** E). This supports the general expectation and our conclusion that ET to BNC facilitates the transfer of chemical proton to the catalytic site. Our calculated rate for transfer of the chemical proton in P_R state is in quantitative agreement with the experimental P_R→ F transition rate (8, 9), and is slower than the D-channel PT rates (see **Table 7.2**). Therefore, we conclude that the PT from E286 to the Cu_B bound hydroxide in the BNC is the rate-

limiting step for the $P_R \rightarrow F$ transition. To summarize, the above thermodynamic and kinetic results suggest that ET from heme *a* to BNC provides a thermodynamic driving force for PT both from E286 to the PLS *and* from E286 to the BNC, while PT from E286 to the BNC is the rate-limiting step for the $P_R \rightarrow F$ transition.

Table 7.1 Calculated time constants (inverse of rate constants*) for PT of the pumped (E286→PLS), chemical (E286→BNC), and back leaked (PLS→E286) protons in the P_M' and P_R states, compared with experimental time constants for A→P_R and P_R →F transitions (8,34). The PT from E286 to BNC in P_R state with PLS protonated is the most physically relevant for the P_R→F transition, and is in quantitative agreement with the experimental time constant. (See main text)

State	E286→PLS (μs)[#]	E286→BNC (μs)[#]		PLS→E286 (μs)
		Deprotonated PLS	Protonated PLS	
P_M'	$(2.6 \pm 0.2) \times 10^{-1}$	$(4 \pm 3) \times 10^{6}$	$(7.7 \pm 0.2) \times 10^{8}$	$(1.5 \pm 0.4) \times 10^{-5}$
P_R	2.1 ± 0.3	$(3 \pm 2) \times 10^{2}$	$(1.7 \pm 0.9) \times 10^{2}$	$(1.2 \pm 0.6) \times 10^{3}$
Exp (A→P_R)	25-50			
Exp (P_R→F)	200			

* See Appendix C for more information on the rate constant calculations. Errors are shown in parenthesis.
[#]For E286→PLS, the initial state has E286 protonated and the PLS deprotonated while the final state has E286 deprotonated and the PLS protonated, and vice versa for PLS→E286. For E286→BNC, the initial state has E286 protonated and a hydroxide bound to Cu_B in the BNC while the final state has E286 deprotonated and a water bound to Cu_B in the BNC.

7.3.4 Comparison to Proposed Mechanisms

Our thermodynamic and kinetic results build upon a the previously proposed mechanism based on the orientation and connectivity of water chains in the HC (9, 23, 27, 29). In this mechanism, PT from E286 to the PLS and ET from heme *a* to the BNC are tightly coupled to each other and occur in concerted fashion during the A→P_R transition. The short-circuiting (i.e., premature PT to BNC before PT to PLS) is avoided by the absence of a water chain leading from protonated E286 to BNC in the P_M' state. However, this mechanistic proposal was based on classical nonreactive MD simulations in the absence

of a shuttling excess proton. Our results support the conclusion that PT to the BNC is unfavourable in P_M' state, but given our treatment of the explicit proton transport they further explain why PT to the PLS is preferred after ET occurs, a question raised but not fully answered in ref (23).

Our results are also in partial agreement with the mechanism proposed by Faxen et al (8) in which PT from E286 to the PLS occurs after the formation of the P_R state, such that the ET and PT events are *sequential*. This mechanism was further discussed later by Lepp et al (38), who concluded that the electrometric signal during $A \rightarrow P_R$ transition is caused by the upward movement of K362 side chain rather than PT from E286 to PLS. However, it has been shown that this electrometric signal is large enough to include both the lysine swing and the PT from E286 to PLS. (23) Moreover, the electrometric signal was shown to be lost in the E286Q mutant (36), resulting in a product that was actually no longer pure P_R, but a P_R/P_M' mixture with the latter dominating (see Section 5.2. and Fig. 10 in ref. (18)). Based on these experimental findings, we suggest that the coupled PT/ET mechanism is at present the best explanation. (23) The results presented herein clearly show that ET drives PT from E286 to PLS, but they do not directly determine whether the PT/ET mechanisms are coupled or sequential. (8, 9)

Table 7.2 Calculated time constants (inverse of the rate constants*) for PT in the D-channel from protonated D132 to deprotonated E286 in the P_R and F states, compared with the experimental time constant for the $P_R \rightarrow F$ transition (8). The PLS is protonated in the simulation.

State	D132→E286 (μs)
P_R	6.7 ± 0.2
F	$(4 \pm 2) \times 10^1$
Exp ($P_R \rightarrow F$)	200

*See Appendix C for more information on the rate constant calculations. Errors are shown in parenthesis.

7.3.5 Proton Transport through the D-Channel

The D-channel is responsible for transporting two protons from the N-side of the membrane to E286 during the $A \rightarrow P_R \rightarrow F$ transition, one for pumping and the other for the chemical reaction. Three highly conserved asparagine residues, N139, N121, and N207, reside roughly one third of the way into the D-channel and form a constricted region (called the asparagine gate) (39). Previous nonreactive classical MD simulations have revealed a gating motion of N139 that controls the hydration state of D-channel (40, 41). However, that work hypothesized that PT through the D-channel follows a sequence in which the proton waits on the proton donor (D132) until a "water wire" is formed, then rapidly dissociates and transports through the pre-formed water wire to the proton acceptor (E286). As discussed above and in ref (13), this historically popular depiction of PT in aqueous systems is both misleading and inconsistent with the dynamically coupled and cooperative nature of the hydration environment and the excess proton migration.

Here and as we did earlier for the PT in the HC, we investigate how the motion of the excess proton is coupled to the change of the hydration level across the asparagine gate by explicitly calculating 2D PMFs in P_R and F states during the A→ F transition (**Figure 7.2** D, **Figure C.4** A and B). In these PMFs the progress of the excess proton CEC through the D-channel (horizontal axis) and the degree of hydration of the asparagine gate region (vertical axis) are used as the two collective coordinates. The minimum free energy pathways are also identified on the 2D PMFs (black lines) and the corresponding 1D PMFs along the minimum free energy pathways (reaction coordinate) are plotted in **Figure C.4** C.

The D-channel PT process starts with protonated D132 and deprotonated E286. Initially, when D132 is protonated the space between N139 and N121 is narrow and dehydrated, forming an effective gate for PT (**Figure C.6** A). As the excess proton transitions to the water above D132 and approaches this gate, the N139 side chain rotates and opens a pathway for solvation and PT past the asparagine residues (**Figure C.6** B). The transition state is reached when the excess proton is in the middle of the asparagine gate (**Figure C.6** C). Once the excess proton shuttles through, the asparagine gate gradually closes and becomes dehydrated again (**Figure C.6** D). The curvy and non-horizontal nature of the minimum free energy pathways on the 2D PMF again indicates that PT and hydration changes are concerted and coupled processes. After traversing the gate region, the excess proton proceeds to the serine zone (S200 and S201), where it forms hydrogen bonds with the hydroxyl groups of the pore lining serine residues in a metastable state. Subsequently, the excess proton protonates E286 rotated in the "down" conformation (x≃5 Å **Figure**

C.4 C), reaching the global free energy minimum. Then, protonated E286 rotates up for either pumping or the chemical reaction (x≃1 Å **Figure C.4** C). The protonated E286 is slightly more energetically favorable in the P_R state than in the F states (**Figure C.4** C), likely due to the more negative charge on the BNC of the P_R state.

The 1D PMFs along the minimum free energy pathways on the 2D PMFs reveal free energy barriers for the proton to pass through the asparagine gate in the D-channel (**Figure C.4** C). For both states the calculated rates are much faster than the overall $P_R \rightarrow F$ transition rate (**Table 7.2**), confirming that forward PT (D132 to E286) through the D-channel is not rate limiting in the $P_R \rightarrow F$ transition. The D-channel PT rates are also faster than the PT back flow rates from the PLS, as discussed above, allowing for the fast reprotonation of E286 and preventing proton back flow from the PLS to deprotonated E286. It is plausible, based on these results, that the decoupling mutants, such as N139T (38) and N139D (42), disable kinetic gating by either slowing down PT through the D-channel (enabling back flow from the PLS to E286 to outcompete the reprotonation of E286; ref (18, 34, 41), or by slowing down the pumping rate (bypassing pumping altogether).

In recent years an alternative mechanism has been suggested (17, 24, 43), in which the excess proton approaches a protonated (neutral) E286 through the D-channel and facilitates proton pumping through a positively charged, biprotonated E286 transition state. In ref (24) the biprotonated mechanism was proposed because the proton pumping free energy barrier from singly protonated E286 calculated with the QM/MM computational methodology in that paper was too high. However, our results show that

the approximate SCC-DFTB method used in ref (24) significantly overestimates the pK_a for aspartic acid deprotonation in bulk water and that the biprotonated E286 transition state is highly energetically unfavorable (see Appendix C). These results suggest that the prohibitive pumping barrier reported in ref (24) could, at least in part, be an artifact of the SCC-DFTB method due to its overestimation of the proton affinity of E286 (see Appendix C for further discussion).

7.4 Conclusions

The quantitative multiscale reactive MD computational analysis of the explicit PT steps in C*c*O presented in this work, combined with previous experimental findings [see refs (18, 23, 25) and references therein], lead us to the following conclusions regarding the most likely sequence of PT and ET events during the $A \rightarrow P_R \rightarrow F$ transition. First, an electron is transferred from heme *a* to the BNC during the 25-50 μs $A \rightarrow P_R$ transition, likely following the chemistry that occurs at the BNC (36, 44). Either coupled with this ET event (during the $A \rightarrow P_R$ transition) or immediately after it (during the $P_R \rightarrow F$ transition), a proton is transferred from a singly protonated E286 to the PLS (PRa_3). Experimental findings suggest that the former (coupled PT/ET) is more likely. (18, 23, 36) This PT induces and is accompanied by an increase in the HC hydration level from approximately four to eight waters. Second, E286 is rapidly reprotonated through the D-channel, and the HC relaxes back to the low hydration state. Third, the uncompensated negative charge in the BNC caused by ET triggers transfer of the chemical proton from E286 to the Cu_B-bound hydroxide in the BNC, forming a water molecule. Fourth, E286 is

reprotonated again through D-channel, accompanied by partial ET from Cu_A to heme *a* and proton ejection from PLS to the P-side.

Based on the above reaction sequence, we combine the PMFs and available experimental data to derive the free energy diagram for the A→F transition (**Figure 7.3**). When the ET is allowed to the BNC without the possibility of protonating PLS (E286Q mutant), only 30% of P_R is formed, the rest remains as P_M' (ref (18), Fig. 10). From this, one can estimate that the ET from heme *a* to BNC alone (without protonation of PLS or BNC) is endergonic by ~ 0.5 kcal/mol. Therefore, starting from the P_M' state (state I, 0 kcal/mol), the PT from E286 to the PLS coupled with ET from heme *a* to heme *a3* leads to P_R state (state II) with energy level of -2.9 + 0.5 = -2.4 kcal/mol. Following this, assuming the pK_a of D132 is similar to that in bulk [which is ~ 3.9 (45)], the proton uptake of D132 from N-side bulk leads to state III (1.9 kcal/mol). Then the proton is transported from D132 to E286, forming state IV (-3.7 kcal/mol). The proton on E286 is subsequently transferred to the BNC, forming the F state (state V, -8.9 kcal/mol). The subsequent proton uptake of D132 from N-side bulk leads to state VI (-4.6 kcal/mol), followed by the second reprotonation of E286 that leads to state VII (-7.7 kcal/mol). Thus, proton release from the PLS to the P-side bulk is estimated to have exergonicity of 5 kcal/mol, completing the A→F transition and leading to state VIII (-12.7 kcal/mol). This 12.7 kcal/mol exergonicity reflects the overall driving force for the A→F transition, as estimated by experimental measurement. (18, 32)

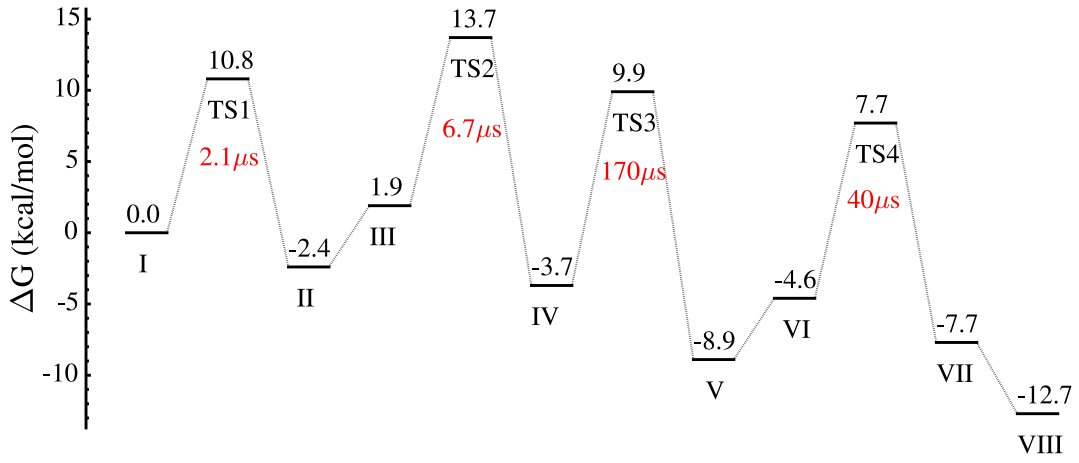


Figure 7.3 Free energy diagram for the reaction sequence during the A→F transition. The resting states are described in the main text. The transition states between them are labeled with "TS", and the activation barriers are obtained from the free energy profiles (PMFs) in **Figure 7.2** C, S3 E, and S4 C. The black numbers indicates the energy levels for each state in kcal/mol. The red numbers indicate the time constants for the forward transitions between the two neighboring resting states.

Our results show that ET from heme *a* to the BNC drives both PT events (E286 to the PLS *and* E286 to the BNC). Among all of these processes, transfer of the chemical proton from E286 to the BNC is rate-limiting and our calculated rate is in quantitative agreement with experiment (8, 9). In this mechanism, the pathways that would decouple pumping from transfer of the chemical proton or damage the directionality of proton flow are avoided by kinetic gating in two ways. First, the fast PT from E286 to the PLS precedes proton transfer to the BNC during the entire A→F transition. Second, the possible proton back flow from the PLS to deprotonated E286 is avoided by fast reprotonation of E286 through the D-pathway combined with a large barrier for proton back flow from the PLS to E286 after ET from heme *a* to BNC (P_R state). Although PT in the D-channel is clearly coupled with solvation changes through the asparagine gate, it is

never found to be rate limiting in the A→F transition. The multiscale reactive MD simulations presented herein have thus contributed to a comprehensive understanding of the functional mechanism of C*c*O, the important final enzyme in the respiratory electron transfer chain. Based on the success of this approach for CcO, we are optimistic that it can also be fruitfully applied to other redox-coupled proton pumping proteins.

7.5 References

1. Knight C & Voth GA (2012) The Curious Case of the Hydrated Proton. *Accounts of Chemical Research* 45(1):101-109.

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

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

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

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

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

7. Lee S, Liang R, Voth GA, & Swanson JMJ (2016) Computationally Efficient Multiscale Reactive Molecular Dynamics to Describe Amino Acid Deprotonation in Proteins. *J. Chem. Theory Comput.* (In Press):DOI: 10.1021/acs.jctc.1025b01109.

8. Faxen K, Gilderson G, Adelroth P, & Brzezinski P (2005) A mechanistic principle for proton pumping by cytochrome c oxidase. *Nature* 437(7056):286-289.

9. Belevich I, Verkhovsky MI, & Wikstrom M (2006) Proton-coupled electron transfer drives the proton pump of cytochrome c oxidase. *Nature* 440(7085):829-832.

10. Svensson-Ek M*, et al.* (2002) The X-ray Crystal Structures of Wild-type and EQ(I-286) Mutant Cytochrome c Oxidases from Rhodobacter sphaeroides. *J. Mol. Biol.* 321(2):329-339.

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

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

13. Peng Y, Swanson JMJ, Kang S-g, Zhou R, & Voth GA (2015) Hydrated Excess Protons Can Create Their Own Water Wires. *The Journal of Physical Chemistry B* 119(29):9212-9218.

14. Kim YC, Wikstrom M, & Hummer G (2007) Kinetic models of redox-coupled proton pumping. *Proc. Natl. Acad. Sci. U. S. A.* 104(7):2169-2174.

15. Belevich I, Bloch DA, Belevich N, Wikstrom M, & Verkhovsky MI (2007) Exploring the proton pump mechanism of cytochrome c oxidase in real time. *P Natl Acad Sci USA* 104(8):2685-2690.

16. Kim YC, Wikstrom M, & Hummer G (2009) Kinetic gating of the proton pump in cytochrome c oxidase. *Proc. Natl. Acad. Sci. U. S. A.* 106(33):13707-13712.

17. Siegbahn PEM & Blomberg MRA (2010) Quantum Chemical Studies of Proton-Coupled Electron Transfer in Metalloenzymes. *Chem Rev* 110(12):7040-7061.

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

19. Hammes-Schiffer S & Stuchebrukhov AA (2010) Theory of Coupled Electron and Proton Transfer Reactions. *Chem Rev* 110(12):6939-6960.

20. Kim YC & Hummer G (2012) Proton-pumping mechanism of cytochrome c oxidase: A kinetic master-equation approach. *Bba-Bioenergetics* 1817(4):526-536.

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

22. Lu J & Gunner MR (2014) Characterizing the proton loading site in cytochrome c oxidase. *P Natl Acad Sci USA* 111(34):12414-12419.

23. Wikström M, Sharma V, Kaila VRI, Hosler JP, & Hummer G (2015) New Perspectives on Proton Pumping in Cellular Respiration. *Chem. Rev.* 115(5):2196-2221.

24. Goyal P, Yang S, & Cui Q (2015) Microscopic basis for kinetic gating in cytochrome c oxidase: insights from QM/MM analysis. *Chem. Sci.* 6(1):826-841.

25. Lepp H, Svahn E, Faxen K, & Brzezinski P (2008) Charge transfer in the K proton pathway linked to electron transfer to the catalytic site in cytochrome c oxidase. *Biochemistry* 47(17):4929-4935.

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

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

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

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

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

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

32. Wikstrom M & Verkhovsky MI (2007) Mechanism and energetics of proton translocation by the respiratory heme-copper oxidases. *Biochimica et biophysica acta* 1767(10):1200-1214.

33. Lee HJ, Ojemyr L, Vakkasoglu A, Brzezinski P, & Gennis RB (2009) Properties of Arg481 Mutants of the aa(3)-Type Cytochrome c Oxidase from Rhodobacter sphaeroides Suggest That neither R481 nor the Nearby D-Propionate of Heme

a(3) Is Likely To Be the Proton Loading Site of the Proton Pump. *Biochemistry* 48(30):7123-7131.

34. Chang HY*, et al.* (2012) Exploring the proton pump and exit pathway for pumped protons in cytochrome ba(3) from Thermus thermophilus. *P Natl Acad Sci USA* 109(14):5259-5264.

35. Wikstrom M*, et al.* (2005) Gating of proton and water transfer in the respiratory enzyme cytochrome c oxidase. *P Natl Acad Sci USA* 102(30):10478-10481.

36. Gorbikova EA, Belevich I, Wikstrom M, & Verkhovsky MI (2008) The proton donor for O-O bond scission by cytochrome c oxidase. *P Natl Acad Sci USA* 105(31):10733-10737.

37. Belevich I*, et al.* (2010) Initiation of the proton pump of cytochrome c oxidase. *P Natl Acad Sci USA* 107(43):18469-18474.

38. Lepp H, Salomonsson L, Zhu JP, Gennis RB, & Brzezinski P (2008) Impaired proton pumping in cytochrome c oxidase upon structural alteration of the D pathway. *Biochimica et biophysica acta* 1777(7-8):897-903.

39. Han D*, et al.* (2006) Replacing Asn207 by Aspartate at the Neck of the D Channel in the aa3-Type Cytochrome c Oxidase from Rhodobacter sphaeroides Results in Decoupling the Proton Pump. *Biochemistry* 45(47):14064-14074.

40. Henry RM, Yu CH, Rodinger T, & Pomes R (2009) Functional Hydration and Conformational Gating of Proton Uptake in Cytochrome c Oxidase. *J Mol Biol* 387(5):1165-1185.

41. Henry RM, Caplan D, Fadda E, & Pomes R (2011) Molecular basis of proton uptake in single and double mutants of cytochrome c oxidase. *J Phys Condens Matter* 23(23):234102.

42. Namslauer A, Pawatet AS, Gennis R, & Brzezinski P (2003) Redox-coupled proton translocation in biological systems: Proton shuttling in cytochrome c oxidase. *P Natl Acad Sci USA* 100(26):15543-15547.

43. Siegbahn PEM & Blomberg MRA (2008) Proton Pumping Mechanism in Cytochrome c Oxidase. *J Phys Chem A* 112(50):12772-12780.

44. Karpefors M, Adelroth P, Namslauer A, Zhen YJ, & Brzezinski P (2000) Formation of the "peroxy" intermediate in cytochrome c oxidase is associated with internal proton/hydrogen transfer. *Biochemistry* 39(47):14664-14669.

