UNIVERSITY OF CALIFORNIA,
IRVINE

Protein Association in the Membrane, Mitochondria, and Cytosol: Results from Combined Brownian and Molecular Dynamics Studies

DISSERTATION

Submitted in partial satisfaction of the requirement
for the degree of

DOCTOR OF PHILOSOPHY

in Physical Chemistry

by

James B. Fields

Dissertation Committee:
Professor Douglas J. Tobias, Chair
Professor Vladimir A. Mandelshtam
Professor Craig C. Martens

2017

DEDICATION

To my loving and supportive parents, Jonathan and Evelyn.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

ACKNOWLEDGEMENTS

I can't thank enough the many people and organizations for which my successful graduate career would not have been possible. First off, my advisor Doug Tobias has been the guiding light throughout my graduate career. His vast knowledge of physics, chemistry, biology, and science in general are evident every time you talk to him. We've had many great conversations and great beers over the years, and I am so much better for that. Jim Hall has served as a second advisor to me throughout the last four years, and I am extremely grateful to have worked with him. He has a way of livening up every meeting, and his passion, excitement, and enthusiasm for scientific research are contagious. Both Doug and Jim have been incredibly generous with their financial and academic support of me. Doug and Jim have sent me to numerous conferences around the US (including Hawaii a couple of times!) to disseminate my research, and without those experiences, I would not be half the scientist I am today.

I would also like to thank the entire Chemistry Department at UCI. In particular, Craig Martens, Vladimir Mandelshtam, and Ioan Andricioaei, who have taught me so much inside and outside of the classroom, whether it's statistical mechanics or important life lessons. I would like to thank my collaborator, Thomas Poulos, for many insightful discussions as well as helping me through the manuscript-writing process, which can be pretty tough for a young scientist. The department has provided me with invaluable assets that have proved vital to my scientific development. We have a wonderful cross-disciplinary, collaborative research environment here, and I wouldn't trade that for anything. The department accommodated my switch from organic to physical chemistry, which I am extremely grateful for and probably owe the completion of my degree to. To the computational staff, particularly Nathan Crawford, Harry Mangalam, Joe Farran, Jim Dao, and Richard North for keeping the computers running as smoothly as possible and for all the helpful advice that made the supercomputing resources so easy to use. Last but not least, the wonderful staff who do such a great job keeping the Chemistry Department in order. In particular, I would like to thank Jamie Albano, Tenley Dunn, Laura Donaldson, Romina Van Alstine, and Terri Short for being so helpful to me throughout my graduate career. You could not ask for more from them and their efforts do not go unnoticed.

My fellow group members have played an enormous part in my scientific development. I am particularly indebted to J. Alfredo Freites and Matthias Heyden, for sharing their knowledge and passion for physical simulations with me. Matthias was my first mentor at UCI and is one of the most talented scientists I have ever met. He taught me some of the fundamentals of scripting, writing code, and performing molecular simulations. Alfredo is also a spectacular scientist and teacher. His knowledge of physics is impeccable and he has always been willing to help me through the tough problems. All of the conversations we had were incredibly insightful. For every success I had here, he was right behind

me. Thanks to my labmate Krista Perry, who has been a wonderful friend to me over the years. We have been through quite a lot, but we made it through. To Stephen White and the TEMPO group, I am indebted to the great meetings I was able to take part in over the years. The monthly meetings were always packed with brilliant scientists and thought-provoking discussions. It was through TEMPO group that I met Jim Hall and Daniel Clemens, who became my collaborators. So in that way, I owe this thesis to the TEMPO group.

I would also like to thank the entire Chemistry Department at my undergraduate institution, the Pennsylvania State University; particularly my academic and research advisors, Dan G. Sykes and Kenneth S. Feldman. Dan helped me tremendously in navigating the dizzying world of an undergraduate chemistry student, allowing me to get the most out of my degree. Ken was a tremendous advisor. He gave me a proper research experience even when he couldn't get much in return. I was allowed to TA for an organic chemistry lecture during my senior year, and for that experience, I am extremely grateful. I also want to thank Ken for bringing UCI's spectacular Chemistry Department to my attention! I would not be here today if it were not for him.

To my best friends, Brian Albee, Mohammed El Makkaoui, Eugene Gutman, and Ian Finn: we have had too many memories to name here. We were all going through graduate school at the same time, and we started a hockey team together. I can't believe how time has flown by, but without you guys, I wouldn't be who I am today. To you guys and the rest of my hockey team, you are my West Coast family. I do not think you know how important you have been to me over the years. To Jeffery Murchet, I am so grateful for all of the early morning surf sessions we had. I learned from you just as much about surfing as I did about life. These memories got me through the best and toughest times.

By far, the most important emotional support I had throughout my graduate school career came from my parents. Even though they live across the country, I could always depend on them. Through all the ups and downs, my father was a rock in my life. Any time of day, I could call him and vent. He would always know what to say, and when things were tough, he always seemed to have a thoughtful solution for every problem. My mother is one of the hardest working, most genuinely caring persons I know. She has accomplished so much in her career all while raising my sister and me. I always try to be as kind as she is, but it is a pretty hard standard to live up to. I love you guys so much. Thank you.

CURRICULUM VITAE

James B. Fields II

PREPARED: June 8, 2017

Education

Doctor of Philosophy, Physical Chemistry 2017
University of California, Irvine Irvine, California

Bachelor of Science, Chemistry 2011
Pennsylvania State University University Park, Pennsylvania

RESEARCH EXPERIENCE

Modeling of Biological Systems
University of California, Irvine
2012-2017
Tobias Research Group

- Application of classical Molecular and Brownian dynamics simulations towards the study of protein dynamics in biological matrices.
- Investigating the structure-function relationships involved in the gating of aquaporin water channels.
- Use of statistical methods for the analysis of classical molecular dynamics trajectories.
- Training and mentoring undergraduate and graduate students in molecular simulations and Linux terminal usage.

Total Synthesis of Prekinamycin Derivatives
Pennsylvania State University, University Park
2008-2011
Feldman Research Group

- Running organic reactions using air-free Schlenck techniques.
- Purification of crude reaction mixtures using recrystallization, chromatography, and short-path distillation.
- Characterization of organic compounds by NMR, IR, and CD spectroscopy.
- Carrying out a complex total synthetic route of multiple non-linear steps.
- Training and mentoring undergraduate and graduate students in organic preparation methods and hazardous materials safety.

PUBLICATIONS

1. Madrona, Y., Hollingsworth, S. A., Stripathi, S.; **Fields, J. B.**, Rwigima, C., Tobias, D. J., and Poulos, T. L. (2014) Crystal Structure of Cindoxin, the P450cin Redox Partner. *Biochemistry*. **53** (9):1435-1446.
2. **Fields, J. B.**, Hollingsworth, S. A., Chreifi, C., Heyden, B. M., Poulos, T. L., Tobias, D. J. (2015) "Bind and Crawl" Association Mechanism of Leishmania major Peroxidase and Cytochrome C Revealed by Brownian and Molecular Dynamics Simulations. *Biochemistry* **54** (49), 7272-7282.
3. **Fields, J. B.**, Németh-Cahalan, K., Vorontsova, I., Freites, J. A., Hall, J. E., and Tobias D. J. (2017) Calmodulin gates Aquaporin 0 permeability through a positively charged cytoplasmic loop. *J. Biol. Chem.* **292**, 185-195.
4. **Fields, J. B.**, Freites, J. A., Hall, J. E., and Tobias D. J. (2017) His-40 protonation modulates the permeability of Aquaporin 0 through the first constriction site. *In preparation.*

CONFERENCE TALKS

1. **James B. Fields**, J. Alfredo Freites, James E. Hall, and Douglas J. Tobias. Effects of Aquaporin 0 phosphorylation on its interaction with calmodulin: atomic-scale details from computer simulation studies. National Foundation for Eye Research International Conference on the Lens (NFER ICL), Kona, HI, **2014**
2. **James B. Fields**, Karin Németh-Cahalan, J. Alfredo Freites, James E. Hall, and Douglas J. Tobias. Phosphorylation of Aquaporin Zero and the Calmodulin Interaction. National Foundation for Eye Research International Conference on the Lens (NFER ICL), Kona, HI, **2015**

CONFERENCE PUBLICATIONS

1. **James B Fields**, Yumay Chen, Yu-Han Chen, Charity Juang, Douglas J. Tobias, and Ping H. Wang. “New Insights into Atomic-Level Interactions between Akt1 and Pyruvate Dehydrogenase Complex in Cardiomyocyte Mitochondria.” American Diabetes Association (ADA) 75th Scientific Session, San Francisco, CA, **2014**
2. **James B. Fields**, Karin Németh-Cahalan, J. Alfredo Freites, James E. Hall, and Douglas J. Tobias. “An Arginine-rich Loop is Critical for the Modulation of the Water Permeability of Aquaporin 0.” Biophysical Society (BPS) 60th Annual Meeting, Los Angeles Convention Center, Los Angeles, CA, **2016**
3. Scott A. Hollingsworth, **James B. Fields**, Georges Chreifi, Matthias Heyden, Anton P. Arce, Hugo I. Magaña-Garcia, Douglas J. Tobias, Thomas L. Poulos. Association Mechanism of Leishmania Major Peroxidase and Cytochrome C Revealed Through Brownian and Molecular Dynamics. Biophysical Society (BPS) 60th Annual Meeting, Los Angeles Convention Center, Los Angeles, CA, **2016.**

TEACHING EXPERIENCE

Undergraduate Teaching Assistant
Pennsylvania State University, University Park
2010-2011

- Chem 211 (Undergraduate Organic Chemistry II Lecture)

Graduate Teaching Assistant
University of California, Irvine
2011-2014

- Chem 1A,B (Undergraduate General Chemistry Lecture)
- Chem 1LA,LB,LC (Undergraduate General Chemistry Laboratory)
- Chem 51LA,LB (Undergraduate Organic Chemistry Laboratory)
- Chem 232B (Graduate Statistical Mechanics)

PROFESSIONAL AFFILIATIONS

- American Chemical Society (ACS)
- Biophysical Society (BPS)
- National Foundation for Eye Research (NFER)

ABSTRACT OF THE DISSERTATION

Protein Association in the Membrane, Mitochondria, and Cytosol:
Results from Combined Brownian and Molecular Dynamics Studies

By

James B. Fields

Doctor of Philosophy in Physical Chemistry

University of California, Irvine, 2017

Professor Douglas J. Tobias, Chair

Protein-protein interactions (PPIs) play a crucial role in the regulation of protein activity, and are therefor of great importance in biology. Interactions between proteins require a certain degree of specificity, as the functional regulation of a protein needs to occur within a particular signal cascade and cannot be triggered by non-native interactions. At the same time, most PPIs are dominated by long-range coulombic interactions, which attract positively and negatively charged proteins, or patches of a protein, towards each other. In cases where the secondary structures of the two interacting proteins in infinite dilution are complementary towards the formation of the active complex, the association rate constant becomes diffusion-limited. However, many active protein complexes require some degree of conformational sampling after the formation of the initial encounter, which adds a second rate-limiting step.

The work presented herein details the molecular association and regulatory mechanisms of three unique protein complexes. The primary methods that I employed in this dissertation include two forms of atomistic dynamics simulations, specifically molecular dynamics (MD) and Brownian dynamics (BD), which I use both separately and in tandem. First, I discuss my BD/MD generated model of the cytochrome P450cin/Cindoxin complex. Very few crystal structures exist of cytochrome P450 protein complexes. Our BD/MD complex allowed us to characterize the electron transfer pathway in the P450cin/Cindoxin complex. In another case where a complex crystal structure was previously solved, the *Leishmania major* peroxidase-cytochrome c complex, BD simulations helped to reveal the first encounter site that plays a role in the formation of the active complex. Finally, I detail the regulation of the water channel protein, aquaporin 0 (AQP0) by external calcium concentration, phosphorylation, and pH.

Chapter 1

Introduction

Molecular association underpins a wide variety of principal biological processes including signal propagation, energy transfer, and protein activity regulation [1-3]. A great deal of effort has gone into the production of theoretical and experimental methods to study protein association [1, 4-6]. In this dissertation, I present data from two kinds of atomistic dynamics simulations, Brownian dynamics (BD) and molecular dynamics (MD). Atomistic dynamics simulations are essential tool for the study of protein interactions, folding, and conformational change. The pioneers of MD simulations were recognized with the 2013 Nobel Prize in Chemistry [7]. In atomistic MD, a microscopic N-body system is propagated using Newton's equations of motion and forces derived from quantum mechanics [8]. MD simulations remain computationally expensive as they are limited by the calculation of the higher frequency modes in the system, such as water motion [9-11]. A typical atomistic MD trajectory of a solvated protein computed on a modern supercomputer can reasonably attain the microsecond timescale, with some special examples pushing this limit by a few orders of magnitude [10, 11]. However, experimental measurements of protein association are taken on the

bench-top timescale (i.e. seconds) where MD simulations currently cannot reach.

Where BD simulations are particularly adept is in the study of protein motion on the diffusional encounter timescales (10^{-3} to 10^{-7} seconds) where a single all-atom MD trajectory becomes computational expensive [12-15]. With the BD method, millions of trajectories can be generated in less time than it takes to run a single MD trajectory. In a bimolecular BD trajectory, two proteins are initiated from a fixed distance away from one another [15]. The proteins are allowed to diffuse until they reach a predefined escape distance or until a bound complex is formed. The diffusional association rate constant can be calculated from the fraction of trajectories that result in binding (see Ch. 3; following derivation from Ref. [16]). Prevalent binding modes can be visualized using a pair-wise probability density (see Ch. 3). Further clustering and decomposition of the high probability modes can indicate which portions of a protein's surface are most likely to interact with an other (see Ch. 2 and Ch. 3). From this data, high-interest targets for mutagenesis are tested experimentally for their effect on activity (see Ch. 2, Ch. 3, and Ch. 4).

The BD method simplifies the calculation of protein-protein interactions by using precomputed, grid-based potentials in an implicit solvent environment. Removing the explicit treatment of solvent (water) molecules substantially reduces the computation time required to sample protein diffusion [15, 17, 18]. A Poisson-Boltzmann treatment is used for the computation of electrostatic interactions, which represent the dominant contribution to the long-range forces between proteins [18-20]. Additionally, most BD simulations use rigid configurations of the

protein atoms [14]. Although fixed atomic coordinates prohibit conformational change, the effects of protein conformation can be modeled with independent sets of BD trajectories for each available configuration (see Ch. 3) [21].

By definition, the association rate of a diffusion-limited process is dependent on the time it takes for the two molecules to collide [15]. Therefor, a diffusion association rate constant (k_a) determined from a protein activity assay can give crucial insight into the amount of conformational change involved in forming the active complex from the initial encounter [22]. A typical value of k_a for a protein complex can range between 1 and 10^{10} M^{-1} s^{-1}, which represents the number of activation events that occur given a fixed concentration [23, 24]. According to transient complex theory, a k_a value of 10^4 M^{-1} s^{-1} or greater signifies a process that is diffusion-limited. For these cases, the rate constant can be accurately predicted by a Boltzmann factor of the electrostatic interaction energy of the protein complex (ΔG_{EP}) (Eq. 4 from Ref. [23]):

$$k_a = k_{a0}\, e^{\frac{-\Delta G_{EP}}{k_B T}} \tag{1.1}$$

where k_{a0} is the so-called basal rate constant for the formation of the encounter complex assuming random diffusion with no interactions. This model works best when there are few to no conformational changes associated with activation. For processes in which conformational flexibility and thermal fluctuations play a significant role, k_a becomes smaller than 10^4 M^{-1} s^{-1} and this simple model (see Eq. 1.1) fails to reproduce experimental rate constants [23]. These protein complexes may have predominant electrostatic interactions that contribute to

initial encounter and dissociation events, but to form the active complex requires a substantial conformational change. For these complexes, the value of k_a decreases substantially as the timescale of the protein conformational change (10^{-2} to 10^{-9} seconds) becomes an additional rate-limiting step that follows the formation of the encounter complex [22, 23]. For these cases, it is crucial to recognize the limitations of the BD method.

Some of the potential drawbacks of MD and BD can be circumvented by employing a combined MD/BD approach [21]. In this approach, BD is employed to model long-range association processes while MD is employed to study the finer atomic motion of the active and transient complexes. For cases where the structure of the active complex is unknown, the individual protein structures can be used to generate a set of BD trajectories. The high-probability configurations from the BD trajectories are used to produce an initial complex model [25, 26]. The two-protein complex is then solvated in explicit solvent and neutralizing ions, after which an equilibrium MD trajectory is computed. The MD simulation can be used to relax the BD complex to generate an equilibrated BD/MD model (see Ch. 2), which can be validated by experimental mutagenesis of the predicted interfacial residues [25]. In examples where the individual and complex structures have been solved, combined BD/MD simulations can be used to identify important transient complexes that are difficult or impossible to observe experimentally (see Ch. 3) [26].

In Ch. 2 and Ch. 3, I describe the results from a collaboration with the lab of Thomas L. Poulos (Chancellor's Professor of Biochemistry, Chemistry, and

Pharmaceutical Sciences, UC Irvine) where this combined approach helped to reveal mechanistic details in the activation pathways of soluble redox-active cytochrome proteins. In Ch. 2, a combined BD/MD approach was used to generate a structural model of a P450-ferrodoxin complex, which proved difficult to solve experimentally. From the model complex, we identified mutants that decreased the electron transfer rate in steady-state assays, validating the model and providing insight into the nature of the electron transfer pathway between the redox partners [25]. In Ch. 3, I detail BD simulations of the *Leishmania major* peroxidase (LmP) cytochrome c (LmCytc) complex. This complex plays an important role in the pathogenesis of Leishmaniasis, an infectious flesh-eating disease. LmP and LymCytc were previously crystalized both individually and in the active electron transfer complex. Our BD simulations revealed an important α-helix on LmP, known as helix A, where the initial encounter complex is formed. From crystallography alone, the importance of helix A would not have been recognized [26]. The examples from Ch. 2 and Ch. 3 illustrate how BD simulations can be employed when crystal structures of the active complex may or may not be available.

In Ch. 4 and Ch. 5, the focus shifts toward a membrane water channel protein, aquaporin 0 (AQP0). The results presented in these chapters stem from a long-standing collaboration with the lab of James E. Hall (Professor of Physiology and Biophysics, School of Medicine, UC Irvine). AQP0 plays a critical role in the pathology of cataracts. Of the identified AQP0 mutations linked to congenital cataract formation, many are known to interfere with the modulation of the

channel water permeabilities by external calcium concentration and pH [27-30]. In Ch. 4, I detail atomistic simulations of the AQP0 complex with the calcium-modulatory protein, Calmodulin (CaM). MD and BD simulations were critical in identifying the AQP0 channel-gating residues and their modulation by CaM [31, 32]. In Ch. 5, I detail the pH-gating mechanism of the aromatic/arginine selectivity filter (Ar/R) of AQP0.

MD simulations remain the most important computational tool for studying protein folding, side chain dynamics, and conformational change. Despite the simplicity of the BD model, BD simulations can provide important insight into the mechanisms of protein association. In one form or another, all of the chapters herein draw conclusions from MD and BD data, revealing details of activation and modulatory mechanisms that neither approach could alone. I have included two appendices that explain the technical details of the BD and MD methods. In appendix A, I explain the BD algorithm and the calculation of the BD potential energy. In appendix B, I detail the MD force field and how a classical MD trajectory is propagated forward in time.

Bibliography

1. Gabdoulline, R. R., and Wade, R. C. (2001) Protein-protein association: investigation of factors influencing association rates by brownian dynamics simulations. *J Mol Biol* **306**, 1139-1155

2. Schreiber, G., and Fersht, A. R. (1996) Rapid, electrostatically assisted association of proteins. *Nat Struct Biol* **3**, 427-431

3. Kerppola, T. K. (2009) Visualization of molecular interactions using bimolecular fluorescence complementation analysis: characteristics of protein fragment complementation. *Chem Soc Rev* **38**, 2876-2886

4. Hu, C. D., Chinenov, Y., and Kerppola, T. K. (2002) Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Mol Cell* **9**, 789-798

5. Northrup, S. H., and Erickson, H. P. (1992) Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc Natl Acad Sci U S A* **89**, 3338-3342

6. Gabdoulline, R. R., and Wade, R. C. (1998) Brownian dynamics simulation of protein-protein diffusional encounter. *Methods* **14**, 329-341

7. Buckingham, A. D. (2017) Fifty Years of Chemical Physics Letters. Elsevier

8. Verlet, L. (1967) Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159**, 98-120

9. Feenstra, K. A., Hess, B., and Berendsen, H. J. (1999) Improving e_ciency of large timescale molecular dynamics simulations of hydrogen-rich systems. *J Comput Chem* **20**, 786-798

10. Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O., and Shaw, D. E. (2009) Long-timescale molecular dynamics simulations of protein structure and function. *Curr Opin Stuc Biol* **19**, 120-127

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

12. Ermak, D. L., and Mccammon, J. A. (1978) Brownian Dynamics with Hydrodynamic Interactions. *J Comput Phys* **69**, 1352-1360

13. Zhou, H. X. (1993) Brownian dynamics study of the influences of electrostatic interaction and diffusion on protein-protein association kinetics. *Biophys J* **64**, 1711-1726

14. Martinez, M., Bruce, N. J., Romanowska, J., Kokh, D. B., Ozboyaci, M., Yu, X. F., Ozturk, M. A., Richter, S., and Wade, R. C. (2015) SDA 7: A modular and parallel implementation of the simulation of diffusional association software. *J Comput Chem* **36**, 1631-1645

15. Gabdoulline, R. R., and Wade, R. C. (1998) Brownian dynamics simulation of protein-protein diffusional encounter. *Method Enzymol* **14**, 329-341

16. Northrup, S. H., and Erickson, H. P. (1992) Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc Natl Acad Sci U S A* **89**, 3338-3342

17. Gabdoulline, R. R., and Wade, R. C. (2009) On the contributions of diffusion and thermal activation to electron transfer between Phormidium laminosum plastocyanin and cytochrome f: Brownian dynamics simulations with explicit modeling of nonpolar desolvation interactions and electron transfer events. *J Am Chem Soc* **131**, 9230-9238

18. Gabdoulline, R. R., and Wade, R. C. (1996) Effective Charges for Macromolecules in Solvent. *J Phys Chem* **100**, 3868-3878

19. Novotny, J., Bruccoleri, R. E., Davis, M., and Sharp, K. A. (1997) Empirical free energy calculations: a blind test and further improvements to the method. *J Mol Biol* **268**, 401-411

20. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-10041

21. Motiejunas, D., Gabdoulline, R., Wang, T., Feldman-Salit, A., Johann, T., Winn, P. J., and Wade, R. C. (2008) Protein-protein docking by simulating the process of association subject to biochemical constraints. *Proteins* **71**, 1955-1969

22. Zhou, H.-X. (1997) Enhancement of protein-protein association rate by interaction potential: accuracy of prediction based on local Boltzmann factor. *Biophys J* **73**, 2441-2445

23. Alsallaq, R., and Zhou, H. X. (2008) Electrostatic rate enhancement and transient complex of protein–protein association. *Proteins* **71**, 320-335

24. Zhou, H.-X. (1993) Brownian dynamics study of the influences of electrostatic interaction and diffusion on protein-protein association kinetics. *Biophys J* **64**, 1711-1726

25. Madrona, Y., Hollingsworth, S. A., Tripathi, S., Fields, J. B., Rwigema, J. C., Tobias, D. J., and Poulos, T. L. (2014) Crystal structure of cindoxin, the P450cin redox partner. *Biochemistry* **53**, 1435-1446

26. Fields, J. B., Hollingsworth, S. A., Chreifi, G., Heyden, M., Arce, A. P., Magana-Garcia, H. I., Poulos, T. L., and Tobias, D. J. (2015) "Bind and Crawl" Association Mechanism of Leishmania major Peroxidase and Cytochrome c Revealed by Brownian and Molecular Dynamics Simulations. *Biochemistry* **54**, 7272-7282

27. Ball, L. E., Garland, D. L., Crouch, R. K., and Schey, K. L. (2004) Post-translational modifications of aquaporin 0 (AQP0) in the normal human lens: spatial and temporal occurrence. *Biochemistry* **43**, 9856-9865

28. Nemeth-Cahalan, K. L., Kalman, K., and Hall, J. E. (2004) Molecular basis of pH and Ca2+ regulation of aquaporin water permeability. *J Gen Physiol* **123**, 573-580

29. Németh-Cahalan, K. L., Clemens, D. M., and Hall, J. E. (2013) Regulation of AQP0 water permeability is enhanced by cooperativity. *J Gen Phys* **141**, 287-295

30. Hu, S., Wang, B., Qi, Y., and Lin, H. (2012) The Arg233Lys AQP0 mutation disturbs aquaporin0-calmodulin interaction causing polymorphic congenital cataract. *PLoS One* **7**, e37637

31. Reichow, S. L., Clemens, D. M., Freites, J. A., Nemeth-Cahalan, K. L., Heyden, M., Tobias, D. J., Hall, J. E., and Gonen, T. (2013) Allosteric mechanism of water-channel gating by Ca2+-calmodulin. *Nat Struct Mol Biol* **20**, 1085-1092

32. Fields, J. B., Németh-Cahalan, K. L., Freites, J. A., Vorontsova, I., Hall, J. E., and Tobias, D. J. (2017) Calmodulin Gates Aquaporin 0 Permeability through a Positively Charged Cytoplasmic Loop. *J Biol Chem* **292**, 185-195

Chapter 2

Predicting the Association Complex of P450cin-Cindoxin

The research in this chapter was originally published in Biochemistry [1]. I performed the Brownian dynamics simulations described in the Brownian dynamics docking methods sections (section 2.2.1). I assisted Scott Hollingsworth (Department of Molecular Biology and Biochemistry, UC Irvine) in running and analyzing the molecular dynamics refinement of the Brownian dynamics complex structure. The biochemical experiments, including the solving and assignment of the Cindoxin crystal structure, were carried out by Yarrow Madrona (Department of Molecular Biology and Biochemistry, UC Irvine), Sarvind Tripathi (Department of Molecular Biology and Biochemistry, UC Irvine), Scott Hollingsworth, and Jean-Christophe Rwigema (Department of Molecular Biology and Biochemistry, UC Irvine) in the laboratory of Thomas Poulos (Chancellor's Professor of Biochemistry, Chemistry, and Pharmaceutical Sciences, UC Irvine).

2.1 Introduction

The Cytochrome P450 (CYP) class of enzymes is a critical component of drug metabolism, capable of activating a wide variety of unreactive saturated alkanes for hydroxylation. The P450s are able to accomplish this by activating molecular oxygen through the addition of an electron. The majority of P450s accept electrons from redox partners containing flavin mononucleotide (FMN) or iron-sulfur cofactors. P450s have traditionally been divided into two classes depending on the nature of their redox partners. The first, aptly named Class I, accept electrons through multiple redox partners, starting from a flavin adenine nucleotide (FAD) flavoprotein, transferring an electron to an iron-sulfur cluster ferrodoxin, which then transfers the electron to the P450 heme. Although Class 1 P450s are predominantly found in prokaryotes, they also exist in some vertebrate mitochondria. However, most mammalian P450s are categorized under Class II, which instead bind a single redox partner, the cytochrome P450 reductase (CPR), which possesses individual FAD and FMN cofactors in a single protein. Class II P450s are highly concentrated in both the liver and adrenal glands, where they are responsible for xenobiotic metabolism and steroid biogenesis, respectively.

The P450cin system is tough to classify, since its redox partner, an FMN-containing flavodoxin known as cindoxin (Cdx) [2], takes the place of the iron-sulfur ferredoxin [3]. The human Cdx domain is ~37% homologous to both human CPR-FMN [4] and bacterial fusion protein XplA [5]. Very few crystal structures exist of a P450 complex with its redox partner. Some examples include the FMN-heme domains of P450BM3 [6], the Adx–CYP11A1 complex [7], and the recent Pdx–P450cam structures [8, 9].

My experimental collaborators in the Poulos lab, including Yaro Madrona *et al*, solved the crystal structure of the Cdx cofactor domain of P450cin (pdb codes 4FMX and 4L77) [1]. In this chapter I will describe my results from Brownian and molecular dynamics simulations that were used to produce an atomistic model of the P450cin-Cdx complex structure. The resulting complex model was used to guide further mutagenesis studies, which elucidated important structural and mechanistic features of the redox partners towards the successful transfer of electrons. Additionally, we were able to use molecular dynamics simulations to determine if the binding of Cdx produces similar structural changes in the P450cin domain that have been observed in the P450cam system.

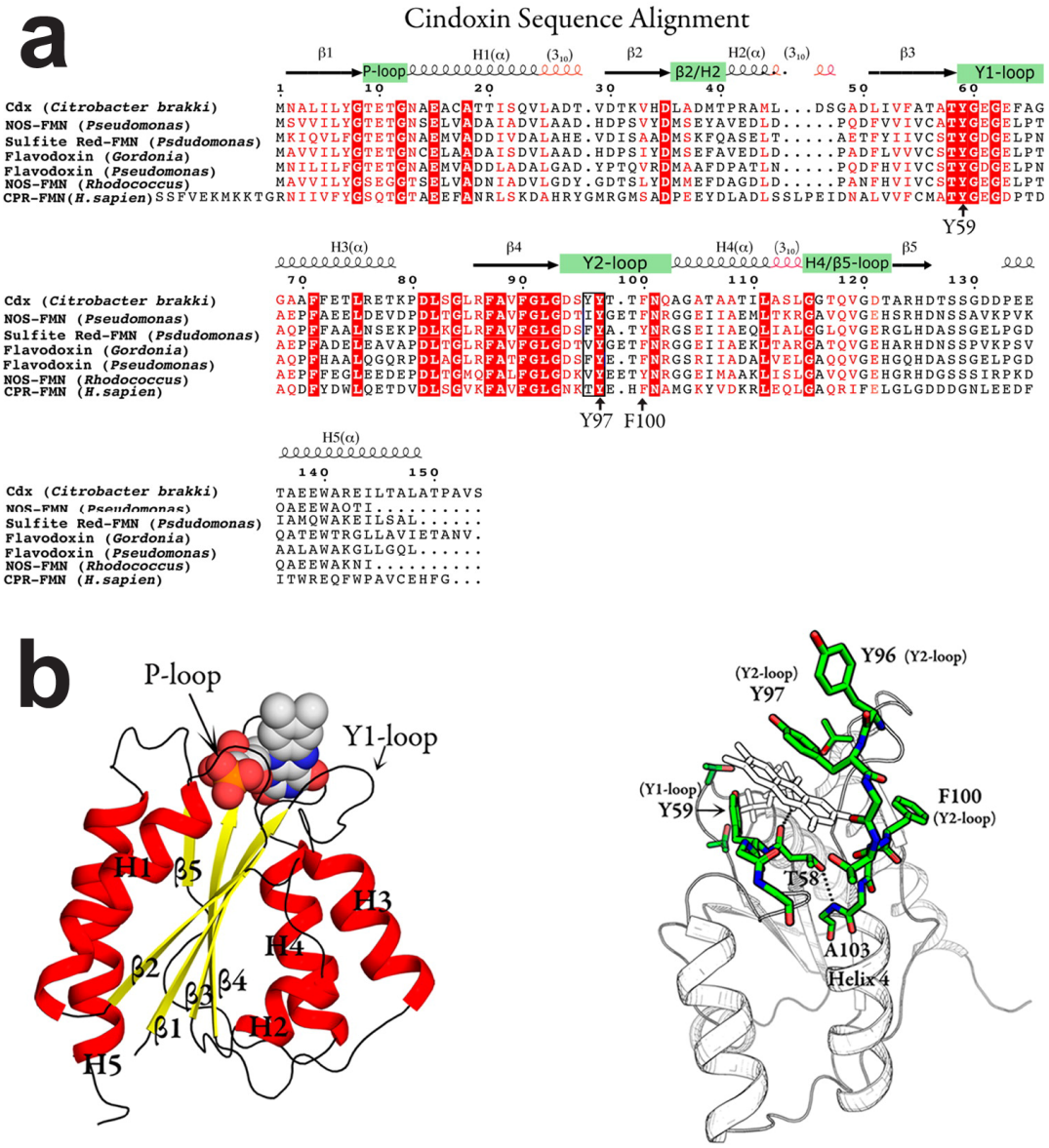

Figure 2.1 Crystal Structure of Cindoxin. a) The major secondary structural elements of the Cdx sequence aligned with other homologous domains. b) Cartoon representation of the Cdx crystal structure with the major secondary structural elements labeled. c) Close-up of the FMN binding pocket. d) Stereo model of the Cdx domain with important interactions labeled which are discussed in the text. Figure adapted from Fig. 1 of Ref. [1].

2.2 Methods

2.2.1 Brownian Dynamics Docking

The bimolecular Brownian dynamics simulation method that we employed is detailed in appendix A. Although experimental electron transfer kinetics will be discussed later in this chapter, we used BD trajectories to determine the structure of the highly occupied encounter complexes.

BD simulations were used to dock Cdx to P450cin. The approach was similar to the one used by Motiejunas *et. al.* [10], with the exception of the included restraints derived from biochemical analysis. The BD method that we employed models proteins as atomically detailed rigid bodies and generates translational and rotational displacements using the Ermak–McCammon algorithm [11]. The SDA program (version 6) was used to conduct the BD simulations [12]. The electrostatic potentials are obtained by finite-difference solution of the nonlinear Poisson–Boltzmann equation using the APBS program package. The solvent dielectric constant was set to 78 with an ionic strength corresponding to 50 mM NaCl.

The crystal structures of both substrate-free and substrate-bound forms of Cdx (PDB: 4L77 and 4FMX, respectively) were used in the BD simulations. Initial configurations were prepared by placing the larger of the two proteins, P450cin, at the origin, where it was kept fixed in place during the BD simulations, and the smaller protein, Cdx, at a random position 300 Å away. The relative diffusion constants for Cdx diffusion around the fixed P450cin were calculated from diffusion constants of Cdx and P450cin, which

were estimated on the basis of the crystal structures using the Hypropro 10 package [13]. Each BD trajectory was prolonged for 500 μs, or until Cdx traveled 400 Å from the origin, at which point the trajectory was terminated. The simulation time step was decreased from 20 to 1 ps when the separation between the two proteins was 20 Å. For each system (substrate-free and substrate-bound Cdx), 20000 BD trajectories were generated.