45. Lide DR (2004) CRC handbook of chemistry and physics.

Chapter 8

Multiscale simulations reveal acid activation mechanism of the influenza A M2 proton channel

8.1 Introduction

The influenza A M2 channel (AM2) is activated in low pH condition and it transports protons into the virus interior, which is a crucial step in viral replication (1-3). The His37 tetrad is located near the center of the M2 transmembrane peptide (M2TM) and serves as a pH sensor of viral exterior as well as a proton selectivity filter. The Trp41 tetrad C-terminal of the His37 is found to function as a pH dependent gate for proton conductance.(4) The interplay between the protonation state change of His37 tetrad and conformational changes of the protein backbone and the Trp41 side chains is believed to be essential for the pH activation mechanism of the channel. (5-7) In high pH condition the Trp41 side chains close the C-terminal pore below His37, forming a gate that blocks proton flow through the channel, whereas in low pH condition the Trp41 gate is opened and allows proton flow through the channel, activating the proton conduction. (4)

In our recent simulation work (8), we reported that during the transition from intermediate to low pH condition, the C-terminal helices of the channel expand and the hydration level in the C-terminal pore is increased. This decreases the His37 deprotonation barrier and increases the proton conductance in low pH condition, which partially explains the acid activation mechanism of the channel. However, the acid activation mechanism for the transition from high to intermediate pH condition was not

investigated. Moreover, in a recent work by Thomaston et al (9), several M2TM crystal structures were resolved in both high and low pH conditions in lipid cubic phase. It was found that the backbones of the AM2 channel adopt similar structures in both low and high pH conditions. In all structures, the C-terminal helices are wide open, allowing sufficient solvation of His37 and Trp41 tetrad. The similarity of the new set of crystal structures in both low and high pH conditions give rise to further questions regarding how the channel conduction is activated upon lowering the pH value.

Building upon our previous simulations (8) and the recently published crystal structures (9), in this work we further investigated the AM2 acid activation mechanism for the transition from high to intermediate pH condition. We performed multiscale simulations starting from the high pH room temperature crystal structure in ref (9) (PDB code 4QKL). Our results show that in membrane environment, the channel closes its C-terminal helices in high pH condition and the resulting large deprotonation barrier of His37 blocks the channel. Lowering the pH value increases the charge state of His37, and as a result the Trp41 gate is gradually opened. This increases the channel conductance and leads to activation of the AM2 channel in intermediate pH condition.

In the paper by Sharma et al (5), the NMR structure of AM2 conductance domain (residues 22-62) was published (PDB code 2L0J). The structure includes both M2TM (residues 26-46) and the amphipathic helices. However, Ma et al (10) concluded that the M2TM helices conserves the essential features of the proton transport in the full length AM2 channel, and it is for this reason that our previous simulations (8, 11) and many other computational and experimental works (6, 12-17) used the M2TM helices to investigate the proton conduction mechanism of AM2. Here in this work we also

investigated the effect of including the amphipathic helices on the proton conduction mechanism of the channel by performing simulations starting from the 2L0J NMR structure. Our results show that the amphipathic helices do not significantly change the proton conduction mechanism of the AM2.

8.2 Method and Simulation details

8.2.1 Classical Molecular Dynamics (MD) Simulations

The crystal structure resolved in room temperature and high pH condition (4QKL) (9) and the NMR structure obtained in intermediate pH condition (2L0J) (18) were used as initial structures for MD simulation. For the simulations starting from crustal structure, the protein was inserted into lipid bilayer membrane consisting of 170 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) molecules and then solvated by a 25-Å thick layer of water molecules on each side. The protein's principal axes were aligned with the z axis, and the lipid bilayer was aligned with the x–y plane. For the equilibration of +0 state system, all the four His37 residues were singly protonated at N_ε atom.(5, 6) In the first stage of the equilibration, 100 kcal·mol^{-1}·Å^{-2} harmonic restraints were first applied to the protein backbone C_α atoms and then gradually reduced to 0.01 kcal·mol^{-1}·Å^{-2} during 30 ns of simulation time, while the lipid and water molecules relaxed to stabilize the protein–lipid bilayer packing. Then the harmonic restraints were removed, allowing all molecules in the system to relax for another 5 ns. Equilibration was followed by ~500 ns of production simulation. The equilibration of +1 and +2 state systems was initiated from the final structure of the +0 state production simulation by protonating of one or two of the His37, respectively. For the +2 state, the two charged

His37 residues were in the diagonal position.(14) For each of the two states, 1 kcal·mol^{-1}·Å^{-2} harmonic restraints were first applied to protein backbone C_α atoms and then gradually reduced to 0.01 kcal·mol^{-1}·Å^{-2} over ~10 ns simulation time. Then all harmonic restraints were released and a production trajectory of ~180 ns was sampled for each system.

For the simulations starting from NMR structure, the protein was inserted into lipid bilayer membrane consisting of 207 POPC molecules, and then solvated by a 30-Å thick layer of water molecules on each side. The protein's principal axes were aligned with the z axis, and the lipid bilayer was aligned with the x–y plane. Only the +2 state was simulated, where the two doubly protonated His37 were in diagonal position (14), and the singly protonated His37 were protonated at N_ε atom.(5, 6). To facilitate the equilibration, the Trp41 side chains were manually rotated from t-105 rotamer to t90 rotamer, according to ref (7) and (14). In the first stage of the equilibration, 100 kcal·mol^{-1}·Å^{-2} harmonic restraints were first applied to the all of the protein backbone's C_α atoms for ~30 ns, while the lipid and water molecules relaxed to stabilize the protein–lipid bilayer packing. In the second stage of the equilibration, the restraints on the C_α atoms of amphipathic helices were gradually reduced to 0.01 kcal·mol^{-1}·Å^{-2} during 20 ns of simulation time, while keeping the restraints on the M2TM C_α at 100 kcal·mol^{-1}·Å^{-2} . In the third stage, the amphipathic helices were allowed to move freely, while the restraints on the C_α atoms of M2TM helices were reduced to 0.01 kcal·mol^{-1}·Å^{-2} over 65 ns. Then all the harmonic restraints were totally removed, allowing all molecules in the system to relax for another 10 ns. Equilibration was followed by ~220 ns of production simulation.

The protein and lipid were modeled using the CHARMM36 force field (19-22) and the water molecules were described using the TIP3P model.(23) The electrostatic interactions were treated with particle mesh Ewald, using a spherical cutoff of 12.0 Å and an accuracy threshold of 10^{-6} (24). The Lennard-Jones (LJ) interaction cutoff was 12.0 Å using a switching function starting at 10 Å. Simulations were performed in the constant number, pressure, temperature (NPT) ensemble at 308 K, 1 atm. For simulations starting from the crystal structure, the classical MD simulations were performed with the NAMD package (25). The temperature and pressure were controlled by the Langevin temperature coupling and Langevin piston pressure coupling schemes (26), respectively. The equations of motion were integrated using the velocity Verlet integrator, with time step of 2.0 fs (27). For the simulation starting from NMR structure, the classical MD simulations were performed with GROMACS 5.0.4 software package (28, 29). The temperature and pressure were controlled by the V-rescale algorithm (30) and the Berendsen pressure coupling algorithm (31), respectively. The equations of motion were integrated using the leap-frog integrator, with time step of 2.0 fs

8.2.2 QM/MM simulation

The free energy profile (or potential of mean force, PMF) for proton transport through the His37-Trp41 tetrad region was calculated with hybrid quantum mechanical/molecular mechanical (QM/MM) umbrella sampling simulations. The PMF is calculated as a function of collective variable (CV) defined as the z coordinate difference between the center of mass of C_α atoms of Gly34 and that of the excess proton center of excess charge

(CEC). The excess proton CEC coordinate used in the QM/MM simulation was defined as (32):

$$\vec{\xi} = \sum_{i=1}^{N_H} \vec{r}^{H_i} - \sum_{j=1}^{N_X} w^{X_j} \vec{r}^{X_j} - \sum_{i=1}^{N_H} \sum_{j=1}^{N_X} f_{sw}\left(d_{X_j H_i}\right)\left(\vec{r}^{H_i} - \vec{r}^{\mathrm{X}_j}\right) + \vec{\xi}_{\mathrm{correct}}, \quad (1)$$

where the X_j's are the histidine nitrogen atoms and the water oxygen atoms in the QM region, and H_i's are the hydrogen atoms bound to those heavy atoms in the QM region. The w^{X_j}'s are the hydrogen coordination number of a molecule in the least protonated state during the PT process. Therefore, for water oxygen atoms they were set to 2, and for His37 nitrogen atoms they were set to 0.50 in +0 PMF (4 hydrogens shared by 8 nitrogens in the +0 state). Similarly, the w^{X_j}'s for the His37 nitrogen atoms were set to 0.625 in +1 PMF (5 hydrogens shared by 8 nitrogens in +1 state) and 0.75 in +2 PMF (6 hydrogens shared by 8 nitrogens in the +2 state). The $d_{\mathrm{X_j H_i}}$ is the distance between atom X_j and atom H_i, and $f_{sw}\left(d_{X_j H_i}\right) = 1/(1 + \exp\left[(d_{X_j H_i} - r_{sw})/d_{sw}\right])$ reflects the coordination number of H_i to X_j, with the parameters chosen as $d_{sw} = 0.04$ Å, $r_{sw} = 1.25$ Å (33).

In addition, a correction term $\vec{\xi}_{\mathrm{correct}}$ was introduced in the excess proton CEC definition in Eq. 1) to correct for the contribution due to the presence of multiple protons around multiple protonatable sites in the His37 tetrad:(8)

$$\vec{\xi}_{\mathrm{correct}} = \frac{1}{8}\sum_{i=1}^{8}\sum_{j=1}^{8} m_i(\vec{r}^{\mathrm{X}_j} - \vec{r}^{\mathrm{X}_i}), \quad (2)$$

where X denotes the 8 nitrogen atoms of the His37 tetrad, and

$$m_i = \sum_{H_j \in \{H\}} f_{sw}(d_{X_i,H_j})^{16} \Big/ \sum_{H_j \in \{H\}} f_{sw}(d_{X_i,H_j})^{15} \quad (3)$$

m_i gradually switches from 1 to 0 as the nitrogen atom i is deprotonated. The excess proton CEC as defined in Eq. 1 describes the delocalized nature of the excess proton and has been shown to accurately locate the position of the excess proton charge defect in previous QM/MM simulations of biological PT channels (32, 33).

The last snapshot of +0, +1 and +2 classical MD simulation trajectories was used as the initial structure for the QM/MM simulations for the +0, +1 and +2 PMFs, respectively. An excess proton was added to a water molecule ~6 Å above the His37 residue. The QM atoms included His37 side chains and up to three solvation shells of water molecules above and below His37. In all simulations, the QM box size was chosen to be 8 Å larger than the actual size of the QM atoms in each dimension. A quadratic confining wall potential was applied to restrain the QM atoms within the QM box, with the wall skin thickness being 2 Å × 2 Å × 2 Å. The QM region was treated by BLYP level density functional theory (34, 35) with empirical dispersion corrections (36), under the Gaussian plane wave (GPW) scheme (37). The Goedecker-Teter-Hutter (GTH) pseudopotentials (38) were used and the Kohn-Sham orbitals were expanded in the Gaussian TZV2P basis set. The electron density was expanded by auxiliary plane wave basis set up to 360 Ry. The Gaussian Expansion of the Electrostatic Potential (GEEP) scheme was used to treat the QM/MM electrostatic coupling with periodic boundary conditions (PBCs) (39, 40), and the spurious QM/QM periodic image interactions were decoupled as described in ref (41). The equation of motion of nuclei was integrated using a time step of 0.5 fs, and the wavefunction was optimized to the Born-Oppenheimer surface by an orbital

transformation method (42) with a convergence criterion of 10^{-6}. The temperature was controlled at 308K by a Nosé-Hoover thermostat with a relaxation time constant of 0.1 ps. The initial configurations of all windows were prepared from the previous equilibrated window with smaller z value. With a ~0.25 Å window spacing and force constants of 40 kcal/mol/Å^2), ~40 windows were simulated for each of the QM/MM PMFs, respectively. All windows had an equilibration time of ~2 ps and production sampling of 5~15 ps. The excess proton CEC position was collected every step (0.5 fs). All QM/MM simulations are performed using the CP2K package (43).

8.2.3 MS-RMD Umbrella Sampling

The PT PMF in the region outside the His37-Trp41 tetrad region was calculated using the multi-scale reactive molecular dynamics (MS-RMD) method (44-47) with the RAPTOR software (48) implemented with the LAMMPS MD package (http://lammps.sandia.gov) (49). Umbrella sampling was employed and the same CV was used, except that in the MS-RMD simulation the excess proton CEC is given by (50):

$$\vec{r}_{\mathrm{CEC}} = \sum_{i}^{N} c_i^2 \vec{r}_{\mathrm{coc}}^{\,i}, \tag{4}$$

where $\vec{r}_{\mathrm{coc}}^{\,i}$ is the coordinates of the center-of-charge of the i^{th} diabatic state. It was verified that the MS-RMD and QM/MM CEC definitions were nearly identical in the overlapping regions (i.e., when the CEC is nearly 3 solvation shells away from the His37-Trp41 tetrad).

The cutoffs for LJ and real space electrostatic interactions were 12 Å, employing switching function starting at 10 Å for the LJ interactions. Long-range electrostatics were treated by Particle-Particle Particle-Mesh (PPPM) method (51) with an accuracy

threshold of 10^{-4}. The integration time step was 1 fs. The temperature was controlled at 308 K by Nosé-Hoover thermostat with a relaxation time constant of 0.1 ps.

The initial configuration for each umbrella window was prepared by replacing a water molecule with a hydronium close to the center of the window. With a ~0.5 Å window spacing and force constants of 10 kcal/mol/Å^2, 75-80 windows were simulated for each PMF. The replica-exchange umbrella sampling (REUS) method (52) was employed to facilitate convergence, with an exchange attempt frequency of 1 ps^{-1}. All windows were equilibrated for ~100 ps and sampled with REUS for 1~2 ns. The CV value was collected every 10 fs.

8.2.4 Calculation of Proton Conductance

The maximum proton conductance g_{max} for proton permeation through M2 in different charge states was estimated from Poisson-Nernst-Planck (PNP) electrodiffusion theory (53-55):

$$g_{max} = \frac{e^2}{k_B T L^2} \langle D(z)^{-1} e^{+F(z)/k_B T} \rangle^{-1} \langle e^{-F(z)/k_B T} \rangle^{-1} \quad (5)$$

where *e* is the elementary charge, k_B is the Boltzmann's constant, *T* is the temperature, *L* is the length of the M2 channel (L = 45 Å for systems setup with crystal structure, L = 54 Å for system setup with NMR structure because the amphipathic helices of the latter extends the proton pathway within the protein), $D(z)$ and $F(z)$ are the diffusion coefficient and the PMF as a function of the CV, respectively. The brackets denote spatial averaging over the length of the channel in the direction of the channel axis (z

axis). The position-dependent diffusion coefficient for the excess proton was evaluated by the Woolf-Roux equation (56):

$$D(z_i) = \lim_{s\to 0} \frac{-\hat{C}(s;z_i)\langle \delta z^2\rangle_{(i)}\langle \dot{z}^2\rangle_{(i)}}{\hat{C}(s;z_i)\left[s\langle \delta z^2\rangle_{(i)} + \langle \dot{z}^2\rangle_{(i)}/s\right] - \langle \delta z^2\rangle_{(i)}\langle \dot{z}^2\rangle_{(i)}} \quad (6)$$

where $C(t;z_i) = \langle \dot{z}(t)\dot{z}(0)\rangle_i$ is the velocity autocorrelation function for the CV calculated based on the data from window *i*, z_i is the reference point for the harmonic restraint potential in window *i*, and $\hat{C}(s;z_i) = \int_0^\infty e^{-st}C(t;z_i)dt$ is the Laplace transform of this function. To estimate the value of the limit as $s \to 0$, we linearly extrapolate from the range $5 \leq s \leq 15$.

8.3 Results

8.3.1 M2 channel has closed C-terminal helices in high pH condition

The M2TM crystal structure 4QKL (residues 22-46) was resolved at room temperature and high pH condition (pH=8.0) in lipidic cubic phase, and it features opened C-terminal helices. In contrast, the M2TM G34A mutant crystal structure 3LBW (residues 25-46) was resolved at intermediate pH condition (pH=6.5) in detergent, and it features more closed C-terminal helices. Previous wild-type simulations starting from 3LBW structure had showed that the protein structure is stable in lipid bilayer at room temperature for +1 and +2 states, which corresponds to the high to intermediate pH conditions.(12) It was also shown by previous simulations that the C-terminal helices start to open in the +3 state, which corresponds to intermediate to low pH condition.(8, 12) Therefore, it is necessary to also test whether or not the 4QKL structure remains stable in the lipid

bilayer from high to intermediate pH conditions. In order to do so, we inserted the 4QKL crystal structure into POPC lipid bilayer and equilibrate the protein structure for long time (~500 ns) for +0 state. Following this, we increase the charge state of the His37 tetrad and equilibrate the system for +1 and +2 state, which mimics the activation of the channel upon lowering the pH to intermediate condition.

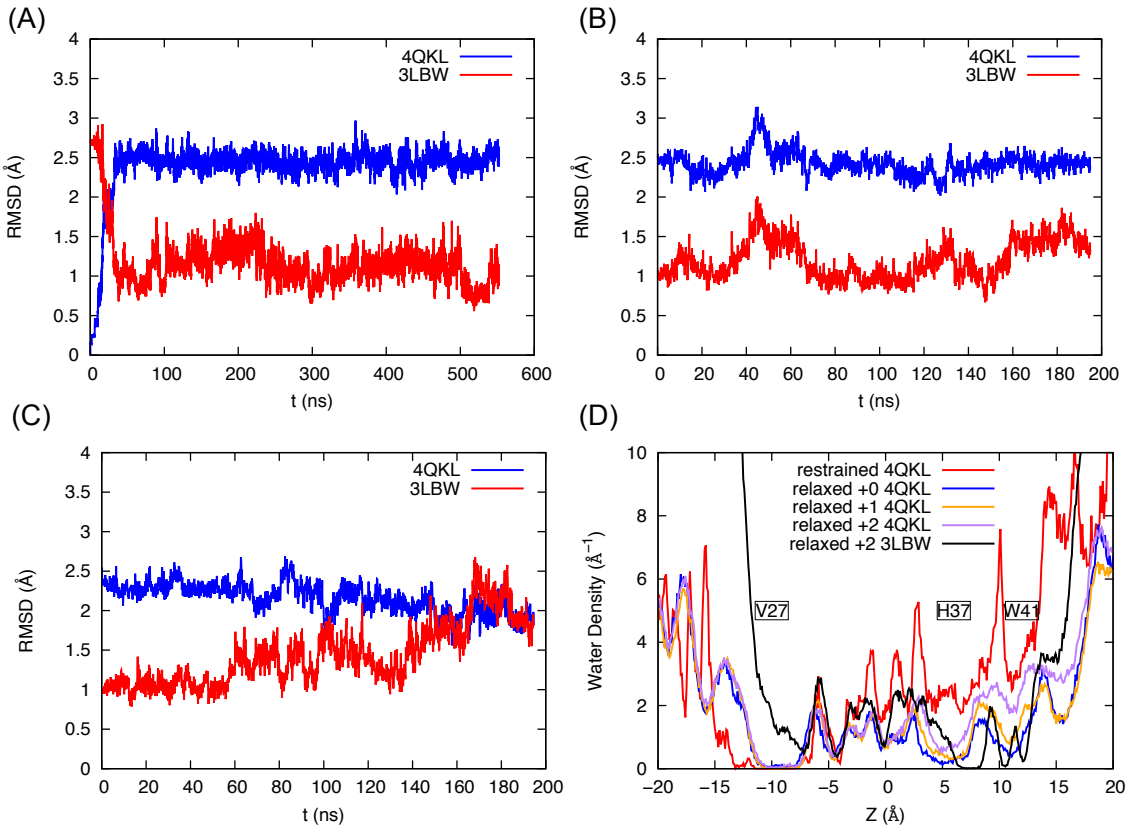


Figure 8.1 Classical MD equilibration of the +0 to +2 state structures. (A-C): RMSD of the C_α atoms of M2TM compared to the starting crystal structure 4QKL (blue) and crystal structure 3LBW for +0, +1 and +2 states (red). (D) Water density profile across the M2 channel for +0 state with protein backbone restrained to 4QKL structure (red), +0, +1 and +2 states for backbones equilibrated in membrane starting from crystal structure 4QKL (blue, yellow and purple, respectively), +2 state for simulations starting from crystal structure 3LBW.

It is found that for the +0 state, after long time equilibration the C-terminal helices are closed, and the protein backbone relaxes towards that of the 3LBW crystal structure,

which has relatively more closed C-terminal helices. (**Figure 8.1** A) The closing of the C-terminal helices brings closer the His37 and Trp41 side chains, and this dehydrates the region between His37 and Trp41. This is evident when we compare the water density profile for the equilibrated +0 state structure with the one having the backbones C_α atom positions restrained to the 4QKL structure (**Figure 8.1** D). In the C-terminal half of the channel (3 Å < z <16 Å), the pore hydration level is significantly reduced for protein structure fully equilibrated in the membrane, as compared to that with protein backbone restrained to the 4QKL structure (blue curve vs. red curve). In +1 and +2 states, the pore hydration is slightly increased due to the increased charge of the His37, and the backbone shows larger conformational flexibility. However, the hydration level in the C-terminal pore is still lower than when the backbone is restrained to 4QKL structure, especially in the His37 and Trp41 region. The low hydration level in the C-terminal pore, as well as the close packing of the Trp41 side chains in the high pH condition contributes to large barriers for His37 deprotonation, as discussed below.

8.3.2 M2 channel is inactivated in high pH condition

The free energy profiles (or potential of mean forces, PMF) for proton transport across the M2 channel in different pH conditions are plotted in **Figure 8.2**. The +0 PMF describes the process where the +0 state His37 tetrad is protonated by an excess proton from viral exterior to form +1 state, followed by deprotonating the excess proton to the viral interior to regenerate the +0 state. This process occurs in high pH condition. It is observed that the barrier for His37 deprotonating the excess proton to the viral interior (CV from 6 Å to 11 Å) is more than 15 kcal/mol. The high barrier of His37 deprotonation makes it the rate-limiting step, and it reduces the proton conductance to values much

lower than the experimental one measured when the M2 is activated in intermediate pH condition (**Table 8.1**). Similar observations are also true for the +1 PMF, where the +1 state His37 tetrad is protonated by an excess proton from viral exterior to form +2 state, followed by deprotonating the excess proton to the viral interior to regenerate the +1 state. Therefore, the +0 and +1 PMFs and conductances indicate that the channel is inactivated in high pH condition.

The large deprotonation barrier of His37 in high pH condition is caused by the low hydration between the His37 and the Trp41. The close packing of the Trp41 side chains in high pH condition (low charge states) reduces the amount of water molecules accessible for His37 deprotonation to the viral interior. The lack of water molecules below the His37 tetrad causes a significant barrier for the deprotonation of His37 to the first solvation shell water molecules below (CV from 6 Å to 8 Å). The Trp41 side chain's steric hindrance further increases the barrier when the excess proton moves from the first solvation shell to the second below His37 (CV from 8 Å to 11 Å). In addition to the high deprotonation barrier, in the +0 and +1 PMFs the free energy minima in the His37 tetrad region are very deep relative to bulk (~ -10 kcal/mol). The deep free energy minima for the excess proton protonating the His37 tetrad indicate that it is energetically favorable for the excess proton to be trapped in the His37 tetrad in high pH conditions, and the His37 deprotonation is thermodynamically unfavorable. This is in agreement with the higher pK_a values for the first and second protonation of the His37 tetrad than the third and fourth (3, 15, 57, 58). The reason for this could be that the His37 tetrad in the low charge state is able to better delocalize the positive charge defect of the excess proton, making the His37 protonation energetically favorable.

Particularly interesting is the asymmetry in the +0 PMF. The barriers for the excess proton to reach His37 from viral exterior and interior are different (**Figure 8.2**). The PT barrier from viral exterior to His37 (~4.9 kcal/mol, CV from -25 Å to 6 Å, denoted as "forward PT barrier" below) is smaller than the PT barrier from viral interior to His37 (~8.9 kcal/mol, CV from 20 Å to 6 Å, denoted as "backward PT barrier" below). The small forward PT barrier is due to lack of significant electrostatic repulsion between the excess proton and the zero positive charge on the +0 state His37 tetrad. The large backward PT barrier is due to the steric hindrance from the tightly packed Trp41 side chains. The asymmetry of the forward and backward PT barriers could explain the rectification of the M2 channel (4, 59), i.e. the M2 channel conducts inward current when the pH in viral exterior (pH_{out}) is low and the pH in viral interior (pH_{in}) is high, but does not conducts outward current when the pH gradient is reversed. The low forward PT barrier allows rapid protonation of His37 from viral exterior when the pH_{out} is low and pH_{in} is high, thereby activating the channel and giving rise to inward current (see below). In contrast, the large backward PT barrier slows down His37 protonation from viral interior when the pH_{in} is low and pH_{out} is high, which hampers channel activation and the formation of outward current. We note that our backward PT barrier in the +0 PMF (~8.9 kcal/mol) is not large enough to completely block the backward PT, and based on our PMF it is possible that the His37 tetrad gets protonated from viral interior when the pH_{in} is low and pH_{out} is high. This could be caused by the limitation of the equilibrated structure we obtained from the simulation, or the force field we are using to simulate the interactions among the Trp41 side chains. A +0 state structure with a more tightly packed Trp41 tetrad could in principle increase the backward barrier to totally block the

backward proton current. Despite the quantitative inaccuray, the qualitative asymmetry in the +0 PMF is evident and is consistent with previous experimental conclusions regarding the rectification.(4, 59)

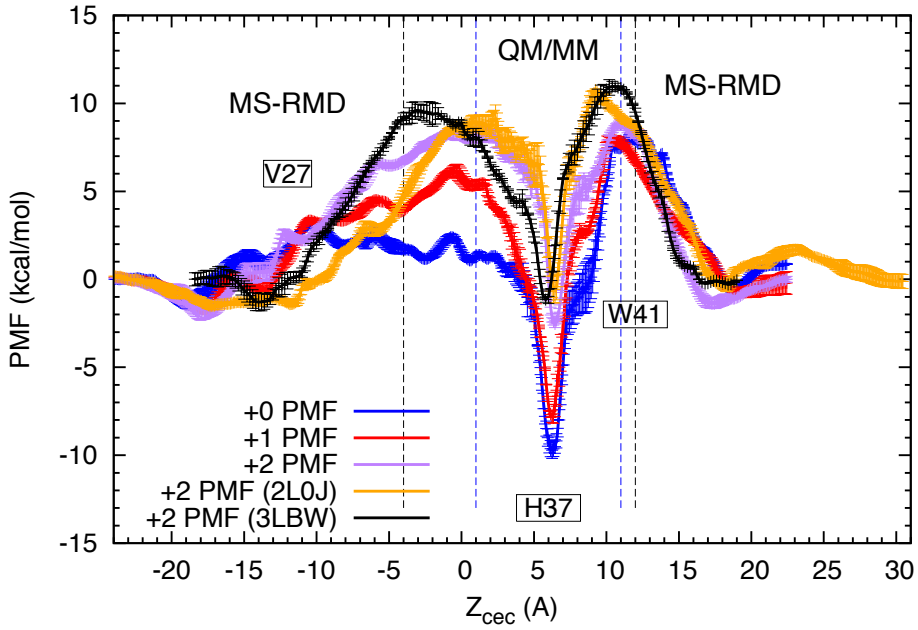


Figure 8.2 Free energy profiles (PMFs) for proton transport through influenza A M2 channel. The blue, red and purple curves are the +0, +1 and +2 PMFs (defined in the text), respectively, calculated from simulations using 4QKL (without amphipathic helices) as initial structure. The yellow curve is the +2 PMF calculated from simulations using 2L0J (with amphipathic helices) as initial structure. The black curve is the +2 PMF calculated from simulations using 3LBW as initial structure and has been reported in ref (8). The His37 deprotonation barrier is lowered in the +2 PMF, and the minimum in the His37 tetrad is raised up, suggesting a pH-dependent activation mechanism. For all of the +2 PMFs, the similarities in the depth of the minima around His37 and the barriers of His37 deprotonation indicate that the proton conduction mechanism is not changed significantly by the amphipathic helices or the initial structure used, as long as the protein is well equilibrated in the membrane.

Table 8.1 Proton conductance comparison between simulations for the +0, +1 and +2 processes (defined in text) and experiments.(60)

State	Conductance (fS)
+0	8.9×10^{-4}
+1	5.9×10^{-2}
+2	20
+2 with amphipathic helices	1.0
+2 from 3LBW	1.2
Experiment (in intermediate pH condition)	0.4 - 4 (5, 60, 61)

8.3.3 M2 channel is activated during +2 to +3 state transition

In the +2 PMF, the +2 state His37 tetrad is protonated by an excess proton from viral exterior to generate +3 state, then deprotonates the excess proton to the viral interior to regenerate the +2 state. The His37 deprotonation barrier is decreased to ~12 kcal/mol, (**Figure 8.2**) and the proton conductance is decreased accordingly (**Table 8.1**). The conductance is in reasonable agreement with our previous result (8) and the experimental values. (5, 60, 61) The decrease in the His37 deprotonation barrier in the +2 PMF could be attributed to increased solvation between His37 and Trp41 (**Figure 8.1** D), as well as increased positive charge on the His37 tetrad that reduces the delocalization of the positive charge defect associated with the excess proton on the His37 residues. On the other hand, the free energy minima in the His37 tetrad region are significantly raised up.

This indicates that the excess proton is no longer deeply trapped in the His37 tetrad during PT. All of the above observations suggest that the M2 channel is activated in the +2 process. The forward PT barrier is increased due to the larger electrostatic repulsion between His37 tetrad and the incoming proton, as compared with the +0 and +1 PMFs. The backward PT barrier remains similar. The Trp41 gets more solvated in the +2 state, slightly reducing the steric hindrance in the backward direction, but the increased electrostatic repulsion between the His37 tetrad and the excess proton compensates for this, resulting in similar backward PT barrier.

Because the full length M2 channel includes both M2TM and amphipathic helices, here we also investigate the influence of the amphipathic helices on the proton conductance mechanism in M2TM helices. For the NMR structure 2L0J, which has both M2TM and amphipathic helices, after manually adjusting the Trp41 to t90 rotamer the M2TM helices remains stable in the membrane and the backbone structure is similar to that of the crystal structure 3LBW. (**Figure 8.3**) The t90 rotamer of Trp41 have been concluded by crystal structures (9, 12, 62), SSNMR measurements (7, 63), as well as all-atom MD simulations.(14) The +2 PMF for the system equilibrated from 2L0J structure is in general similar to that of the system equilibrated from the 4QKL structure, which has only M2TM helices. Specifically, the His37 deprotonation process, which is the rate-limiting step, has similar barrier in both cases, resulting in similar conductances. After the excess proton is transported through the M2TM helices, it only experiences a small (~2 kcal/mol) barrier to be further transported through the amphipathic helices. Therefore, the existence of the amphipathic helices does not significantly influence the proton conductance mechanism in the M2TM helices. This is in line with the experimental

measurements showing that the proton conductances are similar for the construct with M2TM helices only and the construct having both M2TM and amphipathic helices.(10)

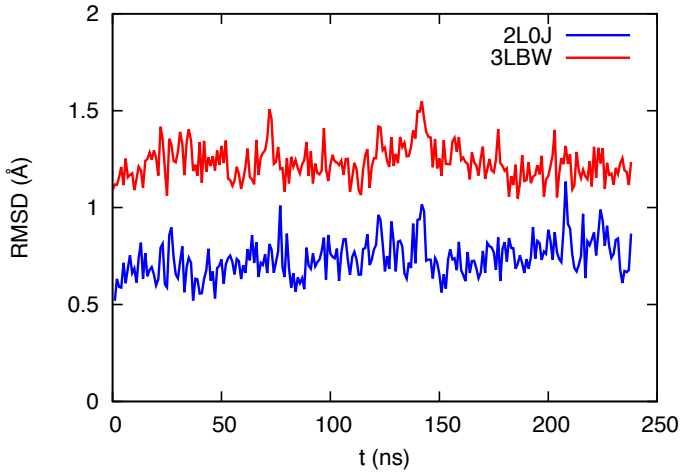


Figure 8.3 RMSD of the C_α atoms in the M2TM helices during the equilibration of NMR structure 2L0J (with amphipathic helices) in lipid bilayer. The two reference structures are 2L0J (blue) and the crystal structure 3LBW (without amphipathic helices, red). The RMSD of M2TM helices indicate that the structure is stable and similar to that of the crystal structure.

8.4 Conclusion

In M2 channel, at high pH condition the C-terminal helices adopt closed conformation in lipid bilayer. The closed Trp41 tetrad dehydrates His37 tetrad and increases the His37 deprotonation barrier, thus blocking proton conduction in high pH condition. The protonation of His37 is easier from viral exterior than from viral interior in low charge state of His37 tetrad, and this asymmetry partly explains the rectification of proton flow observed in experiments(4, 59). As the pH decreases to intermediate value, the positive charges on the His37 increases and the C-terminal helices are more opened and hydrated. This lowers the deprotonation barrier and increases the proton conductance, activating the channel. In combination with our previous work (8), our current results here provide a

complete understanding of the acid activation mechanism of the AM2. Moreover, our results indicate that the inclusion of the amphipathic helices does not influence the proton conductance mechanism of the channel significantly.

8.5 References

1. Wang C, Lamb RA, & Pinto LH (1995) Activation of the M(2) Ion-Channel of Influenza-Virus - a Role for the Transmembrane Domain Histidine Residue. *Biophys. J.* 69(4):1363-1371.

2. Chizhmakov IV*, et al.* (1996) Selective proton permeability and pH regulation of the influenza virus M2 channel expressed in mouse erythroleukaemia cells. *J Physiol-London* 494(2):329-336.

3. Hu J*, et al.* (2006) Histidines, heart of the hydrogen ion channel from influenza A virus: Toward an understanding of conductance and proton selectivity. *Proc. Natl. Acad. Sci. U. S. A.* 103(18):6865-6870.

4. Tang YJ, Zaitseva F, Lamb RA, & Pinto LH (2002) The gate of the influenza virus M-2 proton channel is formed by a single tryptophan residue. *J. Biol. Chem.* 277(42):39880-39886.

5. Sharma M*, et al.* (2010) Insight into the Mechanism of the Influenza A Proton Channel from a Structure in a Lipid Bilayer. *Science* 330(6003):509-512.

6. Hu FH, Luo WB, & Hong M (2010) Mechanisms of Proton Conduction and Gating in Influenza M2 Proton Channels from Solid-State NMR. *Science* 330(6003):505-508.

7. Williams JK, Zhang Y, Schmidt-Rohr K, & Hong M (2013) pH-Dependent Conformation, Dynamics, and Aromatic Interaction of the Gating Tryptophan Residue of the Influenza M2 Proton Channel from Solid-State NMR. *Biophys. J.* 104(8):1698-1708.

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

9. Thomaston JL*, et al.* (2015) High-resolution structures of the M2 channel from influenza A virus reveal dynamic pathways for proton stabilization and transduction. *Proc. Natl. Acad. Sci. U. S. A.* 112(46):14260-14265.

10. Ma CL, *et al.* (2009) Identification of the functional core of the influenza A virus A/M2 proton-selective ion channel. *Proc. Natl. Acad. Sci. U. S. A.* 106(30):12283-12288.

11. Chen HN, Wu YJ, & Voth GA (2007) Proton transport Behavior through the influenza a M2 channel: Insights from molecular simulation. *Biophys. J.* 93(10):3470-3479.

12. Acharya R, *et al.* (2010) Structure and mechanism of proton transport through the transmembrane tetrameric M2 protein bundle of the influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* 107(34):15075-15080.

13. Wang J, *et al.* (2011) Molecular Dynamics Simulation Directed Rational Design of Inhibitors Targeting Drug-Resistant Mutants of Influenza A Virus M2. *J. Am. Chem. Soc.* 133(32):12834-12841.

14. Wei C & Pohorille A (2013) Activation and Proton Transport Mechanism in Influenza A M2 Channel. *Biophys. J.* 105(9):2036-2045.

15. Hu FH, Schmidt-Rohr K, & Hong M (2012) NMR Detection of pH-Dependent Histidine-Water Proton Exchange Reveals the Conduction Mechanism of a Transmembrane Proton Channel. *J. Am. Chem. Soc.* 134(8):3703-3713.

16. Dong H, Fiorin G, DeGrado WF, & Klein ML (2013) Exploring Histidine Conformations in the M2 Channel Lumen of the Influenza A Virus at Neutral pH via Molecular Simulations. *J. Phys. Chem. Lett.* 4(18):3067-3071.

17. Carnevale V, Fiorin G, Levine BG, DeGrado WF, & Klein ML (2010) Multiple Proton Confinement in the M2 Channel from the Influenza A Virus. *J. Phys. Chem. C* 114(48):20856-20863.

18. Vanommeslaeghe K, *et al.* (2010) CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* 31(4):671-690.

19. MacKerell AD, *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102(18):3586-3616.

20. MacKerell AD, Feig M, & Brooks CL (2004) Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* 126(3):698-699.

21. Best RB, *et al.* (2012) Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. *J. Chem. Theory Comput.* 8(9):3257-3273.

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

23. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein ML (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 79(2):926-935.

24. Tom D, Darrin Y, & Lee P (1993) Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98(12):10089-10092.

25. Phillips JC*, et al.* (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26(16):1781-1802.

26. Feller SE, Zhang Y, Pastor RW, & Brooks BR (1995) *Constant Pressure Molecular Dynamics Simulation : The Langevin Piston Method* (American Institute of Physics, Melville, NY, ETATS-UNIS) p 9.

27. Verlet L (1967) Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* 159(1):98.

28. Berendsen HJC, Vanderspoel D, & Vandrunen R (1995) Gromacs - a Message-Passing Parallel Molecular-Dynamics Implementation. *Comput. Phys. Commun.* 91(1-3):43-56.

29. Abraham MJ*, et al.* (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2:19-25.

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

31. Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, & Haak JR (1984) Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* 81(8):3684-3690.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

47. Lee S, Liang R, Voth GA, & Swanson JMJ (2016) Computationally Efficient Multiscale Reactive Molecular Dynamics to Describe Amino Acid Deprotonation in Proteins. *J. Chem. Theory Comput.* 12(2):879-891.

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

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

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

51. Hockney RW & Eastwood JWja (1981) *Computer simulation using particles* (McGraw-Hill International Book Co., New York) p 540.

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

53. Levitt DG (1986) Interpretation of Biological Ion Channel Flux Data--Reaction-Rate versus Continuum Theory. *Annu. Rev. Biophys. Biophys. Chem.* 15(1):29-57.

54. Roux B & Karplus M (1991) Ion-Transport in a Gramicidin-Like Channel - Dynamics and Mobility. *J. Phys. Chem.* 95(12):4856-4868.

55. Levitt DG (1991) General continuum theory for multiion channel. I. Theory. *Biophys. J.* 59(2):271-277.

56. Woolf TB & Roux B (1994) Conformational Flexibility of O-Phosphorylcholine and O-Phosphorylethanolamine - a Molecular-Dynamics Study of Solvation Effects. *J. Am. Chem. Soc.* 116(13):5916-5926.

57. Colvin MT, Andreas LB, Chou JJ, & Griffin RG (2014) Proton Association Constants of His 37 in the Influenza-A M2(18-60) Dimer-of-Dimers. *Biochemistry* 53(38):5987-5994.

58. Liao SY, Yang Y, Tietze D, & Hong M (2015) The Influenza M2 Cytoplasmic Tail Changes the Proton-Exchange Equilibria and the Backbone Conformation of the Transmembrane Histidine Residue to Facilitate Proton Conduction. *J. Am. Chem. Soc.*

59. Chizhmakov IV*, et al.* (2003) Differences in conductance of M2 proton channels of two influenza viruses at low and high pH. *J Physiol-London* 546(2):427-438.

60. Leiding T, Wang J, Martinsson J, DeGrado WF, & Arskold SP (2010) Proton and cation transport activity of the M2 proton channel from influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* 107(35):15409-15414.

61. Mould JA*, et al.* (2000) Mechanism for proton conduction of the M-2 ion channel of influenza A virus. *J. Biol. Chem.* 275(12):8592-8599.

62. Stouffer AL*, et al.* (2008) Structural basis for the function and inhibition of an influenza virus proton channel. *Nature* 451(7178):596-U513.

63. Luo WB, Mani R, & Hong M (2007) Side-chain conformation of the M2 transmembrane peptide proton channel of influenza a virus from (19)F solid-state NMR. *J. Phys. Chem. B* 111(36):10825-10832.

Appendix A
Supporting Information for Benchmark Study of the SCC-DFTB Approach for a Biomolecular Proton Channel

A.1 Comparison to multi-state empirical valence bond simulations

The multi-state empirical valence bond method (MS-EVB) has been applied to studying proton transport in the LS2 channel.(1, 2) This method describes the excess proton Grotthuss shuttling and charge delocalization explicitly by evolving the system on a reactive potential energy surface defined by a linear combination of multiple diabatic basis states.(3-7) In addition to the LS2 channel, the MS-EVB method has been widely applied to studying a variety of other biomolecular PT processes, as reviewed in ref (7). In this section of the supplementary information, we compare the results of the MS-EVB3 method (5) to the SCC-DFTB and DFT methods. Comments on these comparisons are given in the table and figure captions.

A.1.1 System setup

The MS-EVB3 systems were equilibrated in the same way as in SCC-DFTB and DFT simulations, except that the SPC/Fw water model,(8) which is consistent with the MS-EVB3 method, was used instead of the TIP3P water model. For the production runs, the MS-EVB3 model was used to explicitly describe excess proton Grotthuss shuttling and charge delocalization. For the LS2 channel, the production run was simulated in the constant NVT ensemble for 300 ps at 300 K with the temperature maintained by a Nose-Hoover thermostat. The cutoff radius for LJ and real space electrostatic interactions was 12 Å employing a switching function starting at 10 Å for the LJ interactions. The long-range electrostatics were treated by Particle-Particle Particle-Mesh (PPPM) method(9) with an accuracy threshold of 10^{-4}. For the CNT, the production run was simulated in the

constant NVT ensemble for 200 ps at 300 K using Nose Hoover thermostat. The cutoffs for LJ and real space electrostatic interactions were 10 Å, employing switching function starting at 8 Å for the LJ interactions. Long-range electrostatics were treated by Ewald summation with a relative accuracy threshold of 10^{-6}. The integration time step was 1 fs. All of the MS-EVB simulations were carried out by the RAPTOR software.(10)

A.1.2 Supplementary Tables

Table A.1 Average hydrogen bond relaxation times for hydrogen bonds within 4 Å of the excess proton CEC in protonated LS2 channel and CNT. Unlike SCC-DFTB, the MS-EVB3 method predicts slower hydrogen bonding dynamics around the excess proton CEC. The MS-EVB3 hydrogen bond relaxation time for the CNT is comparable to the DFT methods, while that in the LS2 channel is slower due to an apparent over-attraction between the 1st solvation shell around the excess proton CEC and the serine side chains.