2.2.2 MD Refinement of the BD Complex

BD encounter complexes between Cdx and P450cin were clustered by center-of-mass position to identify a configuration for an atomistic MD simulation. Both substrate-free and substrate-bound Cdx formed complexes with P450cin, although the substrate–P450cin produces more encounter configurations, so we chose the open form of P450cin for further analysis. We selected the P450cin–Cdx BD configuration that had the shortest substrate-free Cdx flavin–P450cin heme distance.

The complex was solvated in a box of water with initial dimensions of 127 Å × 118 Å × 149 Å. Following energy minimization, a 300 ps MD trajectory was generated at a constant temperature (300 K) and pressure (1 bar) using NAMD version 2.8 [14]. The lengths of bonds containing hydrogen atoms were fixed using the SHAKE and SETTLE algorithms [15, 16]. The equations of motion were integrated with a reversible, multiple-step algorithm with a time step of 2 fs for electrostatic forces and 1 fs for bonded and short-range nonbonded forces [17]. The smooth particle mesh Ewald method with a fourth-order interpolation scheme was used to calculate the electrostatic interactions [18], and the real-space contributions were truncated at 12 Å. Nosé–Hoover–Langevin

pistons were used for pressure control [19], and a Langevin dynamics scheme was used for temperature control. The CHARMM22 force field was used for the proteins [20], the TIP3P force field for water [21], and the CGenFF force field for the FMN cofactor [22].

2.3 Results

2.3.1 Model of the P450cin-Cdx Complex

After exhaustive efforts, our experimental collaborators were able to crystalize a E-to-A triple mutants of human Cdx. The Surface Entropy Reduction Prediction server (SERp) at the University of California at Los Angeles (Los Angeles, CA) was used to determine the best positions for mutants to generate a crystal [23]. The positions of these glutamate residues are E134, E135, and E138. The structure of the Cdx domain, whose sequence is >39% identical to some flavodoxins, shows a similar alternating pattern of the β-strands (β2−β1−β3−β4−β5 arrangement) with 5 α-helical domains running in opposite direction around the protein. I used this single crystal structure to dock both the substrate-free and substrate-bound forms of Cdx (PDB codes: 4L77 and 4FMX, respectively) to P450cin in bimolecular BD simulations. The substrate-bound structure produced complexes with significantly higher occupancy than those of the substrate-free Cdx.

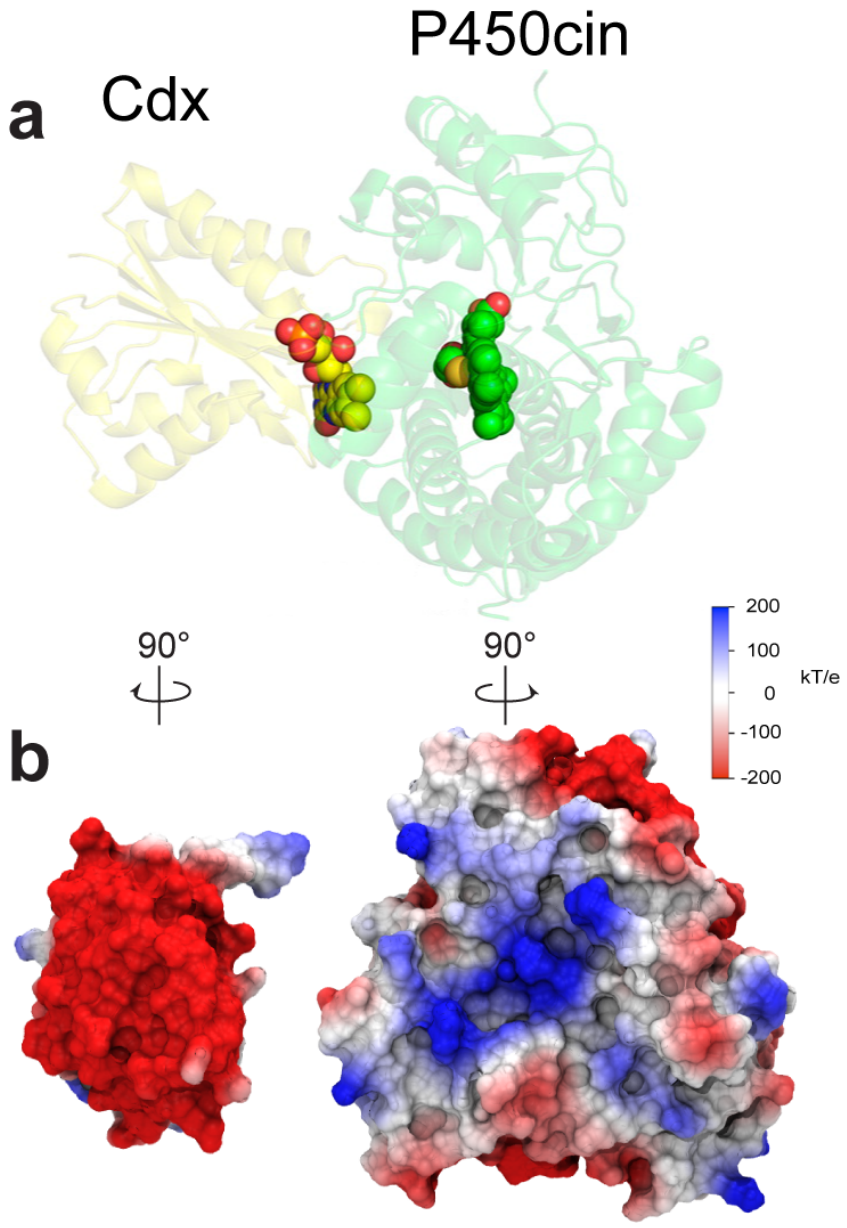

Figure 2.2: Electrostatic Potentials of P450cin and Cindoxin a) A rendering of the P450cin-Cdx MD-BD model complex. The protein domains are labeled and shown in cartoon representations. The heme domain of P450cin is shown in green and atomically colored Van der Waals spheres while the flavin nucleotide of Cdx is shown in yellow and atomically colored Van der Waals spheres. b) Electrostatic potential calculations are mapped on the Van der Waals surfaces. (a) was adapted from TOC Graphic of Ref. [1].

To generate a model of the Cdx-P450cin complex, we utilized a combined BD/MD approach similar to the one used by Motiejunas *et al* [10], which is detailed in the methods section (see section 2.2.1). The initial docked structure was produced from rigid-body bimolecular BD trajectories, where the binding position and orientation was determined by clustering the highest-occupied configurations of Cdx around P450cin. Looking at the interfacial regions of the two proteins along the clustered complex shows substantial hydrophobic and electrostatic complementarity between them. Some of these interactions may be important for the electron transfer reaction. Specifically the salt bridge formed between the R346 of P450cin and the D94 of Cdx forms an interconnected network of charge-polar interactions with the ribityl hydroxyl group of the FMN ligand of Cdx. D94 completes the network with a hydrogen-bond formed between its amide nitrogen to the FMN ring oxygen. Interestingly, D94 in conserved across the bacterial flavodoxins, where it may play a role in the formation of their redox competent complexes. Cdx Y96 protrudes into a hydrophobic cleft created by P450cin V105, A106, and F109 of helix C and loop CD, as well as L348. Y96 forms a π-stacking interaction with P450cin R102. In summary, the major interactions in our model include a combination of electrostatic interactions between R346 and the FMN hydroxyl group/Cdx D94, and hydrophobic interactions between P450cin and Cdx Y96.

2.3.2 Experimental Validation of the P450cin-Cdx Complex Model

To test the legitimacy of our simulation-derived model of the P450cin-Cdx complex, our experimental collaborators expressed a series of Cdx mutants and measured their effect on the catalytic activity of P450cin. The results are summarized in the table above (see Fig. 2.3d). The 15 single point mutations near the FMN-binding site failed to alter

the catalytic activity of P450cin. Coincidently, an earlier publication demonstrated that similar mutations to the FMN-binding domain of FMN-bound CPR showed little to no effect on the catalytic activity [4]. Despite the limited effects of these mutations, the Cdx triple mutant that was used to determine the crystal structure reduced NADPH turnover activity by 92% of that of the wild type (see the EEEA mutant in Fig. 2.3d). Reducing the overall surface charge of a protein by 3 can produce a significant effect on the electrostatic properties of that protein. Since the single mutants E134A, E135A, and E138A produce no significant effect on the catalytic activity of P450cin, we suggest that the triple mutant, EEEA, interferes with the ability of Cdx to interact with P450cin, leading to a reduction in the formation of the ET-competent complex. We hypothesize that this electrostatic patch helps provide coulombic force to drive the formation of the initial encounter complex, since the residues are located outside of the FMN portal away from the P450cin binding site.

The complex model shows that the Y96 of Cdx and the R102 and R346 P450cin sit in the interaction interface. Our initial hypothesis was that the tyrosine hydroxyl group contributes to the electron transfer reaction. However, the Y96F mutant shows activity that is similar to WT (see Fig. 2.3d). The characteristic flavoprotein spectrum, with an A_{280}/A_{456} ratio, shows that Y96F is almost identical to WT, indicating that FMN binding is not compromised by the loss of the hydroxyl group. However, the Y96L mutant exhibits a marked decrease in k_{cat} by greater than eight-fold from 847 to 96 min^{-1}. However, increasing the Cdx-Y96L concentration to 40 μM (a ten-fold increase) increases activity to approximately 40% of the wild-type activity with a k_{cat} of 353 min^{-1}.

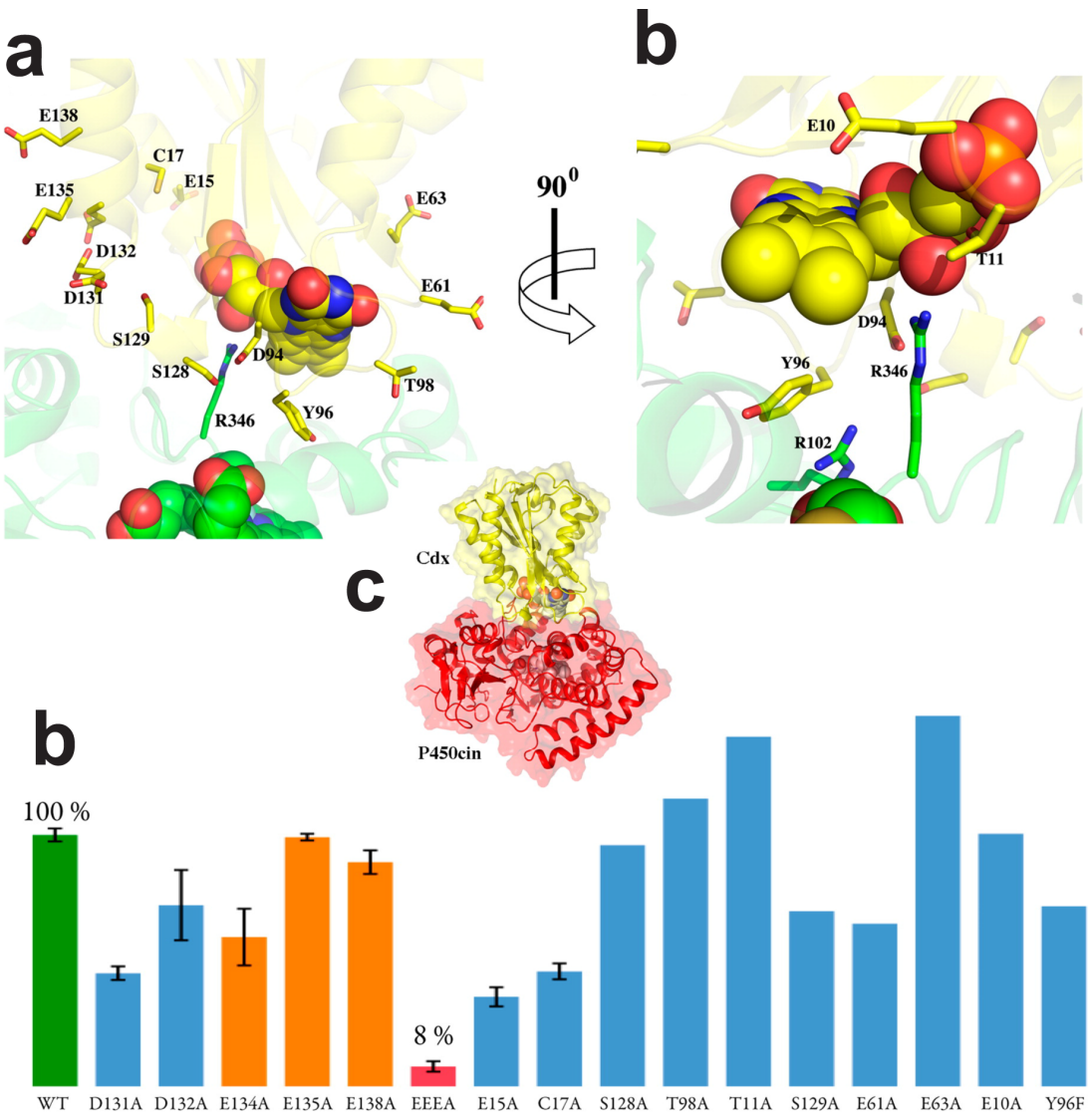

Figure 2.3: The BD/MD Model of the P450cin-Cindoxin Complex. a,b) The binding interaction interface is shown from two different angles. P450cin is colored green and Cdx is colored yellow. The shown structure is the result of clustered BD trajectories relaxed in a 300 ns MD simulation. c) Cartoon representation of the complex. d) P450cin enzyme activity after being solubilized with various Cdx mutants. Figure adapted from Fig. 3 of Ref. [1].

The observed concentration dependence of Cdx-Y96L on the turnover rate indicates that this aromatic residue plays a role in the P450cin-Cdx interaction without completely interfering with the electron transfer mechanism. To further probe the electron transfer activity of Cdx-Y96L, our experimental collaborators performed stopped-flow spectroscopy on WT and Cdx-Y96L towards the first electron transfer reaction with P450cin. WT and Cdx-Y96L gave rates (*k*) of 6.6 s^{-1} ($\chi^2 = 4.4 \times 10^{-5}$) and 2.1 s^{-1} ($\chi^2 = 1.0 \times 10^{-4}$), respectively (see Fig. 2.4a,b,c). Although the rate of the first electron transfer is slowed by the Y96L mutation, the difference is not as large as the eight-fold reduction in NADPH turnover. Because the reduction of the first electron transfer rate is moderate, it is likely that the second electron transfer step where oxy-P450cin is reduced is severely hindered. The P450cin arginine mutants, R102A and R346A, show a substantial decrease in NADPH turnover activity, with decreases in activity at approximately 90% and 95% of WT, respectively. These reductions in NADPH turnover support the hypothesis that these arginine residues form important electrostatic and polar interaction with Y96 and D94 of Cdx, which we observed in the BD/MD complex. Increasing the concentration of P450cin-R102A by ten-fold shows an approximately four-fold increase in the activity. P450cin-R346A does not show the same concentration dependence, suggesting that it is a critical for the formation of the electron transfer complex.

2.4 Discussion

Our crystal structure of Cdx-EEEA closely resembles the structures of related FMN-containing redox-partners of P450. One of the features is the Y96 residue, which is unique to Cdx. In most other ferrodoxin sequences, this residue is a threonine. In the BD/MD complex model, we see this tyrosine residue interacting with the hydrophobic cleft of P450cin, somewhat in between the two cofactors. We initially hypothesized that this residue would be important for the electron transfer reaction. Although Cdx-Y96L has a substantially reduced activity in steady state assays, the activity does increase with an increasing mutant Cdx concentration, indicating that the role of Y96 is primarily in binding and not electron transfer. The arginine residues R346 and R102 of P450cin appear near the interaction interface of the docked complex model. R346A reduces the activity of P450cin to 5% of WT. It is difficult to determine the precise role of R346 in the enzymatic mechanism. Since the Cdx-binding surface of P450cin is located on an electropositive patch of the protein, we hypothesize that this basic residue contributes to the electrostatic interaction with Cdx. However, the R364 side chain interacts with the Cdx FMN domain and Y94. This, coupled with the large reduction in enzymatic activity by R364A, suggests that this arginine might play an additional role in promoting electron transport.

In contrast to the interfacial arginine residues of P450cin, the individual E-to-A mutants (E134A, E135A, or E138A) of Cdx have little to no effect on P450cin activity. Only the triple mutant, dubbed EEEA, has an affect on turnover. Although these residues sit slightly outside of the interaction interface, we believe that this negatively-charged helix containing these glutamate residues serves as a promoter for the formation of the initial

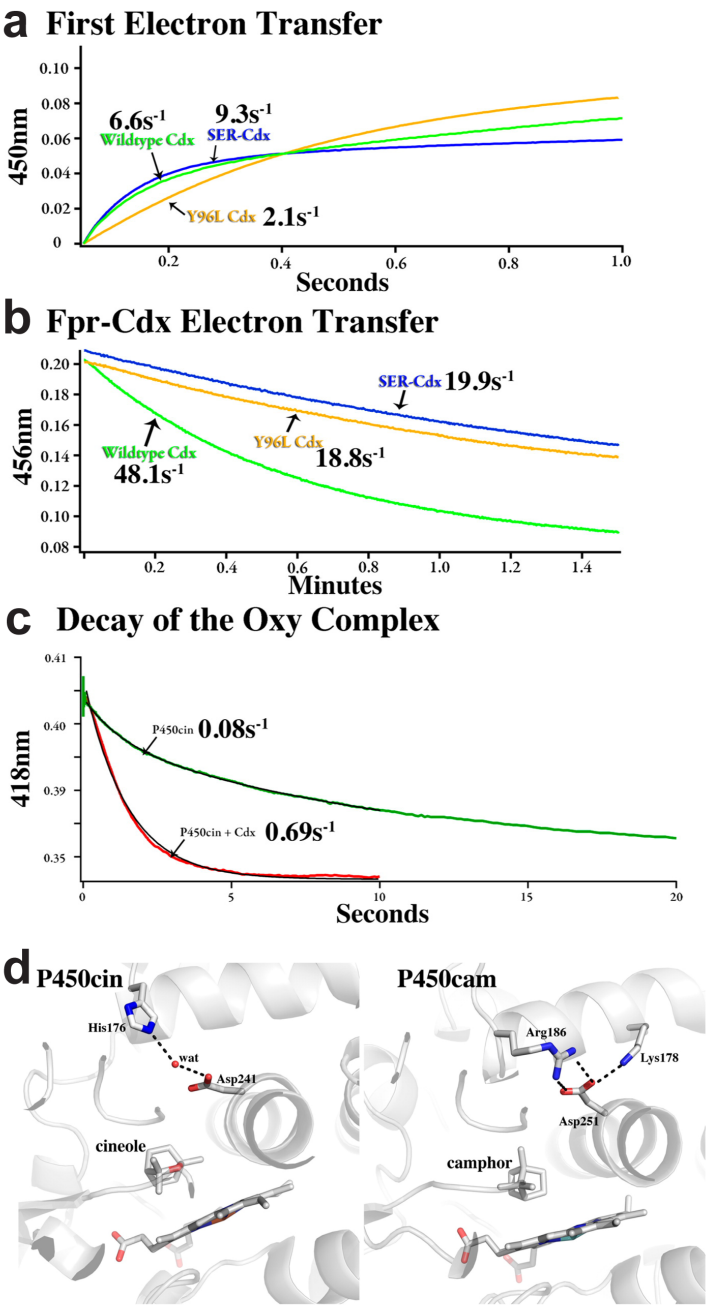

Figure 2.4: P450cin-Cindoxin Kinetics Assays. (a) Reduction rates of P450cin by WT and point mutants of Cdx. (b) Electron transfer rates from 0.125 μM flavodoxin reductase (Fldr) and 20 μM Cdx. (c) Decay rates of oxy-P450cin. (d) Cartoon representations of the P450cin and P450cam substrate-binding portals, with conserved residues labeled and shown in a licorice representation. Figure adapted from Fig. 4 and Fig. 5 of Ref. [1].

encounter complex. Cdx-EEEA delivers the first electron at the same rate as WT Cdx. However, the observed reduction of the steady state oxy-P450cin reduction indicates that the second electron transfer step is greatly hindered by the EEEA mutation.

The small divergence in the secondary structure of the P450cam and P450cin active sites (see Fig. 2.4d) may explain why P450cam requires a larger contribution from its redox partner. Earlier studies of P450cam suggest that D251 (see Fig. 2.4d) plays a critical role in the effector function [24]. We show that the corresponding aspartate residue on P450cin is also important for this function. It has been proposed that this residue plays an important role in delivering protons from solvent to iron-bound molecular oxygen [25], which is important to the formation of the active Compound 1. In the closed form of P450cam, D251 forms salt bridge contacts with K178 and R189, preventing it from shuttling protons to the active site (see Fig. 2.4d). The binding of Pdx is able to break this salt bridge. P450cin is remarkably different, in that the corresponding D241 residue is not paired with any side chains of basic residues, instead forming a hydrogen-bonded network with H176 (see Fig. 2.4d). As a result, proton shuttling can occur in both open and closed forms of P450cin, which may explain why it is not as dependent on its redox partner as P450cam.

2.5 Bibliography

1. Madrona, Y., Hollingsworth, S. A., Tripathi, S., Fields, J. B., Rwigema, J. C., Tobias, D. J., and Poulos, T. L. (2014) Crystal structure of cindoxin, the P450cin redox partner. *Biochemistry* **53**, 1435-1446

2. Hannemann, F., Bichet, A., Ewen, K. M., and Bernhardt, R. (2007) Cytochrome P450 systems--biological variations of electron transport chains. *Biochim Biophys Acta* **1770**, 330-344

3. Hawkes, D. B., Adams, G. W., Burlingame, A. L., Ortiz de Montellano, P. R., and De Voss, J. J. (2002) Cytochrome P450(cin) (CYP176A), isolation, expression, and characterization. *J Biol Chem* **277**, 27725-27732

4. Zhao, Q., Modi, S., Smith, G., Paine, M., McDonagh, P. D., Wolf, C. R., Tew, D., Lian, L. Y., Roberts, G. C., and Driessen, H. P. (1999) Crystal structure of the FMN-binding domain of human cytochrome P450 reductase at 1.93 A resolution. *Protein Sci* **8**, 298-306

5. Rylott, E. L., Jackson, R. G., Edwards, J., Womack, G. L., Seth-Smith, H. M., Rathbone, D. A., Strand, S. E., and Bruce, N. C. (2006) An explosive-degrading cytochrome P450 activity and its targeted application for the phytoremediation of RDX. *Nat Biotechnol* **24**, 216-219

6. Sevrioukova, I. F., Li, H., Zhang, H., Peterson, J. A., and Poulos, T. L. (1999) Structure of a cytochrome P450-redox partner electron-transfer complex. *Proc Natl Acad Sci U S A* **96**, 1863-1868

7. Strushkevich, N., MacKenzie, F., Cherkesova, T., Grabovec, I., Usanov, S., and Park, H. W. (2011) Structural basis for pregnenolone biosynthesis by the mitochondrial monooxygenase system. *Proc Natl Acad Sci U S A* **108**, 10139-10143

8. Tripathi, S., Li, H., and Poulos, T. L. (2013) Structural basis for effector control and redox partner recognition in cytochrome P450. *Science* **340**, 1227-1230

9. Hiruma, Y., Hass, M. A., Kikui, Y., Liu, W. M., Olmez, B., Skinner, S. P., Blok, A., Kloosterman, A., Koteishi, H., Lohr, F., Schwalbe, H., Nojiri, M., and Ubbink, M. (2013) The structure of the cytochrome p450cam-putidaredoxin complex determined by paramagnetic NMR spectroscopy and crystallography. *J Mol Biol* **425**, 4353-4365

10. Motiejunas, D., Gabdoulline, R., Wang, T., Feldman-Salit, A., Johann, T., Winn, P. J., and Wade, R. C. (2008) Protein-protein docking by simulating the process of association subject to biochemical constraints. *Proteins* **71**, 1955-1969

11. Ermak, D. L., and Mccammon, J. A. (1978) Brownian Dynamics with Hydrodynamic Interactions. *J Chem Phy* **69**, 1352-1360

12. Gabdoulline, R. R., and Wade, R. C. (1998) Brownian dynamics simulation of protein-protein diffusional encounter. *Methods-a Method Enzymol* **14**, 329-341

13. Ortega, A., Amoros, D., and Garcia de la Torre, J. (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys J* **101**, 892-898

14. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802

15. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J Comp Phys* **23**, 327-341

16. Miyamoto, S., and Kollman, P. A. (1992) Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J Comp Chem* **13**, 952-962

17. Grubmuller, H., Heller, H., Windemuth, A., and Schulten, K. (1991) Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-Range Interactions. *Mol Sim* **6**, 121-142

18. Sciortino, F., Essmann, U., Stanley, H. E., Hemmati, M., Shao, J., Wolf, G. H., and Angell, C. A. (1995) Crystal stability limits at positive and negative pressures, and crystal-to-glass transitions. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* **52**, 6484-6491

19. Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994) Constant-Pressure Molecular-Dynamics Algorithms. *J Chem Phys* **101**, 4177-4189

20. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102**, 3586-3616

21. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **79**, 926-935

22. Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., and MacKerell, A. D. (2010) CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J Comp Chem* **31**, 671-690

23. Goldschmidt, L., Cooper, D. R., Derewenda, Z. S., and Eisenberg, D. (2007) Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Prot Sci* **16**, 1569-1576

24. Stok, J. E., Yamada, S., Farlow, A. J., Slessor, K. E., and De Voss, J. J. (2013) Cytochrome P450(cin) (CYP176A1) D241N: Investigating the role of the conserved acid in the active site of cytochrome P450s. *BBA-Proteins Proteom* **1834**, 688-696

25. Gerber, N. C., and Sligar, S. G. (1994) A Role for Asp-251 in Cytochrome P-450cam Oxygen Activation. *J Biol Chem* **269**, 4260-4266

Chapter 3

Leishmania Major Peroxidase and Cytochrome C

The research in this chapter was originally published in Biochemistry [1]. I performed the Brownian dynamics simulations described in the Brownian dynamics docking methods section (section 3.2.1). I assisted Scott Hollingsworth (Department of Molecular Biology and Biochemistry, UC Irvine) in running and analyzing the molecular dynamics refinement of the Brownian dynamics complex structure. The biochemical experiments, including the solving and assignment of the Cindoxin crystal structure, were carried out by Yarrow Madrona (Department of Molecular Biology and Biochemistry, UC Irvine), Sarvind Tripathi (Department of Molecular Biology and Biochemistry, UC Irvine), Scott Hollingsworth, and Jean-Christophe Rwigema (Department of Molecular Biology and Biochemistry, UC Irvine) in the laboratory of Thomas Poulos (Chancellor's Professor of Biochemistry, Chemistry, and Pharmaceutical Sciences, UC Irvine).

3.1 Introduction

In Ch. 2, I introduced the cytochrome P450 class of heme-containing enzymes and their importance to drug metabolism. This omnipresent class of enzymes exists in almost all organisms, and has thus become a hot drug target for the treatment of certain illnesses in addition to the potentiation of drugs. In this chapter, we will discuss enzymes involved in leishmaniasis, a tropical disease transmitted by the sand fly, and caused by parasitic protozoa from the genus *Leishmania*. The *Leishmania major* parasite is able to survive host macrophages and the oxidative environment they create by scavenging reactive oxidative species (ROS) such as hydrogen peroxide [2]. *Leishmania major* is able to scavange ROS by secreting a mitochondrial heme peroxidase (*L. major* peroxidase, LmP) and a mitochondrial cytochrome *c* (*L. major* cytochrome *c*, LmCytc), which form an interprotein electron transfer complex [3, 4].

In order for electron transfer (ET) to take place, the redox partners need to interact specifically and efficiently. The redox partners need to contain a high degree of specificity through hydrophobic and electrostatic interactions, while at the same time, they need to associate quickly, transfer an electron or electrons, and dissociate equally quickly in order to precipitously turnover ROS. In this chapter, we will focus on the first step in this process, specifically, the formation of the initial encounter complex. A pertinent question is whether the initial encounter complex is ET-competent, or does the initial encounter take place away from the active site, followed by local sampling to create the ET-active complex? *Saccharomyces cerevisiae* cytochrome *c* peroxidase (yeast CCP) has long served as a standard for studying protein–protein interactions and

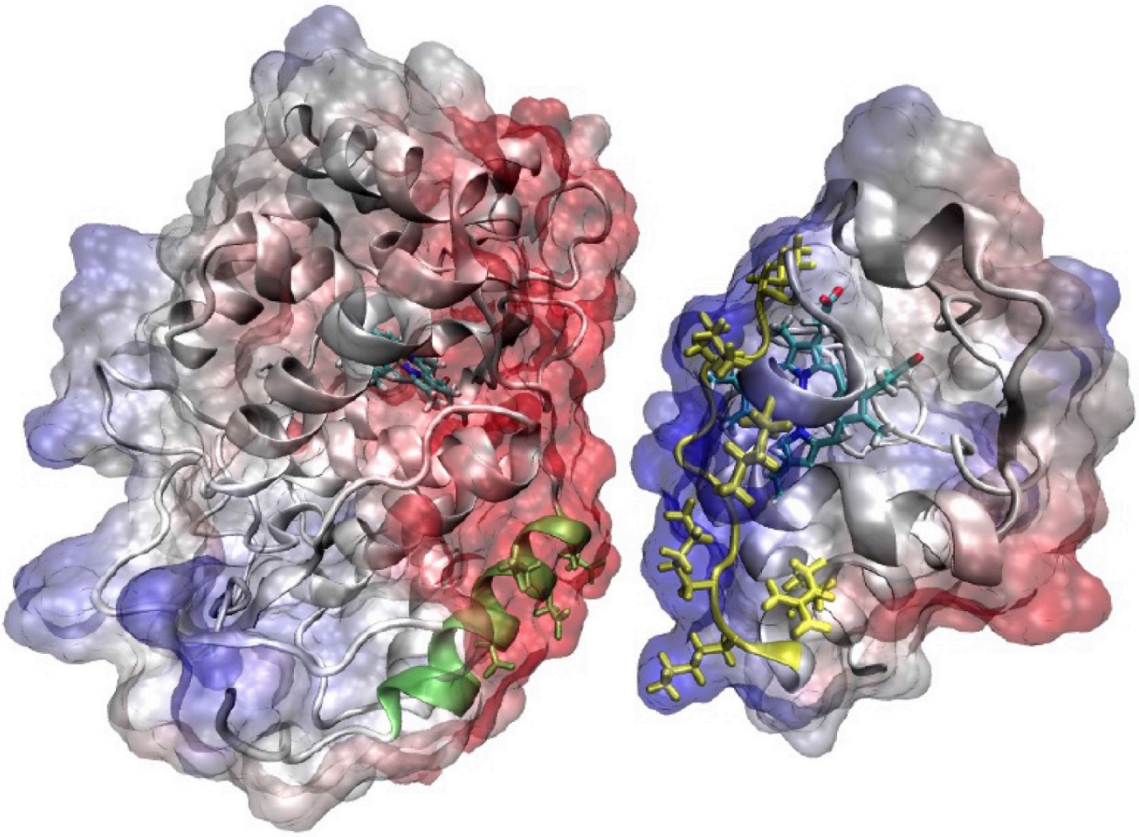

Figure 3.1: The LmP-LmCytc Complex Crystal Structure. The crystal structure was solved by Jasion et al. Ref. [4] and is shown where LmP (left) and LmCytc (right) have been separated to highlight the complimentary electrostatic surfaces. The backbone of each protein is shown in a cartoon representation and electrostatic potentials are displayed on transparent solvent-accessible surfaces where negative potential is red and positive is blue. The backbone of helix A of LmP is colored green. Negatively-charged side chains on helix A of LmP are shown in a green licorice representation and positively-charged side chains of LmCytc at the binding interface are shown in yellow licorice. Figure adapted from Fig. S4 of Ref. [1].

the mechanism of interprotein ET [5]. LmP, which also uses a cytochrome *c* as its substrate [4], possesses a similar ET mechanism to yeast CCP [6-8]. The enzyme is first oxidized by hydrogen peroxide to give the so-called Compound I, which consists of a ferryl iron and Trp radical, Fe^{IV}=O;$Trp^{+\bullet}$. The Trp radical is reduced by cytochrome *c* to give Compound II, Fe^{IV}=O;Trp. This is followed by an intramolecular equilibrium between the iron and Trp to give Fe^{III}–OH;$Trp^{\bullet+}$. Finally, a second molecule of cytochrome *c* reduces the Compound II $Trp^{\bullet+}$ radical to give the resting enzyme.

Despite the fact that the redox chemistry of yeast CCP and LmP is roughly the same, there is strong evidence that the association mechanisms differ significantly [4]. Firstly, LmP obeys Michaelis–Menten kinetics [4, 9], while yeast CCP does not [10]. Secondly, while K_M steadily increases with an increasing ionic strength, k_{cat} is effectively independent of ionic strength in LmP [7]. Yeast CCP is quite different with a k_{cat} steadily increasing with ionic strength until a maximum near 150 mM ionic strength [10]. This has been attributed to a change in the rate-limiting step at low ionic strengths from product dissociation to intramolecular ET at high ionic strengths. The rate-limiting step for LmP remains intramolecular ET at all ionic strengths [8]. We hypothesis that LmP's dependence on ionic strength is due to the electrostatic interactions between LmP and its redox partner, whereas yeast CCP relies more heavily on hydrophobic interactions. This hypothesis is supported by the crystal structures of the respective ET complexes. In the LmP complex, LmP D211 and LmCytc R24 form a salt bridge at the center of the binding interface and LmP E49 forms an ion pair with LmCytc K98 at the periphery of the interface [9]. In contrast, there are close contacts between nonpolar side chains

(e.g., yeast CCP A193 and Cytc V28) but no salt-bridges in the yeast CCP–Cytc interface [11]. Calorimetric studies show that the entropy of association is −2.1 kcal mol^{-1} while the entropy term is −6.0 kcal mol^{-1} [12]. Although the formation of the initial encounter complex may be driven by complementary electrostatic surfaces (see Fig. 3.1), the predominance of the entropic term in the association free energy indicates that the stability of the yeast CCP–Cytc complex is dominated primarily by nonpolar interactions.