Method	LS2 hydrogen bond relaxation time (ps)	CNT hydrogen bond relaxation time (ps)
BLYP-D	12.3	13.7
HCTH	20.4	10.1
B3LYP-D	10.7	NA
SCC-DFTB	1.73	1.4
MS-EVB3	64.9	8.1

A.1.3 Supplementary Figures

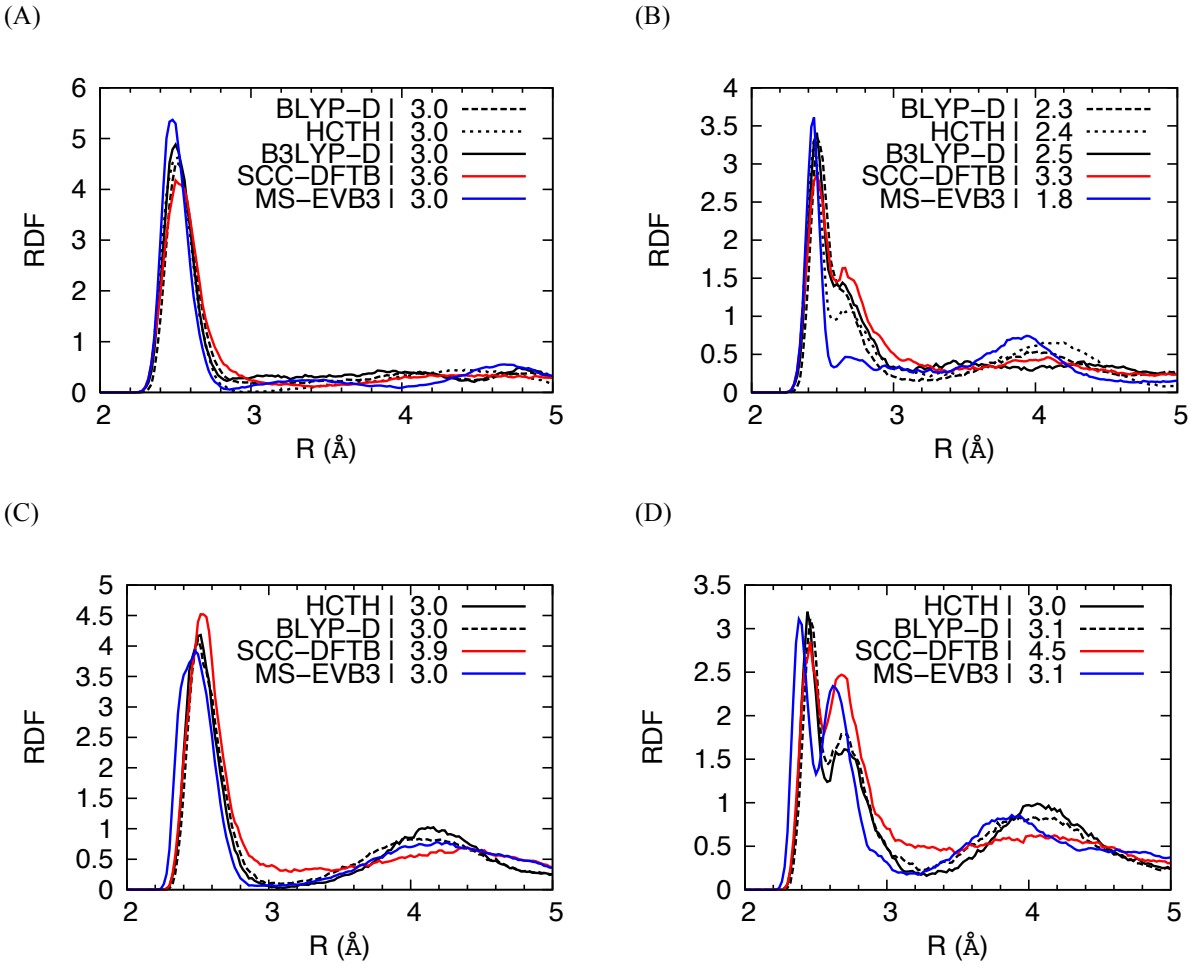


Figure A.1 RDFs of (A) O*-O_w and (B) O_{1x}-O, where O includes both O* and O_w, in the protonated LS2 channel. RDFs of (C) O*-O_w and (D) O_{1x}-O, where O includes both O* and O_w, in protonated CNT. Coordination numbers are indicated in the legend. The MS-EVB3 method produces a more structured solvation shell compared to SCC-DFTB and is in better agreement with the DFT methods. The MS-EVB3 coordination numbers also agree better with the DFT methods, suggesting that the MS-EVB3 does not have the overcoordination issue demonstrated for SCC-DFTB. However, the MS-EVB3 O_{1x}-O RDF in LS2 shows an early drop of the first peak around 2.6 Å. This is caused by an apparent over-attraction between the first solvation shell of the hydronium and serine hydroxyl groups. The oxygen atom in serine side chain mostly replaces one of the water molecules coordinating O_{1x}. The RDF of O_{1x}-O including both O_w and O_s (not shown here) recovers the complete solvation shell of O_{1x}. It should also be noted that the QM/MM boundary may cause errors in the water-serine interactions in the DFT and SCC-DFTB results that are excluded in the MSEVB3 results where the entire system is treated on the same MM footing.

(A) (B)

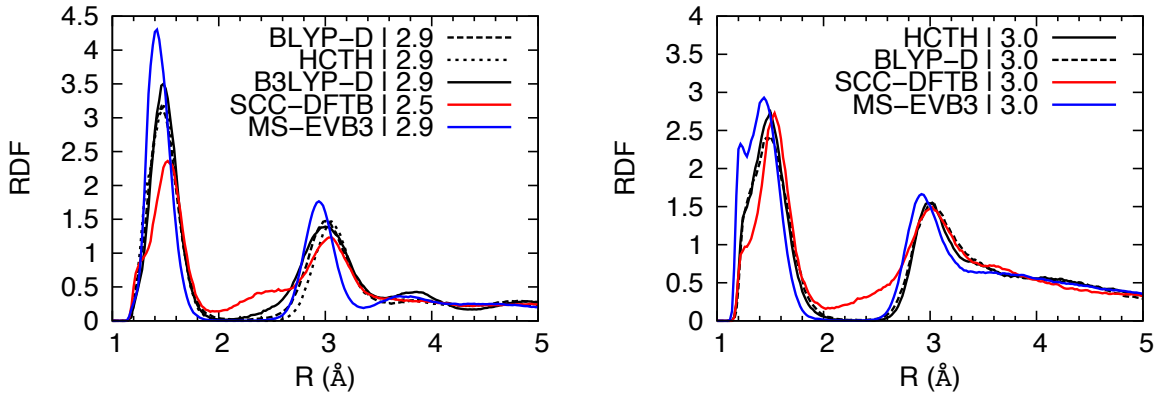


Figure A.2 RDFs of H*-O_w in the protonated (A) LS2 channel and (B) CNT channels, with coordination numbers indicated in the legend. The MS-EVB3 method produces a density depletion region after the first peak in both systems and agrees better with the DFT methods than SCC-DFTB does, as demonstrated by the coordination numbers. The more narrow first solvation peak for MS-EVB3 in the LS2 system is again caused by the over-attraction between the hydronium first solvation shell and serine hydroxyl groups. The bimodal first peak the MS-EVB3 H*-O_w CNT RDF implies that the Zundel structure, in which two water molecules equally share an excess proton, occurs more frequently in MS-EVB3 simulation than in DFT. Note that it occurs much less frequently in the SCC-DFTB simulations. It should also be noted that many aspects of the MS-EVB3 model, including excess proton hydrated cluster energies and internal proton transfer barriers, were parameterized to a higher level of electronic structure than DFT (MP2). It is therefore not clear if these MS-EVB3 results are more or less accurate than the DFT ones.

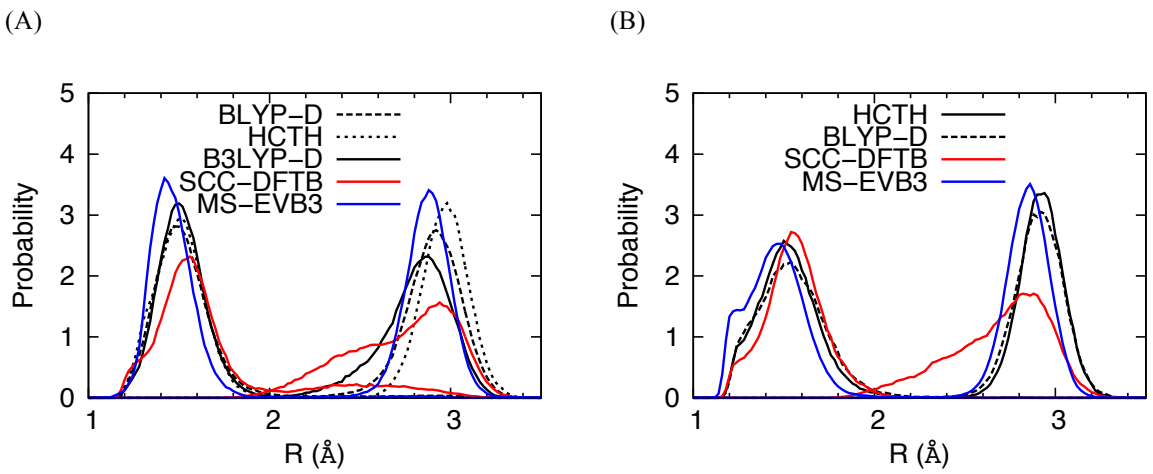


Figure A.3 Distribution of H*-O_w distance in the protonated (A) LS2 and (B) CNT channels. The first and second peaks correspond to the closest and second closest O_w to H*. The MS-EVB3 method shows a similar distribution pattern to those described by the DFT methods: two distinct peaks without overlap in the middle. This indicates the absence of bifurcated hydronium hydrogen bonds and is in contrast to the results from SCC-DFTB in which the two peaks have obvious overlap. Again, the MS-EVB3 method predicts a splitting of the first main peak in the CNT system due to more frequent Zundel structures. See also the comment at the end of Fig. S2 caption.

(A) (B)

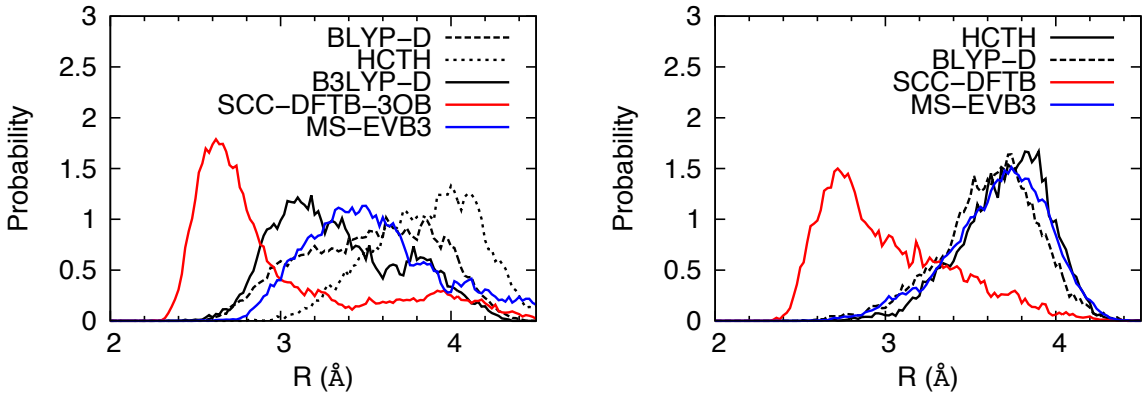


Figure A.4 Distribution of the O*-O distance, where O is the closest atom not hydrogen bonded to O*, for the protonated (A) LS2 and (B) CNT channels. The MS-EVB3 method predicts similar distribution patterns to the DFT methods, indicating that for the majority of simulation the water molecules within the first solvation shell of excess proton are hydrogen bonded to the hydronium O*. This is in contrast to SCC-DFTB result, where the water molecules in the first solvation shell are not hydrogen bonded to O* for a significant amount of time.

(A) (B)

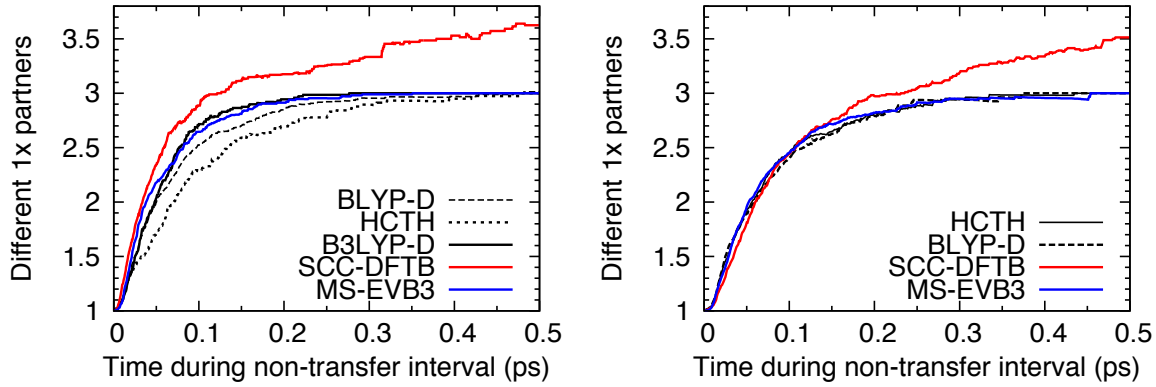


Figure A.5 Average number of different O_{1x} partners to O* as a function of time during non-transfer intervals for the protonated (A) LS2 and (B) CNT channels. The MS-EVB3 method predicts a "special-pair dance" behavior(11) similar to the DFT results, with the special partner alternating among 3 nearby water molecules. This is in contrast with SCC-DFTB results, which predict more than 3 water molecules as special partners due to overcoordination of the excess proton.

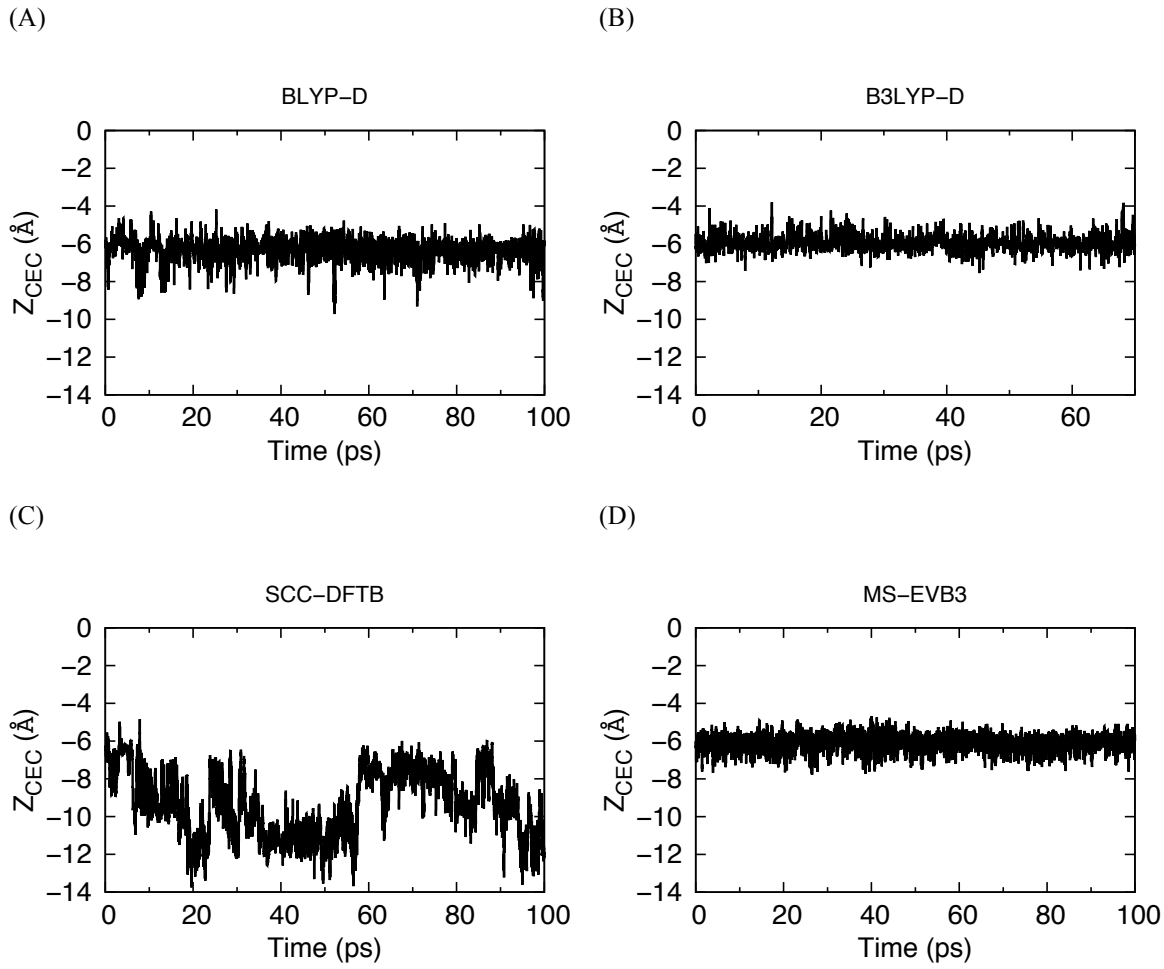


Figure A.6 Z coordinate of the proton CEC as a function of time for unconstrained simulations using the (A) BLYP-D (B) B3LYP-D (C) SCC-DFTB and (D) MS-EVB3 methods. The excess proton CEC definition used in the MS-EVB3 simulation was the same as that used in the DFT and SCC-DFTB simulations, and the unconstrained simulation was run in the same way as described for DFT and SCC-DFTB in the main text. For MS-EVB3, the z coordinate fluctuates slightly around z = -6 Å, in agreement with the DFT simulations and with the previous MS-EVB free energy profile(2) that shows a distinct local minimum at z = -6 Å. In contrast, for SCC-DFTB the z coordinate deviates as far as 6 Å from z = -6 Å.

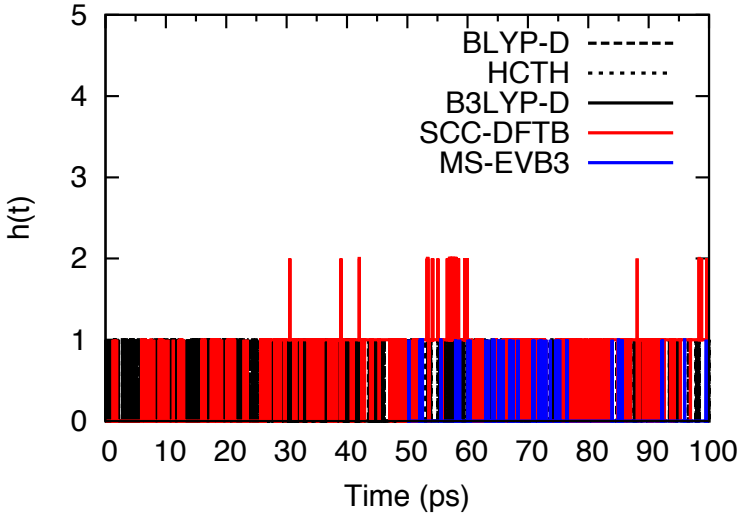


Figure A.7 Proton hopping function for the free excess proton simulation initiated from the constrained case where the center of excess charge is trapped at a wide region. In the MS-EVB3 simulation the hopping function fluctuates between 0 and 1, in agreement with the DFT results and in contrast to the SCC-DFTB result.

(A) (B)

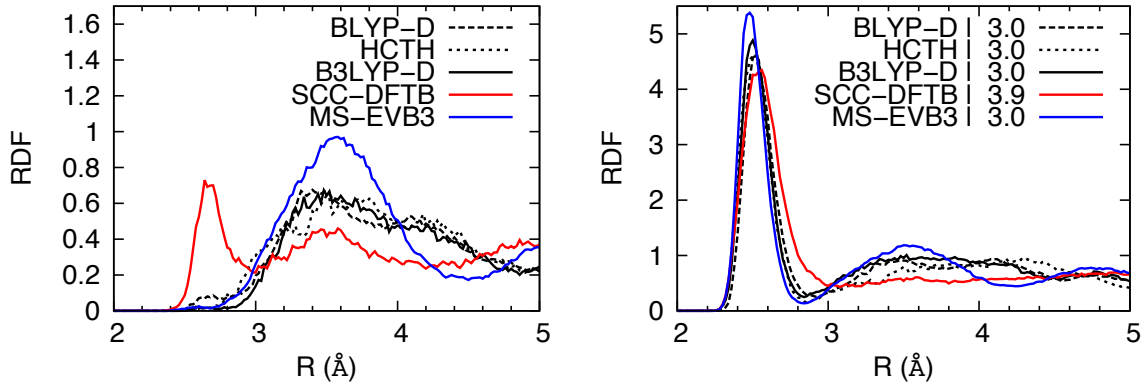


Figure A.8 (A) RDF of O*-O_s in the protonated LS2 channel. The location of first peak predicted by MS-EVB3 agrees better with the DFT methods than the SCC-DFTB peak does, which indicates that the hydronium O* is not directly hydrogen bonded to serine O_s. However, the MS-EVB3 method predicts a more pronounced first peak than the DFT methods, which suggests an over-attraction between O* and O_s. (B) The RDF of O*-O, where O includes both O_w and O_s. The coordination number predicted by MS-EVB3 agrees with the DFT methods whereas that predicted by SCC-DFTB is larger.

A.2 Additional results for proton solvation and transport in a narrower region of the LS2 channel

In this section, additional results for the proton solvation and transport in a narrow region of the LS2 channel (z = - 3.5 Å) are summarized. The results are in line with the conclusions for the wide region of the LS2 channel mentioned above and in the main text, i.e., the SCC-DFTB method yields an overcoordinated excess proton, weak hydrogen bonding around the excess proton CEC, and PT dynamics at odds with the DFT results, whereas the MS-EVB method yields similar results to the DFT methods.