In this chapter, I employed Brownian dynamics (BD) simulations to investigate the mechanism of LmP–LmCytc complex formation. Determining the initial encounter complex can be difficult experimentally because many of these reactions are nearly diffusion-controlled. BD simulations have been employed extensively to model the formation of many ET protein complexes, thus providing molecular level details about the initial association reaction that are often experimentally inaccessible [13-23]. In the case of the LmP–LmCytc system, comparison of the individual and co-crystal structures reveals no major conformational changes and only minimal reorientations of side chains at the interface upon complex formation. Thus, rigid-body BD simulations should be capable of accurately describing the association of LmP and LmCytc. In this chapter, I used BD simulations to compare the binding interaction of both the co-crystal and individual structures to identify the importance of specific residues to the formation of the complex. We validated our BD simulations by comparing computed second-order association rates (k_a) with experimentally measured association rate constants determined from single-turnover experiments (k_1). Finally, we conducted BD simulations of the formation of the yeast CCP–Cytc complex to gain insights into the differences in

the mechanisms of association of LmP and yeast CCP with their respective cytochrome *c* redox partners.

3.2 Methods

3.2.1 Brownian Dynamics Simulations

Simulation Details

The Brownian dynamics method that we employ models proteins as atomically-detailed rigid bodies and generates translational and rotational displacements using the Ermak–McCammon algorithm [24]. The SDA program, versions 6 and 7, was used to conduct the BD simulations [17, 19]. SDA version 6 was used to compute the probability distributions, and version 7 was used to calculate the second-order association rate constants (k_a; both methods are detailed below). The bimolecular BD simulation method that we employed is detailed further in appendix A.

The electrostatic interactions were treated in the BD simulations by using a well-established approach based on the effective charge approximation [25]. The electrostatic potentials were obtained by numerical solution of the finite-difference Poisson–Boltzmann (PB) equation, taking into account the inhomogeneous dielectric medium and the surrounding electrolyte solution, using the APBS 1.4 program package [26]. We utilized a manual multigrid calculation with a single Debye–Hückel boundary condition for each solute on a 200 × 200 × 200 grid of 1 Å spacing. The CHARMM22 force field was used to assign atomic charge and radius parameters for all atoms of the protein and heme groups [27, 28]. The dielectric boundary was defined by the van der

Waals surface of the protein. Harmonic dielectric boundary smoothing and charge-antialiasing were employed as implemented in APBS 1.4 [29]. The solute and protein dielectric constants were set to 78 and 1, respectively, at 298.15 K for all electrostatic potential grid calculations. Ionic strengths of 50 and 150 mM NaCl were used for the LmP–LmCytc complex to match experimental conditions, and 50, 100, and 300 mM NaCl were used for the yeast CCP–Cytc complex to compare probability distributions with those of the LmP–LmCytc complex and to compare to rates experimentally measured and computed from BD simulations. Ionic charges of +1 and −1 and radii of 1.1 and 1.7 Å were used to represent Na^{+} and Cl^{-}, respectively, and the solvent radius was set to 1.4 Å. Charges at effective charge sites on basic and acidic side chains were fit so that, in a shell around the protein, the electrostatic potential that was precomputed with an inhomogeneous dielectric is reproduced with a homogeneous dielectric [26]. The exclusion probe radius and skin thickness were set to 4 Å and 3 Å, respectively. For the sake of consistency, we used the same ionic strengths and solvent dielectric constant that were used for the computed electrostatic potential.

The electrostatic desolvation energy grids (200 × 200 × 200 with a 1 Å spacing) were computed using equation 2 from Ref. [30] with scaling factor α set to 1.67. Ionic strengths were varied to be consistent with the values used to compute the electrostatic potential grids in APBS. A single ionic radius of 1.5 Å was used. The nonpolar desolvation energy grids (200 × 200 × 200 with a 1 Å spacing) were computed (using Eq. 3 from Ref. [30]) with distance parameters *a* and *b* set to 3.10 and 4.35 Å, respectively, and normalization constants *c* and β set to 0.5 and −0.0065 kcal mol^{-1} $Å^{-2}$, respectively. The exclusion grids were calculated using a probe radius of 1.77 Å [20].

In our BD simulations of the formation of ET complexes, the larger of the two proteins (LmP or yeast CCP) is held fixed and the other (LmCytc or yeast Cytc) is allowed to move. The larger protein was placed at the center of a sphere with a radius of 300 Å, while the smaller protein was allowed to diffuse from a random starting position on the surface of the sphere. A single trajectory can result in either a reaction event (defined below) or an escape event, in which the center-of-mass to center-of-mass displacement between the two proteins reaches 400 Å. The integration time step was linearly decreased from 20 to 1 ps over a protein surface–surface distance interval of 90 to 60 Å to allow the simulation to capture the finer details of the protein dynamics at close proximity, while maintaining efficient sampling at large separation distances. The diffusion coefficients that appear in the Ermak–McCammon algorithm were calculated using Hydropro 10 [31]. The translational and rotational diffusion coefficients used in our BD simulations were 8.634×10^{-3} $Å^2$ ps^{-1} and 1.021×10^{-5} rad^2 ps^{-1}, respectively, for LmP, 1.161×10^{-2} $Å^2$ ps^{-1} and 2.439×10^{-5} rad^2 ps^{-1}, respectively, for LmCytc, 8.264×10^{-3} $Å^2$ ps^{-1} and 8.734×10^{-6} rad^2 ps^{-1}, respectively, for yeast CCP, and 1.177×10^{-2} $Å^2$ ps^{-1} and 2.482×10^{-5} rad^2 ps^{-1}, respectively, for yeast Cytc.

Two separate sets of structures were used to initiate BD simulations of the LmP–LmCytc complex: (i) the individual structures of each protein determined separately (PDB code: 3RIV, LmP from Ref. [32]; PDB code: 4DY9, LmCytc from Ref. [4]) and (ii) the structure of each protein in the co-crystal structure of the complex (PDB code: 4GED, LmCytc from Ref. [9]). There are no major conformational differences between the two sets of LmP and LmCytc structures, but rather a series of small rearrangements of side chains at the binding interface, with the largest change being the movement of

R24 of LmCytc into what is believed to be the active conformation for ET. The same BD simulation protocols were applied to both sets of initial structures to investigate the roles that subtle rearrangements of the binding interface play in complex formation. In addition to WT LmP, we considered the following mutants in which negatively charged side chains were changed to neutral side chains that preserve the side chain orientation: D211N, D47N, E49Q, D50N, E54Q, and D47N/D50N/E54Q. We also considered the D47A/D50A/E54A mutant because it provided an additional opportunity to compare computed and experimentally determined association rates. The *in silico* mutations were introduced using the psfgen plugin of VMD starting with the WT LmP structure and manually truncating uncommon atoms of the side chain to preserve the orientation of the mutated residue. This was done for the D to N and E to Q mutations by truncating the terminal oxygen (OD2), while the corresponding D to A and E to A mutations required truncation to the β-carbon (CB). A total of 200000 BD trajectories were generated at 298.15 K for each of the two sets of starting structures for each LmP variant. To investigate the consequences on complex formation by differences in the sequences of cytochrome *c* peroxidases from *L. major* and yeast, we also generated 200000 BD trajectories for the yeast CCP–Cytc system, starting from the protein structures in the ET complex (PDB entry: 2PCC, from Ref. [11]).

Calculation of Second-Order Association Rates

BD association rates are calculated using a method developed by Northrup *et al* [33] according to which the diffusion-controlled reaction rate is given by:

$$k = k_D(b)p \tag{3.1}$$

where p is the probability that the pair of proteins, starting at separation $R = b$, will associate and react (i.e., not diffuse apart, ultimately to infinite separation) and $k_D(b)$ is the rate constant for pairs with $R > b$ to reach $R = b$, regardless of whether they ultimately react. In practice, p is not calculated directly from BD trajectories. Rather, trajectories are terminated when R exceeds a cutoff distance q, with $q > b$, and the following expression for the rate constant, which accounts for the truncation of trajectories that lead to $R > q$, is used:

$$k = \frac{k_D(b)\beta}{1 - (1 - \beta)\Omega} \tag{3.2}$$

where b is the fraction of trajectories in which the complex forms before the proteins diffuse to $R > q$ and $W = k_D(b)/k_D(q)$. This equation is based on the assumption that every trajectory in which an active site collision occurs, as defined by a specified criterion [typically, a set of interprotein distances (see below)], produces a reaction. For fast reactions such as ET, this assumption is justified. In practice, b and q are chosen to be sufficiently large that the potential of mean force (PMF) for the interaction between the two proteins is centrosymmetric, so that $k_D(r)$ $(r = q \text{ or } b)$ may be calculated using the following analytical expression:

$$k_D(r) = \left[\int_r^\infty dR \frac{e^{U(R)/k_BT}}{4\pi R^2 D(R)}\right]^{-1} \tag{3.3}$$

where $U(R)$ is the PMF and $D(R)$ is the relative diffusion constant of the two proteins. In

this chapter, I used a q of 400 Å and a b of 300 Å.

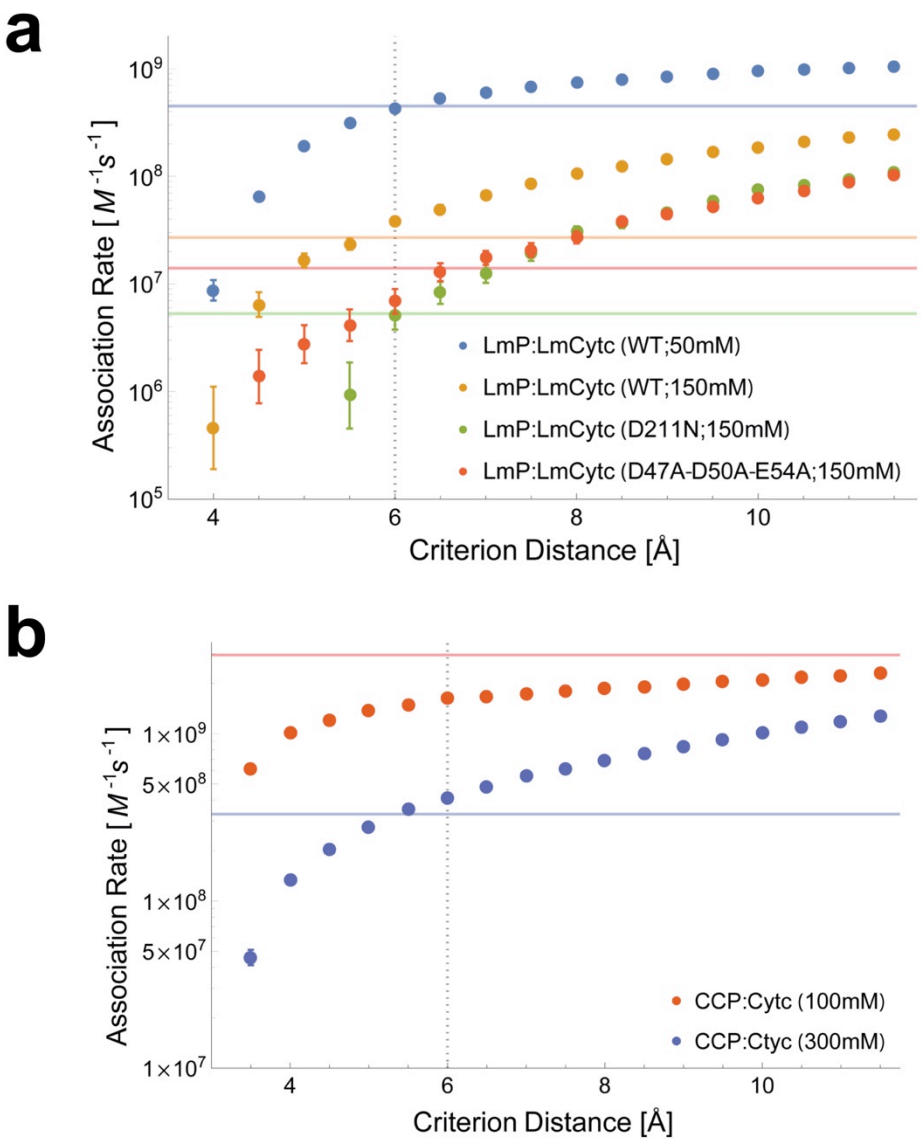

Figure 3.2: Determining the Reaction Criterion for Association Rate Calculations. a) Calculated second order association rate constants for LmP:LmCytc from 200,000 BD trajectories, plotting the occurrence of two crystal contacts as a function of criterion distance. The solid colored lines indicate the experimentally determined association rate constants (k_1) for LmP:LmCytc for the different conditions/mutations indicated in the legend. The dotted vertical line at 6 Å shows the criterion distance that produces the best overall agreement with the experimentally determined rates. b) Calculated second order association rate constants for yeast CCP:Cytc at 100 mM and 300 mM ionic strength, respectively. The dotted vertical line indicates the criterion distance used in the calculation of the second order association rate from our BD trajectories, which is the same distance (6 Å) used for the LmP:LmCytc mutants from (a). Figure adapted from Fig. S1 and Fig. S2 of Ref. [1].

Error estimates on the computed second-order association rates were obtained by using the bootstrapping analysis implemented in SDA 7 [17, 34]. Monitored were those between OE1 and OE2 of E49 and NZ of K98 and between OD1 and OD2 of D211 and NE and NH1 of R24. An association reaction was defined by any configuration where the distances in both pairs of side chains were shorter than a criterion distance, d_{max}. The particular value of d_{max} used, 6 Å, produced second-order association rates that are in good agreement with experimental data on wild-type LmP and its D211N and D47A/D50A/E54A variants (see Fig. 3.2a).

Because the yeast CCP–Cytc complex does not contain interprotein salt bridges, we used the nda-pairs function in SDA version 7 to produce a list of 10 independent, nonspecific contacts (donor–acceptor atom pairs between the two proteins in the complex structure with a maximal distance of 4.5 Å between each pair) to define the association of yeast CCP and Cytc. An association reaction was considered to occur when two atom pairs achieve simultaneously a separation distance that is shorter than d_{max}. The value of d_{max} that we used to define association of yeast CCP and Cytc, 6 Å, provided good agreement with experimental association rates measured at ionic strengths of 100 and 300 mM (see Fig. 3.2b).

Probability Distributions

Probability histograms of the position of the center-of-mass of the mobile protein were produced from the BD trajectories and displayed as three-dimensional isosurfaces (contours), superimposed on a display of the fixed protein. We used a voxel width of 1 Å

on a cubic grid of 200 × 200 × 200 voxels and a sampling sphere radius of 60 Å to capture configurations generated with a time step of 1 picosecond. Probabilities within each voxel were calculated by dividing the number of trajectory points within each voxel by the total number of trajectory points within the sampling sphere. The resulting probability distributions include orientational information implicitly because sharp features correspond to consistently oriented complexes. All of the probability isosurfaces we report are contoured at the same value of 2.2×10^{-3} such that the isosurfaces from different systems may be readily compared to one another.

3.2.2 Molecular Dynamics Details

I helped conduct an atomistic dynamics (MD) simulation (see additional details in appendix B) of the initial encounter complex of the individual structure BD trajectories. We selected a configuration from a BD simulation in which LmCytc was docked to helix A of LmP, and we moved LmCytc approximately 15 Å away from LmP prior to initiating the MD simulation. The individual crystal structures of LmP were used as the initial configurations of the two proteins in the MD simulation. Hydrogen atoms were added to both the protein and the crystallographic water molecules using the psfgen plugin of VMD version 1.9.1 [34]. Patches were used to connect the heme with the coordinating histidine residues in both structures, and an extra bond was added to describe the Met-Fe coordination in LmCytc. The Ca^{2+} and K^{+} metal ions present in each of the crystal structures were modeled by adding extra bonds between the coordinating atoms and the metal ion to preserve the coordination site. The solvent box was built using a 15-Å cushion in all three directions around the proteins. The resulting system contained

219,540 atoms.

The MD simulation was performed using NAMD, version 2.9 [35]. The CHARMM22 force field was used for the proteins and the TIP3P model for water [27, 36]. The system was subjected to 1000 steps of conjugate gradient energy minimization prior to MD simulation. The simulation was conducted for 100 ns at a constant temperature (300 K) and pressure (1 atm) using Langevin dynamics for temperature control and a Nosé–Hoover–Langevin piston pressure control [37, 38]. A multiple-time step algorithm was used to integrate the equations of motion with time steps of 4 fs for the electrostatic interactions and 2 fs for the bonded and short-range nonbonded interactions [39]. The smooth particle mesh Ewald algorithm was used to treat the electrostatic interactions, and the real-space part of the Ewald sum and the Lennard-Jones interactions were smoothly switched off between 10 and 12 Å [40]. The SHAKE algorithm was used to constrain the lengths of bonds to hydrogen atoms [41].

3.2.3 Experimental Methods Details

The experimental methods described below were performed in the laboratory of Thomas L. Poulos. The results of these experiments are discussed in sections 3.3 and 3.4 in the context of my simulation data.

Steady State Activity Assays

In order to validate our observations from the BD trajectories, spectrophotometric steady state activity assays monitoring the oxidation of reduced LmCytc were performed

at 25 °C on a Cary 3E UV–visible spectrophotometer (Varian/Agilent). Reduction of LmCytc was performed by the addition of granules of sodium dithionite followed by an incubation period of 30 minutes on ice. Excess remaining sodium dithionite was then removed by passing the reduced sample through an Econopac 10DG desalting column (Bio-Rad) previously equilibrated with 50 mM potassium phosphate (pH 6.5). The eluate was then concentrated in a 10000 molecular weight cutoff centricon at 4 °C. The concentration of reduced LmCytc was determined using the previously determined molar extinction coefficient ($\Delta\varepsilon_{558}$) of 29 mM^{-1} cm^{-1} [4]. The assay spanned 2 h, during which time negligible amounts of LmCytc spontaneously oxidized. The LmP concentration was determined using the Soret molar extinction coefficient ($\Delta\varepsilon_{408}$ = 113.6 mM^{-1} cm^{-1}) [32]. The hydrogen peroxide concentration was also measured using the molar extinction coefficient ($\Delta\varepsilon_{240}$ = 0.0436 mM^{-1} cm^{-1}).

In each experiment, the reaction buffer consisted of 25 mM potassium phosphate (pH 6.5) and the ionic strength was increased to 150 mM via the addition of KCl. For each experiment, the reaction was initiated by adding 0.18 mM hydrogen peroxide and the absorbance change was monitored at 558 nm for 1 min. The turnover number was calculated using the previously determined $\Delta\varepsilon_{558}$ of 19.4 mM^{-1} cm^{-1} [4]. The data were analyzed by fitting to the Michaelis–Menten equation:

$$v = \frac{V_{max}[\mathrm{LmCytc}]}{K_M + [\mathrm{LmCytc}]} \tag{3.4}$$

Stopped-Flow Kinetics

Stopped-flow kinetic measurements were performed as described previously [8]. Compound I reduction was second-order, and the bimolecular rate constant (k_1) was calculated by fitting the kinetic traces using the software provided with the stopped-flow instrument (Applied Photophysics) according to the following standard single-exponential equation:

$$A_{420} = Ce^{-k_{abs}t} + b \quad (3.5)$$

where C is the amplitude term, k_{abs} is the observed rate constant for the decay of oxidizing LmCytc, and b is an offset value. The association rate constant k_1 was determined as the slope of a fit of the linear concentration dependence of k_{abs}.

Crystallography

The LmP D211N protein sample was prepared for crystallization as previously described [8]. Crystals were grown at room temperature in 10% (w/v) PEG 5000 MME, 0.1 M MES-NaOH (pH 6.5), and 7.5 mM praseodymium (III) acetate hydrate in a hanging drop vapor diffusion setup. Freshly grown crystals were harvested after 24 h and passed stepwise through a cryoprotectant solution containing 30% (v/v) glycerol for 4 h at 4 °C. Crystals were then flash-cooled with liquid nitrogen. Cryogenic (100 K) X-ray diffraction data were collected remotely at the Advanced Light Source (ALS) facility, using the data collection control software Blu-Ice, and a crystal-mounting robot [42]. An ADSC Q315r detector at beamline 12.3.1 was used for data collection. Raw data frames

were indexed, integrated, and scaled using XDS [43]. The LmP D211N structure was determined by molecular replacement using phaser and the wild-type LmP structure (PDB code: 3RIV) [44]. The initial difference Fourier map was then calculated, and the model was refined using REFMAC [45]. Water molecules were also modeled using REFMAC and checked in COOT [46]. The TLS (translation–liberation–screw rotation model) protocol was implemented in the final stage of refinement with each chain as one TLS group, and the refined structure was validated in COOT [47]. Coordinates and structure factors were deposited in the RCSB Protein Data Bank.

3.3 Results

3.3.1 Brownian Dynamics Simulations Divulge the Importance of LmP Side Chain Orientations for LmCytc-Binding

A probability isosurface for the center-of-mass of LmCytc around wild-type LmP obtained from BD trajectories initiated from the co-crystal structures (see Fig. 3.3a). The most prominent feature in the isosurface coincides with the position of the center-of-mass of LmCytc in the co-crystal structure (see Fig. 3.3b). Thus, the BD simulations of the co-crystal structures recapitulate the binding of LmCytc to the LmP active site, as established in the crystal structure of the complex. In the probability isosurfaces (see Fig. 3.3a), there is a second, small feature corresponding to an interaction between LmCytc and helix A of LmP, which will be discussed in more detail below.

The second-order association rate computed from the BD trajectories of the WT co-crystal structures at an ionic strength of 50 mM NaCl, $(4.3 \pm 0.1) \times 10^8$ M^{-1} s^{-1} (see

Table 3.1), is in good agreement with measured association rate constant (k_1), 4.5×10^8 M^{-1} s^{-1}, determined by a single-turnover assay with 50 mM KCl in reaction buffer (see Table 3.1) [8]. While the k_{cat} for LmP is independent of ionic strength, the K_M increases substantially, thereby resulting in a decrease in k_{cat}/K_M with an increase in ionic strength [8]. Consistent with this observation, increasing the ionic strength from 50 to 150 mM in the BD simulations reduces the second-order association rate by 1 order of magnitude (see Table 3.1), in good agreement with the experimental data (see Table 3.2).

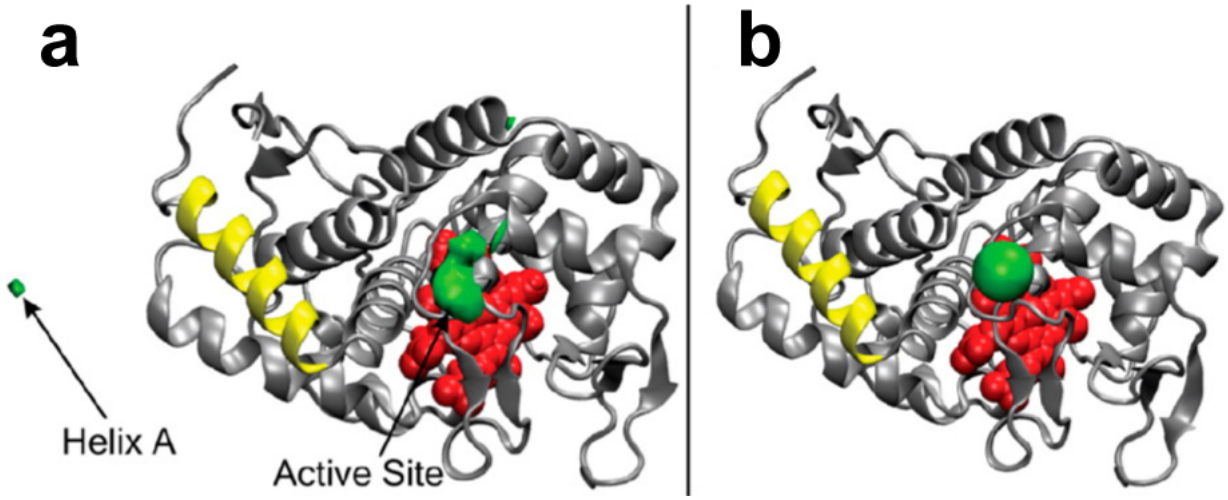

Figure 3.3: Co-Crystal BD Pair Probability Distribution. a) LmP in the LmP-LmCytc complex from BD simulations at 50 mM ionic strength. The backbone is drawn as a gray ribbon, except for helix A, which is drawn as a yellow ribbon. The heme is depicted with red van der Waals spheres. a) The green surface is the probability isosurface for the LmCytc center-of-mass calculated from BD simulations initiated from the co-crystal structures contoured at a value of 2.2×10^{-3}. At this contour level, there are two features in the isosurface, the larger of which corresponds to LmCytc binding to the active site of LmP, and a smaller feature corresponding to LmCytc interacting with helix A of LmP. a) The position of the center-of-mass of LmCytc in the co-crystal structure Ref. [3] is depicted as a green sphere. Figure adapted from Fig. 1 of Ref. [1].

To assess the role of electrostatic interactions, several charge-neutralizing mutations in LmP were considered. When D211, which is located at the active site binding interface, is mutated to N *in silico*, binding is disrupted, as is evident from the lack of a feature at the active site in the probability isosurface (see Fig. 3.4a). The D211N mutation results in a substantial reduction in both the BD association rate (see Table 3.1) and the experimental k_1 (see Table 3.2) at an ionic strength of 150 mM NaCl. The good agreement between the BD predicted and experimentally measured consequences of the D211N mutation provides further validation of the BD simulations.

LmP Variant[a]	Ionic Strength, mM NaCl	BD rate, $M^{-1}s^{-1}$
wild type	50	$4.3 \pm 0.1 \times 10^8$
	150	$3.8 \pm 0.4 \times 10^7$
D211N (AS)	50	$3.3 \pm 0.4 \times 10^7$
	150	$5.1 \pm 1.5 \times 10^6$
D47N (HA)	50	$2.1 \pm 0.1 \times 10^8$
	150	$4.6 \pm 1.4 \times 10^6$
E49Q (HA)	50	$1.3 \pm 0.1 \times 10^8$
	150	$2.8 \pm 1.0 \times 10^6$
D50N (HA)	50	$1.6 \pm 0.1 \times 10^8$
	150	$2.8 \pm 1.0 \times 10^6$
E54Q (HA)	50	$1.8 \pm 0.1 \times 10^8$
	150	$4.1 \pm 1.3 \times 10^6$
D47A-D50A-E54A (HA)	150	$6.9 \pm 1.8 \times 10^6$

Table 3.1: Second Order Association Rates of LmCytc and LmP from Brownian Dynamics. [a]AS denotes that the mutation is in the active site of LmP and [a]HA denotes that the mutation is in helix A.

The most significant interaction between LmCytc and D211N LmP occurs in the vicinity of helix A of LmP (see Fig. 3.4a). Helix A contains several negatively-charged residues that protrude away from the active site interface (see Fig. 3.1). Elimination of negatively-charged side chains on helix A in the D50N, D47N, E49Q, E54Q, and

D47N/D50N/E54Q variants also reduces the level of binding of LmCytc to the active site of LmP (see Fig. 3.4a and Fig. 3.6a). With the exception of E49Q, all of the charge-neutralizing D to N and E to Q mutations also reduce the level of interaction of LmCytc with helix A of LmP. The computed association rates were significantly lower than that of WT for all of the D to N and E to Q mutants of helix A (see Table 3.1).

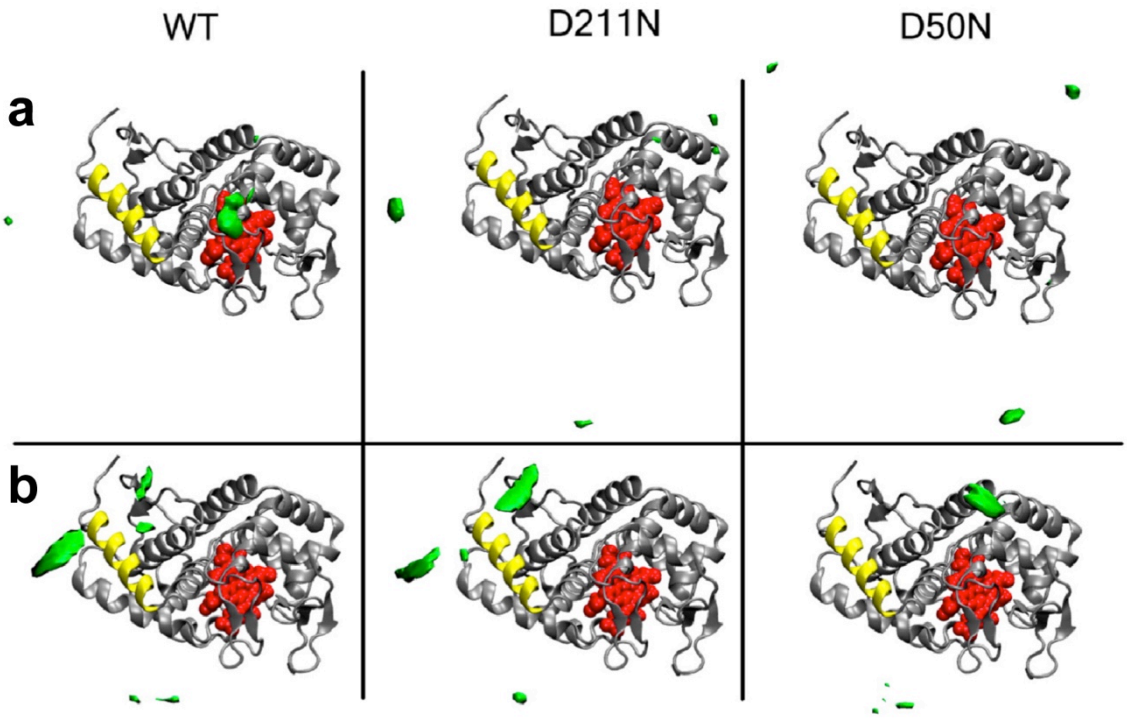

Figure 3.4: Individual Structure BD Pair Probability Distributions. Probability isosurfaces calculated with an isocontour value of 2.2 x 10^{-3} (shown in green) for LymCytc center-of-mass from BD simulations at an ionic strength of 50 mM NaCl, superimposed on the LmP crystal structure (same coloring as seen in Fig. 3.3). a) BD simulations initiated from the co-crystal structures. b) BD simulations initiated from the individual crystal structures. Figure adapted from Fig. 2 of Ref. [1].

The importance of the helix A charges was confirmed by steady state assays of the LmP D47A/D50A/E54A triple mutant: at an ionic strength of 150 mM, the k_1 for the triple mutant is ~19 times lower than that of WT LmP (see Table 3.2). The approximately six-fold reduction in the second-order association rate of the triple mutant versus that of WT computed from the BD trajectories (see Table 3.1) agrees qualitatively with the reduction observed experimentally.

Enzyme	KCl added, mM	k_{cat}, s^{-1}	K_M, μM	k_1, M^{-1} s^{-1}
wild type	0	409 ± 9[b]	4.1 ± 0.3[b]	
	50	531 ± 35[b]	13 ± 2.1[b]	4.5×10^8
	150	531 ± 74[b]	83 ± 17[b]	2.7×10^{7} [b]
D211N	0	34.5 ± 0.8	23 ± 1.3	
	150			5.3×10^6
D47A/D50A/E49A	0	372 ± 10	7 ± 0.6	
	150	162 ± 10	31 ± 8.6	1.4×10^7

Table 3.2: Kinetic Parameters of LymCytc Oxidation by LmP Obtained Through Steady State Assays.

3.3.2 Structure of the D211N Mutant of LmP

LmP shares a common feature with many other peroxidases in that it has a structurally important cation-binding site near the proximal side of the heme (see Fig. 3.5a). In LmP, this is either K^+ or Na^+ and the carbonyl oxygen atom of D211 provides one of the ligands. Our colleagues were concerned that the D211N mutation might cause structural perturbations around the cation site; the crystal structure of the D211N mutant was determined. The cation-binding site (see Fig. 3.5b) remains fully occupied and there are no significant structural perturbations. Therefore, any change in kinetics is due solely to the mutation and not to structural changes. The D211N structure used in the BD simulations, which was prepared by *in silico* mutation before the D211N crystal structure was determined, is essentially identical to the crystal structure (see Fig. 3.5a).

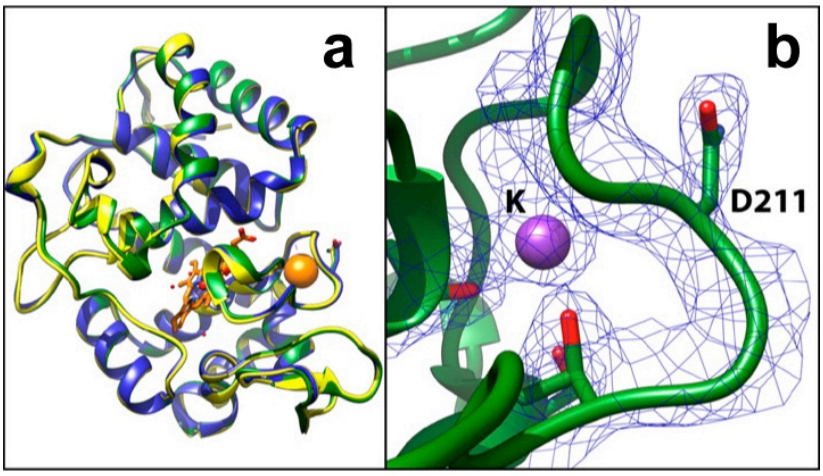

Figure 3.5 LmP-D211N Crystal Structure. a) Superposition of the LmP crystal structures for WT (blue) and D211N (green), as well as the LmP D211N structure that was prepared *in silico* for performed BD simulations (yellow). b) Potassium ion-binding site, with the $2F_o - F_c$ density map contoured at 1.0σ. Figure adapted from Fig. 3 of Ref. [1].

3.3.3 Brownian Dynamics Simulations Indicate the Importance of Helix A of LmP for LmCytc-Binding

In the co-crystal structures, the side chain conformations in the two proteins are optimized at the binding interface for forming a tight ET complex. In the individual structures, the conformations of the side chains at the binding interface are slightly different and are presumably more representative of the conformations of the proteins when they first encounter one another in solution, before the formation of the tight ET complex. Thus, BD simulations initiated from the individual structures permit us to sample configurations of the initial encounter complex and determine which interactions bring the two proteins together initially.

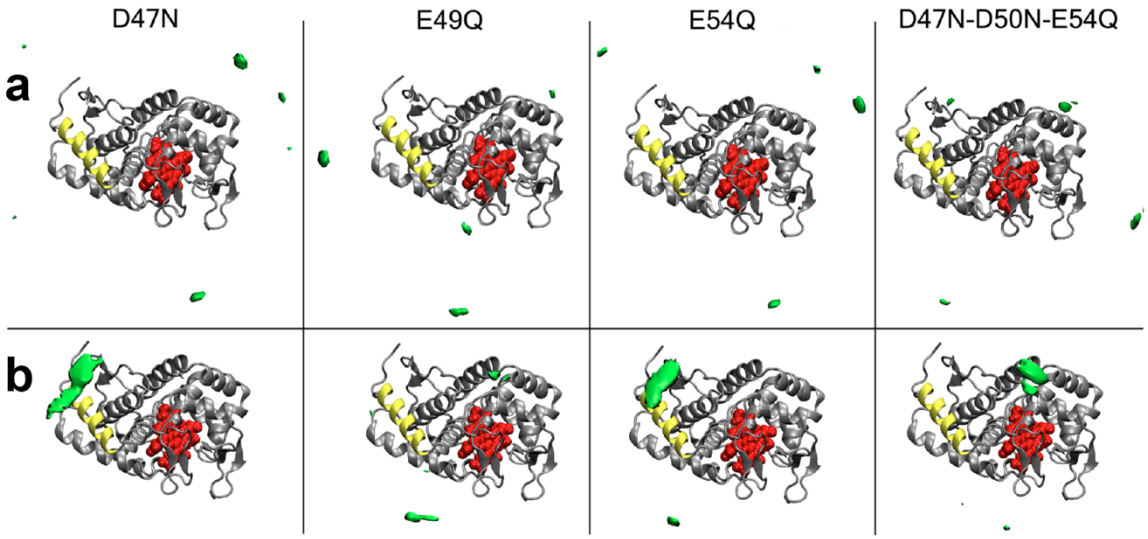

Figure 3.6: Co-Crystal BD Probability Distributions of LmP Mutants. Probability isosurfaces contoured at a value of 2.2×10^{-3} (green) for the LmCytc center-of-mass computed from BD simulations at 50 mM ionic strength superposed on the structure of LmP, drawn and colored as seen in Fig. 3.3 and Fig. 3.4, for four additional helix A mutants of LmP (D47N, E49Q, E54Q and D47N-D50N-E54Q). a) BD simulations initiated from the co-crystal structures; b) BD simulations initiated from the individual structures. Figure adapted from Fig. S5 of Ref. [1].

The probability isosurface obtained from BD trajectories of the individual WT structures show key differences (see Fig. 3.4a). In contrast to the case of the WT co-crystal structures (see Fig. 3.4a), there are no features in the isosurface near the LmP active site for the individual WT structures. This suggests that minor adjustments of the conformations of side chains at the binding interface are required for the formation of the tight ET complex. The tiny feature near the N-terminal end of helix A of LmP in the isosurface from the simulations of the co-crystal structures (see Fig. 3.4a) is significantly enhanced in the isosurface from the simulations of the individual structures (see Fig. 3.4b). The broadness of this feature in the case of the individual structures is indicative of local sampling of LmCytc around helix A of LmP. The existence of the feature near helix A suggests that interactions at the binding interface that are required for the formation of the final complex are not necessary to bring the proteins together initially. This hypothesis is supported by the persistence of the helix A feature in the isosurfaces obtained from both sets of simulations of the D211N mutant (see Fig. 3.4a), which do not display a feature in the isosurface near the active site in either set of simulations (co-crystal or individual structures).

The key role for negative charges on helix A suggested by the BD simulations of the co-crystal structures of mutants in which helix A charges were neutralized is reinforced by simulations of the corresponding mutants of the individual structures. The helix A feature is absent in the isosurfaces from the simulations of the D50N, E49Q, and D47N/D50N/E54Q variants initiated from the individual structures (see Fig. 3.3a and Fig. 3.6a). The helix A feature is present but shifted significantly away from its position in the isosurfaces of both the WT and active site D211N mutant in the D47N and E54Q

variants (see Fig 3.6b). Thus, consistent with the simulations initiated with the co-crystal structures, the simulations initiated from the individual structures predict that interactions of LmCytc with helix A of LmP are altered by elimination of negative charges on helix A.

3.3.4 Two-Step "Bind and Crawl" Mechanism for the Formation of the LmP–LmCytc Complex

The BD simulations initiated from the co-crystal structures exposed two binding sites for LmCytc on wild-type LmP, one at the active site and the other near helix A of LmP (see Fig. 3.3a). BD simulations initiated from the individual structures, in which the side chain conformations are not optimized for active site binding, show persistent interactions between LmCytc and helix A of WT LmP (see Fig. 3.4b).

The observation that LmCytc binds to both the active site and helix A of LmP raises the question of whether helix A competes with the active site for LmCytc or whether binding to helix A is a prelude to binding to the active site. Visual inspection of the BD trajectories reveals that LmCytc first binds to helix A and then travels down helix A toward the co-crystal interface, stopping at the end of the helix before jumping from the end of the helix to the interface where it begins local sampling of the active site region of LmP. A representative BD trajectory initiated from the individual structures shows that LmCytc comes within 3 Å center-of-mass distance of the co-crystal structure (see Fig. 3.7). By tracking the LmCytc center-of-mass in the BD trajectories that were initiated

from the co-crystal structures, we found that LmCytc binding to the active site was preceded by binding to helix A in ~70% of the reactive trajectories.

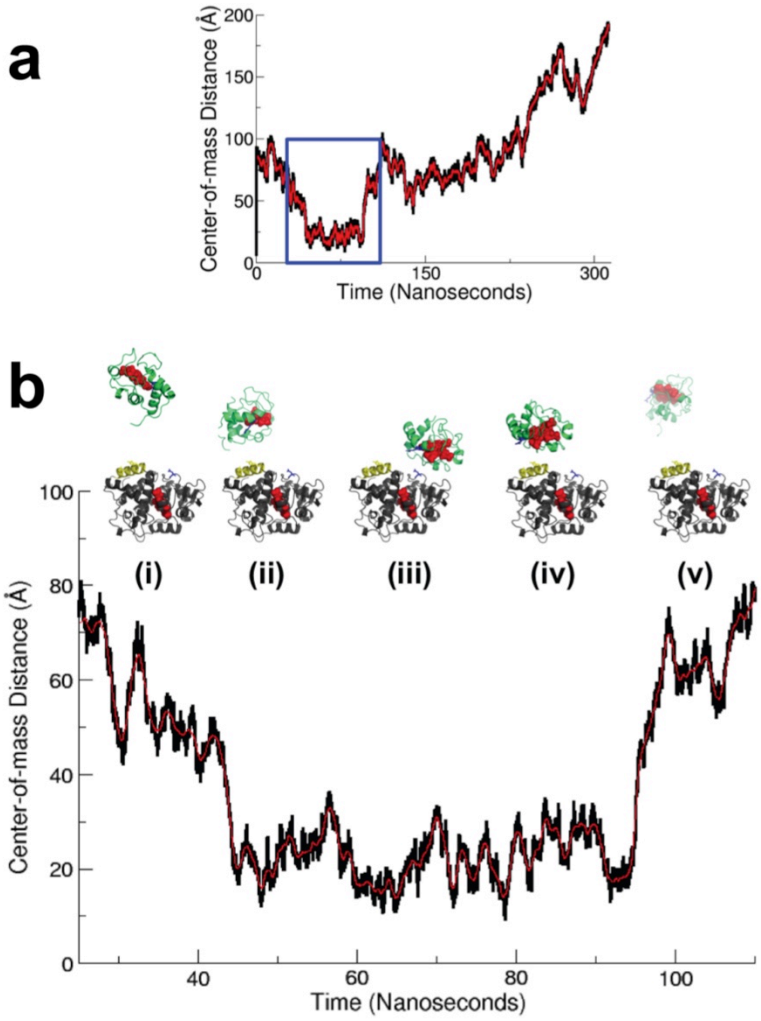

Figure 3.7: "Bind and Crawl" BD Trajectory. Representative Brownian dynamics trajectory, initiated from the individual crystal structures, showing the "bind and crawl" of LmCytc at helix A of LmP. a) Distance between the electron transfer residues of LmP D211(C_γ) and LmCytc R24(C_ζ). The blue box designates the region highlighted in panel (b). a) Zoom of the helix A bind and crawl association of LmCytc to LmP. Snapshots from the trajectory are shown above the plot to visualize the association and disassociation of LmCytc to LmP, where LmP is shown in black with helix A in yellow, LmCytc in green, respective heme cofactors in red, and the catalytically ET active residues LmP D211 and LmCytc R24 in blue sticks. Association can be broken down into five phases: (i) approach of LmCytc to LmP; (ii) binding of LmCytc to helix A of LmP; (iii) movement of LmCytc from helix A to the co-crystal active site of LmP; (iv) movement of LmCytc back from the active site of LmP to helix A; and (v) disassociation of the LmP-LmCytc complex. Figure adapted from Fig. S7 of Ref. [1].

The BD trajectories suggest that helix A plays two roles in the formation of the complex, one as the initial point of contact for LmCytc on LmP and the second as a runway to guide LmCytc to the ET interface. On the basis of the BD trajectories, we propose that the formation of the LmP–LmCytc ET complex proceeds, as suggested previously, via a two-step “bind and crawl” mechanism in which LmCytc first binds to helix A and subsequently migrates to the active site, with concomitant changes in the conformations of side chains at the binding interface allowing the formation of a tight complex [9].

To examine the importance of helix A in more detail, we performed an atomistic MD simulation. A configuration from a BD simulation of the wild-type proteins initiated from the individual structures, in which LmCytc is close to helix A of LmP, was selected as the starting point. The two proteins were then moved farther from each other along the separation vector connecting their geometric centers to the point where their respective surfaces were 15 Å apart. Upon initiation of the simulation, LmCytc returned rapidly (within 2 nanoseconds) to helix A near its location in the configuration selected from the BD simulation. Once bound to helix A, LmCytc began to travel down the helix toward its location in the co-crystal structure. The movement down the helix, depicted in successive snapshots (see Fig. 3.8c), was completed in 20 nanoseconds, reaching the closest approach to the co-crystal structure at approximately 50 nanoseconds. At that point, LmCytc reached the end of the helix and remained there without jumping to the active site interface for the remainder of the 100 nanoseconds MD simulation. The motion of LmCytc relative to LmP was also monitored by measuring the distance between the C_ζ atom of R24 of LmCytc and the C_γ atom of D211 of LmP, which is approximately 6 Å in the co-crystal structure. This distance reaches its maximal value of

approximately 38 Å when LmCytc first binds to helix A of LmP and is gradually reduced to its minimal value of approximately 15 Å as LmCytc is guided toward the active site interface along helix A (see Fig. 3.8a).

The main difference in the complex versus the individual structures is that R24 of LmCytc is in a different rotameric state that allows it to form a salt bridge with D211. Whereas, R24 forms an intramolecular salt bridge with E101 in the individual structure of LmCytc. We posit that this observed change in the rotameric state of R24 is a crucial step along the way from the initial encounter complex, where LmCytc is docked onto helix A of LmP, to the active ET complex. The intramolecular salt bridge remained intact during a large portion of the MD simulation (see Fig. 3.8b), and this likely explains why distance between LmCytc R24 and LmP D211 remained smaller than 15 Å throughout the simulation.

To assess the role of helix A on activity, both steady state and single-turnover kinetics for the helix A triple mutant, D47A/D50A/E54A, were measured. The k_{cat} for the triple mutant is approximately the same as that of WT LmP, and as with WT LmP, k_{cat} is independent of ionic strength and is not limited by formation of the LmP–LmCytc complex or dissociation of the product. Instead, the rate-limiting process remains intramolecular ET from the LmP active site Trp to Fe^{IV}=O. We next used stopped-flow spectroscopy to measure the rate of LmCytc oxidation by LmP Compound I. As in our previous study, an excess of Compound I over LmCytc was used to ensure that only reduction of the Trp radical was measured [9]. The rate of this process is approximately 50% slower in the mutant (see Table 3.2). These results are consistent with the BD and

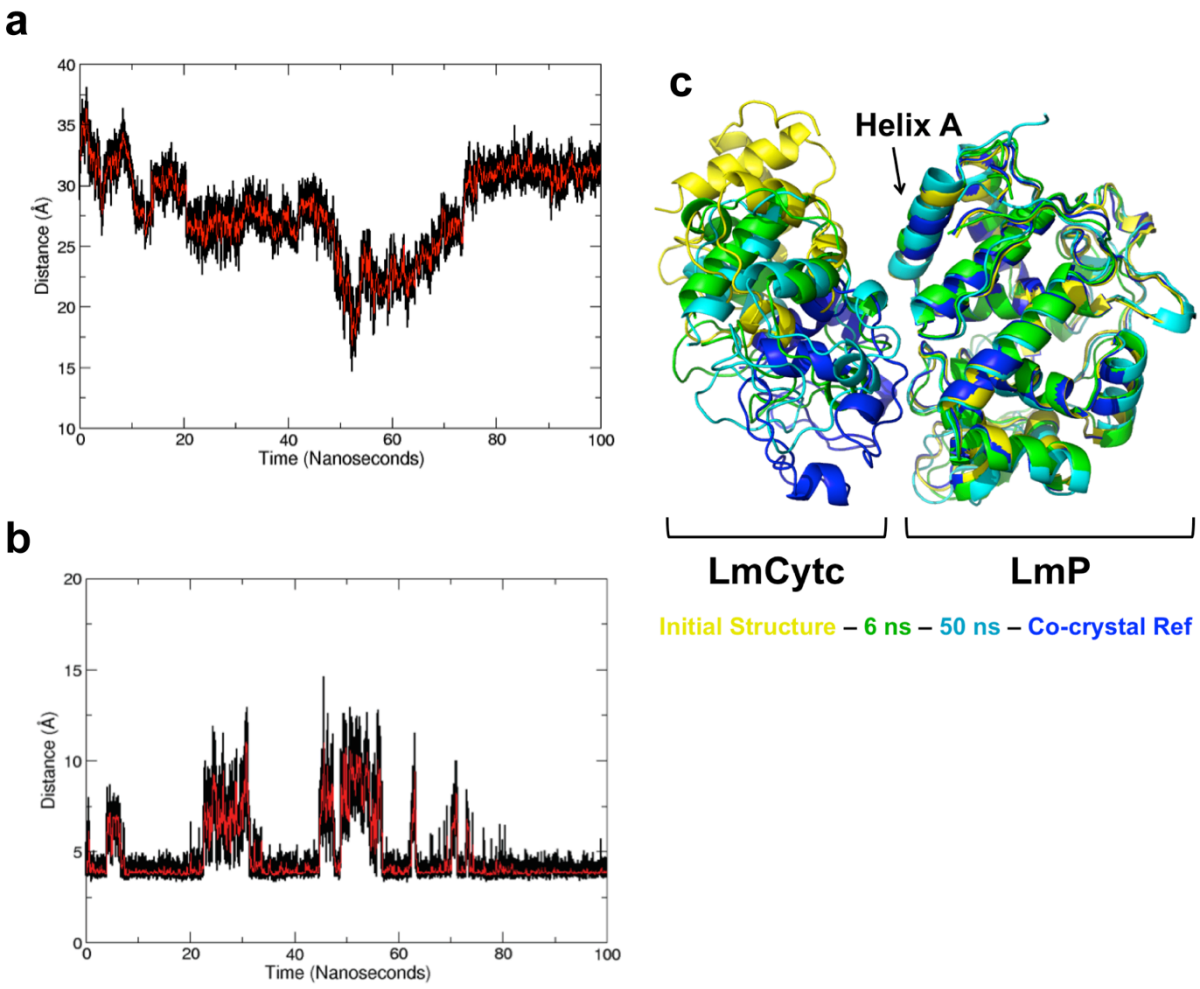

Figure 3.8: MD Trajectory Initiated from Helix A. a) Plot of the distance between LmP D211(C_γ) and LmCytc R24(C_ζ) from the LmP-LmCytc molecular dynamics simulation, shown in black with a 100-picosecond running average shown in red. LmCytc, after binding helix A of LmP, begins to move towards the LmP active site, reaching its closest approach at approximately 52 nanoseconds, before moving back and forming a stable complex with helix A for the remainder of the simulation. b) Plot of the distance between LmCytc R24(C_ζ) and LmCytc E101(C_δ) to track the intermolecular R24-E101 ion pair. A 100-picosecond running average is shown in red. The closest approach of LmP D211 to LmCytc R24 in (a) occurred during an approximately 15-nanosecond period where the intermolecular ion pair was separated. c) Snapshots after 6 nanoseconds (green) and 50 nanoseconds (cyan) from the MD simulation of LmP and LmCytc initiated from a configuration that was prepared by selecting a configuration from a BD simulation in which LmCytc was docked onto helix A of LmP and then displacing LmCytc by 15 Å. For comparison, the positions of the two proteins in the complex (co-crystal structure from Ref. [3]) are shown in dark blue while the initial starting configuration of the system derived from a Brownian dynamics simulation is shown in yellow. Figure adapted from Fig. S8 and Fig. 4 of Ref. [1].

MD simulations and support the view that helix A plays an important role in forming the initial encounter complex.

3.3.5 Yeast Cytc Does Not Associate with Helix A of Yeast CCP

The combination of BD and MD simulations suggests that the charges on helix A play a vital role in the association of LmP and LmCytc. Three of the four charges found on helix A of LmP are not present in helix A of yeast CCP [12]. This suggests that the heme peroxidases of the more complex organisms may have evolved to have a different mechanism of association, a notion that is supported by previous studies showing that the formation of the yeast CCP–Cytc complex is driven by nonpolar rather than electrostatic interactions [9, 12].

The lack of charges on helix A makes the yeast CCP system a useful control for further validation of the role of helix A charges in the formation of LmP–LmCytc encounter complexes. The probability isosurface obtained from BD trajectories of yeast Cytc around yeast CCP was initiated using the structures of the proteins in their ET complex (see Fig. 3.9a). For comparison, the location of the center of mass of Cytc in the structure of the complex is shown side-by-side (see Fig. 3.9b). There are several small features distributed broadly over the surface of CCP, but in contrast to the case of LmP (see Fig. 3.4), there are no features near the N-terminal end of helix A of yeast CCP. Thus, it is apparent that the helix A charges are required for the formation of the initial encounter complex in which LmCytc is docked onto helix A of LmP.

The second-order association rates computed from our BD trajectories of the yeast CCP–Cytc system at ionic strengths of 100 and 300 mM, $(1.65 \pm 0.02) \times 10^9$ and $(0.414 \pm 0.012) \times 10^9$ $M^{-1}s^{-1}$, respectively, are in good agreement with the corresponding experimental values, 2.95×10^9 and $(0.22 \pm 0.01) \times 10^9$ M^{-1} s^{-1}, respectively [20, 48]. Our rates are lower than the values 6.71×10^9 and 0.856×10^9 M^{-1} s^{-1}, respectively, which were computed previously by Gabdoulline and Wade using a similar BD approach, and also agree well with the experimental values [20]. Moreover, at first glance, the probability isosurface (see Fig. 3.9a) is dramatically different from an isosurface of the Boltzmann factor calculated on the basis of electrostatic interaction energies (see Fig. 5b from [20]). According to the latter, Cytc has a broad distribution of configurations on the side of yeast CCP that contains the active site, which is clearly not evident in the probability isosurface (see Fig. 3.9a). However, when we lower the value at which we contour the probability isosurface by a factor of ten (i.e., to 2.2×10^{-4}). The difference between our rates and those computed by Gabdoulline and Wade presumably reflects differences in the details of the BD simulation protocols used in the two studies.

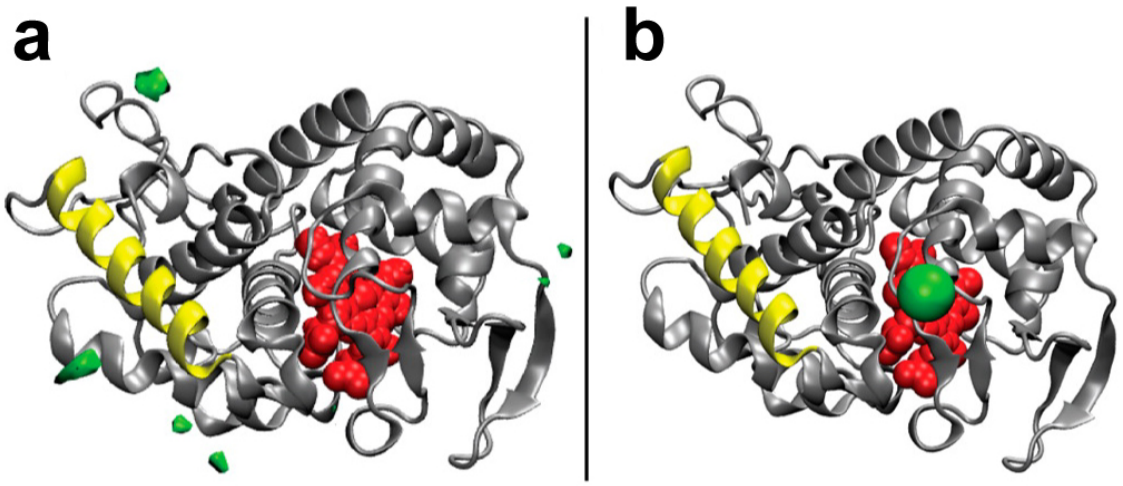

Figure 3.9: Yeast CCP-Cytc BD Pair Probability Distribution. Structure of yeast CCP in the CCP–Cytc complex from BD simulations at an ionic strength of 50 mM (drawn and colored as in Fig. 3.3). a) The green surface is the probability isosurface for the yeast Cytc center-of-mass calculated from BD simulations initiated from the co-crystal structures contoured at a value of 2.2×10^{-3}. At this contour level, there are several features distributed over the surface of yeast CCP, but none indicating a strong interaction of Cytc with helix A of CCP. b) The position of the center of mass of yeast Cytc in the structure of the complex is depicted as a green sphere. Figure adapted from Fig. 5 of Ref. [1].

3.4 Discussion

This study is an example of how computational approaches can provide novel insights into the mechanism of an important protein–protein interaction that were not at all obvious from previous experimental work, such as crystal structures. My computational effort guided the experimental work reported herein, and together, the simulations and experiments revealed new aspects of the dynamics of the initial encounter complex formed by LmP and LmCytc. The calculated rates for both WT and mutant LmP agree very well with experimental rates, thereby validating the BD approach. The most important new insight into the LmP–LmCytc system, predicted by the simulations and confirmed by the experiments, is the unexpected role of helix A of LmP. The dynamic picture that emerges from these studies is that LmCytc rapidly forms an initial encounter complex with helix A of LmP and then “crawls” toward the ET active site. The encounter complex is transient, with no specific intermolecular ion pairing, but rather an interaction between two large and electrostatically complementary surfaces that ensures very rapid complex formation. However, the final ET active complex does require specific interactions such as the ion pair between D211 of LmP and R24 of LmCytc. The probability of this ion pair forming upon initial complex formation is low, which implies that the system is designed to rapidly form a nonspecific interaction with helix A that reduces the search for the ET-active complex from a relatively slow three-dimensional search to a more rapid two-dimensional search. This ensures that complex formation and dissociation are not rate-limiting, which is exactly what we observe experimentally. The interplay between computational and experimental approaches used here should

prove useful in applications to many other systems in which complementary electrostatic surfaces play a major role in complex formation.

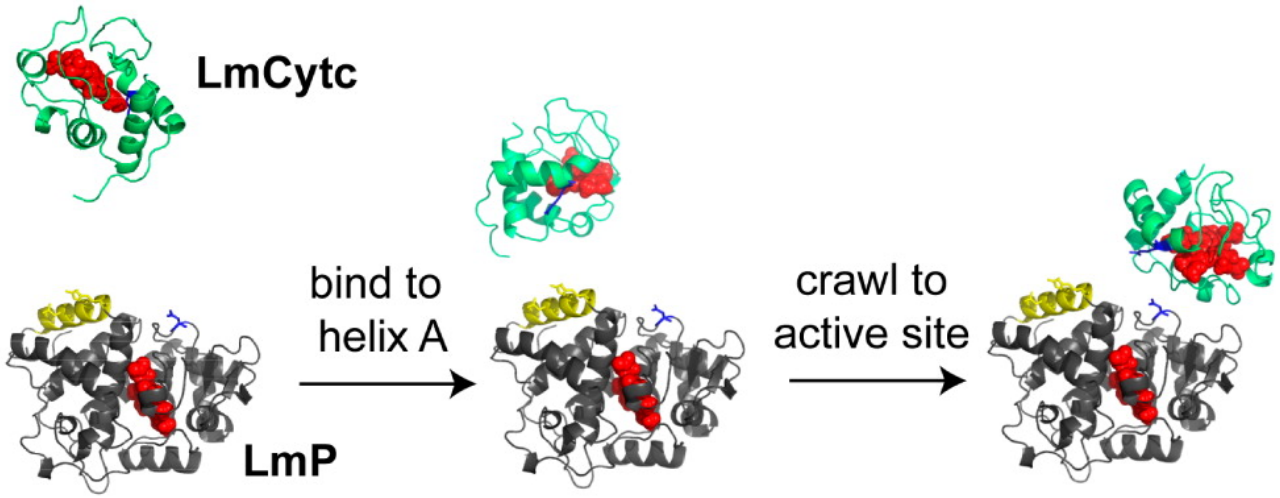

Figure 3.10: “Bind and Crawl” Association Mechanism. A schematic representation of the “Bind and Crawl” mechanism by which LmCytc associates with LmP, where binding of LmCytc to helix A of LmP precedes local sampling of the cofactor towards the active site. Our experimental and simulation results suggest that this mechanism is important to the turnover rate of LmP. The reverse process, where LmCytc moves from the active site back towards helix A, may play also play a role in the turnover rate. Figure adapted from the TOC Graphic of Ref. [1].

3.5 Bibliography

1. Fields, J. B., Hollingsworth, S. A., Chreifi, G., Heyden, M., Arce, A. P., Magana-Garcia, H. I., Poulos, T. L., and Tobias, D. J. (2015) "Bind and Crawl" Association Mechanism of Leishmania major Peroxidase and Cytochrome c Revealed by Brownian and Molecular Dynamics Simulations. *Biochemistry* **54**, 7272-7282

2. Pal, S., Dolai, S., Yadav, R. K., and Adak, S. (2010) Ascorbate peroxidase from Leishmania major controls the virulence of infective stage of promastigotes by regulating oxidative stress. *PLoS One* **5**, e11271

3. Adak, S., and Datta, A. K. (2005) Leishmania major encodes an unusual peroxidase that is a close homologue of plant ascorbate peroxidase: a novel role of the transmembrane domain. *Biochem J* **390**, 465-474

4. Jasion, V. S., and Poulos, T. L. (2012) Leishmania major peroxidase is a cytochrome c peroxidase. *Biochemistry* **51**, 2453-2460

5. Volkov, A. N., Nicholls, P., and Worrall, J. A. (2011) The complex of cytochrome c and cytochrome c peroxidase: the end of the road? *Biochim Biophys Acta* **1807**, 1482-1503

6. Wang, K., Mei, H., Geren, L., Miller, M. A., Saunders, A., Wang, X., Waldner, J. L., Pielak, G. J., Durham, B., and Millett, F. (1996) Design of a ruthenium-cytochrome c derivative to measure electron transfer to the radical cation and oxyferryl heme in cytochrome c peroxidase. *Biochemistry* **35**, 15107-15119

7. Miller, M. A. (1996) A complete mechanism for steady-state oxidation of yeast cytochrome c by yeast cytochrome c peroxidase. *Biochemistry* **35**, 15791-15799

8. Chreifi, G., Hollingsworth, S. A., Li, H., Tripathi, S., Arce, A. P., Magana-Garcia, H. I., and Poulos, T. L. (2015) Enzymatic Mechanism of Leishmania major Peroxidase and the Critical Role of Specific Ionic Interactions. *Biochemistry* **54**, 3328-3336

9. Jasion, V. S., Doukov, T., Pineda, S. H., Li, H., and Poulos, T. L. (2012) Crystal structure of the Leishmania major peroxidase-cytochrome c complex. *Proc Natl Acad Sci U S A* **109**, 18390-18394

10. Kang, C. H., Ferguson-Miller, S., and Margoliash, E. (1977) Steady state kinetics and binding of eukaryotic cytochromes c with yeast cytochrome c peroxidase. *J Biol Chem* **252**, 919-926

11. Pelletier, H., and Kraut, J. (1992) Crystal structure of a complex between electron transfer partners, cytochrome c peroxidase and cytochrome c. *Science* **258**, 1748-1755

12. Pielak, G. J., and Wang, X. (2001) Interactions between yeast iso-1-cytochrome c and its peroxidase. *Biochemistry* **40**, 422-428

13. Northrup, S. H., Luton, J. A., Boles, J. O., and Reynolds, J. C. (1988) Brownian dynamics simulation of protein association. *J Comput Aided Mol Des* **1**, 291-311

14. Northrup, S. H., and Erickson, H. P. (1992) Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc Natl Acad Sci U S A* **89**, 3338-3342

15. Northrup, S. H., Boles, J. O., and Reynolds, J. C. (1988) Brownian dynamics of cytochrome c and cytochrome c peroxidase association. *Science* **241**, 67-70

16. Pearson, D. C., Jr., and Gross, E. L. (1998) Brownian dynamics study of the interaction between plastocyanin and cytochrome f. *Biophys J* **75**, 2698-2711

17. Martinez, M., Bruce, N. J., Romanowska, J., Kokh, D. B., Ozboyaci, M., Yu, X. F., Ozturk, M. A., Richter, S., and Wade, R. C. (2015) SDA 7: A modular and parallel implementation of the simulation of diffusional association software. *J Comp Chem* **36**, 1631-1645

18. Gabdoulline, R. R., and Wade, R. C. (1998) Brownian dynamics simulation of protein-protein diffusional encounter. *Methods-a Method Enzymol* **14**, 329-341

19. Gabdoulline, R. R., and Wade, R. C. (1998) Brownian dynamics simulation of protein-protein diffusional encounter. *Methods* **14**, 329-341

20. Gabdoulline, R. R., and Wade, R. C. (2001) Protein-protein association: investigation of factors influencing association rates by brownian dynamics simulations. *J Mol Biol* **306**, 1139-1155

21. Gross, E. L. (2007) A Brownian Dynamics computational study of the interaction of spinach plastocyanin with turnip cytochrome f: the importance of plastocyanin conformational changes. *Photosynth Res* **94**, 411-422

22. Grove, T. Z., Ullmann, G. M., and Kostic, N. M. (2012) Simultaneous true, gated, and coupled electron-transfer reactions and energetics of protein rearrangement. *J Inorg Biochem* **106**, 143-150

23. Cruz-Gallardo, I., Diaz-Moreno, I., Diaz-Quintana, A., and De la Rosa, M. A. (2012) The cytochrome f-plastocyanin complex as a model to study transient interactions between redox proteins. *FEBS Lett* **586**, 646-652

24. Ermak, D. L., and Mccammon, J. A. (1978) Brownian Dynamics with Hydrodynamic Interactions. *J Chem Phys* **69**, 1352-1360

25. Gabdoulline, R. R., and Wade, R. C. (1996) Effective Charges for Macromolecules in Solvent. *J Phys Chem* **100**, 3868-3878

26. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-10041

27. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102**, 3586-3616

28. Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* **30**, 1545-1614

29. Novotny, J., Bruccoleri, R. E., Davis, M., and Sharp, K. A. (1997) Empirical free energy calculations: a blind test and further improvements to the method. *J Mol Biol* **268**, 401-411

30. Gabdoulline, R. R., and Wade, R. C. (2009) On the contributions of diffusion and thermal activation to electron transfer between Phormidium laminosum plastocyanin and cytochrome f: Brownian dynamics simulations with explicit modeling of nonpolar desolvation interactions and electron transfer events. *J Am Chem Soc* **131**, 9230-9238

31. Ortega, A., Amoros, D., and Garcia de la Torre, J. (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys J* **101**, 892-898

32. Jasion, V. S., Polanco, J. A., Meharenna, Y. T., Li, H., and Poulos, T. L. (2011) Crystal structure of Leishmania major peroxidase and characterization of the compound i tryptophan radical. *J Biol Chem* **286**, 24608-24615

33. Northrup, S. H., Allison, S. A., and Mccammon, J. A. (1984) Brownian Dynamics Simulation of Diffusion-Influenced Bimolecular Reactions. *J Chem Phy* **80**, 1517-1526

34. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38

35. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802

36. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **79**, 926-935

37. Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994) Constant-Pressure Molecular-Dynamics Algorithms. *J Chem Phys* **101**, 4177-4189

38. Feller, S. E., Zhang, Y. H., Pastor, R. W., and Brooks, B. R. (1995) Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *J Chem Phys* **103**, 4613-4621

39. Grubmuller, H., Heller, H., Windemuth, A., and Schulten, K. (1991) Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-Range Interactions. *Mol Sim* **6**, 121-142

40. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A Smooth Particle Mesh Ewald Method. *J Chem Phys* **103**, 8577-8593

41. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J Comp Phys* **23**, 327-341

42. McPhillips, T. M., McPhillips, S. E., Chiu, H. J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M., and Kuhn, P. (2002) Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J Synch Rad* **9**, 401-406

43. Kabsch, W. (2010) Xds. *Acta Crystallog D* **66**, 125-132

44. Mccoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J App Crystallog* **40**, 658-674

45. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog D* **53**, 240-255

46. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallog D* **60**, 2126-2132

47. Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallog D* **57**, 122-133

48. Mei, H. K., Wang, K. F., McKee, S., Wang, X. M., Waldner, J. L., Pielak, G. J., Durham, B., and Millett, F. (1996) Control of formation and dissociation of the high-affinity complex between cytochrome c and cytochrome c peroxidase by ionic strength and the low-affinity binding site. *Biochemistry* **35**, 15800-15806

Chapter 4

Calmodulin Modulates AQP0 Water Permeability through a Positively-Charged Cytoplasmic Loop

The research in this chapter was originally published in the Journal of Biological Chemistry [1]. I performed all of the simulations described in the methods section below (see Section 4.2). J. Alfredo Freites (Department of Chemistry, UC Irvine) assisted me in the analysis of the atomistic trajectories. The oocyte swelling assays were performed by Karin L. Németh-Cahalan (Department of Physiology and Biophysics, UC Irvine) and the Western blot experiments were conducted by Irene Vorontsova (Department of Physiology and Biophysics, UC Irvine). These experiments were carried out in the laboratory of James E. Hall (Professor of Physiology and Biophysics, School of Medicine, UC Irvine). The study was conceived and designed by James E. Hall and Douglas J. Tobias.