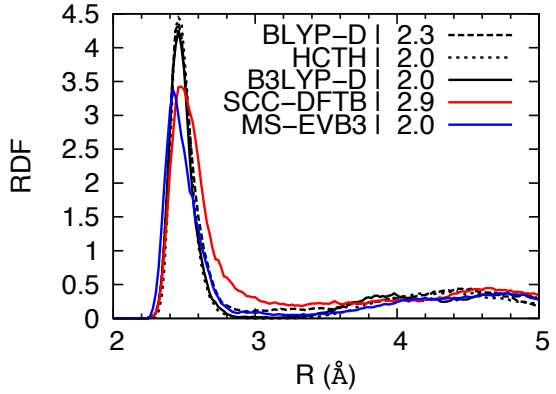
A.2.1 Supplementary Tables

Table A.2 Average hydrogen bond relaxation time for hydrogen bonds within 4 Å of the excess proton CEC in the narrow region of the protonated LS2 channel.

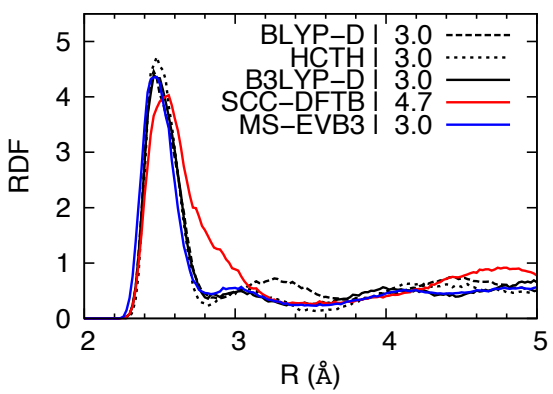
Method	LS2 HB relaxation time (ps)
BLYP-D	21
HCTH	22
B3LYP-D	9
SCC-DFTB	0.8
MS-EVB3	6

A.2.2 Supplementary Figures

(A)



(B)



(C)

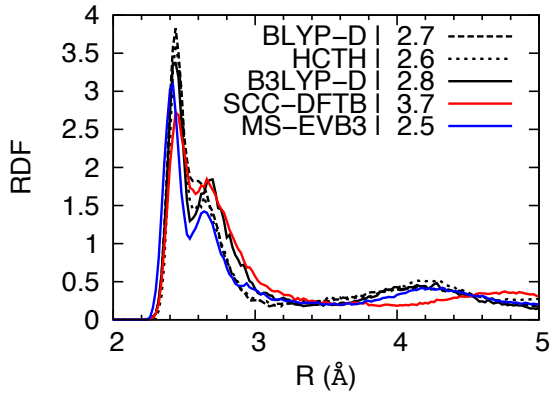


Figure A.9 RDFs of (A) O*-O_w, (B) O*-O , and (C) O_{1x}-O , where O includes both O_w and O_s, in a narrow region of the protonated LS2 channel.

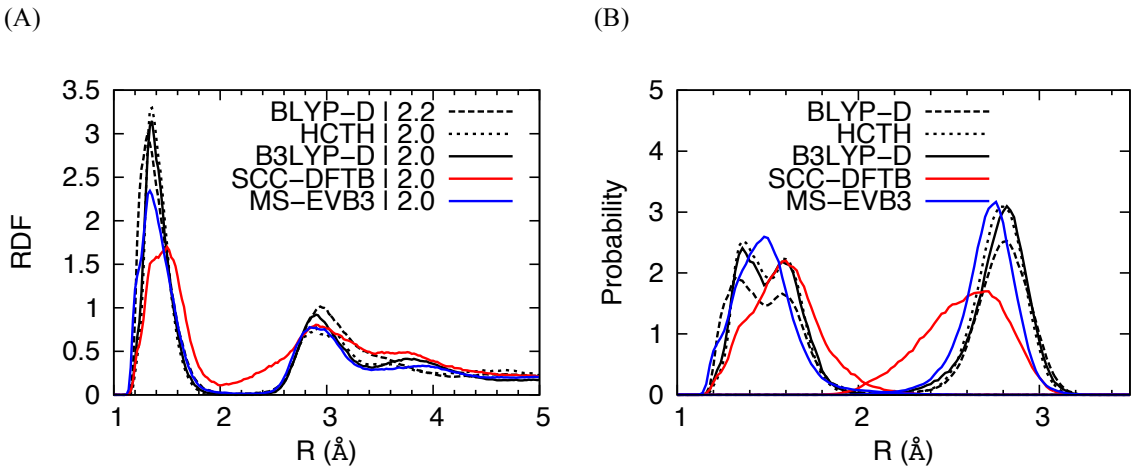


Figure A.10 (A) RDF of H*-O_w in the narrow region of the protonated LS2 channel. (B) Distribution of H*-O distances, where O includes both O_w and O_s, in the narrow region of the protonated LS2 channel.

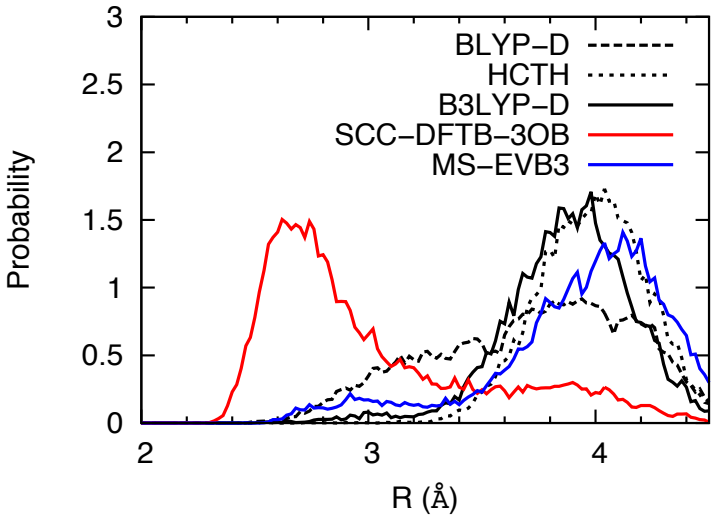


Figure A.11 Distribution of the O*-O distance, where O is the closest atom not hydrogen bonded to O*, for a narrow region of the protonated LS2 channel.

A.3 B3LYP-D/GTH-DZVP improves description of forces and energies over BLYP-D/GTH-TZV2P

To illustrate the advantage of employing B3LYP-D/GTH-DZVP over BLYP-D/GTH-TZV2P, 10 uncorrelated frames were chosen from 80 ps of the BLYP-D/MM simulation, and the forces on each QM oxygen atom was calculated with BLYP-D/GTH-TZV2P and B3LYP-D/GTH-DZVP and plotted against that calculated with B3LYP-D and a large basis set (6-31++G**; **Figure A.12**). The B3LYP-D/GTH-DZVP leads to a root mean squared error (RMSE) that is 6.56 kcal/(mol Å) less than that of BLYP-D/GTH-TZV2P, suggesting that B3LYP-D/GTH-DZVP does indeed offer an improved description of the forces on QM atoms over BLYP-D/GTH-TZV2P. In addition, using B3LYP-D/GTH-TZV2P as benchmark we calculated the energy error for BLYP-D/GTH-TZV2P and B3LYP-D/GTH-DZVP, averaged over the same set of configurations used in the force comparison. The B3LYP-D/GTH-DZVP reduces the energy error of BLYP-D/GTH-TZV2P from 0.65 hartree to 0.27 hartree, and therefore provides significant improvement over the latter. The TZV2P basis set used in the benchmark was shown to predict converged structural properties of bulk water. (12)

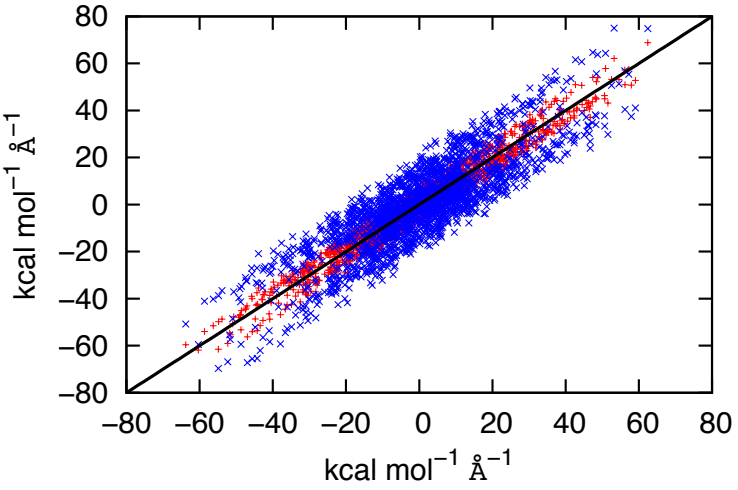


Figure A.12 Scatter plot of forces from B3LYP-D/6-31++G** vs B3LYP-D/GTH-DZVP (red) and BLYP-D/GTH-TZV2P (blue). The B3LYP-D/GTH-DZVP provides more accurate forces than BLYP-D/GTH-TZV2P.

A.4 Supplementary References

1. Wu YJ & Voth GA (2003) A computer simulation study of the hydrated proton in a synthetic proton channel. *Biophys. J.* 85(2):864-875.

2. Wu YJ, Ilan B, & Voth GA (2007) Charge delocalization in proton channels, II: The synthetic LS2 channel and proton selectivity. *Biophys. J.* 92(1):61-69.

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

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

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

6. Voth GA (2006) Computer simulation of proton solvation and transport in aqueous and biomolecular systems. *Acc. Chem. Res.* 39(2):143-150.

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

8. Wu YJ, Tepper HL, & Voth GA (2006) Flexible simple point-charge water model with improved liquid-state properties. *J. Chem. Phys.* 124(2).

9. Hockney RW & Eastwood JWja (1981) *Computer simulation using particles* (McGraw-Hill International Book Co., New York) pp xix, 540 p.

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

11. Markovitch O, *et al.* (2008) Special pair dance and partner selection: Elementary steps in proton transport in liquid water. *J. Phys. Chem. B* 112(31):9456-9466.

12. VandeVondele J, *et al.* (2005) The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water. *J. Chem. Phys.* 122(1).

Appendix B
Supporting Information for Multiscale Simulation Reveals a Multifaceted Mechanism of Proton Permeation through the Influenza A M2 Proton Channel

B.1 Classical Molecular Dynamics (MD) Simulations.

Starting from the crystal structures (PDB codes 3LBW and 3C9J), which are both G34A mutants, the classical MD simulations were performed by first changing the Ala34 residues back to wild type glycines. The AM2/TM tetramers of distinct protonation states were then embedded in an 80×80 $Å^2$ 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer, and solvated by 25 Å thick layer of TIP3P water on each side (1). The protein principal axes were aligned with the *z*-axis, and the lipid bilayer was aligned with the *x*-*y* plane. All interactions were described by the CHARMM27 force field (2). During the first stage of equilibration, the protein heavy atoms were fixed for 20 ns while the lipid and water molecules relaxed to stabilize the protein-lipid bilayer packing. In the second stage of equilibration, 100 kcal/mol/$Å^2$ harmonic restraints were first applied to the protein backbone C_α atoms, and then gradually reduced to 1 kcal/mol/$Å^2$ during 5 ns of simulation time. Finally, the harmonic restraints were removed, allowing every particle in the system to relax for another 5 ns. Equilibration was followed by 15-20 ns of production simulation. The electrostatic interactions were treated with particle mesh Ewald, using a spherical cutoff of 12.0 Å and an accuracy threshold of 10^{-6} (3). The Lennard-Jones interaction cutoff was 12.0 Å employing a switching function starting at 10 Å. Simulations were performed in the constant NPT ensemble at 308 K, 1 atm using the Langevin temperature coupling and Langevin piston pressure coupling schemes (4). The equations of motion were integrated using the velocity Verlet integrator, with time step of 1.0 fs (5). All classical MD

simulations were performed with the NAMD package (6). For the V27A, V27R and S31N mutants, the Q2 state (with 2 biprotonated His37 residues) was studied. The mutated structures were equilibrated for at least 20ns starting from the equilibrated wild type structure of D2 (originating from the 3LBW crystal structure).

The two crystal structures 3C9J and 3LBW were tested to be sure that the simulations captured the correct structural ensemble for each His37 charge state, and to minimize the bias due to starting conformation. The 3C9J and 3LBW structures were crystallized at different pH values (5.3 and 6.5, respectively), suggesting that they might represent different protonation states of the His37 tetrad. The structures differ significantly in the C terminal region where 3C9J is more open due to a larger helix tilt angle. It was found that the Q2 state converged to a structure very similar to 3BLW, regardless of the starting structure and in agreement with other work suggesting that 3BLW is in the Q2 state (7). Hence, the equilibrated Q2 structure from the 3LBW simulations was used as our initial configuration for the Q2 MS-EVB and QM/MM simulations. In fact, both the Q1 and Q2 starting from 3LBW adopt a similar backbone conformation with the crystal structure. This is in agreement with the simulation result of Acharya et al. [ref. (12) of the main text], which shows that starting from 3LBW the RMSD values for both Q1 and Q2 are less than unity. However, in the Q1 equilibration we observed that on average there is one fewer water molecule between His37 and Trp41 that accepts a hydrogen bond from the His37 tetrad, providing evidence of Q1 being more closed than Q2 in terms of the barrier for His37 deprotonation.

In contrast, the Q3 and Q4 states resulted in more open and solvated structures, but none were consistent with 3C9J and both states showed more dependence on the starting

structure. The Q3 structure equilibrated from 3C9J remained more splayed than that from 3BLW. Since it was also more splayed than other available PDB structures, we chose the 3LBW equilibrated structure for our Q3 MS-EVB and QM/MM simulations. This choice is supported by recent work (8), which reported that the more splayed 3C9J Q3 structure converged to 3LBW Q3 structure during microsecond MD simulation.

Based on this classical MD equilibration results, we emphasize that the Q1 to Q2 transition still results in a relatively closed state channel, whereas the Q2 to Q3 transition results in an opened channel with more hydrated His37, thereby lowering the deprotonation barrier and increasing the proton conductance. This is reason why we started from Q2 and studies the Q2 -> Q3 -> Q2 process (our +2 PMF) for the purpose of investigating the A M2 proton conduction. In other words, it is the large difference between Q2 and Q3 structures, as well as the small difference between Q1 and Q2 structures, that motivated us to select Q2 as the initial structure for investigating the A M2 proton conduction process.

B.2 QM/MM Umbrella Sampling.

The potential of mean force (PMF) for an excess proton passing the His37-Trp41 tetrad region was calculated with hybrid quantum mechanical/molecular mechanical (QM/MM) MD simulations using the CP2K package (9). Umbrella sampling was performed with the collective variable (CV) defined as the difference between the z coordinate of the center of mass of C_α atoms of Gly34 and that of the excess proton center of excess charge (CEC). The excess proton CEC coordinate was defined as (10):

$$\vec{\xi} = \sum_{i=1}^{N_H} \vec{r}^{H_i} - \sum_{j=1}^{N_X} w^{X_j} \vec{r}^{X_j} - \sum_{i=1}^{N_H} \sum_{j=1}^{N_X} f_{sw}\left(d_{X_j H_i}\right)\left(\vec{r}^{H_i} - \vec{r}^{X_j}\right) + \vec{\xi}_{\text{correct}}, \quad \text{[S1]}$$

where the "X" denotes heavy atoms in the QM region (specifically the water oxygen atoms and the eight histidine nitrogen atoms) and "H" denotes the hydrogens bound to those heavy atoms. The weighting factors w^{X_j}'s of the heavy atoms reflect the hydrogen coordination number of a molecule in the unprotonated state. Therefore, for water oxygen atoms they were set to 2, and for His37 nitrogen atoms they were set to 0.75 in +2 PMF (6 hydrogens shared by 8 nitrogens in the unprotonated state) and 0.875 in +3 PMF (7 hydrogens shared by 8 nitrogens). The term $d_{X_j H_i}$ is the distance between heavy atom X_j and atom H_i, and $f_{sw}\left(d_{X_j H_i}\right) = 1/(1 + \exp\left[(d_{X_j H_i} - r_{sw})/d_{sw}\right])$ is the switching function describing the coordination number of H_i to X_j, with the parameters chosen as $d_{sw} = 0.04$ Å, $d_{sw} = 0.125$ Å (11).

To correct for the contribution to the excess proton CEC position due to the presence of multiple protons around multiple protonatable sites in the His37 tetrad, a correction term $\vec{\xi}_{\text{correct}}$ was introduced in a manner similar to Eq. 8 of ref (10):

$$\vec{\xi}_{\text{correct}} = \frac{1}{8} \sum_{i=1}^{8} \sum_{j=1}^{8} m_i (\vec{r}^{X_j} - \vec{r}^{X_i}), \quad \text{[S2]}$$

where X denotes the 8 nitrogen atoms of the His37 tetrad, and

$$m_i = \sum_{H_j \in \{H\}} f_{sw}(d_{X_i,H_j})^{16} \Big/ \sum_{H_j \in \{H\}} f_{sw}(d_{X_i,H_j})^{15} \quad \text{[S3]}$$

m_i gradually switches from 1 to 0 as the nitrogen atom i is deprotonated. The above excess proton CEC definition (Eq. S1) reflects the delocalized nature of the excess proton and has been shown to adequately locate the position of the excess proton charge defect in previous QM/MM simulations of biological PT channels (10, 11).

The delocalized nature of the CEC has special advantages for this system. First, it allows the excess proton to be transported across a continuous water wire (if there is one) via the Grotthuss mechanism, and at the same time it also allows the proton to protonate the His37 tetrad. Therefore, the CV does not introduce an *a priori* bias towards either the "shutter" mechanism or "shuttle" mechanism. Second, the CV definition does not specify the formation or breakage of a particular N-H bond when the His37 tetrad is protonated and deprotonated. Thus, the CV definition allows us to study such a complicated multi-proton and multi-protonatable site problem with minimum preliminary assumptions.

The last snapshot of D2 and D3 after classical equilibration was used as the initial structure for the QM/MM simulations for the +2 and +3 PMFs, respectively. A water molecule ~5 Å above the His37 residue was replaced by a hydronium. For the +2 PMF, different windows had different QM regions to save computational cost. For all windows in +2 PMF, the QM region included His37 side chains. In the -4 Å < z < 2 Å windows, the water molecules within 9 Å above the His37 and those between His37 and Trp41 were included as QM atoms. In the 2 Å < z < 7 Å windows, the water molecules within 6 Å above the His37 and those between Trp41 and His37 were included as QM atoms. In the 7 Å < z < 12 Å windows, the water molecules within 3 Å above the His37, those between Trp41 and His37, and one solvation shell below Trp41 were included as QM atoms. The Trp41 side chains were also included as QM atoms to take into account the effect of His37-Trp41 cation–π interaction on the deprotonation barrier for the relatively closed D2 backbone structure (12). For the +3 PMF, the His37 side chains, and all water molecules within 0 Å < z < 12 Å were included in QM region for all windows. In all

simulations, the QM box size was chosen to be 6-8 Å larger than the actual size of the QM atoms in each dimension. A quadratic confining wall potential was used to restrain the QM atoms within the QM box, with the wall skin thickness being 1 Å × 1 Å × 2 Å. The QM box size was chosen such that the quadratic wall potential did not affect the QM atoms for the vast majority of the simulation. The QM region was treated at DFT level using the BLYP functional (13, 14) with empirical dispersion corrections (15), under the Gaussian plane wave (GPW) scheme (16). The Goedecker-Teter-Hutter (GTH) pseudopotentials (17) were used and the Kohn-Sham orbitals were expanded in the Gaussian TZV2P basis set. The electron density was expanded by auxiliary plane wave basis set up to 360 Ry. The Gaussian Expansion of the Electrostatic Potential (GEEP) scheme was used to treat the QM/MM electrostatic coupling with periodic boundary conditions (PBCs) (18, 19), and the spurious QM/QM periodic image interactions were decoupled as described in ref (20). The motion of nuclei was integrated using a time step of 0.5 fs, and the wavefunction was optimized to the Born-Oppenheimer surface by an orbital transformation method (21) with a convergence criterion of 10^{-6}. The temperature was controlled at 308K by a Nosé-Hoover thermostat with a relaxation time constant of 0.1 ps. The initial configurations of all windows were prepared from the previous equilibrated window with smaller z value. With a ~0.25 Å window spacing and force constants of 40 kcal/mol/Å^2), 64 and 36 windows were simulated for +2 and +3 PMFs, respectively. All windows had an equilibration time of ~5 ps and production sampling of 15~20 ps. The excess proton CEC position was collected every step (0.5 fs).

B.3 MS-EVB Umbrella Sampling.

The PMF for an excess proton transported in the region outside the His37-Trp41 tetrad region was calculated using the multi-state empirical valence bond (MS-EVB) method (22) with the RAPTOR software (23) interfaced with the LAMMPS MD package (http://lammps.sandia.gov) (24). Similar to QM/MM MD simulations, umbrella sampling was employed and the same CV was used, except that in the MS-EVB simulation the excess proton CEC is given by (25):

$$\vec{r}_{\mathrm{CEC}} = \sum_{i}^{N} c_i^2 \vec{r}_{\mathrm{coc}}^{\,i}, \qquad [S4]$$

where $\vec{r}_{\mathrm{coc}}^{\,i}$ is the position vector of the center-of-charge of the hydronium cation of the i^{th} EVB state. It was verified that the MS-EVB and QM/MM CEC definitions were nearly identical in the overlapping regions (i.e., when the CEC is more than 3 solvation shells away from the His37-Trp41 tetrad).