4.1 Introduction

In Ch. 2 and Ch. 3, I described the results of Brownian Dynamics simulations studies on the binding of water-solvated P450 redox partners. This chapter will focus on the binding of a soluble protein Calmodulin (CaM) to a membrane-embedded water channel, Aquaporin 0 (AQP0). The Aquaporins (AQPs) are a group of highly-conserved transmembrane proteins that facilitate the selective transport of water across plasma membranes [2]. The individual monomers contain a single-file water pore housed in a six-helix bundle domain [3]. Critical to the water selectivity and permeation rate are a series of residues that line the pore axis of the AQPs. Starting from the extracellular face, a region known as the aromatic arginine selective filter (ar/R) creates an approximately 3 Å-wide hole which excludes larger molecules from the outside while easily accommodating water [3]. This conserved selectivity filter, also known as the first constriction site (CS-I), contains a positively-charged arginine residue which helps to exclude protons from permeating. The width of the CS-I varies from isoform to isoform, and accounts for the large variation in the observed osmotic water permeability constants (P_f) of these channels, which can differ by as much as two orders of magnitude [4, 5]. Further along the pore axis sits a pair of Asn-Pro-Ala (NPA) motifs which order water molecules into a single file and flip the orientation of the waters such that their oxygen atoms point towards the center of the channel [3, 6]. Additional constriction sites can be found in the cytosolic side of some of the AQP channels. The existence of these constrictions accounts for some of differences in the cytosolic modulation of the channel P_f [7]. In particular, AQP0 and AQP6 bind calmodulin (CaM) at their C termini in a Ca^{2+}-dependent manner [8-10].

Unlike the other AQPs, which are expressed ubiquitously throughout eukaryotic and prokaryotic cells, AQP0 is localized almost exclusively to the eye lens, where it plays an essential role in water transport and possibly cell-cell adhesion [5, 11]. AQP0 water permeability is modulated by CaM, which interacts with its C-terminal helical domain [12] in a noncanonical fashion. CaM interacts simultaneously with two C-terminal helices of AQP0 in antiparallel fashion [7, 9], giving rise to a 2:1 CaM:AQP0 tetramer stoichiometry. This type of CaM-bound configuration is also found in the petunia glutamate decarboxylase A complex [13, 14]. Because increases in intracellular calcium lead to an approximately two-fold reduction in water permeability, it has been hypothesized that CaM physically “plugs” two of the four AQP0 monomers. Previous studies have shown that CaM does not occlude the pore permeation pathway [7]. Instead, AQP0 possesses a modifiable second constriction site (CS-II) near its cytoplasmic opening, whose dynamics and configuration are influenced by the presence of bound-CaM. Molecular dynamics (MD) simulations of CaM-bound AQP0 revealed a reduction of the root mean square fluctuations (RMSF) of the CS-II gating residues, specifically Y149, reducing the size of the pore opening to the cytosol [7]. In this study we investigate the poorly understood molecular mechanism of the CaM-induced modulation of the AQP0 channel.

The interaction between AQP0 and CaM is of considerable interest because the structure of the complex between CaM and the holo protein is known, albeit at relatively low resolution [7]. Most CaM-protein structures include only the CaM-binding peptide portion of the protein and thus may leave out critical interactions with the remainder of the protein, interactions that may well be critical for the modulation of protein function by CaM.

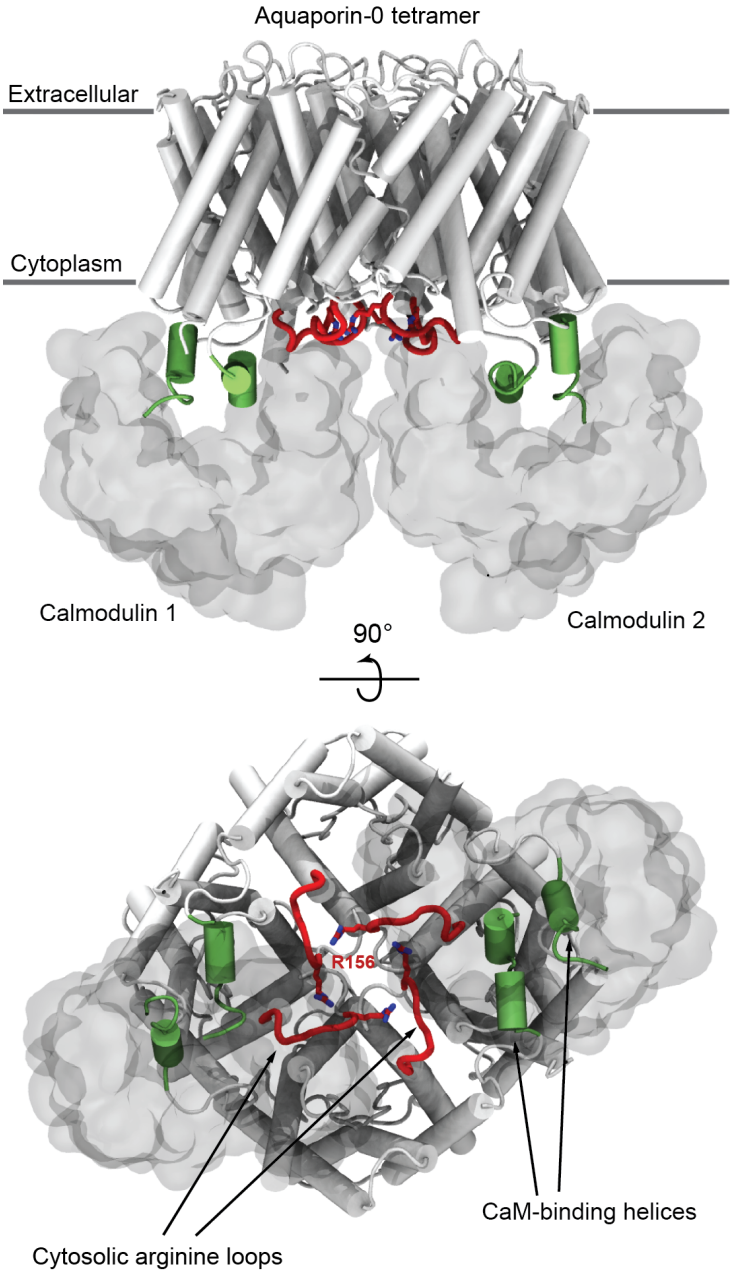

Figure 4.1: Cytosolic Arginine Loop of Aquaporin 0. Shown is the atomistic model of the AQP0-CaM complex (PDB code 3J41 from Ref. [7]), highlighting the important structural features of AQP0 for CaM-sensing. CaM (shown in a white molecular surface representation) binds AQP0 (shown in white secondary structure representation) at its CaM-binding domain (helical regions are shown in green) in antiparallel fashion. CaM reduces the water permeability of AQP0 by interacting with the R156 residue (shown in red licorice representation) of the cytosolic arginine loop (shown in red) located at the center of the AQP0 tetramer. Figure adapted from Fig. 1 of Ref. [1].

The structure of Reichow *et al.* [7] strongly suggests that the electrostatic interface between predominantly negatively-charged CaM and predominantly positively-charged AQP0 may well play such a role. In this chapter, I detail the experimental and theoretical evidence for the role of the electrostatic interactions between CaM and AQP0 in modulating the water permeability of AQP0. We found that the electrostatic interface between the two proteins interacts with the more conventional CaM-binding site to adjust the position of Y149 at the CS-II and thus modulate the P_f of AQP0. In particular, we demonstrate a role for the previously overlooked amino acids in the arginine-rich loop adjacent to Y149, namely, their coupling to the CaM molecule and their allosteric effects on the water permeability of AQP0.

Many natural modifications of AQP0 occur at a C-terminal helical domain, which is often referred to as the AQP0 CaM-binding domain ($AQP0^{CBD}$) [5, 15, 16]. These modifications include mutations that are known to cause cataracts in humans, such as R233K, as well as serine phosphorylation at S229, S231, and S235 [5, 15, 16]. Additionally, the phosphorylation state of the $AQP0^{CBD}$ is dependent on its distance from the lens periphery and is involved in the translocation of AQP0 into lens fiber cell membranes [11, 17, 18]. Hypotonic swelling experiments in *Xenopus laevis* oocytes have shown that $AQP0^{CBD}$ modifications reduce the sensitivity of AQP0 P_f to calcium [19]. However, the effects of these modifications on the AQP0-CaM interaction remain unknown. NMR and mass spectroscopy experiments have revealed that phosphorylation of the $AQP0^{CBD}$ peptide fragment reduces its binding affinity for CaM [14, 20]. Adding phosphates to AQP0 would, in principle, reduce its ability to associate with the negatively-charged CaM. However, some $AQP0^{CBD}$ mutations retain partial

calcium sensitivity [19], suggesting that CaM is still able to bind the phosphorylated AQP0 tetramer.

In this chapter, I used Brownian dynamics (BD) simulations to study the effects of charge mutation on the formation of the AQP0-CaM complex and atomistic MD simulations to elucidate the effects of these mutations on the structure and function of AQP0. Our collaborators then expressed both wild type (WT) and mutated forms of AQP0 into oocytes and assessed their permeabilities to validate insights from the MD simulations with experimental measurements and to guide further analysis and observations of our MD trajectories. This data revealed that an arginine-rich cytosolic loop (see Fig. 4.1) of AQP0 allows CaM to allosterically control the dynamics of the CS-II, whereas the $AQP0^{CBD}$ functions primarily as the CaM-binding site and as a modulatory feature of AQP0.

4.2 Methods

4.2.1 Brownian Dynamics Docking of CaM to AQP0

Brownian dynamics trajectories were generated using the Ermak-McCammon algorithm [21] implemented in the SDA 7 software package [22, 23] described in Appendix A. The electrostatic potentials were computed by finite-difference solution of the Poisson-Boltzmann equation implemented in the APBS 1.4 program package [24] and described in Appendix A.2. We utilized a manual multigrid calculation with a single Debye-Hückel boundary condition for each solute on a 200 × 200 × 200 grid of 1 Å spacing. Atomic charge and radii parameters were assigned to CaM and AQP0 using the CHARMM36

force field [25]. The dielectric boundary was defined by the Van der Waals surface of the protein. Harmonic dielectric boundary smoothing and charge anti-aliasing were implemented using the method developed by Novotny *et al.* [26]. The solute and protein dielectric constants were set to 78 and 2, respectively, at 298.15 K for all electrostatic potential grid calculations. An ionic strength of 50 mM NaCl, which was represented by charges of +1 and −1 with radii of 1.1 and 1.7 Å, respectively, was used for the electrostatic potential calculations. The solvent radius was set to 1.4 Å. The exclusion probe radius and skin thickness were set to 4 Å and 3 Å, respectively. For consistency, we used the same ionic strengths and solvent dielectric constant as the computed electrostatic potential. The electrostatic desolvation energy grids (200 × 200 × 200 with a 1 Å spacing) were computed using Eq. 2 in Gabdoulline and Wade [27] with the scaling factor α set to 1.67. A single ionic radius of 1.5 Å at 50 mM concentration was used. The nonpolar desolvation energy grids (200 × 200 × 200 with a 1 Å spacing) were computed using Eq. 3 from Martinez *et al* [23] with the distance parameters *a* and *b* set to 3.10 Å and 4.35 Å, respectively, and the normalization constants *c* and β set to 0.5 and −0.0065 kcal·mole^{-1}·Å^{-2}, respectively. The exclusion grids were calculated as described in Gabdoulline and Wade [28] using a probe radius of 1.77 Å.

In our BD simulations of the formation of AQP0-CaM complexes, the trajectory of CaM relative to AQP0 is computed by holding AQP0 fixed at the center of a sphere of radius 250 Å, while CaM was allowed to diffuse from a random starting position on the surface of the sphere. The escape sphere radius was set to 350 Å. The integration time step was linearly decreased from 20 ps to 1 ps over a protein surface-to-surface distance of 90 Å to 60 Å. This allows the trajectories to capture the finer details of the motion close

to the formation of the encounter complex. The diffusion coefficients for CaM, which appear in the Ermak-McCammon algorithm, were calculated using the Hydropro 10 software package [29]. The rotational and translational diffusion coefficients were calculated to be 5.424×10^{-2} Å^2 ps^{-1} and 2.453×10^{-6} rad^2 ps^{-1}, respectively.

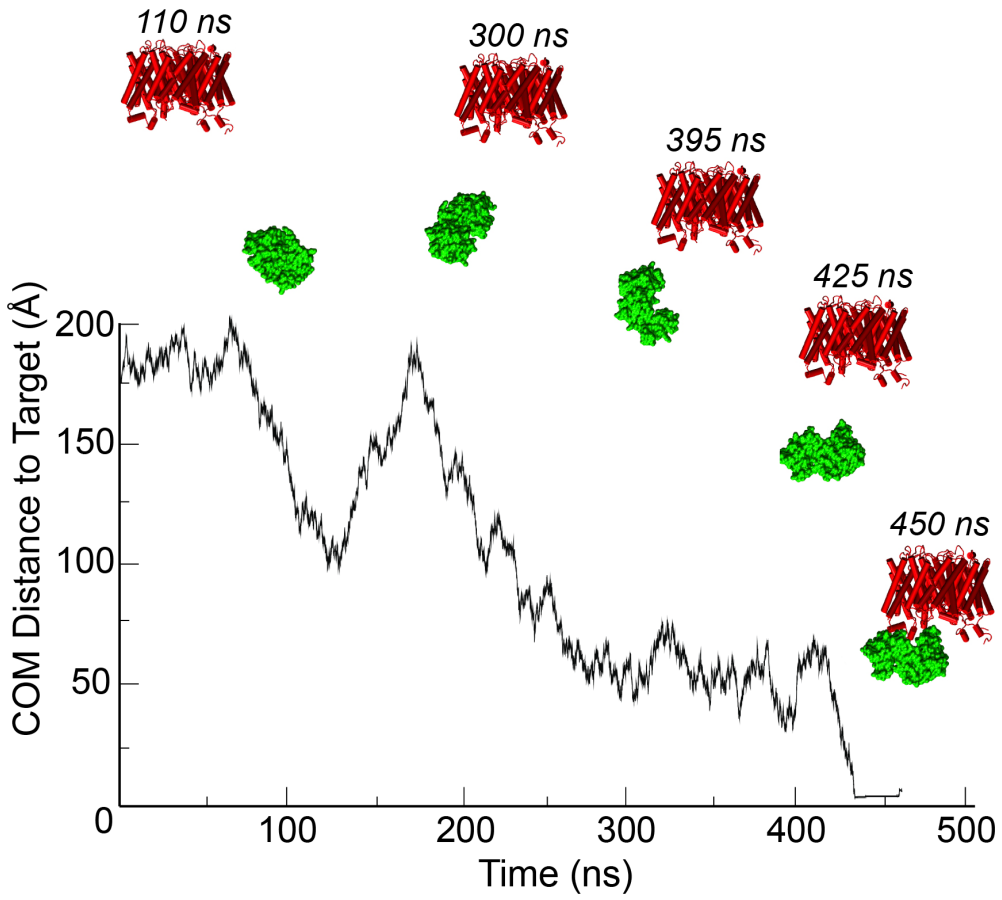

Figure 4.2: BD Docking of the AQP0-CaM Complex. A representative Brownian dynamics trajectory shows CaM docking to WT AQP0. The center-of-mass distance between the mobile CaM protein (shown in green molecular surface representation) and its docked position on AQP0 (shown in red secondary structure representation) shows CaM diffusing to its bound configuration at approximately 450 ns. The initial configurations used for BD simulations were taken from an equilibrated MD simulation of CaM-bound AQP0 [7]. Figure adapted from Fig. 2 of Ref. [1].

A total of 500,000 BD trajectories were generated at 298.15 K for each charge variant (see Table 4.1). A list of 14 donor-acceptor atom pairs between CaM and AQP0 in the complex structure with a maximum distance of 4.5 Å between each pair was generated using SDA 7. A binding trajectory occurs when any 3 of these atom pairs achieves a 5 Å separation simultaneously. The error estimates in the binding fraction were obtained by using the Bootstrap_multiCPU tool in SDA 7 [22, 23]. Serine pseudo-phosphorylation was modeled by removing the hydroxyl hydrogen of the side chain and adding a single negative charge to the hydroxyl oxygen. *In silico* arginine to alanine mutations were made by truncating atoms of the side chain beyond the β-carbon followed by the creation of new structure and parameter files using the psfgen tool in VMD, version 1.9.1 [30].

4.2.2 Atomistic MD Setup

To better understand the effects of phosphorylation on the structure and function of AQP0, I performed atomistic MD simulations of AQP0 and the AQP0-CaM complex embedded in a lipid bilayer in excess water. The model of Reichow *et al* [7] (PDB code 3J41) with S229 and/or S235 modified to a monoanionic phosphoserine was used as the initial configuration [25]. Each system was composed of an AQP0 tetramer (residues 5–239) complexed with 2 CaM monomers bound to 8 Ca^{2+} ions, embedded in a membrane of 410 POPC lipid molecules and 56,701 waters. Sodium counterions were added to neutralize the system. The CaM-free systems were constructed from the complex structure model through the deletion of the CaM monomers.

All MD simulations were performed using the NAMD 2.9 software package [31]. The

CHARMM36 force field was used for proteins [32-34] and lipids [35], and the TIP3P model was used for water [25]. The system was subjected to 1500 steps of conjugate gradient energy minimization before MD simulation. The simulations were carried out for approximately 600 ns each at a constant temperature of 300 K and a constant pressure of 1 atm using Langevin dynamics for temperature control and a Nosé-Hoover-Langevin piston for pressure control [36, 37]. In addition, we extended the trajectories of the CaM-free and CaM-bound WT AQP0 from Reichow *et al.* [7] by approximately 450 ns. A multiple time step algorithm was used to integrate the equations of motion with time steps of 4 fs for the electrostatic interactions and 2 fs for the bonded and short-ranged nonbonded interactions [38]. The smooth particle mesh Ewald algorithm [39] was used to treat the electrostatic interactions, and the real-space part of the Ewald sum and the Lennard-Jones interactions were smoothly switched off between 10 Å and 12 Å. The SHAKE algorithm was used to constrain all bonds lengths involving hydrogen atoms [40]. The VMD 1.9.1 software package was used for visualization and analysis [30]. The statistical uncertainties in the interaction energies were computed using the block transformation method for correlated data [41].

4.2.3 Graph Representation and Contact Analysis

To analyze the contacts of the AQP0-CaM interaction as well as the intraprotein interactions of AQP0, we utilized the chemical group graph representation developed by Benson and Daggett [42]. In this scheme, each amino acid residue is parsed into small molecular moieties constituting the graph nodes. The node types include nonpolar, dipolar, negatively-charged, and positively-charged moieties. Graph edges are defined

by a contact function, $\chi(u_i,v_j)$, which equals 1 when atom i of node u is within 4.5 Å of atom j of node v, and 0 otherwise. If atoms i and j are both carbon, a cutoff distance of 5.4 Å is used instead of 4.5 Å. The edge weight (w) between nodes u and v gives a quantification of the strength of the contact:

$$w(u,v) = \sum_{i=1}^{n}\sum_{j=1}^{m}\chi(u_i,v_j) \tag{4.1}$$

I took the time average:

$$\langle w(u,v)\rangle = \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n}\sum_{j=1}^{m}\chi_t\left(u_{i,t},v_{j,t}\right) \tag{4.2}$$

where the MD timesteps is T. We then summed $\langle w(u,v)\rangle$ for all independent nodes of each residue pair to get a contact score that weighs the contact by both strength and frequency. A rectangular matrix was constructed using residue contact scores for each pair of residues between AQP0 and CaM. The CBD contact matrices were produced using the same scheme to generate the intraprotein residue contact scores between the AQP0^{CBD} region and the remainder of the AQP0 tetramer. All contact matrix plot images were rendered using the Mathematica software package, version 10.1 [43].

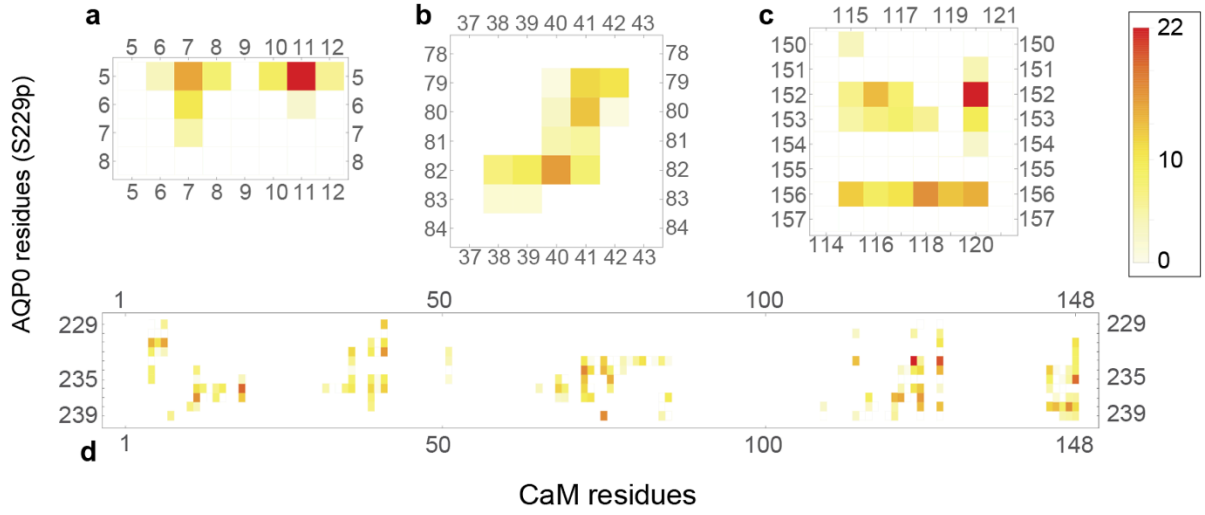

Figure 4.3: Contact Matrix From MD Simulation of the AQP0-CaM Complex. A matrix representation of the graph-based interprotein contacts formed between residues of AQP0-S229P (on the vertical axis) and CaM (on the horizontal axis). The color scales indicate the number of atomic-level contacts averaged over the MD trajectory (see Section 4.2.2). White indicates that the two residues do not interact, whereas a darker red indicates a strong interaction. Four plots show each of the four regions of AQP0 that make contact with CaM: a) the N terminus, b) the 78–84 loop, c) the arginine-rich cytosolic loop, d) and the C-terminal AQP0^{CBD} helix. Figure adapted from Fig. 4 of Ref. [1].

4.2.4 Path Analysis

We utilized a modified Floyd-Warshall algorithm to compute a shortest path between all AQP0 nodes for individual MD configurations of the AQP0-CaM complex [44]. Shortest paths connecting the Y149 hydroxyl and R156 guanidine nodes for each configuration were computed by ignoring edge weights. The resulting interaction networks were subsequently characterized and visualized using VMD 1.9.1 [30].

4.2.5 Xenopus Oocyte Permeability Assays

Forms of AQP0 containing mutations R152A, R153A, R156A, or R153A/S229D were cloned into a transcription vector (pXbG) containing BglII cloning site flanked by 5′- and 3′-untranslated regions (UTRs) of the *X. laevis* β-globin gene driven by the T3 transcription promoter. Oocytes from *X. laevis* were obtained from Ecocyte (Austin, TX) and injected typically with 10 ng of RNA encoding wild type, R152A, R153A, R156A, or R153A/S229D AQP0 generated using the mMessage mMachine T3 kit (Ambion/Life Technologies) as described previously [9] (when different amounts of RNA were injected, they are noted in the text or figure legends). The oocytes were incubated in 100% ND96 (100% ND96: 96 mM NaCl 96, 2 mM KCl, 5 mM HEPES, 1.8 mM $CaCl_2$, 1 mM $MgCl_2$, pH 7.5) with the desired test Ca^{2+} concentration for 5 min before the swelling assay. Swelling assays were performed at room temperature (20–21 °C) by transferring oocytes from a 200 mosmol to a 70 mosmol (30% (v/v) ND96) solution adjusted to the desired calcium concentration. Water permeability (P_f) was calculated from optical measurements of the increase in cross-sectional area of the oocyte with time in response to diluted ND96 using the formula:

$$P_{\mathrm{f}} = \left[\frac{d(V/V_0)}{dt}\right]\left[\frac{V_0}{S_0}\right]/[\Delta_{osm} V_w] \tag{4.3}$$

where V is the volume as a function of time, V_0 is the initial volume, *S* is the geometric surface area, Δ_{osm} is the osmotic gradient, and V_w is the molar volume of water. Each data point is the average of at least six measurements (at least two different batches of oocytes, three oocytes from each batch unless otherwise noted). Error bars are shown as ± S.D.

Two-way analysis of variance was performed on the data sets (see Fig. 4.6) using the statistical package R (R-project) showed that the differences between means were highly significant ($p < 10^{-15}$) [45]. We performed pairwise t tests on all of the data in each data set using the R function “pairwise.t.test.” We show *p* values for appropriate conditions in the legend.

4.2.6 Western Blot Assays

Total membrane protein was isolated using the ProteoExtract^{R} native membrane protein extraction kit (Calbiochem) according to the manufacturer's directions. Two oocytes worth of membrane protein in each lane was separated on a 4–12% SDS NuPAGE gel (Invitrogen) and transferred onto a nitrocellulose membrane. The blot was blocked in 5% milk in Tris-buffered saline with 0.1% Tween (TBS-T), then incubated with anti-AQP0 (H-44, Santa Cruz Biotechnology) antibody diluted at 1:500 in blocking solution (1% BSA, 2 mM EDTA in TBS-T). The blot was then washed and incubated with goat anti-rabbit HRP conjugated antibody (Pierce) for 1 h at room temperature and visualized

by ECL Prime Blotting Detection reagent (Amersham Biosciences). Blots were then stripped and re-probed with an anti-β actin (ab8227 Abcam) antibody diluted at 1:500 in blocking solution followed by secondary antibody labeling and development as above.

4.3 Results

4.3.1 Critical Electrostatic Interactions Between AQP0 and CaM Occur Outside of the AQP0 CaM-Binding Domain

We performed BD simulations of the association of AQP0 with CaM using the atomistic model reported by Reichow *et al.* [7] as the initial configuration by separating the CaM proteins from AQP0. BD simulations are often used to study the effects of charge mutation on protein association [46-50]. BD docking of a globular protein, such as CaM, to a membrane protein is not unprecedented, but it is challenging due to the potential burial of interprotein contacts in the membrane environment, which may be inaccessible when the membrane-embedded protein is modeled as a rigid body [51, 52]. Thus, we chose to remove the membrane environment surrounding the AQP0 structure. Our BD trajectories were able to reproduce the correct AQP0-CaM binding orientation but only when MD-equilibrated input configurations of AQP0 and CaM, taken from the simulated AQP0-CaM complex reported in Reichow *et al.* [7], were used (see Fig. 4.2). The atomic structure of the AQP0-CaM complex (Reichow *et al.* [7] (PDB code: 3J41) was produced using the petunia glutamate decarboxylase calmodulin complex (PDB code: 1NWD) as the template for the interaction between CaM and the $AQP0^{CBD}$. Consequently, interactions outside of the CaM-binding domain were not explicitly

included in the interpretation of the experimental electron microscopy maps used to solve the structure [7, 53]. These results suggest that regions of AQP0, other than the $AQP0^{CBD}$, form contacts with CaM (which we will elaborate on below) that are important both for stabilizing the complex and modulating the CS-II.

AQP0 Variant	Binding fraction (%)
WT	2.846 ± 0.019
S229p	2.143 ± 0.027
S235p	2.475 ± 0.025
S229-235p	1.246 ± 0.021
R152A	1.734 ± 0.021
R153A	2.009 ± 0.025
R156A	1.039 ± 0.013
R153A-S229p	0.413 ± 0.009
R153A-S229p-S235p	0.143 ± 0.004

Table 4.1: CaM-binding Statistics for AQP0 Mutants from Brownian Dynamics Simulations.

To study the effects of serine phosphorylation on the association of the AQP0-CaM complex, we performed additional BD simulations of the association of CaM and AQP0, introducing negative charges to the C-terminal serine residues, S229 and S235. We deleted the S229 and S235 hydroxyl hydrogens to leave a negatively-charged side chain, thereby mimicking experimental pseudo-phosphorylation without adding additional steric bulk that could interfere with the formation of the complex. Using nonspecific contacts to define the association of the complex (see “Experimental Procedures”), we compared the BD complex formation in a total of 500,000 trajectories for S229, S235, and S229/S235 variants versus WT. The binding fractions for each AQP0 variant are reported in Table 4.1. It is evident that serine phosphorylation of the $AQP0^{CBD}$ does not alter the frequency of AQP0-CaM complex formation significantly in our BD simulations. We also ran BD simulations of arginine-to-alanine mutations (R152A, R153A, and R156A) of AQP0 (see Table 4.1). These mutations reduce CaM-

binding only marginally but more than mutations of the $AQP0^{CBD}$. The compound mutants AQP0-R153A/S229P and AQP0-R153A/S229P/S235P produced the largest effect, around an order of magnitude reduction of the binding fraction. It is important to contrast the results of our BD simulations of CaM-binding to the AQP0 holo protein and the experimental NMR measurements of CaM-binding to the AQP0 C-terminal peptide fragment. Experiments have shown that S235 phosphorylation of the AQP0 C-terminal peptide reduces its binding affinity to CaM by more than an order of magnitude [12, 14]. Our BD simulations, which take advantage of the full structure of the AQP0-CaM complex, suggest that serine phosphorylation does not inhibit CaM-binding to the whole protein.

4.3.2 AQP0 Phosphorylation Modifies the Contacts Between CaM and the Arginine-Rich Cytosolic Loop

To study the effects of $AQP0^{CBD}$ serine phosphorylation on the AQP0-CaM interaction, we performed atomistic MD simulations of the AQP0-CaM complex structure (PDB code 3J41) by converting the standard S229 and/or S235 side chain to a monoanionic phosphoserine (henceforth denoted as S229P and/or S235P). This model was chosen to mimic experimental pseudo-phosphorylation, which utilizes a serine-to-aspartate mutation. A small weakening of the attractive electrostatic interactions was expected due to the excess of exposed acidic residues on CaM and was consistent with our BD results (see Table 4.1).

To assess the effect of serine phosphorylation on the pore opening, we monitored the distance between the CS-II residues F75 and Y149 of each subunit (see Fig. 4.4c). Previous studies suggest that the corresponding distance distribution accurately describes the conformational landscape of the CS-II between an open and closed state [7, 54]. In particular, for distances of >8 Å, Y149 rotates away from the pore toward the cytosol, allowing pore waters previously constrained by the aromatic side chain to access the bulk. The distance distribution for all four of the pores (see Fig. 4.4c) shows CS-II open-state populations for both CaM-free WT AQP0 and CaM-bound AQP0-S229P [7] but not in CaM-bound WT AQP0. Thus, serine phosphorylation in the CBD modifies the CS-II configuration and leads to a similar open-state probability in CaM-bound AQP0 as in CaM-free AQP0. We hypothesize that this change in the CS-II configuration of AQP0-S229P locks the channel in a high P_f state [9]. To gain insight into the underlying molecular mechanism, we performed a detailed analysis of the atomic-level interactions between CaM and AQP0 by mapping the atomistic MD configurations onto a chemical group graph representation [42].

In the AQP0-S229P system, CaM formed significant interactions with four distinct regions of AQP0 (see Fig. 4.3). First, the CaM N-terminal lobe interacted with the N terminus of AQP0 (see Fig. 4.3a). When S229 was phosphorylated, the N-terminal CaM lobes interacted with the cytosolic loop around residues 78–84 (see Fig. 4.3b). These interactions occurred between polar side chains of the AQP0 loop and acidic side chains of the CaM N-terminal lobe. As expected, the most extensive interaction region was between CaM and its canonical binding site $AQP0^{CBD}$ (see Fig. 4.3d).