For the +2 PMF of the wild type channel, both 2nd (25) and 3rd (22) generations of MS-EVB (MS-EVB2 and MS-EVB3) were compared. All other PMFs were performed with MS-EVB2. As shown in **Figure B.1**, the MS-EVB2 and MS-EVB3 methods lead to nearly the same PMF. The cutoffs for LJ and real space electrostatic interactions were 12 Å, employing switching function starting at 10 Å for the LJ interactions. Long-range electrostatics were treated by Particle-Particle Particle-Mesh (PPPM) method (26) with an accuracy threshold of 10^{-4}. The integration time step was 1 fs. The temperature was controlled at 308 K by Nosé-Hoover thermostat with a relaxation time constant of 0.1 ps.

The initial configuration for each umbrella window was prepared by replacing a water molecule with a hydronium close to the center of the window from the equilibrated wild type or mutant structures. With a ~0.5 Å window spacing and force constants of 10

kcal/mol/Å^2, 50 windows were simulated for each PMF. In the channel mouth regions, where the pore radius profiles change rapidly, additional flat bottom cylindrical restraints on the x-y plane were applied to allow the hydrated proton to explore a similar area as that within the water pore. The replica-exchange umbrella sampling (REUS) method (27) was employed to facilitate convergence, with an exchange attempt frequency of 1 ps^{-1}. All windows were equilibrated for 500 ps and sampled with REUS for 1.5~2 ns. The excess proton CEC position was collected every 10 steps (10 fs).

B.4 Multiscale bridging of MS-EVB and QM/MM PMFs.

The final PMFs for PT across the entire channel were generated by combining the MS-EVB and QM/MM zone PMFs in a multiscale fashion. (Here, the phrase "multiscale" means a connection of the QM influenced scale of the QM/MM to the more empirical MS-EVB defined scale of the nuclear motion. This terminology is consistent, for example, with the use of the phrase multiscale in the 2013 Nobel Prize in Chemistry citation). The bridging of these PMFs has similar characteristics to the bridging of QM and MM zones in a QM/MM simulation in that a certain "transition" or "buffer" region is required. In the present work, this zone is defined by a boundary between statistical regions of the QM/MM and MS-EVB PMFs. The latter PMF cannot be taken to be too close to zone where the proton binding and hopping in the His tetrad occurs, because the present MS-EVB model has not been developed to describe that process. Likewise, the QM/MM zone cannot be taken to be too far outward from the His37 zone because the QM/MM simulation is unable to adequately sample the proton water-mediated hopping and transport through the wider water filled regions of the channel, as well as the

additional protein motions in those regions. The MS-EVB simulations, which are several orders of magnitude more computationally efficient than QM/MM, are required for this task. The connecting points between the two PMF regions are therefore chosen based on these considerations and also so that the slopes of the QM/MM and MS-EVB PMFs match as closely as possible. The statistics are accumulated in those regions (within 1.5 Å of the connection points) to achieve better statistical certainty at the connection points, and the combined PMFs are then bridged via interpolated using an interpolation order of 2. We note that all combined PMFs for the mutants have the same connecting points between the QM/MM and MS-EVB PMFs as in WT system.

While the above multiscale bridging scheme is approximate, as is the nature of most such multiscale bridging (e.g., between QM and MM in QM/MM), we emphasize that the main conclusions of this work are not sensitive to the position of the boundaries between the QM/MM and MS-EVB regions. The following points are important to note in this regard: (1) The qualitative description of the energetics for an excess proton to translocate across the M2 channel is not affected by this multiscale bridging procedure. (2) The PMF increase from V27 to H37 is larger in +2 PMF than in +3 PMF. This occurs in a region far away from the left connecting point and is entirely captured by the MS-EVB simulations alone, independent of the position of the QM/MM and MS-EVB boundaries. Therefore, the explanation for the pH-dependent activation mechanism is not changed by the exact position of the connecting points. (3) Because the same boundary regions are chosen for the mutants as in WT, and because these mutations are away from the boundary regions, the relative conductance trends are not influenced by the boundary region choices. (4) For quantitative results such as absolute conductance, the errors

caused by a small change in the boundary region will not be larger than the other error sources described in the "**Limitations in Conductance Calculations"** section below.

B.5 Calculation of Proton Conductance

The maximum proton conductance g_{max} can be estimated from Poisson-Nernst-Planck (PNP) electrodiffusion theory (28-30):

$$g_{max} = \frac{e^2}{k_B T L^2} \langle D(z)^{-1} e^{+F(z)/k_B T} \rangle^{-1} \langle e^{-F(z)/k_B T} \rangle^{-1} \quad [S5]$$

where *e* is the elementary charge, $k_B$is the Boltzmann's constant, *T* is the temperature, *L* is the length of AM2/TM (L = 35 Å), $D(z)$ and $F(z)$ are the diffusion coefficient and the PMF as a function of the CV, respectively. The brackets denote spatial averaging over the length of the channel in the direction of the channel axis (z axis). The position-dependent diffusion coefficient for the excess proton was calculated according to the Woolf-Roux equation (31):

$$D(z_i) = \lim_{s \to 0} \frac{-\hat{C}(s; z_i) \langle \delta z^2 \rangle_{(i)} \langle \dot{z}^2 \rangle_{(i)}}{\hat{C}(s; z_i) \left[s \langle \delta z^2 \rangle_{(i)} + \langle \dot{z}^2 \rangle_{(i)} / s \right] - \langle \delta z^2 \rangle_{(i)} \langle \dot{z}^2 \rangle_{(i)}} \quad [S6]$$

where $C(t; z_i) = \langle \dot{z}(t) \dot{z}(0) \rangle_i$ is the velocity autocorrelation function for the CV calculated based on the data from window i, z_i is the reference point for the harmonic restraint potential in window *i*, and $\hat{C}(s; z_i) = \int_0^\infty e^{-st} C(t; z_i) dt$ is the Laplace transform of this function. To estimate the value of the limit as $s \rightarrow 0$, we linearly extrapolate from the range $5 \leq s \leq 15$.

B.6 Limitations in Conductance Calculations.

Although the quantitative agreement between experimental and calculated conductances in this paper is encouraging, it is important to note two possible sources of error in these calculations. The first source is the underlying uncertainty of the PMF inside the QM/MM region due to limited statistical sampling. It is possible that once the 3rd His37 is protonated, the channel will become more opened and hydrated. However, currently the QM/MM simulation cannot capture these changes in the limited amount of sampling time. As compensation for this, we used QM/MM simulations to calculate the deprotonation barrier starting from the classically equilibrated structure of D3, which is more opened and hydrated near C-terminal region (see **Figure B.2**). The result indicates only a small (~1 kcal/mol) decrease of deprotonation barrier when the channel is more opened and hydrated near C-terminal region, in agreement with the hypothesis of Hu et. al (32). In addition, the slow motions of Trp41 indole ring may not be fully sampled, and this would also introduce possible error for the deprotonation barrier (33). A second source of error could be the application of PNP theory for the conductance calculation because that theory has inherent approximations (34-36). Furthermore, in this study in order to obtain converged diffusion constants for the His37-Trp41 region the computational more efficient MS-EVB simulation was employed, but the current model does not allow His37 residues to change protonation states. However, the uncertainty in diffusion constants is expected to have a relatively minor influence on the final conductance value from PNP theory compared to the uncertainty in the PMF.

B.7 Supplementary Figures

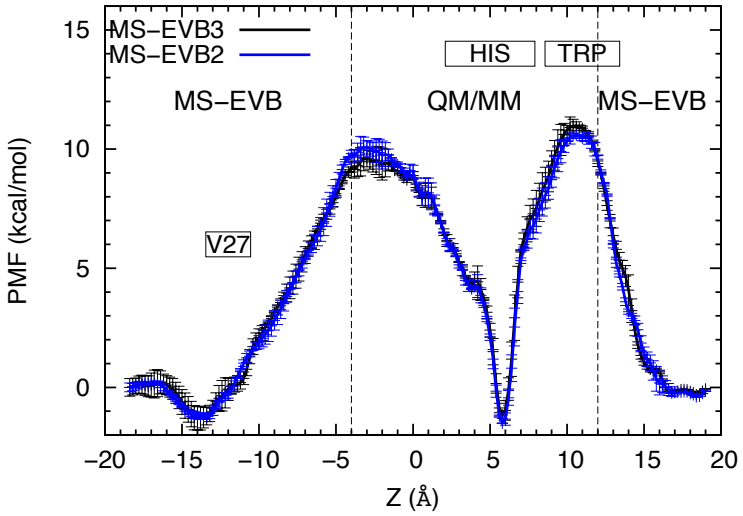


Figure B.1 Proton transport +2 free energy profile (PMF). The blue and black lines are the MS-EVB2 and MS-EVB3 PMF, respectively. The origin is the center of mass of 4 Cα of Gly34. The positions of Val27, His37 and Trp41 are labeled by the text boxes. The regions sampled by MS-EVB and QM/MM are labeled and separated by black dashed vertical lines.

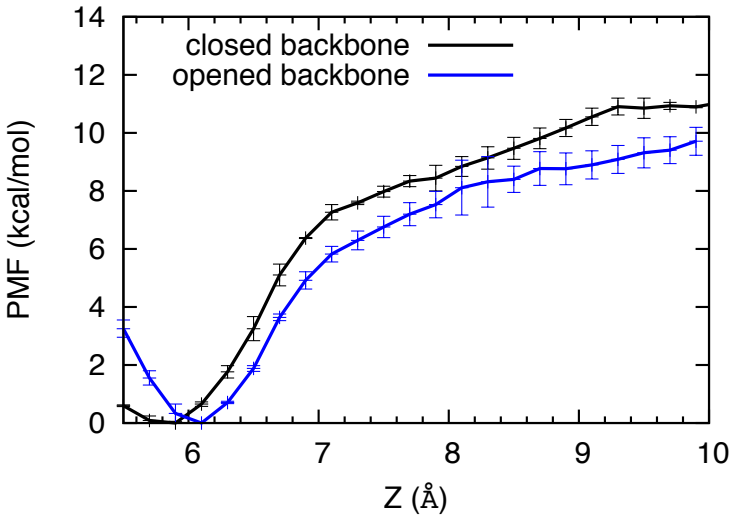


Figure B.2 Comparison for the deprotonation barrier of a Q3 state His37 tetrad under different backbone conformations.The black line is the QM/MM PMF starting from the classically equilibrated D2 backbone structure, where the backbone is relatively closed and the His37 residues are less hydrated. The blue line is the QM/MM PMF starting from the classically equilibrated D3 backbone structure, where the backbone is more opened and the His37 residues are more hydrated.

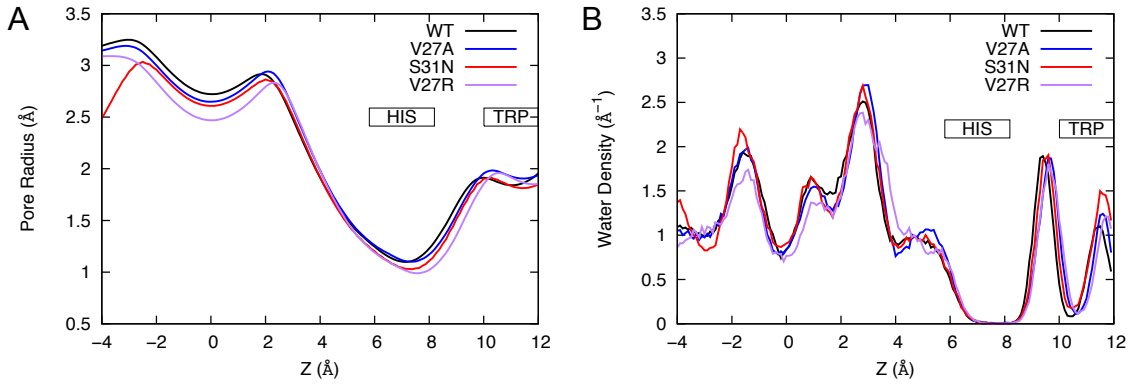


Figure B.3 (A) Pore radius profile and (B) water density for WT AM2/TM and mutants.The WT, V27A, S31N and V27R results are shown in black, blue, red and purple, respectively. The pore radius and water density profile in the His37-Trp41 tetrad for the mutants are quite similar to those of WT, justifying our use of WT QM/MM PMF for mutant PMF calculation.

B.8 Supplementary References

1. Jorgensen WL & Jenson C (1998) Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: seeking temperatures of maximum density. *J. Comput. Chem.* 19(10):1179-1186.

2. MacKerell AD, *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102(18):3586-3616.

3. Tom D, Darrin Y, & Lee P (1993) Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98(12):10089-10092.

4. Feller SE, Zhang Y, Pastor RW, & Brooks BR (1995) *Constant Pressure Molecular Dynamics Simulation : The Langevin Piston Method* (American Institute of Physics, Melville, NY, ETATS-UNIS) p 9.

5. Verlet L (1967) Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* 159(1):98.

6. Phillips JC*, et al.* (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26(16):1781-1802.

7. Acharya R*, et al.* (2010) Structure and mechanism of proton transport through the transmembrane tetrameric M2 protein bundle of the influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* 107(34):15075-15080.

8. Wei C & Pohorille A (2013) Activation and Proton Transport Mechanism in Influenza A M2 Channel. *Biophys. J.* 105(9):2036-2045.

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

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

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

12. Bankura A, Klein ML, & Carnevale V (2013) Proton affinity of the histidine-tryptophan cluster motif from the influenza A virus from ab initio molecular dynamics. *Chem. Phys.* 422:156-164.

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

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

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

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

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

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

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

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

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

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

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

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

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

26. Hockney RW & Eastwood JWja (1981) *Computer simulation using particles* (McGraw-Hill International Book Co., New York) p 540.

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

28. Levitt DG (1986) Interpretation of Biological Ion Channel Flux Data--Reaction-Rate versus Continuum Theory. *Annu. Rev. Biophys. Biophys. Chem.* 15(1):29-57.

29. Roux B & Karplus M (1991) Ion-Transport in a Gramicidin-Like Channel - Dynamics and Mobility. *J. Phys. Chem.* 95(12):4856-4868.

30. Levitt DG (1991) General continuum theory for multiion channel. I. Theory. *Biophys. J.* 59(2):271-277.

31. Woolf TB & Roux B (1994) Conformational Flexibility of O-Phosphorylcholine and O-Phosphorylethanolamine - a Molecular-Dynamics Study of Solvation Effects. *J. Am. Chem. Soc.* 116(13):5916-5926.

32. Hu FH, Luo WB, & Hong M (2010) Mechanisms of Proton Conduction and Gating in Influenza M2 Proton Channels from Solid-State NMR. *Science* 330(6003):505-508.

33. Williams JK, Zhang Y, Schmidt-Rohr K, & Hong M (2013) pH-Dependent Conformation, Dynamics, and Aromatic Interaction of the Gating Tryptophan Residue of the Influenza M2 Proton Channel from Solid-State NMR. *Biophys. J.* 104(8):1698-1708.

34. Furini S, Zerbetto F, & Cavalcanti S (2006) Application of the Poisson-Nernst-Planck theory with space-dependent diffusion coefficients to KcsA. *Biophys. J.* 91(9):3162-3169.

35. Noskov SY, Im W, & Roux B (2004) Ion permeation through the alpha-hemolysin channel: Theoretical studies based on Brownian dynamics and Poisson-Nernst-Plank electrodiffusion theory. *Biophys. J.* 87(4):2299-2309.

36. Kurnikova MG, Coalson RD, Graf P, & Nitzan A (1999) A lattice relaxation algorithm for three-dimensional Poisson-Nernst-Planck theory with application to ion transport through the gramicidin A channel. *Biophys. J.* 76(2):642-656.

Appendix C
Supporting Information for Multiscale simulations reveal key features of the proton pumping mechanism in cytochrome c oxidase

C.1 System Setup and Structural Equilibration.

In the current work, three intermediate redox states during the A→P_R→F transition were simulated: P_M', P_R, F (**Table C.1** and main text). The charges for the heme and copper groups in different redox states were obtained from Johansson et al F(1). Depending on the PT process under study, different protonation states for the key protonatable groups were assigned in a classical MD simulation for structural equilibration. For the study of PT from E286 to the PLS and BNC in the P_M' and P_R states (**Table C.1**), the initial system has E286 protonated while D132, PRDa_3 and PRAa_3 are deprotonated, and Cu_B has a hydroxide ligand. Note that D132 was deprotonated in these simulations because it has a pK_a value close to that of aspartate in bulk. To investigate the two possible conditions under which the PT to BNC occurs, i.e., with and without the PLS loaded (protonated), two setups were prepared: one with a hydronium cation shared between PRDa_3 and PRAa_3 (i.e., the PLS protonated) and one without (i.e., the PLS deprotonated). For the study of PT in the D-channel in the P_R and F states (**Table C.1**), the initial system has E286 deprotonated, D132 protonated, and a hydronium cation shared between the deprotonated PRDa_3 and PRAa_3. Thus, the setup correctly simulates the reprotonation of E286 after PT from E286 to PLS, which is shown to be kinetically favored over PT to the BNC (see main text).

The full structure of CcO from *Rhodobactor sphaeroides* was used for the simulations (PDB code 1M56 (2)). The protein was embedded in a dimyristoylphosphatidylcholine (DMPC) lipid bilayer composed of 310 lipid molecules

and solvated by TIP3P water molecules on each side of the membrane (3). Periodic boundary conditions were applied in all three dimensions. The full system consisted of ~160,000 atoms. The CHARMM22 (4) and CHARMM36 (5) force fields were used for the protein and lipids, respectively. The electrostatic interactions were treated with particle mesh Ewald, (6) using a spherical cutoff of 12.0 Å and an accuracy threshold of 10^{-6}. The Lennard-Jones (LJ) interaction cutoff was 12.0 Å using a switching function starting at 10 Å. The system was equilibrated in the constant number, pressure, and temperature (NPT) ensemble at 310 K and 1 atm using the Langevin temperature coupling and Langevin piston pressure coupling schemes (7). During the first stage of the equilibration, harmonic restraints with a force constant of 20 kcal/mol/Å^2 were imposed on the protein backbone and side chain atoms and was gradually reduced to zero over 10 ns of simulation. Then the system was equilibrated without harmonic restraints on the protein for 50 ns. Harmonic restraints with a force constant of 100 kcal/mol/Å^2 were imposed on the heavy atoms of heme *a*, heme a_3, Cu_A, Cu_B and their ligands throughout the simulations. The equations of motion were integrated using the velocity Verlet integrator with a time step of 1.0 fs (8). After equilibration, the simulation box size was 118 Å ×106 Å ×123 Å and was then fixed in the subsequent simulations in constant number, volume and temperature ensemble (NVT) at 310 K. The structural equilibration was carried out with the NAMD package. (9)

C.2 Multiscale reactive molecular dynamics (MS-RMD).

The MS-RMD (and MS-EVB) method (10) describes electronic delocalization of the excess proton as a linear combination of diabatic states with different chemical bonding

topologies. The Hamiltonian is defined as:

$$\mathbf{H} = \sum_{ij} |i(\mathbf{r})\rangle h_{ij}(\mathbf{r}) \langle j(\mathbf{r})| \quad [S1]$$

where $\mathbf{r}$ are the nuclear coordinates, h_{ii} is the potential energy for state i described by molecular mechanics force field, and h_{ij} is the coupling between states $|i\rangle$ and $|j\rangle$. The diagonalization of the Hamiltonian matrix gives the energy and eigenvector of the ground state:

$$\mathbf{Hc} = \mathrm{E}_0\mathbf{c} \quad [S2]$$

The forces are evaluated by the Hellmann-Feynman theorem in every MD simulation step and are used to propagate the trajectory. The excess proton charge defect delocalization, Grotthuss shuttling, and polarization are explicitly treated in this method, and the resulting MD trajectory is continuous and deterministic to within numerical error.

The diagonal elements h_{ii} of the MS-RMD Hamiltonian represents the potential energy for state i. The h_{ii} corresponding to the state of protonated amino acid (HA) is described as:

$$h_{ii}^{HA} = V_{surrounding}^{intra} + V_{HA}^{intra} + V_{surrounding,HA}^{inter} + V_{ii} \quad [S3]$$

where the first three terms are the inter- and the intra-molecular potentials of protonated amino acid and all other surrounding molecules, such as waters, other protein residues, lipids, and ions in the system. They are computed with CHARMM22 force field,(4) with the exception of the O-H bond in the carboxyl (-COOH) group of the amino acid. To

properly represent the O-H bond breaking during deprotonation, the harmonic bond stretch potential is replaced with a standard Morse potential, $U^{Morse}(r)$:

$$U^{Morse}(r) = D\left[1 - \exp\left(-\alpha(r - r_0)\right)\right]^2 \quad \text{[S4]}$$

where r is the O-H bond length, and D, α, and r_0 are parameters from our previous work.(11) The V_{ii} is added to the state with protonated amino acid as a constant energy shift to compensate for the difference in the energy origin between states with protonated and deprotonated amino acid.

In order to correct the over-attraction between opposite charges on a hydronium and deprotonated amino acid at a short distance described by the classical force field, two repulsive terms, $V_{OO_k}^{rep}$ and $V_{H_jO_k}^{rep}$, are introduced in h_{ii} corresponding to the state with deprotonated amino acid:

$$V_{OO_\epsilon}^{rep}\left(R_{OO_\epsilon}, R_{H_jO_\epsilon}\right) = B\,\exp\left[-b\left(R_{OO_\epsilon} - d_{OO}^0\right)\right] \cdot \sum_j^3 \exp\left[-b'\left(R_{H_jO_\epsilon}\right)^2\right] \quad \text{[S5]}$$

$$V_{H_jO_\epsilon}^{rep}\left(R_{HO_\epsilon}\right) = C\,\exp\left[-c\left(R_{H_jO_\epsilon} - d_{OH}^0\right)\right] \quad \text{[S6]}$$

where R_{OO_ϵ} is the distance between the hydronium oxygen, O, and the carboxyl oxygen of amino acid, O_ϵ (OE1 and OE2 in the PDB), and $R_{H_jO_\epsilon}$ is the distance between each of three hydronium hydrogen atoms, H_j, and the carboxyl oxygen of the amino acid. The functional forms for the repulsive terms are the same as those used in the MS-EVB3.1 model. (11) B, b, b', C, and c are fitted parameters, and d_{OO}^0 and d_{OH}^0 are fixed to the same value used in MS-EVB3.1, which are 2.4 and 1.0 Å, respectively.

The off-diagonal coupling element h_{ij} between the amino acid and water is given by:

$$h_{ij}^{HA} = c_1 \exp[-c_2(r_{OH} - c_3)^2] \quad \text{[S7]}$$

where r_{OH} is the distance from the donor oxygen of the carboxyl group of the amino acid to the hydrogen atom transferred between the two states. c_1, c_2, and c_3 are fitted parameters. h_{ij} for the coupling between hydronium and water is the same as the one used in the MS-EVB3.1 model.

C.3 Finding the Proton Transport Pathways.

MS-RMD simulations using metadynamics (MTD) (12) were performed to identify the PT pathways in both the D-channel and HC. MTD allows the system to repeatedly attempt barrier crossings so that in regions with high barriers and slow dynamics (such as the asparagine gate region) the system can cross repeatedly until the lowest free energy pathway is found. The MS-RMD model (10) has been demonstrated to be capable of accurately and efficiently simulating the structure and dynamics of hydrated excess protons transporting through both water molecules and protonatable groups inside proteins (11, 13, 14). The combination of the MS-RMD model with metadynamics thus can provide accurate curvilinear pathways for the PT events inside a protein (15). The aspartic acid model in bulk water (11) was taken as a trial model for D132, E286, the propionate groups of heme a_3, and the hydroxide bound to the BNC. The resulting PT pathways were not found to be sensitive to the parameters in the trial model compared to the final MS-RMD models of the protonatable sites. To define the CVs used in the MTD,

first the starting and ending point of a pathway was determined based on the average position of the protonatable groups during the classical MD equilibration. In particular, for PT from E286 to the PLS, they were E286 and PRAa_3 protonatable groups, respectively; for PT from E286 to the BNC, they were E286 and hydroxide bound to Cu_B protonatable groups, respectively; for the D-channel, they were the D132 and E286 protonaable groups, respectively. The CV was then defined to be the projection of the vector from the starting point to the proton center of excess charge (CEC) position (see Eq. [S8] [S9] for definition) on the vector from the starting point to the end point. Gaussian hills with width of 0.3 Å and height of 0.3 kcal/mol/Å were deposited on the free energy surface at a rate of 1/500 ps. The MTD simulations were run for 1-2 ns until the both forward and backward barrier crossings were observed. The coordinates of CEC during MTD simulation were collected and sorted into bins based on the CV. In each bin the CEC coordinates were averaged and the line connecting the average coordinates in each bin was smoothed into a curved PT pathway.

C.4 Developing MS-RMD Models for Protonatable Groups.

The FitRMD method developed by Nelson et al (11) and extended by Lee et al (16) was used to parameterize the MS-RMD models from QM/MM data (see below) for protonatable sites in CcO, including the D132, E286, PRDa_3, PRAa_3 and hydroxide bound to BNC. Umbrella sampling simulations were first carried out along the PT pathway identified by the MTD simulation, and approximately 100 configurations were selected from the trajectories that were near the transition state of proton dissociation for each protonatable site. QM/MM calculations were then performed for each configuration

using B3LYP level density functional theory. The MM models were the CHARMM22 and CHARMM36 force fields for the protein and lipids, respectively. The QM region included the side chain of each protonatable amino acid, the excess proton and water molecules within 3 solvation shells of the carboxyl group. In all calculations, the QM box size was chosen to be 6-8 Å larger than the actual size of the QM atoms in each dimension. The Gaussian Expansion of the Electrostatic Potential (GEEP) scheme was used to treat the QM/MM electrostatic coupling with periodic boundary conditions (PBCs) (17, 18), and the spurious QM/QM periodic image interactions were decoupled as described in ref (19). The $C_\alpha - C_\beta$ chemical bond that cross the QM/MM boundary was capped with hydrogen atoms, the forces on which were calculated following the IMOMM scheme (20) with a scaling factor of 1.50. The forces generated by the QM/MM calculations were used to parameterize the MS-RMD parameters for the protonatable sites using an iterative variational methodology. (11) The QM/MM calculation was performed with the CP2K *ab initio* MD package (21), and FitRMD was carried out with in-house software. Due to the significant changes of the electrostatic environment upon ET from heme *a* to BNC, different MS-RMD models for PRDa_3, PRAa_3 and BNC bound hydroxide were parameterized for the P_M' and P_R states. The MS-RMD model for D132, which is relatively far away from the redox centers, was parameterized based on the configurations sampled in the P_R state. The E286 side chain exhibits three different conformations during the pumping cycle, as observed along the PT pathways (see above). For PT through the D-channel, E286 points downward toward D132 for proton uptake. For PT from E286 to the PLS, E286 points upward and moves closer to PRDa_3. For PT from E286 to the BNC, E286 points upward but rotates horizontally with respect to the

membrane and moves closer to the BNC. These three conformations have significantly different surrounding electrostatic and solvation environments. The MS-RMD model for E286 was indeed found to be sensitive to the conformation. Therefore, independent E286 models were parameterized for the three different conformations. However, the E286 model was found to be insensitive to redox changes in the heme-copper groups, and thus for the same E286 conformation, the same model was used for different redox states.

For the MS-RMD state that has water bound to the BNC, the partial charges for the Cu_B complex were parameterized using the same method as described in ref (1). The partial charges for the Cu_B complex with water ligand are listed in **Table C.4** and **Table C.5**. The bond and angle parameters for the Cu_B bound water were parameterized to the potential energy surface calculated for the Cu_B complex in gas phase from B3LYP. They are listed in **Table C.6**. The parameterization for the Cu_B complex was carried out with the NWChem package (22). The MS-RMD model parameters are listed in **Table C.2** and **Table C.3**.

C.5 MS-RMD Free Energy Profiles in the HC and D-Channel.

The MS-RMD umbrella sampling simulations in the HC were carried out by restraining the excess proton CEC position (see Eq. [S8] [S9] for definition) along the PT pathway defined from the MTD procedure defined earlier and the water density in a predefined box that encompasses the HC (see ref (23) for definition). It is important to note that the box used in these calculations was large enough to overlap with the solvation regions peripheral to the HC. This is essential to ensure proper solvation in the boundary regions and to converge sampling. We note that 6 waters in our box is roughly equivalent to 4

waters in the HC defined in previous studies. The force constant for the harmonic umbrella sampling restraint potential was 10 kcal/mol/Å^2 on the proton migration CV and 20 kcal/mol on the water density. The window spacing was ~ 0.5 Å for the CEC and ~ 0.5 for the water density. For each umbrella window the MS-RMD simulation length was ~500 ps. The 2D PMF was constructed by the weighted histogram analysis method (WHAM). (24-26) The minimum free energy path was calculated using the string method (27) at 0 K on the 2D PMF, and the PT rate constant was estimated based on the 1D-PMF along the minimum free energy path with transition state theory (see Eq [S10] [S11]). The MS-RMD umbrella sampling simulation in the D-channel was carried out restraining both the excess proton CEC position along the PT pathway and the water density in a predefined box that encompasses the asparagine gate, with the same force constants for the umbrella potentials as for PT in HC. The window spacing was ~ 0.5 Å for the CEC and ~ 0.2 for the water density. For each umbrella window the MS-RMD simulation length was ~500 ps. The 2D PMF was then constructed by WHAM. The minimum free energy path and the PT rates constants were calculated in the same way as for the PT in HC. The MS-RMD simulations were performed with the in-house RAPTOR (28) software package.

C.6 Definition of the PMF Collective Variables and Rate Calculation.

To track and bias the charge defect associated with the excess proton in the MS-RMD umbrella sampling simulations, we use the center of excess charge (CEC), defined as (13):

$$\mathbf{r}_{\mathrm{CEC}} = \sum_{i}^{N_{\mathrm{EVB}}} c_i^2(\mathbf{r})\mathbf{r}_i^{\mathrm{COC}} \qquad [S8]$$

where the N_{EVB} is the total number of EVB states, $c_i^2(\mathbf{r})$ is the population of EVB state i contributing to the MS-RMD ground state, and $\mathbf{r}_i^{\mathrm{COC}}$ is the position of the center of charge, defined as

$$\mathbf{r}_i^{\mathrm{COC}} = \frac{\sum_k^{\{i\}} |q_k| \, \mathbf{r}_k}{\sum_k^{\{i\}} |q_k|} \qquad [S9]$$

where the summations are over the atoms with partial charge $\mathbf{q_k}$ of the ith EVB complex. Based on the definition of CEC, the CV that describes the proton migration along the curved PT pathway is defined as follows: For a particular umbrella window, the window center is chosen to be one of the discrete set of points on the curve. The projection of the vector from the window center to the CEC coordinate on the tangent vector at the window center is calculated. Then the projection is added to the distance along the curved pathway from the starting window center to the current window center, which defines the value of the CV. The collective variable for water density in a predefined box is defined in the same way as described in ref (23).

Based on the 1D PMFs, the rate constants for PT were calculated using transition state theory (29):

$$k_{\mathrm{rxn}} = \frac{\omega_0}{2\pi} \exp\left(-\frac{\Delta F^{\ddagger}}{k_{\mathrm{B}} T}\right) \qquad [S10]$$

where $\Delta F^{\ddagger}$ is the free energy barrier height, k_{B} is Boltzmann's constant, T is the simulation temperature. The pre-exponential factor ω_0 is the fundamental frequency of the collective variable (CV) around its minimum of reactant well of the 1D-PMF, which is evaluated as:

$$\omega_0 = \sqrt{\left(\frac{\partial^2 \mathrm{PMF}(r)}{\partial r^2}\right)/m_{eff}} \qquad [S11]$$

where the second derivative of the PMF is evaluated at the minimum of the reactant well, and $m_{eff} = k_B T/\langle v^2 \rangle$ by virtue of equipartition theorem. The $\langle v^2 \rangle$ are evaluated from the CV trajectories of the umbrella windows near the minimum of reactant well of the PMF. The reported rates in **Table 7.1** and **Table 7.2** of the main text are the inverse of the corresponding rate constants calculated from Eq. [S10]. The error in the calculated rates was estimated by computing the variance of the block averages.

C.7 Justification for the choice of collective variables in the 2D PMF calculations.

The presented 2D free energy profiles sample the progress of the excess protonic charge as well as the hydration along the PT pathway. One might ask if other protein motions are coupled to PT in a rate-determining manner. However, one of us (GAV) proved in ref (30) that in the process of proton transfer if one samples the PMF of the proton transfer reaction coordinate then slow coupled modes, such as the solvent activation (reorganization) term, will be included in the final calculated activation free energy (Eqs 19 and 20 in ref (30)), just as in the traditional Marcus/Holstein picture of charge transfer in polar media. In addition, there are no known results on CcO that indicate there exist other very slow protein motions that would limit PT. On the contrary, the motions described have been found to be on a picosecond to nanosecond timescale, such as dissociative motion of the ARG-PRDa$_3$ salt bridge as observed in classical MD simulations of ref (31) and also the "up/down" motion of the E286 as observed in ref

(32). In the current work, we explicitly sampled the Asn139 rotation and found it was very fast relative to the formation of a water wire across the asparagine gate. We also explicitly sampled the Arg-PRDa_3 salt bridge dissociative motion and found it was fast relative to the protonation of PRDa$_3$ and increase in hydration of the hydrophobic cavity. These findings support the why in which the CcO mechanism has been described - as a "solid state electrostatic mechanism", in which protein motions are scarce, very small, and fast, and the direct electrical charge-charge coupling is essential.(33, 34) Finally, it should be noted that the MS-RMD methodology is very computationally efficient so that extensive statistical sampling can be carried out over multiple MD runs, thus limiting the risk of inadequately sampled degrees of freedom.

C.8 Aspartic Acid Deprotonation in Bulk Water Modeled by the SCC-DFTB Method.

Figure C.1 shows the free energy profiles (or potentials of mean force, PMFs) for aspartic acid deprotonation in bulk water calculated by the SCC-DFTB/3OB model (35) and, for comparison, the MS-RMD model developed in ref (11), which was fit to forces from QM/MM calculations with B3LYP level density functional theory (36). The pK$_a$ values are then calculated from the PMFs using the following equation:

$$pK_a = -\log\left[C^0\left(\int_0^{\dagger} 4\pi r^2 \exp[-\beta w(r)]\, dr\right)^{-1}\right] \qquad \text{[S12]}$$

where $w(r)$ is the PMF and β is $(k_B T)^{-1}$, with k_B Boltzmann's constant and T is the simulation temperature. The symbol "†" denotes the reaction coordinate (CV) value at the transition state during the deprotonation and is set as 3.0 Å for both DFTB/3OB and MS-RMD models. The pK$_a$ value is not sensitive to the choice of † in the deprotonated

region of the PMF ($CV > 2.0$ Å). The constant C^0 is the standard state concentration with a value of 1660 $Å^3$/molecule (11).

The aspartic acid pK_a predicted by the DFTB/3OB model and the MS-RMD model are 9.02 and 3.82, respectively. The experimental range of aspartic acid's pK_a value in water is 3.71 to 3.90. (37, 38). The SCC-DFTB method is thus seen to significantly overestimate the pK_a of aspartic acid in bulk water, perhaps related to the method's intrinsic tendency to overestimate the proton affinity of carboxyl groups in amino acids. This behavior likely leads to the overestimation of the proton pumping barrier for singly protonated E286 in C*c*O (39), as discussed in the main text.

C.9 The Biprotonated E286 Mechanism.

In order to test the biprotonated E286 mechanism, we took a similar multiscale approach to that used in ref (40) and calculated the PMF for PT starting from protonated D132 and a singly protonated E286, and ending with deprotonated D132 and a doubly protonated E286 (**Figure C.2**). No proton was loaded in the PLS in this calculation, which is consistent with the proposed mechanism. The QM/MM simulations were performed using BLYP level density functional theory (41, 42) with empirical dispersion corrections (43), and the distributions of collective variable are statistically combined with those from MS-RMD simulations.(40) Based on the calculated PMF, a pumping mechanism involving a positively charged, biprotonated E286 transition state is highly energetically unfavorable by more than 20 kcal/mol. The high energy of a biprotonated E286 can be attributed to three factors. The first is energetic cost of a hydrated excess proton passing through the asparagine gate when E286 is neutral. The second is the cost for a neutral

E286 – which favors the "up" orientation in the P_M' state (as discussed in the main text) and is disconnected from the D-channel waters – to rotate down and reconnect to a retracted water distribution. And the third is that even after the excess proton reaches the first solvation shell of E286, it is still an endergonic process to transfer the proton from hydronium to the neutral E286. Taken together, these factors point to the biprotonated E286 as being an energetically prohibitive intermediate and therefore unlikely to contribute to the pumping mechanism.

Similar to an early computational analysis of PT from D132 to a protonated E286 residue (44), the PMF in **Figure C.2** shows a local minimum around the Ser200/Ser201 region (11 Å $< z <$ 16 Å). This minimum motivated the "proton reservoir" hypothesis originally proposed in ref (45), which suggested that a second excess proton could in principle be trapped in this serine zone when the E286 is singly protonated in the P_M' state. However, as subsequently shown (46), the original minimum was overly exergonic due to the use of a reduced system that included only subunit I and allowed additional water molecules to flow into the D-channel [Fig. 10, ref (46)]. As predicted in that work, and shown herein, simulating the entire system (subunits I-III, membrane and surrounding water and counter ions) and sampling the coupled hydration changes across the asparagine gate was required to get a quantitative and complete PT PMF through the D-channel. The PMF in **Figure C.2** shows that the serine zone has free energy level of ~5 kcal/mol. In physiological pH condition (pH ~ 7), assuming the pKa of D132 is similar to bulk (3.9), the free energy cost for moving a second proton to the serine zone from N-side bulk is ~ 9 kcal/mol, which is thermodynamically unfavorable. Thus, the "proton reservoir" may

transiently occupy an excess proton under certain conditions, but it is not likely to be stably protonated during WT turnover.

C.10 Supplementary Figures

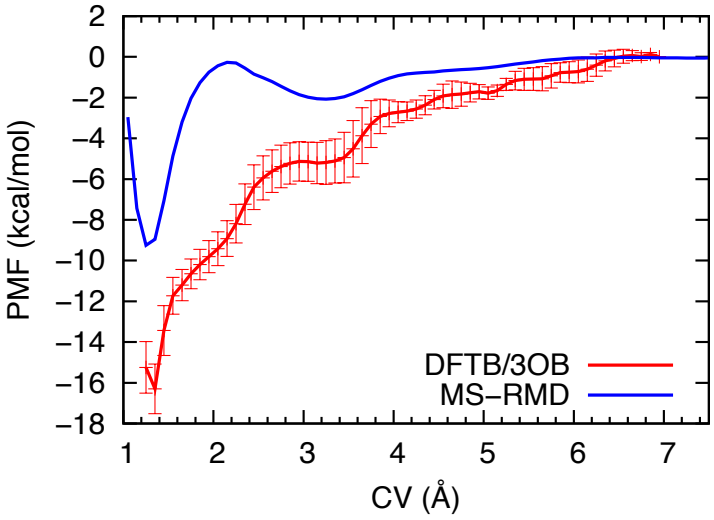


Figure C.1 Free energy profiles (PMFs) for deprotonation of aspartic acid in bulk water. Shown are results from the DFTB/3OB model (red curve) and the MS-RMD model (blue curve) fit to B3LYP QM/MM forces using the FitRMD method (11).

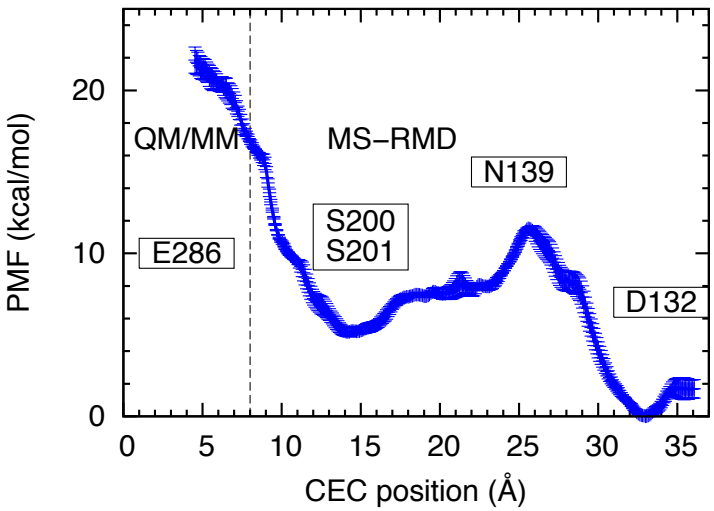


Figure C.2 Proton transfer free energy profile (PMF) in the D-channel for the biprotonated E286 mechanism in the P_M' state. The PT starts with D132 protonated, E286 singly protonated, and PLS deprotonated (CEC ≃ 33 Å). The positions of E286, N139, and D132 are labeled by text boxes. The regions sampled with QM/MM and MS-RMD are labeled and separated by the dashed vertical line.