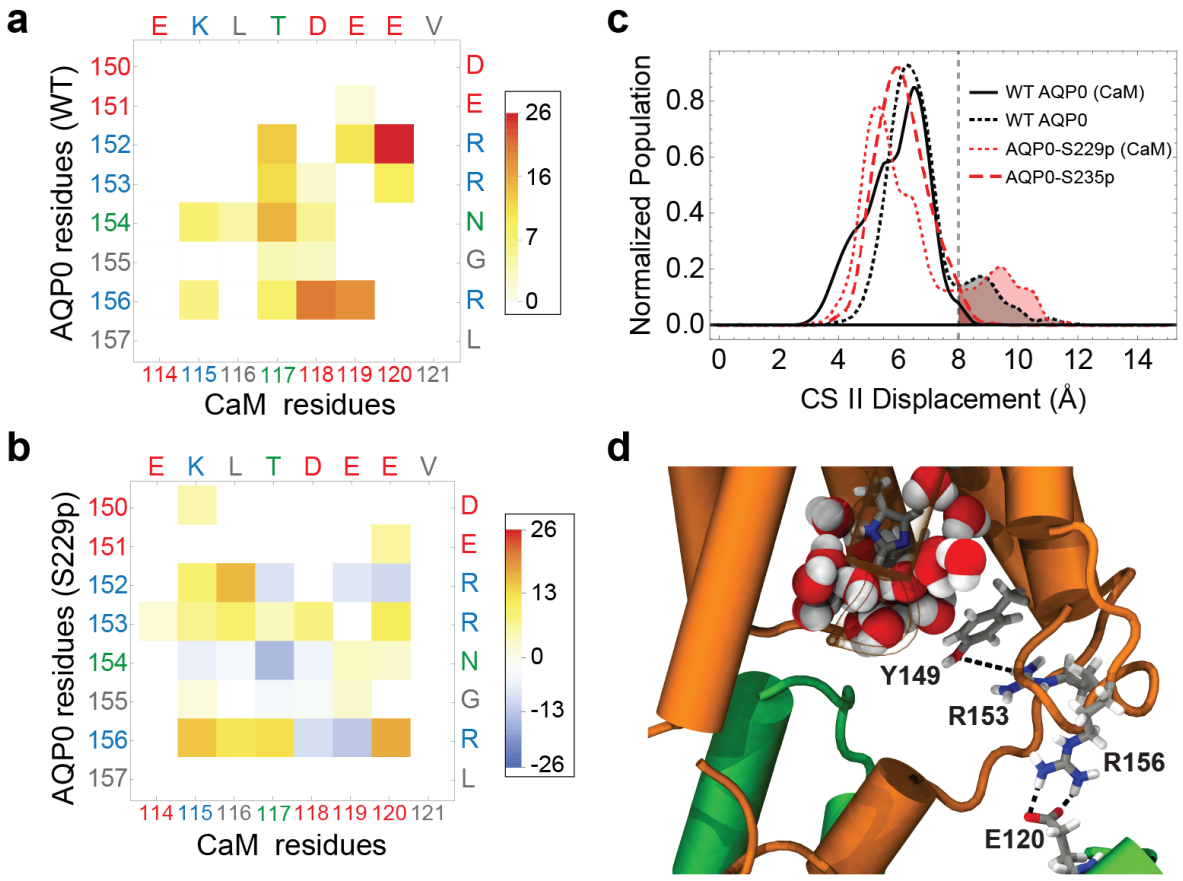

Figure 4.4: Phosphorylation of AQP0 Modifies CaM Contacts. a) Matrix representation of the interprotein contacts formed between WT AQP0 residues and CaM residues at the cytosolic arginine loop. The color scale indicates the average number of atomic-level contacts over a MD trajectory (see Section 4.2.3). Residues are color-labeled on the axis by type (red: negative, blue: positive, green: polar, grey: nonpolar). b) Matrix plot of contact differences between CaM-bound AQP0-S229p and CaM-bound WT AQP0 (subtraction of the CaM-bound WT AQP0 from the CaM-bound AQP0 S229p matrix). Darker red and darker blue indicate, respectively, contact formation and contact elimination upon S229 phosphorylation. In the color scale, positive values correspond to a net increase in the average number of contacts and negative values correspond to a net decrease in the number of contacts upon S229 phosphorylation. c) CS-II distance distributions for different AQP0 variants, indicating the size of the pore opening to the cytosolic vestibule. The different AQP0 variants consist of WT and phosphorylated AQP0, with and without CaM. d) Configuration snapshot from an MD simulation of CaM-bound AQP0-S229p showing R153 interacting with the Y149 hydroxyl group, which displaces Y149 away from the water pore. This is caused by the AQP0 R156 interaction with CaM shifting from D118 to E120. An increase in the density of the AQP0 pore lumen waters (shown in filled-sphere representation colored by atom type) can be seen near the cytoplasmic opening in response to the Y149 displacement. Figure adapted from Fig. 3 of Ref. [1].

The interactions outside the $AQP0^{CBD}$ pertinent to the modulation of CS-II occurred between the C-terminal lobe of CaM and the cytosolic loops that connect the fourth and fifth transmembrane helices (residues 150–156; see Fig. 4.3c). The CaM C-terminal lobes (residues 115–120) formed electrostatic interactions with three arginine residues (R152, R153, and R156) (see Fig. 4.4a). We showed previously that these arginine residues play an important role in the formation of the AQP0-CaM complex in BD simulations (see Table 4.1). The WT AQP0/CaM contact matrix was subtracted from the AQP0-S229P/CaM matrix (see Fig. 4.4d). Vanishing contacts as a result of the phosphorylation appeared in the difference matrix as negative-valued features. These changes were concentrated in the arginine loop region and corresponded to a shift in the interaction of AQP0 R156 with CaM D118 to CaM E120 (see Fig. 4.4b). The increased CS-II distance (see Fig. 4.4c) resulted from this contact shift, which caused R153 to interact with Y149, displacing Y149 away from the pore opening to allow pore waters to access the bulk (see Fig. 4.4d). This mechanism is entirely consistent with the experimentally observed locked-high P_f of AQP0-S229P, opening the pore in the presence of bound-CaM [9].

The behavior of AQP0-S235P is quite different from AQP0-S229. S235 phosphorylation locks the channel in a P_f low irrespective of calcium concentration [19], suggesting that AQP0 is able to reduce its permeability in the absence of bound-CaM. The CS-II distance distribution from our CaM-free AQP0-S235P is consistent with CaM-bound WT AQP0, where the distribution shifts toward smaller distance (see Fig. 4.4c). We observed that S235 phosphorylation in CaM-free AQP0 induced additional contacts between the $AQP0^{CBD}$ and the arginine loop, specifically with E151 (see Fig. 4.5a,b). A

salt bridge between R233 and E151 reduced the fluctuations in the side chain of the proximal Y149 in a similar way as CaM-bound WT AQP0, lowering the P_f. This suggests that AQP0 phosphorylation not only regulates channel permeability by modifying the CaM-interaction but also through intraprotein interactions with the AQP0 arginine loop.

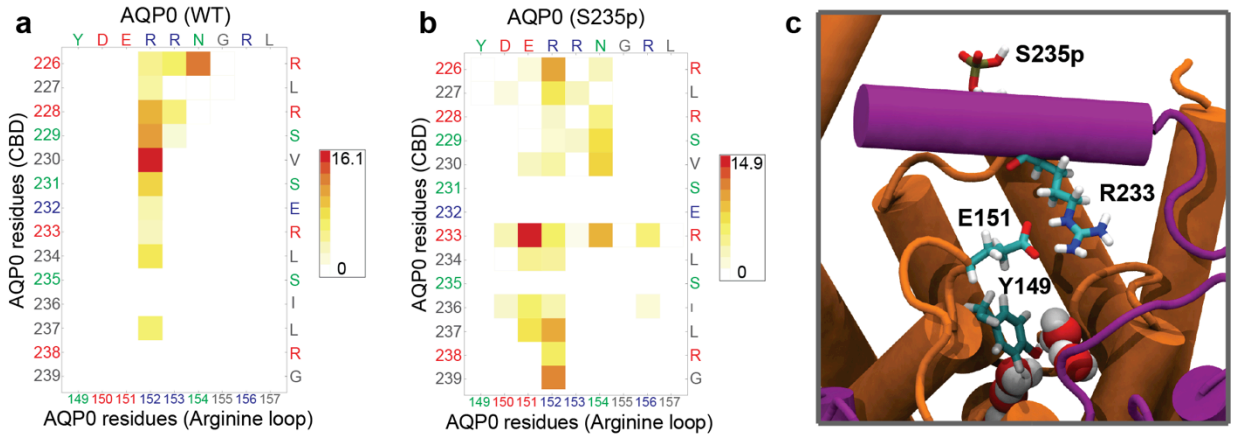

Figure 4.5: Phosphorylation Modifies the Inter-Protein Contacts of AQP0. a,b) Matrix representation of the intraprotein contacts formed between $AQP0^{CBD}$ (on the vertical axis) and the arginine-rich cytosolic loop (on the horizontal axis) in CaM-free simulation systems. The color scale indicates the number of atomic-level contacts averaged over a MD simulation trajectory (see Methods). White indicates that the two residues do not interact while a darker red indicates a strong interaction. Residues are color labeled on the axis according to residue type (red: acidic, blue: basic, green: polar, grey: nonpolar). The $AQP0^{CBD}$ in CaM-free WT AQP0 (a) interacts primarily with R152 of the cytosolic loop, while CaM free AQP0-S235p (b), which displays a smaller CS-II opening (Figure 3c), develops an additional strong interaction between R233 and E151, a residue that we have identified as being part of the CaM gating mechanism (Figure 3). c) A snapshot from an MD simulation of CaM-free AQP0-S235p showing the salt-bridge interaction between R233 and E151. This interaction decreases the distance between Y149 and F75 of CS-II, breaking the single file of water molecules in the pore lumen (shown in filled-sphere representation). Figure adapted from Fig. 6 of Ref. [1].

4.3.3 AQP0 Arginine Loop Mutations Disrupt Calcium Sensitivity

Our MD simulations suggested that CaM modulates AQP0 through the arginine-rich cytosolic loop proximal to the gating Y149 residues. Prior research on the CaM-Ca^{2+} modulation of AQP0 has focused on the C-terminal $AQP0^{CBD}$ helix due to its known interaction with CaM, its similarity to other CaM-binding domains, and the accessibility of solution state NMR and mass spectroscopy experiments on the soluble peptide fragment [7, 20]. However, oocyte-swelling assays allowed us to experimentally measure the permeability of the entire AQP0 tetramer. My collogues expressed various arginine mutants of AQP0 into *X. laevis* oocytes, measuring the water permeabilities (see Fig. 4.6a) by monitoring cell swelling in response to hypotonic stress at differing external calcium concentrations. A Western blot assay shows that the AQP0 mutants are all expressed (see Fig. 4.6b).

Previous experimental results have shown that oocytes expressing WT AQP0 exhibit a two-fold increase in the water permeability rates when calcium is removed from the buffer [7, 9, 19]. We observed that the Ca^{2+} response of R152A is precisely that of WT; the water permeability at 2 mM Ca^{2+} is approximately 21 μm/s and doubles to 36 μm/s in 0 mM Ca^{2+} (see Fig. 4.6a). In contrast, Ca^{2+}-sensitivity is eliminated for both R153A and R156A. R153A locks the P_f low, whereas the P_f of R156A is locked high (see Fig. 4.6a).

Previous studies show that the Y149 residue of the CS-II acts as a dynamic gate for the control of channel water permeability by the action of bound CaM [7]. The different effects caused by these Arg-to-Ala mutations give clues to the specific actions of each of the three arginines. The locked-low permeability of R153A is consistent with our

observation that the R153 side chain sits near Y149 where it can open the CS-II. MD simulations show no substantial contacts formed between R153 and CaM like those formed between CaM and R152 and R156 (see Fig. 4.4b). But R153 interacted directly with the hydroxyl group of Y149 in our CaM-bound AQP0-S229P trajectory (see Fig. 4.4d). This accounts for the experimental result that the double mutant AQP0-R153A/S229D is no longer locked in a high P_f state like AQP0-S229D because R153 is no longer present to exert its influence on Y149. Thus AQP0-R153A/S229D exhibited a low P_f in 2 mM Ca^{2+}, similar to AQP0-R153A (see Fig. 4.6a). But the double mutant added a surprising twist; at 2 mM Ca^{2+}, its P_f displayed a novel zero-permeability state (see Fig. 4.6a).

Because R156A is locked in a high P_f state, we hypothesize that the interaction between R156 and CaM is critical for decreasing pore permeability. MD simulations revealed that residues R152 and R156 form strong salt bridge interactions with CaM (see Fig. 4.4a,b). However, R152A had no effect on the P_f of AQP0 (see Fig. 4.4b), suggesting that R152 contributes to the electrostatic interaction with CaM but does not facilitate CaM-Ca^{2+} modulation. Because R152 and R153 are closer in amino acid sequence to the gating Y149, it is surprising that the interaction between CaM and R156, however indirectly mediated, appeared to be essential to lower the permeability of AQP0-S229D (see Fig. 4.4a,b).

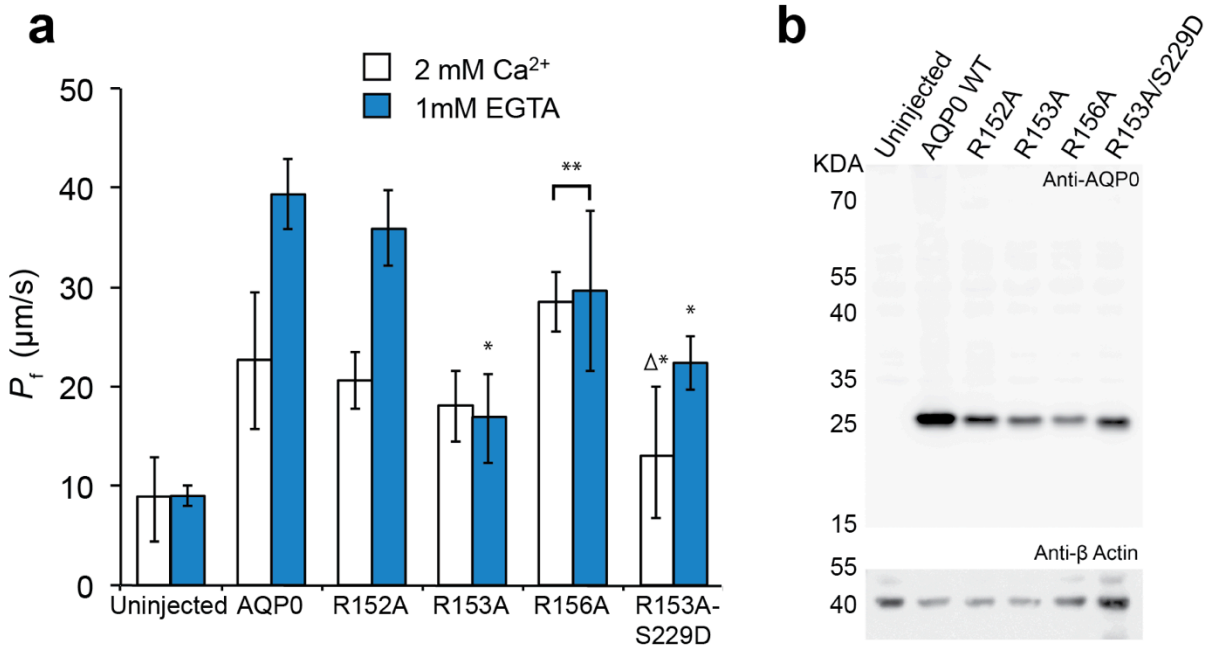

Figure 4.6: Oocyte Permeabilities of Arginine-Loop Mutants of AQP0. a) Water permeability rates (P_f in $\mu m \cdot s^{-1}$) obtained from oocytes expressing WT and various arginine loop mutants of AQP0 (R152A, R153A, R156A, and the double mutant R153A-S229D). Calcium buffer conditions of 2 mM Ca^{2+} (white) and 0 mM Ca^{2+} with 1 mM EGTA (blue) were used to probe the calcium-sensitivity of AQP0 P_f for each mutant and WT AQP0. R152A displays WT behavior, while the R153A and R156A mutants lock the pore in a low and elevated P_f state, respectively. The compound mutant R153A-S229D has an unusual, zero permeability state at 2 mM Ca^{2+}. b) Western blot analysis confirmed AQP0 protein expression resulting from RNA injected of WT AQP0, as well as all the mutants tested – R152A, R153A, R156A and R153A/S229D – at the expected molecular weight of 26 kDa, but not in uninjected oocytes (top panel). Protein loading in all lanes was confirmed by stripping the blot and re-probing for β-actin (bottom panel). Figure adapted from Fig. 7 of Ref. [1].

4.3.4 E151 Couples the Gating Y149 to CaM through R156

Our experimental and simulation results indicate that R156 mechanically couples CaM to the AQP0 CS-II. However, it remains unclear how R156 is able to control the CS-II dynamically when CaM forms significant interactions with residues closer to the gating Y149, such as R152, and it is also puzzling that the AQP0-R152A mutant displays water permeabilities similar to WT AQP0 (see Fig. 4.6a). To resolve this conundrum, we identified potential interaction networks by determining the shortest paths between R156 and Y149 in our chemical group graph representation (see Section 4.2.4). Interestingly, we find that the shortest path between R156 and Y149 of a single subunit is the trace of the backbone. Because CaM also contacts R152, it is unlikely that R156 directly couples CaM to Y149 through the backbone. When looking at intersubunit contacts, we find that R156 forms a salt bridge with the E151 side chain of an adjacent AQP0 subunit (see Fig. 4.7). The R156 interaction with E151, which is a single amino acid residue upstream of Y149, is dependent on the CaM D118 side chain to properly orient R156 (see Fig. 4.7). The breaking of the single file of water molecules in the lumen of the water pore coincides with the formation of this salt bridge (see Fig. 4.7). We hypothesize that the conformation of the gating Y149 residue is restricted by the formation of salt bridge interactions between E151 and R156 of an adjacent subunit, which are facilitated by CaM D118.

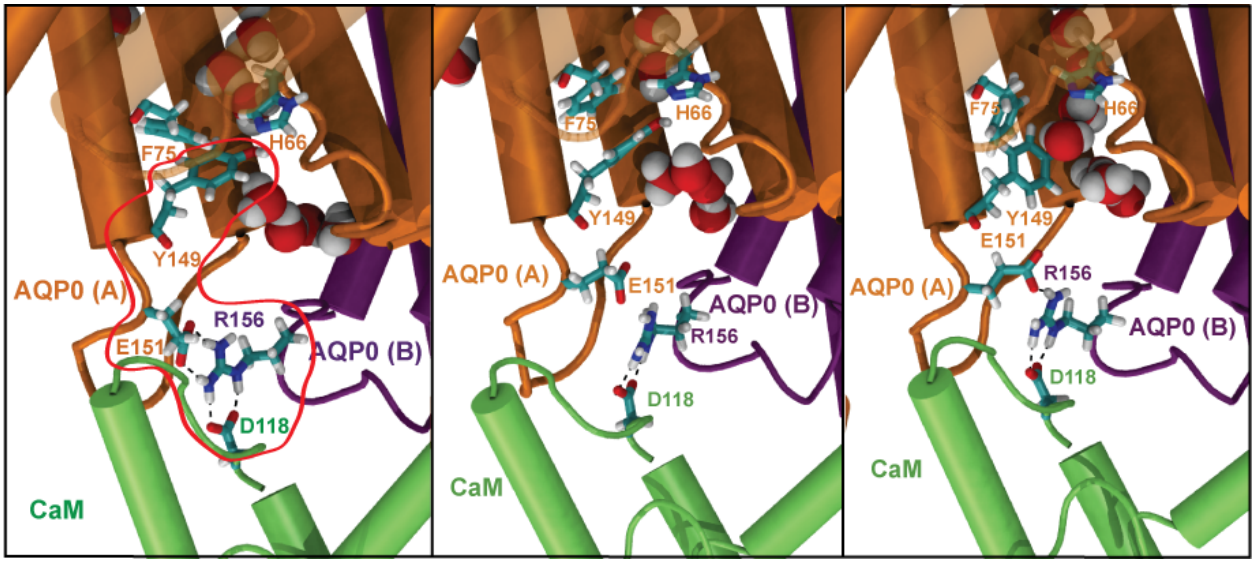

Figure 4.7: Electrostatic Interactions Couple CaM to the CS-II of AQP0. A chain of electrostatic interactions mechanically couples CaM to the pore-gating residues of AQP0. Snapshots from an MD simulation of CaM-bound WT AQP0 show an interaction chain, from D118 of CaM to R156 of AQP0 to E151 of an adjacent AQP0 subunit (highlighted by the red enclosure in the first panel), connects CaM to the gating Y149 residue of the CSII. Movement of the Y149 side chain in the third panel follows the breaking of the salt bridge between R156 and E151 in the second panel, resulting in the reformation of the single file of water molecules through cytosolic portion of the lumen. Figure adapted from Fig. 9 of Ref. [1].

4.4 Discussion

CaM reduces the water permeability of AQP0 in response to increases in the intracellular calcium concentration [9]. However, little is known of the molecular mechanisms by which CaM regulates AQP0 P_f or for that matter how it modulates the functions of other membrane proteins. We employed molecular simulations and biophysical permeability assays to elucidate the molecular mechanism by which CaM modulates AQP0 water permeability as well as how modifications (such as phosphorylation) are able to allosterically modify calcium sensitivity. Our data suggest critical and hitherto unexpected features of the AQP0-CaM complex that may also be important in CaM interactions with other regulated proteins. First, electrostatic interactions between CaM and the cytoplasmic face of AQP0 play a critical role. Second, a series of side chain interactions couples CaM to the CS-II of AQP0 through a cytosolic arginine-rich loop located near the center of the tetramer.

AQP0^{CBD} phosphorylation interferes with the calcium-sensitivity of AQP0, locking the P_f either high or low [19]. It has been hypothesized that phosphorylation reduces calcium-sensitivity by inhibiting or reducing CaM-binding to AQP0. Phosphorylation has been shown to inhibit CaM-binding to the AQP0^{CBD} peptide fragment [20], which lends credence to this hypothesis. However, our BD simulations, which utilize the full-length structures of AQP0 and CaM, suggest that the interfacial arginine residues R153 and R156 at the interface play a dominant role in the formation of AQP0-CaM complex, whereas phosphorylation alone produced little effect on the formation of the complex (see Table 4.1). The AQP0 charge mutants R152A and R153A/S229D maintain calcium sensitivity in oocyte permeability assays (see Fig. 4.6a) even though they can reduce

the positive-charge of the tetramer by as much as eight elementary charges. This means that CaM-binding occurs in both AQP0-R152A and AQP0-R153A/S229D and implies that phosphorylation of AQP0 disrupts calcium-sensitivity by altering the AQP0-CaM interaction interface and not by inhibiting CaM-binding. CaM is negatively-charged and has been experimentally shown to bind to neutral and positively-charged peptides but there appear to be no examples of charge mutations inhibiting CaM-binding to a full-length protein.

Our MD simulations predict that phosphorylation of the $AQP0^{CBD}$ changes the interactions that CaM makes with the arginine residues of the 150–156 cytosolic loop of AQP0. Particularly AQP0-S229P, which is locked in a high permeability state, disrupts the interaction between D118 of CaM with R156 of AQP0, shifting it toward E120. This shift causes the R153 residue to interact with the gating Y149 residues via its hydroxyl group, opening the CS-II of one of the four pores. Experimental oocyte swelling assays confirm that AQP0-R156A is locked in the high P_f state (see Fig. 4.6a). This suggests that the electrostatic interactions between the R156 side chain and CaM are critical for closing the CS-II in response to calcium, and AQP0-S229P has a locked-high P_f because it is able to disrupt this interaction (see Fig. 4.6a). Using a shortest-path algorithm to analyze our MD trajectories of CaM-bound AQP0, we found that R156 simultaneously forms a salt bridge with D151 of an adjacent AQP0 subunit and D118 of CaM. We hypothesize that, through this interaction network, CaM mechanically couples to the backbone atoms of AQP0 D151, restricting the conformational dynamics of the nearby Y149 side chain to block the channel waters from accessing the cytosolic opening.

Conversely, R153A exhibits a locked-low P_f, suggesting that the positively-charged side chain is critical for maintaining high water permeability states at low calcium concentrations. MD simulations of WT AQP0 (CaM) show that CaM rarely forms salt bridges with R153 (see Fig. 4.3 and Fig. 4.4). Instead, the guanidine group of R153 sits near the pore opening, where it can interact with the hydroxyl oxygen of the Y149 side chain. MD simulations show that S229P causes R156 to shift its salt bridge from D118 to E120 of CaM, in turn causing R153 to interact with the hydroxyl group of Y149. This is remarkably consistent with our experimental result that AQP0-R153A/S229D is water-impermeable at normal calcium concentrations. The alanine side chain of this mutant can no longer interact with the Y149 hydroxyl group to move it out of the pore opening, which is likely why AQP0-S229D exhibits locked-high permeability even in the presence of unmodified R153. Additionally, R153 and E151 play a role in the reduced P_f of AQP0-S235D at 0 mM Ca^{2+}. MD simulations of CaM-free AQP0-S235P revealed additional contacts formed between the $AQP0^{CBD}$ and the cytosolic loop residues E151 and R153. These contact changes coincide with an increase in the CS-II opening (see Fig. 4.4b and Fig. 4.5). This suggests that $AQP0^{CBD}$ modifications can alter P_f in a CaM-independent manner, similar to CaM-bound AQP0, in which the $AQP0^{CBD}$ interacts with the cytosolic loop, decreasing the size of the CS-II opening (see Fig. 4.5c and Fig. 4.7).

Calmodulin is able to bind and modulate the functions of a diverse set of proteins in remarkably different ways. Our study has shown that the 150–156 arginine-rich loop region of AQP0 couples CaM to the pore-opening, modulating the channel permeability. The arginine-rich loop region is not only conserved among fish and mammalian AQP0s, but similar loops can be found in other AQPs [5, 55-59]. Our results emphasize the

importance of characterizing regulatory proteins like CaM with the complete target protein and not just a peptide derived from the target protein. As we have shown, whole protein-CaM studies can lead to some unexpected surprises. Future studies will likely expand the roles of these charged loops and electrostatic interactions in the cytosolic modulation of AQP water permeability by CaM and perhaps discover similar interactions between CaM and other types of proteins.

4.5 Bibliography

1. Fields, J. B., Nemeth-Cahalan, K. L., Freites, J. A., Vorontsova, I., Hall, J. E., and Tobias, D. J. (2017) Calmodulin Gates Aquaporin 0 Permeability through a Positively Charged Cytoplasmic Loop. *J Biol Chem* **292**, 185-195

2. Gonen, T., and Walz, T. (2006) The structure of aquaporins. *Q Rev Biophys* **39**, 361-396

3. Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J. B., Engel, A., and Fujiyoshi, Y. (2000) Structural determinants of water permeation through aquaporin-1. *Nature* **407**, 599-605

4. Hashido, M., Ikeguchi, M., and Kidera, A. (2005) Comparative simulations of aquaporin family: AQP1, AQPZ, AQP0 and GlpF. *FEBS Lett* **579**, 5549-5552

5. Gonen, T., Sliz, P., Kistler, J., Cheng, Y., and Walz, T. (2004) Aquaporin-0 membrane junctions reveal the structure of a closed water pore. *Nature* **429**, 193-197

6. de Groot, B. L., Frigato, T., Helms, V., and Grubmuller, H. (2003) The mechanism of proton exclusion in the aquaporin-1 water channel. *J Mol Biol* **333**, 279-293

7. Reichow, S. L., Clemens, D. M., Freites, J. A., Nemeth-Cahalan, K. L., Heyden, M., Tobias, D. J., Hall, J. E., and Gonen, T. (2013) Allosteric mechanism of water-channel gating by Ca2+-calmodulin. *Nat Struct Mol Biol* **20**, 1085-1092

8. Bloemendal, H., Zweers, A., Vermorken, F., Dunia, I., and Benedetti, E. L. (1972) The plasma membranes of eye lens fibres. Biochemical and structural characterization. *Cell Differ* **1**, 91-106

9. Nemeth-Cahalan, K. L., Kalman, K., and Hall, J. E. (2004) Molecular basis of pH and Ca2+ regulation of aquaporin water permeability. *J Gen Physiol* **123**, 573-580

10. Rabaud, N. E., Song, L., Wang, Y., Agre, P., Yasui, M., and Carbrey, J. M. (2009) Aquaporin 6 binds calmodulin in a calcium-dependent manner. *Biochem Biophys Res Commun* **383**, 54-57

11. Golestaneh, N., Fan, J., Zelenka, P., and Chepelinsky, A. B. (2008) PKC putative phosphorylation site Ser235 is required for MIP/AQP0 translocation to the plasma membrane. *Mol Vis* **14**, 1006-1014

12. Girsch, S. J., and Peracchia, C. (1991) Calmodulin interacts with a C-terminus peptide from the lens membrane protein MIP26. *Curr Eye Res* **10**, 839-849

13. Yuan, T., and Vogel, H. J. (1998) Calcium-calmodulin-induced dimerization of the carboxyl-terminal domain from petunia glutamate decarboxylase. A novel calmodulin-peptide interaction motif. *J Biol Chem* **273**, 30328-30335

14. Reichow, S. L., and Gonen, T. (2008) Noncanonical binding of calmodulin to aquaporin-0: implications for channel regulation. *Structure* **16**, 1389-1398

15. Lin, H., Hejtmancik, J. F., and Qi, Y. (2007) A substitution of arginine to lysine at the COOH-terminus of MIP caused a different binocular phenotype in a congenital cataract family. *Mol Vis* **13**, 1822-1827

16. Gold, M. G., Reichow, S. L., O'Neill, S. E., Weisbrod, C. R., Langeberg, L. K., Bruce, J. E., Gonen, T., and Scott, J. D. (2012) AKAP2 anchors PKA with aquaporin-0 to support ocular lens transparency. *EMBO Mol Med* **4**, 15-26

17. Garland, D. L., Duglas-Tabor, Y., Jimenez-Asensio, J., Datiles, M. B., and Magno, B. (1996) The nucleus of the human lens: demonstration of a highly characteristic protein pattern by two-dimensional electrophoresis and introduction of a new method of lens dissection. *Exp Eye Res* **62**, 285-291

18. Ball, L. E., Garland, D. L., Crouch, R. K., and Schey, K. L. (2004) Post-translational modifications of aquaporin 0 (AQP0) in the normal human lens: spatial and temporal occurrence. *Biochemistry* **43**, 9856-9865

19. Kalman, K., Nemeth-Cahalan, K. L., Froger, A., and Hall, J. E. (2008) Phosphorylation determines the calmodulin-mediated Ca2+ response and water permeability of AQP0. *J Biol Chem* **283**, 21278-21283

20. Rose, K. M., Wang, Z., Magrath, G. N., Hazard, E. S., Hildebrandt, J. D., and Schey, K. L. (2008) Aquaporin 0-calmodulin interaction and the effect of aquaporin 0 phosphorylation. *Biochemistry* **47**, 339-347

21. Ermak, D. L., and Mccammon, J. A. (1978) Brownian Dynamics with Hydrodynamic Interactions. *J Chem Phys* **69**, 1352-1360

22. Gabdoulline, R. R., and Wade, R. C. (1998) Brownian dynamics simulation of protein-protein diffusional encounter. *Methods-a Method Enzymol* **14**, 329-341

23. Martinez, M., Bruce, N. J., Romanowska, J., Kokh, D. B., Ozboyaci, M., Yu, X. F., Ozturk, M. A., Richter, S., and Wade, R. C. (2015) SDA 7: A modular and parallel implementation of the simulation of diffusional association software. *J Comput Chem* **36**, 1631-1645

24. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-10041

25. Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* **30**, 1545-1614

26. Novotny, J., Bruccoleri, R. E., Davis, M., and Sharp, K. A. (1997) Empirical free energy calculations: a blind test and further improvements to the method. *J Mol Biol* **268**, 401-411

27. Gabdoulline, R. R., and Wade, R. C. (2009) On the contributions of diffusion and thermal activation to electron transfer between Phormidium laminosum plastocyanin and cytochrome f: Brownian dynamics simulations with explicit modeling of nonpolar desolvation interactions and electron transfer events. *J Am Chem Soc* **131**, 9230-9238

28. Gabdoulline, R. R., and Wade, R. C. (2001) Protein-protein association: investigation of factors influencing association rates by brownian dynamics simulations. *J Mol Biol* **306**, 1139-1155

29. Ortega, A., Amoros, D., and Garcia de la Torre, J. (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys J* **101**, 892-898

30. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38

31. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802

32. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102**, 3586-3616

33. MacKerell, A. D., Jr., Feig, M., and Brooks, C. L., II. (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-

phase quantum mechanics in reproducing conformational distributions in molecular dynamics simulations. *J Comput Chem* **25**, 1400-1415

34. Best, R. B., Mittal, J., Feig, M., and MacKerell, A. D., Jr. (2012) Inclusion of many-body effects in the additive CHARMM protein CMAP potential results in enhanced cooperativity of alpha-helix and beta-hairpin formation. *Biophys J* **103**, 1045-1051

35. Klauda, J. B., Venable, R. M., Freites, J. A., Connor, J. W. O., Tobias, D. J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell Jr., A. D., and Pastor, R. W. (2010) Update of the CHARMM All-Atom Additive Force Fields for Lipids: Validation on Six Lipid Types. *J Phys Chem* **114**, 7830-7843

36. Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994) Constant-Pressure Molecular-Dynamics Algorithms. *J Chem Phys* **101**, 4177-4189

37. Feller, S. E., Zhang, Y. H., Pastor, R. W., and Brooks, B. R. (1995) Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *J Chem Phys* **103**, 4613-4621

38. Rankin, W. T., and Board, J. A. (1995) A portable distributed implementation of the parallel multipole tree algorithm. in *High Performance Distributed Computing, 1995., Proceedings of the Fourth IEEE International Symposium on*

39. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A Smooth Particle Mesh Ewald Method. *J Chem Phys* **103**, 8577-8593

40. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J Comput Phys* **23**, 327-341

41. Flyvbjerg, H., and Petersen, H. G. (1989) Error estimates on averages of correlated data. *J Chem Phys* **91**, 461-466

42. Benson, N. C., and Daggett, V. (2012) A chemical group graph representation for efficient high-throughput analysis of atomistic protein simulations. *J Bioinform Comput Biol* **10**, 1250008

43. Wolfram Research, I. (2015) Mathematica. 10.1 Ed., Wolfram Research, Inc., Champaign, Illinois

44. Cormen, T. H. (2009) *Introduction to algorithms*, 3rd ed., MIT Press, Cambridge, Mass.

45. Team, R. D. C. (2008) R: A Language and Environment for Statistical Computing. Vienna, Austria

46. Frembgen-Kesner, T., and Elcock, A. H. (2010) Absolute protein-protein association rate constants from flexible, coarse-grained Brownian dynamics simulations: the role of intermolecular hydrodynamic interactions in barnase-barstar association. *Biophys J* **99**, L75-77

47. Getzoff, E. D., Cabelli, D. E., Fisher, C. L., Parge, H. E., Viezzoli, M. S., Banci, L., and Hallewell, R. A. (1992) Faster superoxide dismutase mutants designed by enhancing electrostatic guidance. *Nature* **358**, 347-351

48. Zhou, H. X. (1993) Brownian dynamics study of the influences of electrostatic interaction and diffusion on protein-protein association kinetics. *Biophys J* **64**, 1711-1726

49. Schilder, J., and Ubbink, M. (2013) Formation of transient protein complexes. *Curr Opin Struct Biol* **23**, 911-918

50. Fields, J. B., Hollingsworth, S. A., Chreifi, G., Heyden, M., Arce, A. P., Magaña-Garcia, H. I., Poulos, T. L., and Tobias, D. J. (2015) “Bind and Crawl” Association Mechanism of Leishmania major Peroxidase and Cytochrome c Revealed by Brownian and Molecular Dynamics Simulations. *Biochemistry* **54**, 7272-7282

51. Gorba, C., Geyer, T., and Helms, V. (2004) Brownian dynamics simulations of simplified cytochrome c molecules in the presence of a charged surface. *J Chem Phys* **121**, 457-464

52. Flock, D., and Helms, V. (2004) A Brownian dynamics study: the effect of a membrane environment on an electron transfer system. *Biophys J* **87**, 65-74

53. Yap, K. L., Yuan, T., Mal, T. K., Vogel, H. J., and Ikura, M. (2003) Structural basis for simultaneous binding of two carboxy-terminal peptides of plant glutamate decarboxylase to calmodulin. *J Mol Biol* **328**, 193-204

54. Jensen, M. O., Dror, R. O., Xu, H., Borhani, D. W., Arkin, I. T., Eastwood, M. P., and Shaw, D. E. (2008) Dynamic control of slow water transport by aquaporin 0: implications for hydration and junction stability in the eye lens. *Proc Natl Acad Sci U S A* **105**, 14430-14435

55. Gonen, T., Cheng, Y., Sliz, P., Hiroaki, Y., Fujiyoshi, Y., Harrison, S. C., and Walz, T. (2005) Lipid-protein interactions in double-layered two-dimensional AQP0 crystals. *Nature* **438**, 633-638

56. Yu, Y., Yu, Y., Chen, P., Li, J., Zhu, Y., Zhai, Y., and Yao, K. (2014) A novel MIP gene mutation associated with autosomal dominant congenital cataracts in a Chinese family. *BMC Med Genet* **15**, 6

57. Frick, A., Eriksson, U. K., de Mattia, F., Oberg, F., Hedfalk, K., Neutze, R., de Grip, W. J., Deen, P. M., and Tornroth-Horsefield, S. (2014) X-ray structure of

human aquaporin 2 and its implications for nephrogenic diabetes insipidus and trafficking. *Proc Natl Acad Sci U S A* **111**, 6305-6310

58. Sui, H., Han, B. G., Lee, J. K., Walian, P., and Jap, B. K. (2001) Structural basis of water-specific transport through the AQP1 water channel. *Nature* **414**, 872-878

59. Kitchen, P., Oberg, F., Sjohamn, J., Hedfalk, K., Bill, R. M., Conner, A. C., Conner, M. T., and Tornroth-Horsefield, S. (2015) Plasma Membrane Abundance of Human Aquaporin 5 Is Dynamically Regulated by Multiple Pathways. *PLoS One* **10**, e0143027

Chapter 5

Protonation of H40 of Aquaporin 0 opens the Aromatic-Arginine Selectivity Filter

In this chapter, I will describe the results of unpublished work on the pH-sensitivity mechanism of Aquaporin 0 (AQP0).