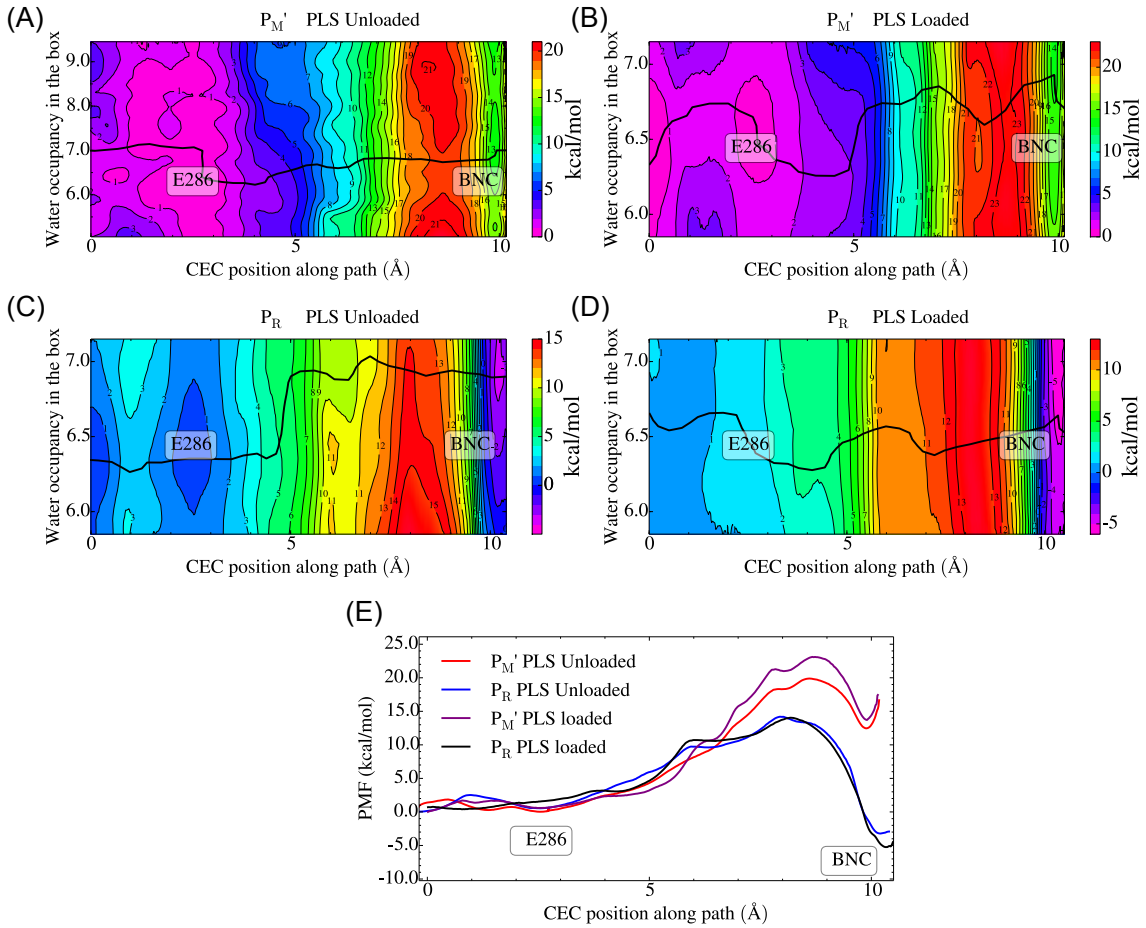


Figure C.3 Free energy profiles for PT through hydrophobic cavity. (A-D) Two-dimensional free energy profiles (2D-PMFs) for the chemical PT from singly protonated E286 to the hydroxide bound to BNC in the (A) P_M' state without PLS preloaded, (B) P_M' state with PLS preloaded, (C) P_R state without PLS preloaded, and (D) P_R state with PLS preloaded. The PMFs are a function of the excess proton center of excess charge (CEC) coordinate through the hydrophobic cavity (HC) as the horizontal axis and the water hydration in the HC as the vertical axis. The positions of E286 and BNC are labeled by text boxes. The statistical errors of the 2D-PMFs in (A-D) are in the range ~ 0.1-3 kcal/mol. The nearly horizontal 1D minimum free energy pathways (black lines) indicate that the PT and HC water hydration are not strongly coupled for the chemical PT. (E) 1D free energy profiles (PMFs) for chemical PT, in P_M' state with and without proton preloaded at PLS (purple and red, respectively), and P_R state (black and blue, respectively). The positions of E286 and BNC are labeled by text boxes. The statistical errors of the 1D-PMFs in (E) are in the range ~ 0.1-1 kcal/mol. PT to the BNC with PLS preloaded (D and black curve in E) is the most physically relevant for the $P_R \rightarrow F$ transition (see main text).

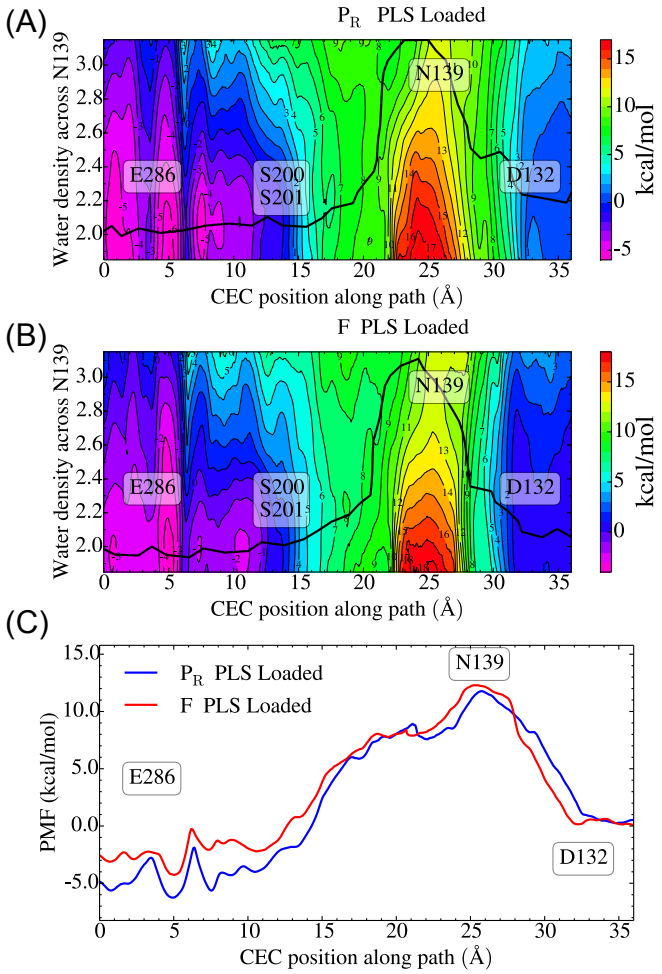


Figure C.4 Free energy profiles for PT through D-channel. (A) and (B): Two-dimensional free energy profiles (2D-PMFs) for PT through the D-channel in the (A) P_R and (B) F states, with PLS preloaded. The PMFs are a function of the excess proton center of excess charge (CEC) coordinate through the D-channel as the horizontal axis and the water hydration in the asparagine gate region as the vertical axis. The minimum free energy pathway for each state is depicted by a black line. The positions of E286, S201, S200, N139, and D132 are labeled by text boxes. The PT starts from a protonated D132 and deprotonated E286 ($x \simeq 35$ Å), and ends with deprotonated D132 and protonated E286 ($x \simeq 0$ Å). The strongly coupled behavior of the PT and the water hydration in the asparagine gate region along the 1D minimum free energy path (black line) is clearly evident. The statistical errors of the 2D-PMFs in (A) and (B) are in the range ~ 0.1-3 kcal/mol. (C) 1D PMFs for PT through the D-channel along the minimum free energy pathway for P_R (blue) and F (red) states. The positions of E286, N139, and D132 are labeled by text boxes. The statistical errors of the 1D-PMFs in (C) are in the range ~ 0.1-1 kcal/mol.

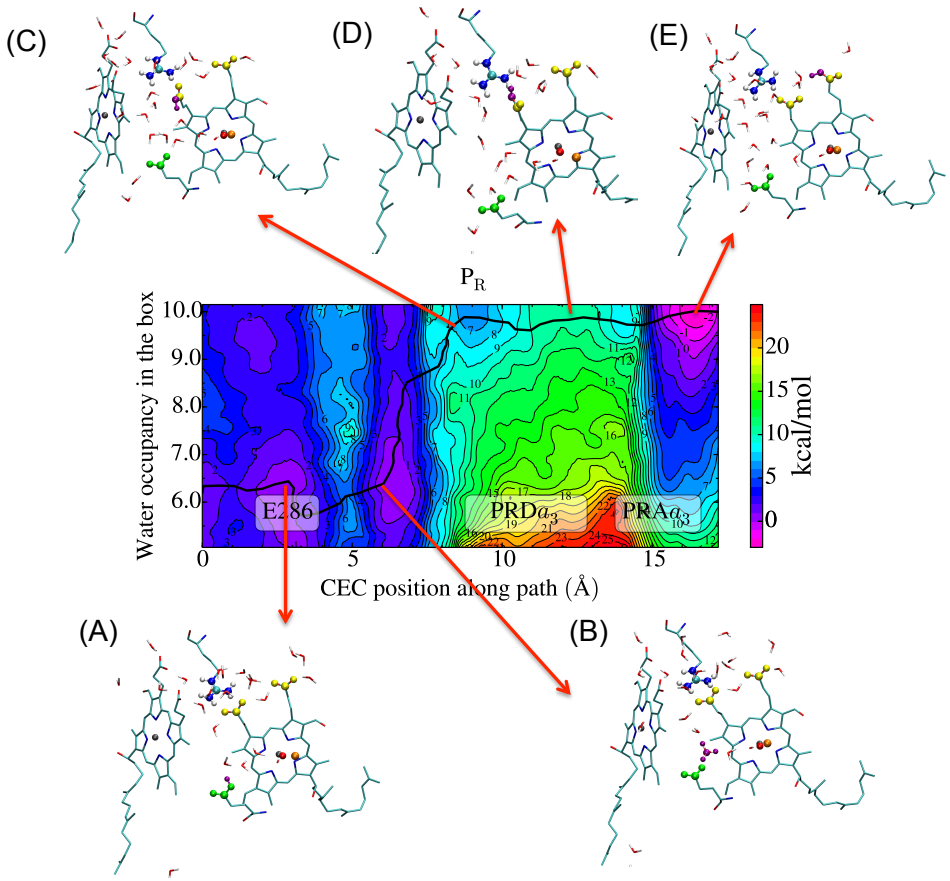


Figure C.5 (A-E) Molecular structures during the PT from E286 to the PLS in the P_R state along the minimum free energy pathway (black line on the 2D-PMF). E286 is shown in green, the propionate groups of the PLS in yellow, the excess proton in purple, the iron atom of the heme groups in gray, and the copper atom in the BNC in orange. The heme groups are shown as sticks. The transition state structure is shown in (D).

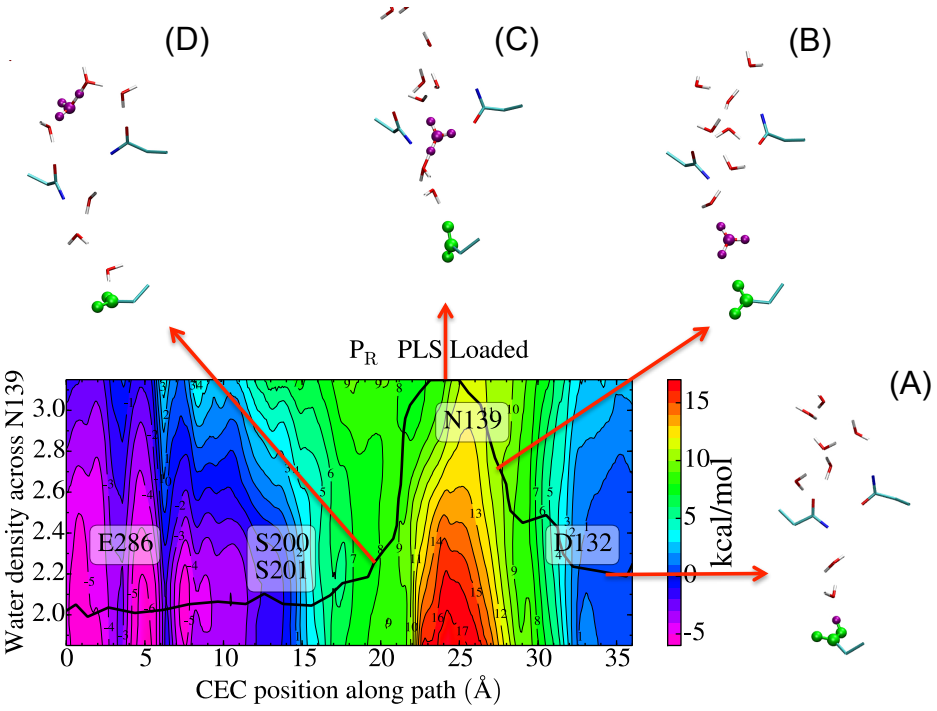


Figure C.6 (A-D) Molecular structures during the PT from D132 to E286 in the P_R state along the minimum free energy pathway (black line on the 2D-PMF).E286 is shown in green and the excess proton in purple. The N139 and N121 residues are shown as sticks. The transition state structure is shown in (C).

C.11 Supplementary Tables

Table C.1 Description of the CcO redox and protonation states in the MS-RMD simulations.

State	Cu_A	Heme *a*	Heme a_3	Cu_B	Y288
P_M'	Cu(I+I)	Fe[II]	Fe[IV]=O^{2-}	Cu[II]-OH^-	TyrO•
P_R	Cu(I+I)	Fe[III]	Fe[IV]=O^{2-}	Cu[II]-OH^-	$TyrO^-$
F	Cu(I+I)	Fe[III]	Fe[IV]=O^{2-}	Cu[II]-H_2O	$TyrO^-$

Table C.2 The MS-RMD model parameters of E286, D132 and BNC in C*c*O. The redox states of the system and the PT processes for which the MS-RMD models are parameterized are labeled in the parenthesis under the protonatable groups.

	E286 (pump)	E286 (react)	E286 (D-channel)	D132 (D-channel)	BNC (P_M')	BNC (P_R)
B	0.04	0.04	5.18	3.0	0.20	1.18
b	0.21	0.21	0.22	8.70	1.98	1.96
b'	0.53	0.53	0.51	8.32	0.00	0.00
d^0_{OO}	2.4	2.4	2.4	2.4	2.4	2.4
C	4.93	4.93	4.93	0.02	0.20	0.01
c	1.98	1.98	1.98	9.78	0.00	2.00
d^0_{OH}	1.0	1.0	1.0	1.0	1.0	1.0
r_s^l	3.5	3.5	3.5	3.5	3.5	3.5
r_s^h	4.0	4.0	4.0	4.0	4.0	4.0
V_{ii}	-136	-141	-126	-136	-110	-119
c_1	-21.6	-25.7	-21.6	-23.0	-25.5	-22.0
c_2	2.8	2.8	2.8	4.6	6.9	1.5
c_3	1.3	1.3	1.3	1.2	1.2	1.3
D	143	143	143	143	143	143
α	1.8	1.8	1.8	1.8	1.8	1.8
r_0	0.98	0.98	0.98	0.98	0.98	0.98

Table C.3 The MS-RMD model parameters of PRDa_3, PRAa_3 in C*c*O. The redox states of the system for which the MS-RMD models are parameterized are labeled in the parenthesis under the protonatable groups.

	PRDa_3 (P_M')	PRAa_3 (P_M')	PRDa_3 (P_R)	PRAa_3 (P_R)
B	1.00	0.99	0.40	1.00
b	2.00	2.00	1.80	2.00
b'	0.00	0.01	0.74	0.01
d_{OO}^0	2.4	2.4	2.4	2.4
C	0.99	0.32	0.57	0.95
c	1.99	1.98	0.66	1.94
d_{OH}^0	1.0	1.0	1.0	1.0
r_s^l	3.5	3.5	3.5	3.5
r_s^h	4.0	4.0	4.0	4.0
V_{ii}	-150	-146	-148	-164
c_1	-31.9	-38.7	-30.6	-40.3
c_2	2.7	1.9	1.2	4.3
c_3	1.1	1.2	1.3	1.1
D	143	143	143	143
α	1.8	1.8	1.8	1.8
r_0	0.98	0.98	0.98	0.98

Table C.4 Charges for the Cu_B complex with water ligand. The total charge of the complex is +1 and the charges represent the F state.

Residue/Atom	Charge	Residue/Atom	Charge
Cu	0.102891	His333	
His284		CB	-0.044220
CB	-0.035004	HB1	0.072991
HB1	0.025432	HB2	0.072991
HB2	0.025432	ND1	-0.287283
ND1	-0.327117	HD1	0.313759
CG	0.240668	CG	0.109448
CE1	0.114118	CE1	-0.005637
HE1	0.145211	HE1	0.127801
NE2	0.015633	NE2	-0.073717
CD2	-0.291661	CD2	-0.206675
HD2	0.185949	HD2	0.194861
Tyr288		His334	
CB	-0.032381	CB	0.055548
HB1	0.046442	HB1	0.042470
HB2	0.046442	HB2	0.042470
CG	0.355739	ND1	-0.222780
CD1	-0.320302	HD1	0.303707
HD1	0.169652	CG	0.059376
CE1	-0.090454	CE1	0.013947
HE1	0.142073	HE1	0.128737
CZ	0.395685	NE2	-0.045397
OH	-0.508863	CD2	-0.187061
CD2	-0.568165	HD2	0.067823
HD2	0.322658	WAT	
CE2	0.207278	O	-0.600232
		H	0.349859
		H	0.349859

Table C.5 Charges for the Cu_B complex with water ligand. The total charge of the complex is +2 and the charges represent the hypothetical state when water is formed at the BNC without electron transferred from heme *a* to the BNC.

Residue/Atom	Charge	Residue/Atom	Charge
Cu	0.367305	His333	
His284		CB	-0.010826
CB	0.076426	HB1	0.097798
HB1	0.034795	HB2	0.097798
HB2	0.034795	ND1	-0.357884
ND1	-0.335539	HD1	0.358854
CG	0.139477	CG	0.160031
CE1	0.086723	CE1	0.083244
HE1	0.154940	HE1	0.122523
NE2	-0.004625	NE2	-0.106667
CD2	-0.177937	CD2	-0.214235
HD2	0.179207	HD2	0.192588
Tyr288		His334	
CB	-0.025228	CB	0.014208
HB1	0.139386	HB1	0.108557
HB2	0.139386	HB2	0.108557
CG	0.452569	ND1	-0.295508
CD1	-0.457841	HD1	0.360531
HD1	0.224088	CG	0.063177
CE1	0.039711	CE1	0.075023
HE1	0.149669	HE1	0.164318
CZ	0.287133	NE2	-0.064106
OH	-0.463121	CD2	-0.103335
CD2	-0.777407	HD2	0.032519
HD2	0.329613	WAT	
CE2	0.335019	O	-0.674438
		H	0.429365
		H	0.429365

Table C.6 Bond and angle parameters for the water ligand in the Cu_B complex. For the O-H bond Morse potential is used. For the H-O-H angle the harmonic potential is used.

O-H bond	D (kcal/mol)	148.777
	a ($Å^{-1}$)	1.873
	r_0 (Å)	0.976
H-O-H angle	k (kcal/radian^2)	103.9719
	θ_0 (degrees)	54.8198

C.12 Supplementary References

1. Johansson MP, Kaila VR, & Laakkonen L (2008) Charge parameterization of the metal centers in cytochrome c oxidase. *J. Comput. Chem.* 29(5):753-767.

2. Svensson-Ek M*, et al.* (2002) The X-ray Crystal Structures of Wild-type and EQ(I-286) Mutant Cytochrome c Oxidases from Rhodobacter sphaeroides. *J. Mol. Biol.* 321(2):329-339.

3. Jorgensen WL & Jenson C (1998) Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: seeking temperatures of maximum density. *J. Comput. Chem.* 19(10):1179-1186.

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

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

6. Tom D, Darrin Y, & Lee P (1993) Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98(12):10089-10092.

7. Feller SE, Zhang Y, Pastor RW, & Brooks BR (1995) *Constant Pressure Molecular Dynamics Simulation : The Langevin Piston Method* (American Institute of Physics, Melville, NY, ETATS-UNIS) p 9.

8. Verlet L (1967) Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* 159(1):98.

9. Phillips JC*, et al.* (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26(16):1781-1802.

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

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

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

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

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

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

16. Lee S, Liang R, Voth GA, & Swanson JMJ (2016) Computationally Efficient Multiscale Reactive Molecular Dynamics to Describe Amino Acid Deprotonation in Proteins. *J. Chem. Theory Comput.* (In Press) DOI: 10.1021/acs.jctc.5b01109.

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

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

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

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

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

22. Valiev M*, et al.* (2010) NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* 181(9):1477-1489.

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

24. Roux B (1995) The Calculation of the Potential of Mean Force Using Computer-Simulations. *Comput. Phys. Commun.* 91(1-3):275-282.

25. Allen MP & Tildesley DJ (1990) *Computer Simulation of Liquids* (Oxford University Press New York, NY).

26. Kumar S, Bouzida D, Swendsen RH, Kollman PA, & Rosenberg JM (1992) The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules .1. The Method. *J. Comput. Chem.* 13(8):1011-1021.

27. Weinan E, Ren WQ, & Vanden-Eijnden E (2005) Finite temperature string method for the study of rare events. *J. Phys. Chem. B* 109(14):6688-6693.

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

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

30. Li DH & Voth GA (1991) Feynman Path Integral Approach for Studying Intramolecular Effects in Proton-Transfer Reactions. *J. Phys. Chem.* 95(25):10425-10431.

31. Wikstrom M*, et al.* (2005) Gating of proton and water transfer in the respiratory enzyme cytochrome c oxidase. *Proc. Natl. Acad. Sci. U. S. A.* 102(30):10478-10481.

32. Kaila VR, Verkhovsky MI, Hummer G, & Wikstrom M (2008) Glutamic acid 242 is a valve in the proton pump of cytochrome c oxidase. *Proc. Natl. Acad. Sci. U. S. A.* 105(17):6255-6259.

33. Wikstrom M & Verkhovsky MI (2007) Mechanism and energetics of proton translocation by the respiratory heme-copper oxidases. *Biochim. Biophys. Acta* 1767(10):1200-1214.

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

35. Gaus M, Goez A, & Elstner M (2013) Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* 9(1):338-354.

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

37. Lide DR (2005) Properties of Amino Acids. *CRC handbook of chemistry and physics, Internet Version*, (CRC Press, Boca Raton, FL), p 7−1.

38. Lide DR (2005) Dissociation Constants of Organic Acids and Bases. *CRC handbook of chemistry and physics, Internet Version*, (CRC Press, Boca Raton, FL), pp 8-46.

39. Goyal P, Yang S, & Cui Q (2015) Microscopic basis for kinetic gating in cytochrome c oxidase: insights from QM/MM analysis. *Chem. Sci.* 6(1):826-841.

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

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

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

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

44. Xu J & Voth GA (2006) Free energy profiles for H+ conduction in the D-pathway of Cytochrome c Oxidase: a study of the wild type and N98D mutant enzymes. *Biochim. Biophys. Acta* 1757(7):852-859.

45. Xu J & Voth GA (2005) Computer simulation of explicit proton translocation in cytochrome c oxidase: the D-pathway. *Proc. Natl. Acad. Sci. U. S. A.* 102(19):6795-6800.

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