5.1 Introduction

In the previous chapter, I described results of simulations of the AQP0-CaM complex that elucidated the molecular mechanism of the regulation of AQP0 water permeability (P_f) by calcium and phosphorylation. In this chapter, I will turn the focus on the pH-dependence of the AQP0 P_f. Of the 13 AQP isoforms, AQP0, AQP3, AQP4, AQP6, and AQP10 exhibit pH-dependent permeabilities [1-8]. Results from Xenopus oocyte swelling assays show that the P_f of AQP0 increases two-fold at a slightly acidic pH of 6.5 [1, 2]. Our collaborators have identified a critical histidine residue, H40, which is responsible for this observed pH-dependence of AQP0 [2]. The amino acid histidine contains a pH-sensitive imidazole-functionalized side chain [9, 10]. H40 caps the C-terminal end of the

first extracellular loop (loop A), which connects the first and second trans-membrane helices. Histidine has two pKa values near physiological-pH, meaning that under varying proton concentrations, solvated histidine residues can change from positively-charged to neutral to negatively-charged [11-13]. Mutation of H40 to arginine, which removes this pH-sensitive imidazole and replaces it with a positively-charged guanidine, kills the pH-sensitivity, locking the P_f high [2]. An H to C mutation shifts the P_f increase from acidic to basic pH (8.5), presumably by deprotonation of the side chain thiol [2]. Furthermore, incorporating a histidine residue on the corresponding Loop A region of other AQPs results in novel pH-sensitivity [2]. This body of evidence suggests that the charge of the C-terminal end of the Aquaporin loop A plays a critical role in modulating the water permeability in response to changes in extracellular pH.

To better understand how H40 contributes to AQP0 pH-sensitivity, I computed multi-microsecond equilibrium MD simulations to calculate the single channel P_f values of AQP0. Three separate trajectories of the AQP0-CaM complex (PDB code: 3J41 from Ref. [14]) embedded in a POPC bilayer were produced using the Anton 2 supercomputer (Pittsburgh Supercomputing Center) [15]. The trajectories were computed using two different H40 protonation states; the diprotonated, positively-charged state (which I will from here on refer to as the protonated state) and the monoprotonated, uncharged state. The third trajectory incorporated H40K mutations to the AQP0 tetramer, which should, in principle, behave similarly to protonated H40. Even though there are other modifiable histidine residues on the extracellular face of AQP0, I left those unmodified

because those residues have little influence on the pH-dependence of AQP0 P_f in experiment [2].

In the previous chapter, I discussed the importance of interfacial arginine residues to the CaM-mediated modulation of the second constriction site (CS-II). The results I will discuss in this chapter highlight a critical arginine residue, R187, which makes up the narrower first constriction site (CS-I) [16, 17]. I observed that the position of R187 can be manipulated by H40 protonation through repulsive electrostatic interactions [18]. The CS-I is sometimes referred to as the aromatic arginine selectivity filter (Ar/R). It is highly conserved amongst AQPs, containing an arginine and two aromatic residues, which help to exclude larger molecules or positively-charged species such as metallic ions and hydronium from entering the pore [19]. We observe that when H40 is positively-charged, it can electrostatically repel R187 away from its aromatic neighbor, opening the pore. Hydrogen-bonding interactions between the side chains of R187 and N119 help to stabilize the open state, a principle supported by earlier findings on the loop C side chain [18].

5.2 Methods

5.2.1 Atomistic MD Setup

To better understand the effects of the H40 protonation state on the structure and function of AQP0, I performed atomistic, equilibrium MD simulations of the AQP0-CaM complex (PDB code: 3J41 from Ref. [14]) embedded in a solvated

POPC bilayer (the same MD system from Ch. 4). Two trajectories were run using different protonation states on H40. The HSP residue was used for H40 of the AQP0-H40$^+$ (CaM) system and the HSD residue was used for H40 of the WT AQP0 (CaM) system [20]. An additional system containing AQP0-H40K was also produced. Each system was composed of an AQP0 tetramer (residues 5–239) complexed with 2 CaM monomers bound to 8 Ca^{2+} ions, embedded in a membrane of 410 POPC lipid molecules and 56,701 waters. Chloride counterions were added to neutralize the AQP0-H40$^+$ system.

The MD simulations were initialized using the NAMD 2.9 software package [21]. The CHARMM36 force field was used for proteins [22-24] and lipids [25], and the TIP3P model was used for water [20]. The system was subjected to 2000 steps of conjugate gradient energy minimization before 2 nanoseconds of equilibrium MD simulation. The simulations were transferred to the Anton2 supercomputer (PSC supercomputing resources) to compute multi-microsecond trajectories for both systems. The r-RESPA algorithm was employed to integrate the equations of motion with a time step of 6 femtoseconds for the long-range non-bonded interactions, and a 2 femtoseconds time step for short-range non-bonded and bonded interactions [26]. The k-Gaussian split Ewald method was used for the long-range and electrostatic interactions [27]. All bonds involving hydrogen were constrained using the SHAKE algorithm [28]. The simulations were performed using the NPT ensemble, using a constant pressure of 1 atm and temperature of 300 K, using Nose-Hoover chains and the Martyna-Tobias-Klein barostat [29, 30]. The RESPA algorithm and temperature and pressure controls were

implemented using the Multigrator scheme [31]. The simulations of WT, $H40^+$, and H40K systems were run for approximately 5 μs. The VMD 1.9.1 software package was used for visualization and analysis of the trajectories [32]. This included the calculation of the side chain dihedral angles (χ_1 and χ_2), which were calculated using the "measure dihed" function in VMD.

5.2.2 Collective Diffusion Model of Water Permeability

The single-channel osmotic permeability constants (P_f) were calculated using the well-established collective diffusion model [33]. The derivation I describe below follows the one described in Ref. [33]. The flux of water molecules across a channel (J) is related to the osmotic pressure gradient (ΔP) by P_f [34]:

$$J = P_f \frac{\Delta P}{RT} \tag{5.1}$$

In order to calculate P_f, the collective diffusion model defines a collective variable (n) that describes the translocation of water molecules through the single-file region (or lumen) of the pore. I used the AQP0 lumen definitions set forth by Jensen *et al* who define the lumen as the region between the N_ε atom of R187 and O atom of A65 [35]. The unit-less collective variable (n) can be solved by integrating (Eq. 3 from Ref. [33]):

$$dn = \sum_{i \in S(t)} \frac{dz_i}{L} \tag{5.2}$$

where the instantaneous displacements (dz_i) of the atoms in the lumen ($S(t)$) are

summed and divided by the lumen length (L). Note that $S(t)$ fluctuates with time. This means that $i \in S(t)$ only contributes to dn if atom i remains in the lumen after an incremental time. The corresponding diffusion coefficient of the collective variable n (D_n) obeys the following Einstein relation (Eq. 5 from Ref. [33]):

$$\langle n^2(t) \rangle = 2D_n t \tag{5.3}$$

where $\langle n^2(t) \rangle$ is the mean square displacement (MSD) of n. The P_f is simply the diffusion coefficient (D_n; seen in Eq. 5.3) multiplied by the average volume of a single water molecule (v_W) (Eq. 9 from Ref. [33]):

$$P_\mathrm{f} = v_w D_n \tag{5.4}$$

Although water flux (J) varies with the osmotic pressure gradient ($\Delta\mu$), P_f is, in principle, invariant between equilibrium and nonequilibrium conditions alike.

5.3 Results

5.3.1 H40 Protonation Does Not Affect the Second Transmembrane Helix

Experiments have shown that the pH- and calcium-gating mechanisms of AQP0 are coupled. Some evidence suggests that these mechanisms are coupled dynamically through the second transmembrane helix (TM2) [2, 3, 36]. The H40K and H61A mutations lock the AQP0 permeability (P_f) high with respect to calcium and pH. These residues cap the N-terminal and C-terminal ends of TM2,

respectively, leading to the hypothesis that this helix mediates the observed modification. In order to distinguish the TM2 helix in the two trajectories, I measured the transmembrane position, RMSD, RMSF per residue, and helix tilt angle of the TM2 (see Fig. 5.1). If the TM2 helix plays a dynamic role in the pH-modulation mechanism of AQP0, the helix position and orientation would likely by modified by protonation of H40. The first thing that is apparent from the WT backbone RMSD (see Fig. 5.1a) is that the TM2 helix is stable. The TM2 backbone RMSD in both the $H40^+$ and the WT systems remains low (<1-Å) throughout the multi-microsecond trajectories. Because the TM2 helix is embedded deep in the protein with significant inter-domain contacts, it is unlikely that a terminal point mutation could drastically change its secondary structure.

Protonation of H40 does not alter the depth or tilt angle of the TM2 helix relative to the lipid bilayer (see Fig. 5.1d,e). Additionally, the RMSF per residue shows a negligible difference between WT and $H40^+$ in the TM2 region (see Fig. 5.1b). One interesting observation from the RMSF is that the region around residues 185 to 200 seems to be stabilized by H40 protonation (see Fig. 5.1b). We will discuss the importance of this region to the pH-gating mechanism later (see section 5.3.3). Thus there is no obvious mechanism for the TM2 helix to couple the pH- and calcium-sensitivity mechanisms of AQP0.

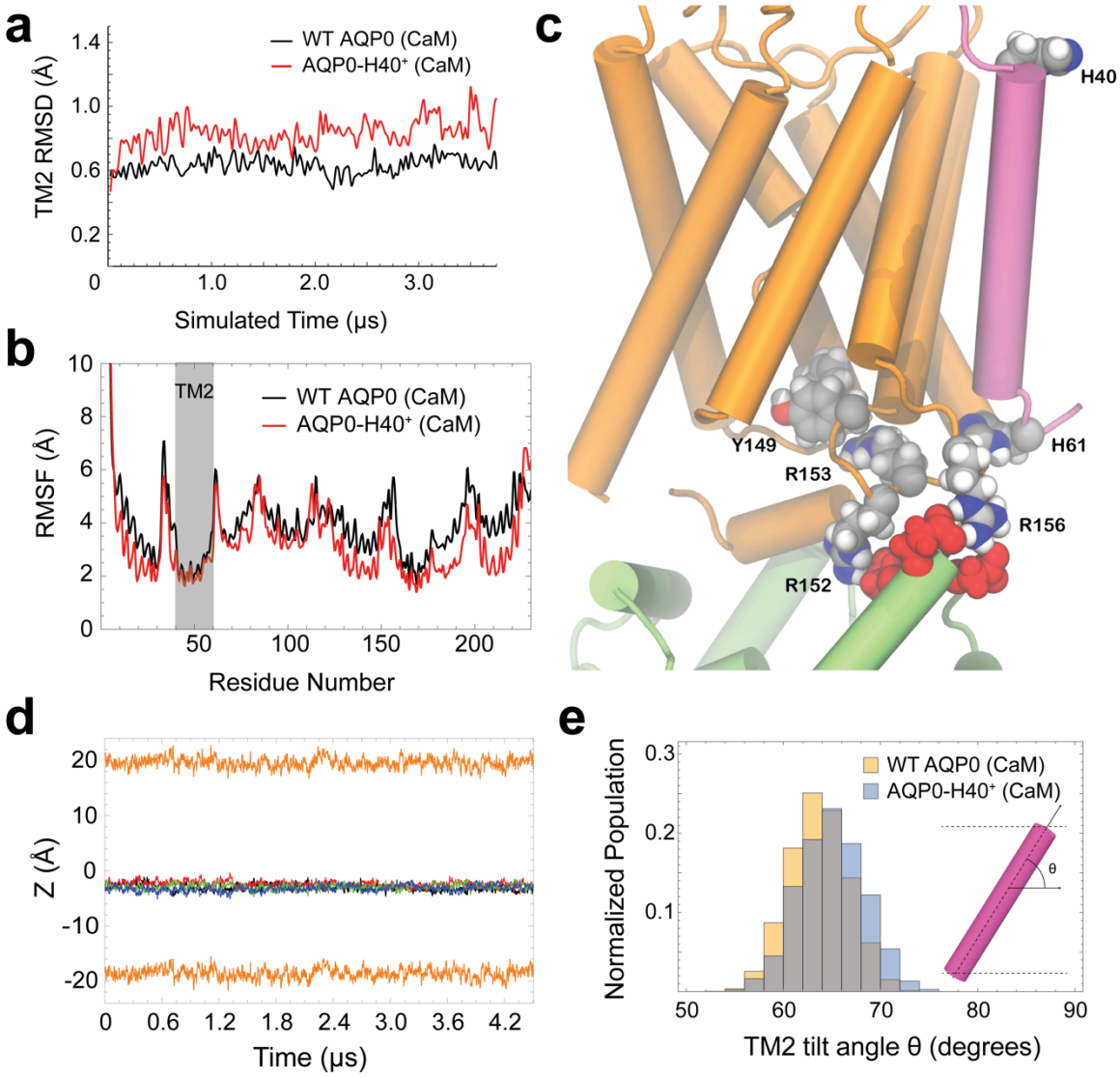

Figure 5.1: Position and Dynamics of the TM2 Helix are Not Modified by H40 Protonation. a) The backbone RMSD of the TM2 for both WT and $H40^+$ systems (shown in black and red, respectively) shows little difference between them. b) The per residue RMSF calculated for both WT and $H40^+$ AQP0 (CaM) systems (shown in black and red, respectively), again showing little difference in the TM2 region (shown in grey, residue numbers 40 to 60). c) A cartoon representation of an AQP0 subunit (shown in orange) with bound-CaM (shown in green) and the TM2 helix on an adjacent AQP0 monomer (shown in purple). H61 interacts with important residues on CaM, which were thought to be important to the pH-sensitivity mechanism through the TM2. d) The z-coordinate of the center-of-mass of the second transmembrane helix (TM2) from the four monomer of AQP0-$H40^+$ (CaM) system (shown by the black, blue, green, and red curves). The average z-coordinate of lipid phosphate atoms of the lower and upper leaflets is shown (the orange curves) for reference. e) The helix tilt angle (θ) of the TM2 for both WT AQP0 (CaM) and AQP0-$H40^+$ (CaM) (shown in yellow and blue, respectively; gray denotes overlapping areas of the histograms).

5.3.2 H40 Protonation Increases the Water Permeability and Water Density in the AQP0 Pores

Osmotic water permeability calculations give us a critical tool for comparing our atomistic simulations with experiment. I used the collective diffusion method developed by Zhu *et al* [33] to calculate the osmotic permeability coefficients (P_f) for each monomeric pore in the AQP0 tetramer. The mean permeability values for the four pores are given for WT and $H40^+$ AQP0 (see Table 5.1), with standard deviation to show the convergence. Because AQP0 has a very low permeability relative to other AQPs, it can be difficult to get convergent values, as small differences in the dynamics of the gating residues can lead to large permeability differences [18, 35, 37]. The reported standard deviations (see Table 5.1) in the calculated permeabilities are below 35%, indicating that the multi-microsecond trajectories were long enough to attain reasonable sampling.

AQP0 Variant	Calculated P_f (10^{-15}cm^3 s^{-1})
WT (CaM)	1.182 ± 0.179
$H40^+$ (CaM)	1.706 ± 0.419
H40K (CaM)	1.642 ± 0.526

Table 5.1: Calculated Osmotic Permeability Constants (P_f) from AQP0-CaM MD Trajectories.

My calculations show that protonation of H40 increased the water permeability of AQP0 (see Table 5.1). This increase is quantitatively similar to the observed increase in the experimental AQP0 P_f at pH 6.5. Additionally, the H40K (CaM) system displays a higher P_f than WT, which is consistent with experimental measurements [2]. In both the $H40^+$ (CaM) and H40K (CaM) systems, the

residue at position 40 has a positively-charged side chain. In an attempt to understand how this positive-charge increases P_f, I calculated the free energy profile of water molecules in the pore by taking the natural log of the water density through the pores (see Fig. 5.2), comparing the profiles of the WT AQP0, WT AQP0 (CaM), and AQP0-H40$^+$ (CaM) systems. A cylindrical volume of a 5-Å radius that encapsulates the pore waters was defined to make the water selection for the density profile. Comparing WT AQP0 and WT AQP0 (CaM), a small increase in the free energy barrier of water molecules in the CS-II is correlated with bound-CaM (see Fig. 5.2). This is consistent with previous results (see Ch. 4) that point to the CS-II as the CaM-dependent gate of P_f [14]. When comparing the free energy profiles of WT and protonated H40 AQP0, we see very little difference in the CS-II region (see Fig. 5.2). However, a softening of the larger CS-I barrier can be seen in the AQP0-H40$^+$ (CaM) energy profile. This suggests that the H40 protonation state influences the density of water molecules in the channel by manipulating the state of the CS-I from the extracellular-side of the channel.

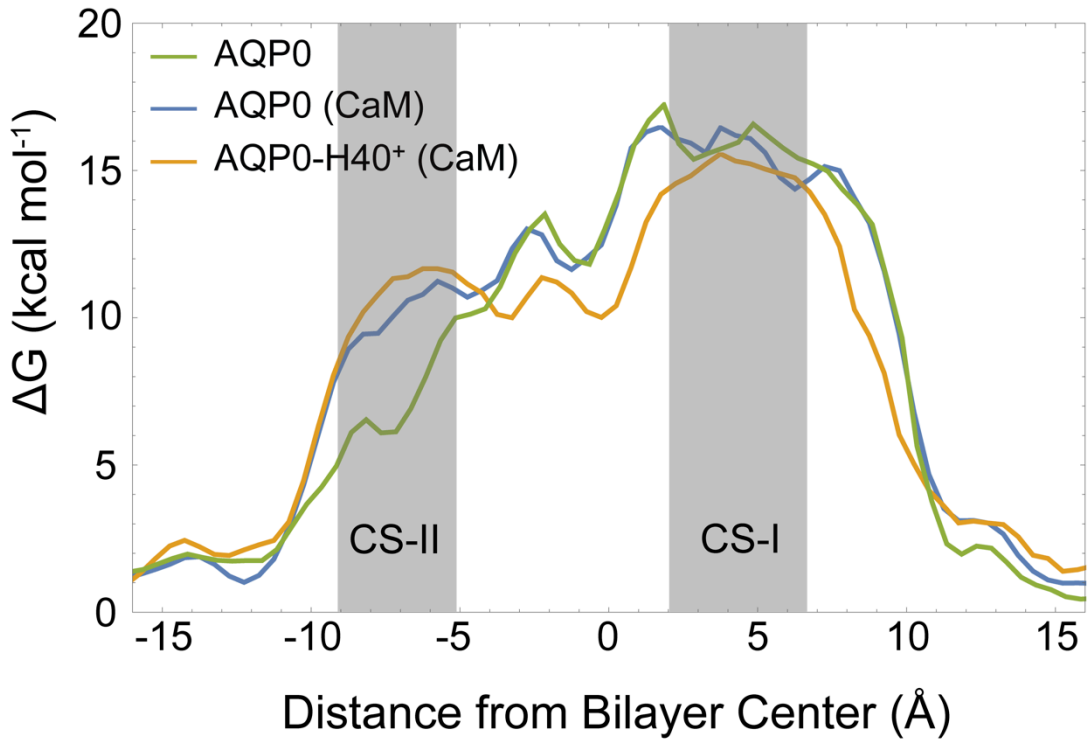

Figure 5.2: The Free Energy Profile of Water Molecules in the AQP0 Pore. The free energy (ΔG) is calculated by taking the negative log of the water density profile through a cylindrical (radius of 5-Å) volume along the pore axis. The density profiles are centered so that the bilayer center is fixed at the origin. The legend in the upper right corner shows the three systems measured, which include WT AQP0 (in green), WT AQP0 (CaM) (in blue), and AQP0-H40^{+} (CaM) (in gold). The first and second constriction site (CS-I and CS-II) regions of the pore axis are indicated by the grey-labeled boxes.

5.3.3 Electrostatic Repulsion Between H40 Opens the First Constriction Site through R187

Previous studies have proposed certain residues that could play a role in the observed pH-dependence of AQP0, including R33, W34, H172, and R187 [18, 35]. In my WT-AQP0 trajectory, R33 and H40 form hydrogen-bonding interactions that are depleted by H40 protonation. This is caused by repulsive electrostatic interactions between the positively-charged side chains of R33 and H40. However, these residues are far removed from the CS-I where we observe an increased water density resulting from H40 protonation (see Fig. 5.2).

AQP0 Variant	R187-H(K)40 Electrostatic PE (kcal mol^{-1})	R187-R33 Electrostatic PE (kcal mol^{-1})	R187-Loop A Electrostatic PE (kcal mol^{-1})
WT (CaM)	-0.41 ± 0.26	1.25 ± 1.42	0.84 ± 1.68
H40^{+} (CaM)	1.27 ± 0.65	0.60 ± 0.71	1.87 ± 1.36
H40K (CaM)	0.30 ± 0.37 (K)	1.44 ± 1.20	1.74 ± 1.57

Table 5.2: The Mean and Standard Deviation of the R187-Loop A Electrostatic Potential Energy.

To compare the CS-I of the WT and H40^{+} systems, I calculated the first side chain dihedral angles (χ_1 and χ_2; see Fig. 5.3) of the residues that make up the CS-I. The residues F48, H172, and R187 form the CS-I of AQP0. Of these residues, the H172 dihedrals are modified most significantly in the H40^{+} trajectory, showing two distinct peaks that are not present in WT (see Fig. 5.3). This corresponds to the rotation of the χ_1 towards the membrane center. There is no obvious cause other than the increased water density in H40^{+} (see Fig. 5.2), as the H172 side chain is uncharged and would not interact significantly with the

H40 of AQP0-H40$^+$. The side chains of F48 and R187 exhibit only small angle changes (see Fig. 5.3b,d). Out of the three residues of the CS-I, R187 is the only one that has a positively-charged side chain that can interact with H40$^+$. It should be noted that arginine has additional dihedrals χ_3 and χ_4 that are not shown in these simple Ramachandran plots. Thus, the confirmations of R187 warranted further examination.

A previous computational study examined the effects of mutation of proximal residues to R187. They showed that the fluctuations in the position of the R187 side chain make the largest contributions to the pore radius of the CS-I [18]. They observed that the AQP0 extracellular Loop C sits proximal to R187 when it was in a more open configuration. The Loop C residues were systematically mutated *in silico* to either aspartate or glutamate with the intention of stabilizing open configurations of R187. It should be noted that the only residue of Loop C that interacts with R187 is D119. The authors reported that an N119D mutation locks the CS-I gate in an open state [18].

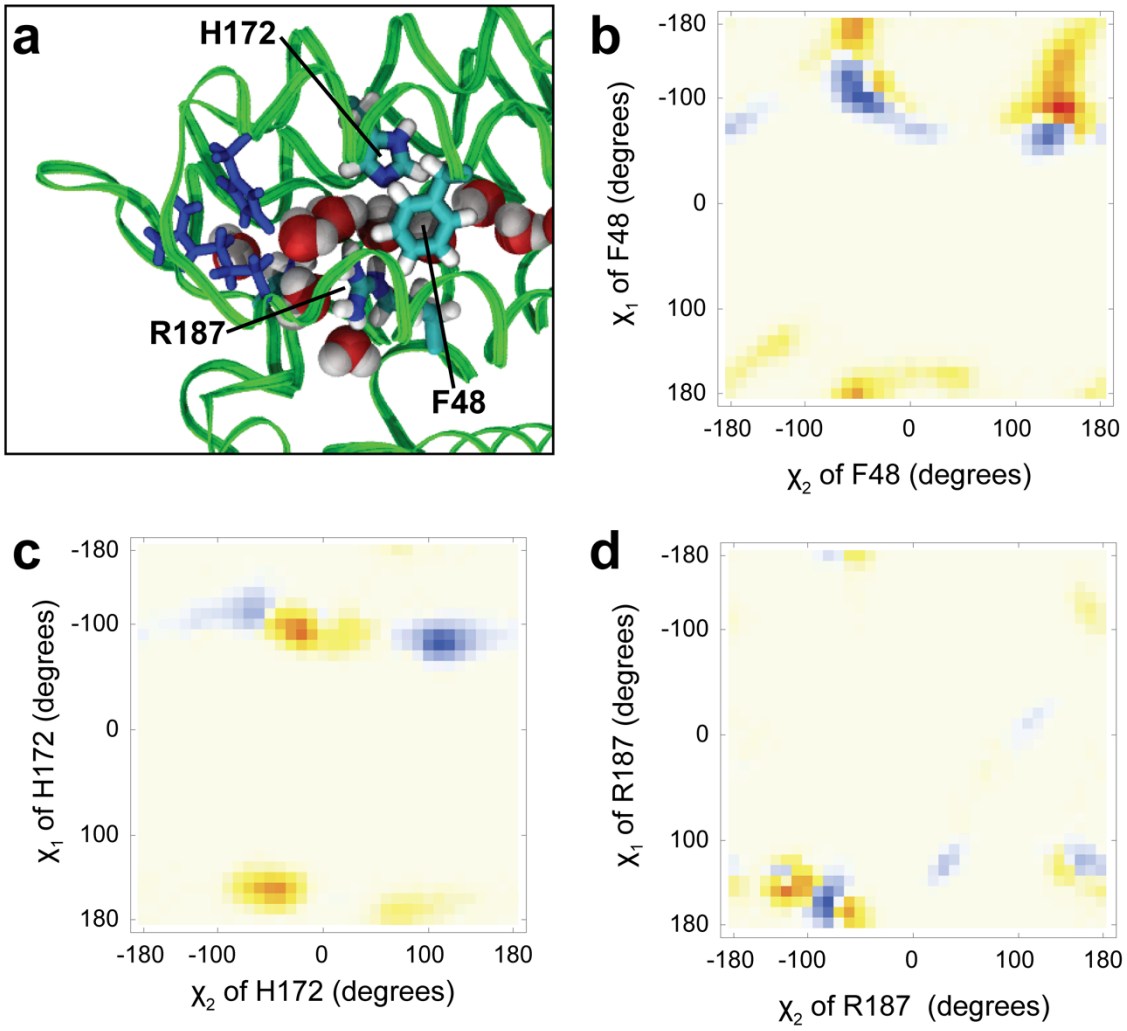

Figure 5.3: Side Chain Ramachandran Plots of the CS-I Residues. a) The residues of the CS-I are shown in the licorice representation and colored by atom type (see included labels). The AQP0 secondary structure is shown in the ribbon representation (colored green) and the Loop A charged residues (H40 and R33) are shown in blue. In the Ramachandran plots (b,c,d), confirmations of the WT AQP0 (CaM) system are shown by varying degrees of blue and confirmations of the AQP0-H40$^+$ (CaM) system are shown in orange/red.b) A Ramachandran plot of the first two side chain dihedral angles (χ_1 and χ_2) of F48. c) A Ramachandran plot of the first two side chain dihedral angles (χ_1 and χ_2) of H172. d) A Ramachandran plot of the first two side chain dihedral angles (χ_1 and χ_2) of R187.

To quantify the side chain hydrogen-bonding interactions between R187 and D119, I measured the distance between the R187(C_ζ) and D119(O_γ) atoms in the three trajectories (WT, H40K, and $H40^+$) and generated comparable histograms (see figure 5.4a). H40 protonation stabilizes configurations of the channel where the R187 and D119 side chains form a hydrogen bond. This corresponds with the enhanced feature below the dotted line (~4.5 Å; see figure 5.4a). To see if the distance accurately defines the open and closed states of the CS-I R187 gate, I partitioned the trajectories by the formation of this hydrogen bond (distance is either greater than or less than 4.5 Å) and calculated the permeability of the two sets of trajectories (see Fig. 5.4c). The concatenated 'open' and 'closed' trajectories were broken into 4 equal parts each to estimate the standard error. This open state model shows a marked three-fold increase in water permeability versus the closed, which qualitatively agrees with previously reported models of the open and closed states of R187. The positive-charge on $H40^+$ increases the overall interaction energy between H40/R33 and R187 by 2 kcal/mol (see Table 5.2). Loop A, which contains R33 and H40, sits directly across the pore from R187 and Loop C (see Fig. 5.4b). Our evidence suggests that $H40^+$ repels R187 towards D119 of loop C, which stabilizes the configurations of R187 that do not occlude the permeation pathway (see Fig. 5.4d).

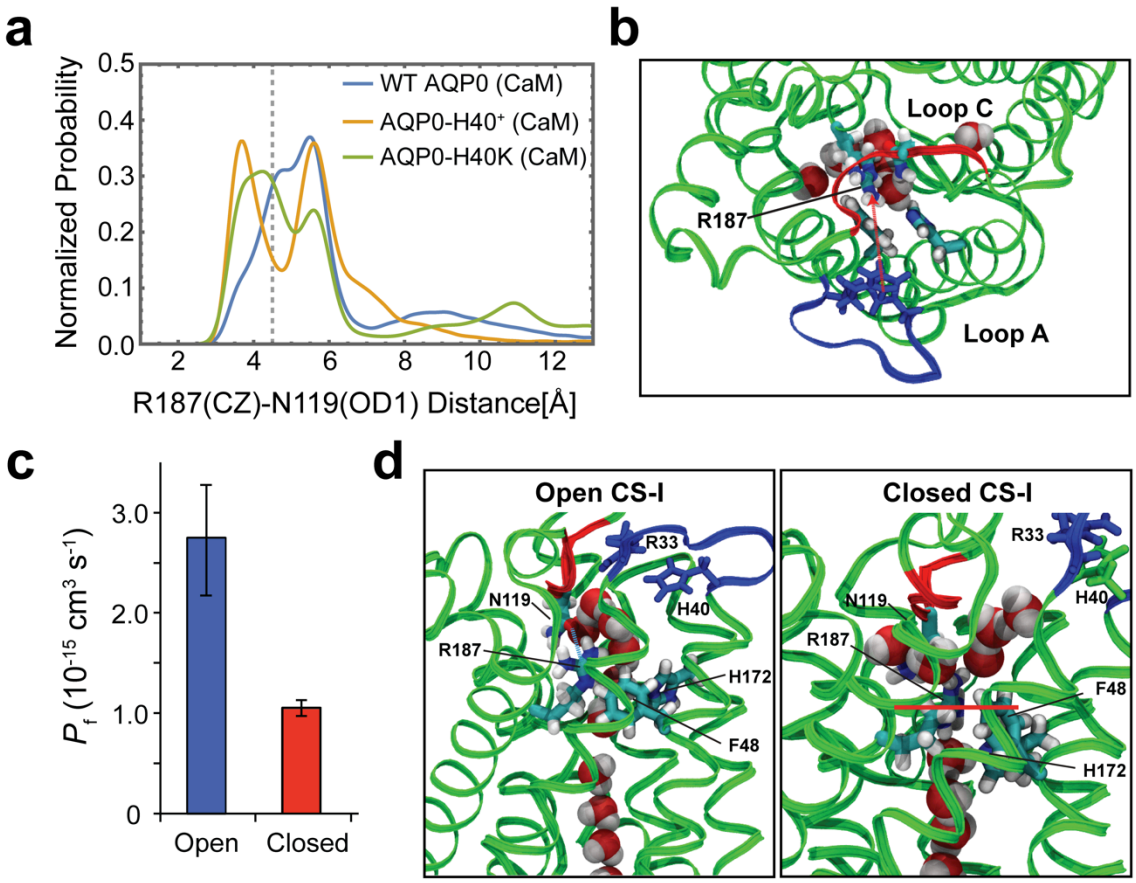

Figure 5.4: H40 Protonation Stabilizes Open Configurations of the R187-Gate. a) A histogram of the R187(CZ)-N119(OD1) distance which quantifies the hydrogen-bonding interactions between their side chains. Specifically, AQP0-H40$^+$ (CaM) and AQP0-H40K (CaM) show an enhancement of the interaction between this side chains. b) A top-down view of the R187-gate showing the extracellular loops, loop A and loop C (colored in blue and red, respectively). Note that R187 sits directly across the permeation pathway from the positively-charged loop A. c) Calculated P_f's from two sets of trajectories of the AQP0-H40$^+$ (CaM) system parsed by the R187(CZ)-N119(OD1) distance (a). Specifically, open states correspond to distances shorter than 4.5-Å and closed states correspond to distances greater than 4.5-Å. d) Two configurations from MD simulations of AQP0-H40$^+$ (CaM) and WT AQP0 (CaM), respectively, using the R187(CZ)-N119(OD1) distance to find the configurations. Note that the H40$^+$ open configuration (left) has an interaction between R187 and N119 and the single-file of water molecules remains unbroken. In the WT closed CS-I configuration (right), R187 no longer interacts with N119, and the R187 occludes the single-file permeation pathway, resulting in a three-fold reduction in the calculated P_f.

5.4 Discussion

The water permeability of AQP0 has long been known to be pH-dependent. Decreasing the pH of the external medium in Xenopus oocyte swelling assays shows an approximate two-fold increase in the water permeability of AQP0 [1-3]. Mutagenesis experiments have isolated two residues on the extracellular loop A region of AQP0, H40 and R33, which contribute to its pH-dependence [2]. Additionally, these residues are conserved in other pH-sensitive aquaporin channels. Previous attempts have been made to determine the molecular mechanism by which acidic pH increases AQP0 P_f [2, 18, 35, 36]. Many of the crucial details remain undetermined. In this chapter, I examine three prominent hypotheses surrounding the AQP0 pH-gating mechanism. One theory states that the titrateable H40 manipulates the position and dynamics of the second transmembrane helix (TM2) to open and close the pore-gate [36]. The other two theories propose that H40 manipulates the positions of bulky residues, either W34 or R187/H172, so that the degree to which they occlude the permeation pathway is changed [18, 35]. The results posited herein suggest that R187 functions as the major pH-gating residue of the CS-I.

The theory that the TM2 couples the pH- and calcium-sensitivity mechanism of AQP0 is partially supported by mutagenesis data. The histidine residues H40 and H61 sit at the terminal ends of the TM2 and have been connected to both pH- and calcium-sensitivity. I did not observe any significant modifications to the backbone, side chains, or tilt angle of the TM2 helix resulting from H40 protonation. Additionally, per residue RMSF calculations (see Fig. 5.1b) show no

significant differences in the TM2 region (residues 40 to 60) associated with the $H40^+$ trajectory. H40 protonation does appear to decrease fluctuations in the residues around R187, which is part of the first constriction site (CS-I) (see Fig. 5.1b). The TM2 helix sits at the center of the AQP0 tetramer where native interactions likely prohibit large displacements by a single point mutation. Therefor, it is unlikely that H40 and H61 mutagenesis modulate the permeability of AQP0 in a concerted mechanism involving the TM2 helix.

Changes in the CS-I configuration were observed as a result of H40 protonation. In particular, the side chain orientations of H172 an R187 were different in the two trajectories (see Fig. 5.3c and Fig. 5.4a). H172's proximity to R187 likely makes the pKa of its imidazole ring too high for protonation in the physiological range. H172 appears to be pushed towards the pore center as a result of the increased water density observed in the $H40^+$ trajectory (see Fig. 5.3c). However, there is no obvious way that this could result directly from $H40^+$. Electrostatic repulsions between H40 and R187 are, not surprisingly, increased by H40 protonation. The loop A region of AQP0 sits directly across the pore opening from R187. Therefore, the repulsive electrostatic interactions between loop A and R187 push the side chain away from the pore opening (see Fig. 5.4b), no longer occluding the passage of water (see Fig. 5.4d). Previous studies identified N119 as a residue that stabilizes open configurations of the R187 gate [18]. We observe an increase in hydrogen-bonding interactions between the side chains of R187 and N119 associated with H40 protonation (see Fig. 5.4a). These configurations keep the R187 side chain from occluding the permeation pathway

and display an approximately three-fold increase in the calculated osmotic permeability constant (see Fig. 5.4c,d). The gating arginine residue R187 is manipulated in a similar way in both the $H40^+$ and H40K systems, suggesting that the electrostatic interactions between loop A and R187 can open or close the CS-I and are responsible for the observed changes in the experimental P_f.

5.5 Bibliography

1. Nemeth-Cahalan, K. L., and Hall, J. E. (2000) pH and calcium regulate the water permeability of aquaporin 0. *J Biol Chem* **275**, 6777-6782

2. Nemeth-Cahalan, K. L., Kalman, K., and Hall, J. E. (2004) Molecular basis of pH and Ca2+ regulation of aquaporin water permeability. *J Gen Physiol* **123**, 573-580

3. Nemeth-Cahalan, K. L., Kalman, K., Froger, A., and Hall, J. E. (2007) Zinc modulation of water permeability reveals that aquaporin 0 functions as a cooperative tetramer. *J Gen Physiol* **130**, 457-464

4. Zelenina, M., Bondar, A. A., Zelenin, S., and Aperia, A. (2003) Nickel and extracellular acidification inhibit the water permeability of human aquaporin-3 in lung epithelial cells. *J Biol Chem* **278**, 30037-30043

5. Zeuthen, T., and Klaerke, D. A. (1999) Transport of water and glycerol in aquaporin 3 is gated by H(+). *J Biol Chem* **274**, 21631-21636

6. Kaptan, S., Assentoft, M., Schneider, H. P., Fenton, R. A., Deitmer, J. W., MacAulay, N., and de Groot, B. L. (2015) H95 Is a pH-Dependent Gate in Aquaporin 4. *Structure* **23**, 2309-2318

7. Matsuki-Fukushima, M., Fujita-Yoshigaki, J., Murakami, M., Katsumata-Kato, O., Yokoyama, M., and Sugiya, H. (2013) Involvement of AQP6 in the Mercury-sensitive osmotic lysis of rat parotid secretory granules. *J Membr Biol* **246**, 209-214

8. Virkki, L. V., Franke, C., Somieski, P., and Boron, W. F. (2002) Cloning and functional characterization of a novel aquaporin from Xenopus laevis oocytes. *J Biol Chem* **277**, 40610-40616

9. Zong, X., Stieber, J., Ludwig, A., Hofmann, F., and Biel, M. (2001) A single histidine residue determines the pH sensitivity of the pacemaker channel HCN2. *J Biol Chem* **276**, 6313-6319

10. Rajan, S., Wischmeyer, E., Xin Liu, G., Preisig-Muller, R., Daut, J., Karschin, A., and Derst, C. (2000) TASK-3, a novel tandem pore domain acid-sensitive K+ channel. An extracellular histiding as pH sensor. *J Biol Chem* **275**, 16650-16657

11. Yu, L., and Fesik, S. W. (1994) pH titration of the histidine residues of cyclophilin and FK506 binding protein in the absence and presence of immunosuppressant ligands. *Biochim Biophys Acta* **1209**, 24-32

12. Loewenthal, R., Sancho, J., and Fersht, A. R. (1992) Histidine-aromatic interactions in barnase. Elevation of histidine pKa and contribution to protein stability. *J Mol Biol* **224**, 759-770

13. Edgcomb, S. P., and Murphy, K. P. (2002) Variability in the pKa of histidine side-chains correlates with burial within proteins. *Proteins* **49**, 1-6

14. Reichow, S. L., Clemens, D. M., Freites, J. A., Nemeth-Cahalan, K. L., Heyden, M., Tobias, D. J., Hall, J. E., and Gonen, T. (2013) Allosteric mechanism of water-channel gating by Ca2+-calmodulin. *Nat Struct Mol Biol* **20**, 1085-1092

15. Shaw, D. E., Grossman, J., Bank, J. A., Batson, B., Butts, J. A., Chao, J. C., Deneroff, M. M., Dror, R. O., Even, A., and Fenton, C. H. (2014) Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. in *Proceedings of the international conference for high performance computing, networking, storage and analysis*, IEEE Press

16. Gonen, T., Sliz, P., Kistler, J., Cheng, Y., and Walz, T. (2004) Aquaporin-0 membrane junctions reveal the structure of a closed water pore. *Nature* **429**, 193-197

17. Gonen, T., and Walz, T. (2006) The structure of aquaporins. *Q Rev Biophys* **39**, 361-396

18. Saboe, P. O., Rapisarda, C., Kaptan, S., Hsiao, Y. S., Summers, S. R., De Zorzi, R., Dukovski, D., Yu, J., de Groot, B. L., Kumar, M., and Walz, T. (2017) Role of Pore-Lining Residues in Defining the Rate of Water Conduction by Aquaporin-0. *Biophys J* **112**, 953-965

19. Tajkhorshid, E., Nollert, P., Jensen, M. O., Miercke, L. J., O'Connell, J., Stroud, R. M., and Schulten, K. (2002) Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science* **296**, 525-530

20. Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* **30**, 1545-1614

21. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802

22. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* **102**, 3586-3616

23. MacKerell, A. D., Jr., Feig, M., and Brooks, C. L., II. (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing conformational distributions in molecular dynamics simulations. *J Comput Chem* **25**, 1400-1415

24. Best, R. B., Mittal, J., Feig, M., and MacKerell, A. D., Jr. (2012) Inclusion of many-body effects in the additive CHARMM protein CMAP potential results in enhanced cooperativity of alpha-helix and beta-hairpin formation. *Biophys J* **103**, 1045-1051

25. Klauda, J. B., Venable, R. M., Freites, J. A., Connor, J. W. O., Tobias, D. J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell Jr., A. D., and Pastor, R. W. (2010) Update of the CHARMM All-Atom Additive Force Fields for Lipids: Validation on Six Lipid Types. *J Phys Chem* **114**, 7830-7843

26. Humphreys, D. D., Friesner, R. A., and Berne, B. J. (1994) A multiple-time-step molecular dynamics algorithm for macromolecules. *J Phys Chem* **98**, 6885-6892

27. Shan, Y., Klepeis, J. L., Eastwood, M. P., Dror, R. O., and Shaw, D. E. (2005) Gaussian split Ewald: A fast Ewald mesh method for molecular simulation. *J Chem Phys* **122**, 054101

28. Kräutler, V., Van Gunsteren, W. F., and Hünenberger, P. H. (2001) A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J Comp Chem* **22**, 501-508

29. Martyna, G. J., Klein, M. L., and Tuckerman, M. (1992) Nosé–Hoover chains: the canonical ensemble via continuous dynamics. *J Chem Phys* **97**, 2635-2643

30. Martyna, G. J., Tobias, D. J., and Klein, M. L. (1994) Constant pressure molecular dynamics algorithms. *J Chem Phys* **101**, 4177-4189

31. Lippert, R. A., Predescu, C., Ierardi, D. J., Mackenzie, K. M., Eastwood, M. P., Dror, R. O., and Shaw, D. E. (2013) Accurate and efficient integration for molecular dynamics simulations at constant temperature and pressure. *J Chem Phys* **139**, 10B621_621

32. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J Mol Graph* **14**, 33-38, 27-38

33. Zhu, F., Tajkhorshid, E., and Schulten, K. (2004) Collective diffusion model for water permeation through microscopic channels. *Phys Rev Lett* **93**, 224501

34. Finkelstein, A. (1987) *Water movement through lipid bilayers, pores, and plasma membranes*, Wiley, 56-72

35. Jensen, M. Ø., Dror, R. O., Xu, H., Borhani, D. W., Arkin, I. T., Eastwood, M. P., and Shaw, D. E. (2008) Dynamic control of slow water transport by aquaporin 0: implications for hydration and junction stability in the eye lens. *Proc Nat Acad Sci U S A* **105**, 14430-14435

36. Németh-Cahalan, K. L., Clemens, D. M., and Hall, J. E. (2013) Regulation of AQP0 water permeability is enhanced by cooperativity. *J Gen Phys* **141**, 287-295

37. Klepeis, J. L., Lindorff-Larsen, K., Dror, R. O., and Shaw, D. E. (2009) Long-timescale molecular dynamics simulations of protein structure and function. *Curr Opin Struct Biol* **19**, 120-127

Chapter 6

Concluding Remarks

In this dissertation, I detailed the results of several projects, all focused on the regulation of protein function. In Ch. 2 and Ch. 3, I detailed the interactions between two redox-active cytochrome c complexes. In Ch. 4 and Ch. 5, I discuss the regulation of aquaporin 0 water permeability by external calcium concentration and pH. All of the chapters of this dissertation feature results from molecular dynamics (MD) and Brownian dynamics (BD) simulations.

In Ch. 2, I described the P450cin-Cindoxin complex model that I produced using a combined BD/MD approach [1]. Very few crystal structures exist of electron-transfer competent cytochrome c complexes. Prior to this study, an individual crystal of P450cin was published. Our collaborators were able to solve the structure of the redox partner, Cindoxin (Cdx) [2]. However, the P450cin-Cdx complex remained elusive experimentally. I first generated a model of the complex from BD simulations initiated from the individual crystal structures. The complex was relaxed in a short MD simulation. In our model, the Y96 residue of Cdx sits in the interaction interface. Since aromatic residues are known to

facilitate electron transfer reactions, our collaborators tested the enzymatic activity of P450cin with various Y96 mutants of Cdx. The Cdx-Y96L mutant, where the aromatic tyrosine is replaced with a non-aromatic leucine, shows an eight-fold reduction in activity, suggesting that our BD/MD complex successfully recapitulated the interaction interface [1].

In Ch. 3, I reported on a collaborative effort to disseminate the association mechanism of the *Leishmania major* peroxidase (LmP) redox complex with cytochrome c (LmCytc). The LmP-LmCytc interaction is a rare example of a complex with a solved co-crystal structure [3]. In my generated BD trajectories, clear differences emerged in the association patterns of the trajectories initiated from the individual versus the co-crystal structures. Particularly, the dominant association complex from the individual structure BD trajectories was very far removed from the active site. However, the co-crystal BD simulations recapitulated the active complex. This observation led to the discovery of the negatively-charged helix A of LmP which serves as the first transient complex site for LmCytc-binding. Steady-state kinetics assays confirmed that mutagenesis of the charged residues of helix A reduced the enzymatic activity of LmP down by more than an order of magnitude. LmCytc therefor first binds to helix A of LmP, followed by local sampling and side chain reorientations that lead to the formation of the active complex, which we refer to as a “bind and crawl” association mechanism [4].

In Ch. 4, I detail the mechanism by which Calmodulin (CaM), in response to changes in the calcium concentration, modulates the water permeability (P_f) of

the water channel protein, Aquaporin 0 (AQP0). In a previous publication, my colleagues solved a structure of the AQP0-CaM complex [5]. Through MD simulations, they revealed that the aromatic residues of the second constriction site (CS-II) of AQP0 constitute the CaM-dependent pore-gate [5]. I extended these MD simulations and divulged a positively-charged arginine-rich cytosolic loop of AQP0 that couples the dynamics of the pore-gating CS-II residues to bound-CaM [6]. My BD and MD simulations revealed that this cytosolic loop is also critical for the modulation of the AQP0 permeability by phosphorylation. Peptide-fragment binding assays suggested that phosphorylation inhibits CaM-binding [7, 8]. However, our combined experimental and simulation studies revealed that AQP0 phosphorylation modifies the contacts between CaM and the AQP0 cytosolic loop, modulating the P_f allosterically without inhibiting CaM-binding.

In Ch. 5, I describe my most recent results on the pH-gating mechanism of AQP0. It has long been known that AQP0 permeability is modulated by external pH. Specifically, a moderately acidic pH of 6.5 increases the AQP0 P_f by approximately two-fold [9, 10]. Previous experimental results point to the extracellular loop A region of AQP0 being the principle pH-modulatory site [10]. Mutation of residues H40 and R33 locks the P_f with respect to external pH [10]. Three prominent theories suggest that the pH-gate is the second transmembrane helix (TM2), the W34 residue, or the R187 residue [5, 11, 12]. My simulation results suggest that the R187 side chain of the first constriction site (CS-I) constitutes the pH-gate. In response to H40 protonation, which occurs at slightly

acidic pH, the charge-repulsion between R187 and H40 moves the arginine side chain away from occluding the pore. The side chain of N119 stabilizes open configurations of the R187 gate, leading to an approximately three-fold increase in the channel P_f.

Protein-protein interactions are necessarily complex processes, as biology requires high specificity in order to carry out particular functions within a signal cascade. The results presented in this dissertation highlight the importance of untangling the molecular mechanisms involved in protein association. A repeating theme that emerges is the importance of specific electrostatic interactions, which can act over large distances, for both protein association and allosteric regulation (see Ch. 3 and 4). From the data presented herein, I hope that additional studies will continue to disseminate information on these mechanisms. I also hope that some of these concepts will prove to be pertinent to other protein complexes.

Bibliography

1. Madrona, Y., Hollingsworth, S. A., Tripathi, S., Fields, J. B., Rwigema, J. C., Tobias, D. J., and Poulos, T. L. (2014) Crystal structure of cindoxin, the P450cin redox partner. *Biochemistry* **53**, 1435-1446

2. Hawkes, D. B., Adams, G. W., Burlingame, A. L., Ortiz de Montellano, P. R., and De Voss, J. J. (2002) Cytochrome P450(cin) (CYP176A), isolation, expression, and characterization. *J Biol Chem* **277**, 27725-27732

3. Jasion, V. S., Doukov, T., Pineda, S. H., Li, H., and Poulos, T. L. (2012) Crystal structure of the Leishmania major peroxidase-cytochrome c complex. *Proc Natl Acad Sci U S A* **109**, 18390-18394

4. Fields, J. B., Hollingsworth, S. A., Chreifi, G., Heyden, M., Arce, A. P., Magana-Garcia, H. I., Poulos, T. L., and Tobias, D. J. (2015) "Bind and Crawl" Association Mechanism of Leishmania major Peroxidase and Cytochrome c Revealed by Brownian and Molecular Dynamics Simulations. *Biochemistry* **54**, 7272-7282

5. Németh-Cahalan, K. L., Clemens, D. M., and Hall, J. E. (2013) Regulation of AQP0 water permeability is enhanced by cooperativity. *J Gen Physiol* **141**, 287-295

6. Fields, J. B., Németh-Cahalan, K. L., Freites, J. A., Vorontsova, I., Hall, J. E., and Tobias, D. J. (2017) Calmodulin Gates Aquaporin 0 Permeability through a Positively Charged Cytoplasmic Loop. *J Biol Chem* **292**, 185-195

7. Reichow, S. L., and Gonen, T. (2008) Noncanonical binding of calmodulin to aquaporin-0: implications for channel regulation. *Structure* **16**, 1389-1398

8. Rose, K. M., Wang, Z., Magrath, G. N., Hazard, E. S., Hildebrandt, J. D., and Schey, K. L. (2008) Aquaporin 0-calmodulin interaction and the effect of aquaporin 0 phosphorylation. *Biochemistry* **47**, 339-347

9. Nemeth-Cahalan, K. L., and Hall, J. E. (2000) pH and calcium regulate the water permeability of aquaporin 0. *J Biol Chem* **275**, 6777-6782

10. Nemeth-Cahalan, K. L., Kalman, K., and Hall, J. E. (2004) Molecular basis of pH and Ca2+ regulation of aquaporin water permeability. *J Gen Physiol* **123**, 573-580

11. Jensen, M. Ø., Dror, R. O., Xu, H., Borhani, D. W., Arkin, I. T., Eastwood, M. P., and Shaw, D. E. (2008) Dynamic control of slow water transport by aquaporin 0: implications for hydration and junction stability in the eye lens. *Proc Natl Acad Sci U S A* **105**, 14430-14435

12. Saboe, P. O., Rapisarda, C., Kaptan, S., Hsiao, Y. S., Summers, S. R., De Zorzi, R., Dukovski, D., Yu, J., de Groot, B. L., Kumar, M., and Walz, T. (2017) Role of Pore-Lining Residues in Defining the Rate of Water Conduction by Aquaporin-0. *Biophys J* **112**, 953-965

Appendices

A Generating Sets of Bimolecular Brownian Dynamics Trajectories

In this dissertation, I presented results from Brownian dynamics (BD) simulations of bimolecular protein encounter processes. Because of the relatively high computational cost of classical molecular dynamics (MD) simulations, many researchers have turned to BD simulations to study diffusional encounters of biomolecules on the microsecond timescale.

A.1 SDA and the Ermak-McCammon Algorithm

Versions 6 and 7 of the SDA package were used to compute the BD trajectory data presented in this dissertation [1, 2]. The SDA program models proteins as atomically-detailed rigid bodies and calculates translational and rotational displacements using the Ermak–McCammon algorithm (derived in Ref. [3]). The translational (Δr) and rotational (Δw) displacement vectors imparted on the proteins are calculated by:

$$\Delta r = \frac{\Delta t}{k_B T} D^T \cdot \boldsymbol{F} + \boldsymbol{R}(\Delta t) \tag{A.1}$$

and:

$$\Delta w = \frac{\Delta t}{k_B T} D^R \cdot \boldsymbol{T} + \boldsymbol{W}(\Delta t) \tag{A.2}$$

Both terms in Eq. A.1 and Eq. A.2 are dependent on the choice of the integration time step (Δt). Δt is chosen to be sufficiently large such that thermal fluctuations can be neglected yet small enough such that the potential energy is accurately sampled. The translational diffusion coefficient (D^T) is the sum of the translation diffusion coefficients of proteins 1 and 2, which are calculated under the assumption of infinite dilution. The rotational displacements (Δw) are computed separately for the two proteins. Therefore, each trajectory is initialized with one translational diffusion coefficient (D^T) and two rotational diffusion coefficients (D^R). The systematic forces ($\boldsymbol{F}$) and torques ($\boldsymbol{T}$) (from Eq. A.1 and Eq. A.2, respectively) are computed using grid-based potentials to decrease the required computation time. The 'Brownian' component of the BD algorithm comes from the Boltzmann-distributed displacement vectors ($\boldsymbol{R}$ and $\boldsymbol{W}$), which randomly displace the proteins, mimicking collisions with solvent particles.

A.2 Calculating the Interaction Energy in SDA

To calculate the systematic forces and torques between proteins 1 and 2, the SDA program uses an additive potential energy function that assumes the following form (Eq. A.1 from Ref. [2]):

$$\begin{aligned} U = {} & U_{EP}^{1-2} + U_{EP}^{2-1} + U_{ED}^{1-2} + U_{ED}^{2-1} + U_{HD}^{1-2} + U_{HD}^{2-1} \\ = {} & \frac{1}{2}\sum_{i_2} \Phi_{EP_1} q_{i_2} + \frac{1}{2}\sum_{i_1} \Phi_{EP_2} q_{i_1} + \frac{1}{2}\sum_{i_2} \Phi_{ED_1} q_{i_2} \\ & + \frac{1}{2}\sum_{i_1} \Phi_{ED_2} q_{i_2} + \frac{1}{2}\sum_{j_2} \Phi_{HD_1} S_{k_2} + \frac{1}{2}\sum_{j_1} \Phi_{HD_2} S_{k_1} \end{aligned} \tag{A.3}$$

The interaction potential energy (U) is computed from the terms U_{EP}, U_{ED}, and U_{HD} that represent long-range electrostatic pair interactions, short-range electrostatic desolvation interactions, and short-range hydrophobic desolvation interactions, respectively. The potential energy terms are calculated twice by multiplying the atomic surface charges (q_i) or surface accessible atoms (S_k) of one protein with the closest grid point of the potentials (Φ) of the other, giving a total of 6 terms.

The electrostatic potential grids (Φ_{EP}) are obtained using an adaptive Poisson–Boltzmann solver, such as the APBS program package [4]. This program numerically solves a linearized finite-difference Poisson–Boltzmann (PB) equation, taking into account the inhomogeneous dielectric medium and the ionic strength of the surrounding electrolyte solution. The linearized form of the PB equation used in the APBS package is shown here (Eq. 1 from Ref. [4]):

$$-\nabla\varepsilon(x)\cdot\nabla\Phi_{EP}(x)+\bar{\kappa}^2(x)\sinh\phi(x)=f(x) \tag{A.4}$$

which is dependent on the dielectric of the medium (ε), the Debye screening length (κ), and the distribution of solute partial charges (f). The CHARMM force field was used to assign the charge and radius parameters for all solute atoms [5]. The force due to the electrostatic potential is computed using the well-established effective charge model [6]. The charge of effective charge sites on the surface-exposed basic and acidic side chains are fit so that, in a shell around the protein, the electrostatic potential that was precomputed with an inhomogeneous dielectric is reproduced with a homogeneous dielectric. For each BD time step, the forces (and torques) acting on each protein (see Eq. A.1) are computed by placing a protein's effective charges (q_i) over the other protein's electrostatic potential grid [1]. The energetic penalty associated with charge-burial in a protein cavity is computed as (using Eq. 2 in Ref. [7]):

$$\begin{aligned} U_{ED} &= \sum_i \Phi_{ED}\,(i) q_i^2 \\ &= \alpha\frac{\varepsilon_s-\varepsilon_p}{(2\varepsilon_s-\varepsilon_p)}\sum_i\sum_j {q_i}^2 {a_j}^3\frac{(1+\kappa r_{ij})^2}{r_{ij}^4}e^{-2\kappa r_{ij}} \end{aligned} \tag{A.5}$$

where $\varepsilon_s$and ε_p are the solvent and protein dielectric permittivities, respectively. This potential creates a penalty for charge q_i of a protein inserting into a low dielectric cavity j of radius a_j on the other protein. The scaling factor (α) is set to 0.36 as to reproduce a finite-difference data of the electrostatic desolvation energy at 0 mM ionic strength.

The nonpolar desolvation potential decreases the energy of nonpolar, surface-accessible atoms (S_k) when they become buried into another protein's surface. The nonpolar desolvation grids (seen in equation A.3) are computed with a step function that estimates the buried area of the surface accessible atoms (Eq. 3 in Ref. [7]):

$$\Phi_{HD}(r) = \beta c \begin{cases} 1 & r < a \\ \dfrac{b-r}{b-a} & a < r < b \\ 0 & r > b \end{cases} \tag{A.6}$$

where β is a negative scaling term which converts the buried area to an energy. The constants a, b, and c are the complete burial length, partial burial length, and scaling term, respectively. When a solvent accessible atom comes within distance b of the other protein's surface, the potential is linearly decreased from βc to 0 until it reaches a distance of a away from the other protein surface. The parameters a, b, and c are are typically set to 3.1 Å, 4.3 Å, and 0.5, respectively. Last but not least, an exclusion volume surrounds the protein surfaces at a radius of 1.77 Å [8]. This exclusion volume is added to reject Brownian moves that result in the overlap of the two protein surfaces. Once the potential energy grids, the exclusion volume grids, and the diffusion coefficients are computed, one can produce large sets of BD trajectories [7]. Once a large enough set of trajectories is generated, binding constants and association rates can be calculated and compared with experimental values.

A.3 Bibliography

1. Gabdoulline, R. R., and Wade, R. C. (1998) Brownian dynamics simulation of protein-protein diffusional encounter. *Methods-a Method Enzymol* **14**, 329-341

2. Martinez, M., Bruce, N. J., Romanowska, J., Kokh, D. B., Ozboyaci, M., Yu, X. F., Ozturk, M. A., Richter, S., and Wade, R. C. (2015) SDA 7: A modular and parallel implementation of the simulation of diffusional association software. *J Comp Chem* **36**, 1631-1645

3. Ermak, D. L., and Mccammon, J. A. (1978) Brownian Dynamics with Hydrodynamic Interactions. *J Chem Phys* **69**, 1352-1360

4. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-10041

5. Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* **30**, 1545-1614

6. Gabdoulline, R. R., and Wade, R. C. (1996) Effective Charges for Macromolecules in Solvent. *J Phys Chem* **100**, 3868-3878

7. Gabdoulline, R. R., and Wade, R. C. (2009) On the contributions of diffusion and thermal activation to electron transfer between Phormidium laminosum plastocyanin and cytochrome f: Brownian dynamics simulations with explicit modeling of nonpolar desolvation interactions and electron transfer events. *J Am Chem Soc* **131**, 9230-9238

8. Gabdoulline, R. R., and Wade, R. C. (2001) Protein-protein association: investigation of factors influencing association rates by brownian dynamics simulations. *J Mol Biol* **306**, 1139-1155

B Generating a Classical Molecular Dynamics Trajectory

In the presented work, I showed several results derived from classical molecular dynamics (MD) simulations. MD simulations remain a powerful tool for probing the motions of individual atoms in a microscopic system. In this dissertation, I created virtual molecular systems derived from atomistic protein structures, which were subsequently solvated in water and/or lipids and neutralized with monovalent counterions. A trajectory is generated using a force field, which is not a typically physical 'field;' i.e. a negative gradient of a scalar potential. Instead, a force field in the context of molecular mechanics is a basic additive form of the potential energy that is solved independently for each particle in a given configuration. The potential energy is used to compute the force on each particle as to generate a forward trajectory in time. Numerous MD methods, such as Car-Parrinello MD, compute the forces from first-principle, *ab-initio* electronic structure calculations, rather than using an empirical potential [1, 2]. Conversely, course-grained simulations, in which a pseudoatom replaces a cluster of heavy atoms, have been implemented to reduce the computational power required to generate long MD trajectories [3]. For the work presented in this dissertation, I employed fully atomistic systems where the classical approximation is appropriate, as we do not consider chemical reactions (i.e. covalent bond-breaking or bond-forming reactions).

B.1 The CHARMM Force Field

To calculate a classical Newtonian trajectory, one must first calculate the forces imparted on each particle in the system. In the presented work, I utilized the latest implementation of the CHARMM potential, specifically CHARMM36, for the lipid and protein atoms [4, 5]. The potential assumes the following form:

$$U(\vec{R}) = \sum_{bonds} k_b\,(b - b_0)^2 + \sum_{UB} k_S\,(S - S_0)^2 + \sum_{angles} k_\theta\,(\theta - \theta_0)^2$$
$$+ \sum_{dihedrals} k_\chi\,(1 - \cos(n\chi - \delta))^2 + \sum_{impropers} k_\phi\,(\phi - \phi_0)^2 \qquad \text{(B.1)}$$
$$+ \sum_{nonbond} \epsilon\left[\left(\frac{R_{min_{ij}}}{r_{ij}}\right)^{12} + \left(\frac{R_{min_{ij}}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\epsilon_1 r_{ij}}$$

where the constants k_b, k_S, k_θ, k_χ, and k_ϕ are the bond, Urey-Bradley, angle, dihedral, and improper force constants, respectively. The values of b, S, θ, χ, and ϕ represent the instantaneous bond length, Urey-Bradley 1-3-distance, dihedral angle, and improper torsion angle, respectively; where the subscript 0 represents the equilibrium value of the particular term. The Lennard-Jones potential given by the 6th term is solved as a function of r_{ij}, which is the distance between atoms i and j. The Lennard-Jones well depth and distance at the energy-minimum are given by ϵ and $R_{min_{ij}}$, respectively. Electrostatic interactions are computed by the 7th term, where q_i is the partial charge of atom i and ϵ_1, which is typically set to 1 for protein simulations, is the effective dielectric constant.

B.2 The Velocity Verlet Algorithm

The velocity Verlet algorithm is a popular method for integrating Newton's equations of motion [6]. The MD simulations employed in this dissertation were run using the NAMD v1.9 simulation package [7], which uses a velocity Verlet algorithm to generate a classical trajectory, given a set of initial conditions and velocities (generated randomly from a Boltzmann distribution) as well as a parameterized force field. The derivation I describe below follows the one described in Ch. 4 of Frenkel & Smit [8].

A Taylor expansion is used to solve for the coordinates of the particles at an incremental time (Δt) in the past or future from the current time (t):

$$\begin{aligned} r(t+\Delta t) &= r(t) + \dot{r}(t)\Delta t + \frac{\ddot{r}(t)}{2}\Delta t^2 + \cdots \\ &= r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 + \cdots \end{aligned} \tag{B.2}$$

and:

$$r(t-\Delta t) = r(t) - v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 + \cdots \tag{B.3}$$

where the higher order terms can be neglected. The first and second derivatives of the position as a function of time, $r(t)$, are the velocity of and force on each particle, $v(t)$ and $f(t)$, respectively. Summing these two expressions and rearranging gives the updated position independent of the particle velocity:

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{f(t)}{m}\Delta t^2 \quad \text{(B.4)}$$

To calculate the velocity of each particle at an incremental time Δt in the future, the expression of the position of the particle at time $t + 2\Delta t$ can be rearranged to obtain an approximate expression for $v(t + \Delta t)$:

$$v(t + \Delta t) \approx v(t) + \frac{f(t + \Delta t) + f(t)}{2m}\Delta t \quad \text{(B.5)}$$

Equations B.4 and B.5 are used to generate a reversible molecular dynamics trajectory from which thermodynamics properties can be calculated.

B.3 Bibliography

1. Car, R., and Parrinello, M. (1985) Unified approach for molecular dynamics and density-functional theory. *Phys Rev Lett* **55**, 2471-2474

2. Koppel, H., Domcke, W., and Cederbaum, L. S. (1984) Multimode Molecular-Dynamics Beyond the Born-Oppenheimer Approximation. *Adv Chem Phys* **57**, 59-246

3. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., and de Vries, A. H. (2007) The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B* **111**, 7812-7824

4. Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* **30**, 1545-1614

5. Klauda, J. B., Venable, R. M., Freites, J. A., Connor, J. W. O., Tobias, D. J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell Jr., A. D., and Pastor, R. W. (2010) Update of the CHARMM All-Atom Additive Force Fields for Lipids: Validation on Six Lipid Types. *J Phys Chem* **114**, 7830-7843

6. Verlet, L. (1967) Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys Rev* **159**, 98-133

7. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802

8. Frenkel, D., and Smit, B. (1997) Simulation of phase coexistence for complex molecules. *Computers in Physics* **11**, 246-255