

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Multiscale analysis and visualization of biophysical structure and biochemical function with computational microscopy

Permalink

<https://escholarship.org/uc/item/30j8t943>

Author

Hirakis, Sophia P

Publication Date

2019

Supplemental Material

<https://escholarship.org/uc/item/30j8t943#supplemental>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Multiscale analysis and visualization of
biophysical structure and biochemical function with
computational microscopy

A dissertation submitted in partial satisfaction of the
requirements for the degree of Doctor of Philosophy

in

Chemistry
with a Specialization in
Multiscale Biology

by

Sophia P. Hirakis

Committee in charge:

Professor Rommie E. Amaro, Chair
Professor J. Andrew McCammon, Co-Chair
Professor Katja Lindenberg
Professor Charles L. Perrin
Professor Terrence J. Sejnowski
Professor Elizabeth Villa

2019

Copyright

Sophia P. Hirakis, 2019

All rights reserved.

The Dissertation of Sophia P. Hirakis is approved and is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California San Diego

2019

DEDICATION

I am fortunate to have been blessed with two angels,
who each provided me with one wing so that I could fly..

I dedicate this work to my parents

To my beloved father, my baba,
πατέρα μου

Petros Emmanuel Hirakis

Thank you for being an example of humility and hard work.

To beautiful mother, my reason,
μανούλα μου

Marina (Prearis) Hirakis

Thank you for believing in me when I did not believe in myself.

This work is also dedicated to the eternal memory of the people that I loved and lost during this journey

My grandmother, *Sofia Sotiriou XHPAKH*

My grandmother, *Kalliopi Mastromanolis Prearis*

My grandfather, *Minas Emmanuel Prearis*

My aunt, *Theodosia Prearis Dargakis*

My uncle, *George Dargakis*

My koumparo, *Kostas N. Protopapas*

My cousin, *Kostas N. Dais*

My fellow villager, *Georgios Nikitas*

My heroic fighter, *Anthony “Tony” Chester Miller*

My father-figure, *George J. Tsimiklis*

My favorite musician, *Antonios Nicolaidis*

My Johnny boy, *John LaBarbera*

My dear friend, *Vasilios N. Diakokomninos*

My spiritual brother, *Matthew G. Walker*

My former roommate, *Hollis Joseph Hamilton Crittendon*

My sea sister, *Sophia Tiaré Bartlow*

My beloved chemistry professor, *Brother Andrew Joseph Winka*

My fearless microscopy professor, *Gina Sosinsky*

and My almost-colleague, *Anushka Michailova*

Finally, I dedicate my thesis to the youth of Greece,
whose future remains bright despite our tribulations
and to the everlasting life of my village,

Olympos, Karpathos

EPIGRAPH

I believe there exists, and I feel within me,
an instinct for the truth, or knowledge or discovery,
of something of the same nature as the instinct of virtue,
and that our having such an instinct is **reason enough**
for scientific researches without any practical
results ever ensuing from them.

Charles Darwin

Scientists may seek Truth, but
Science is the Art of Approximation.
Just as artists use their media to represent the truths they see about the world,
we scientists use our equations, models and case studies
to represent the truth as it has been **revealed** to us.
But in those dark weary hours in the dead of night, we acknowledge to
ourselves that we have provided only a representation of reality,
an approximation of a deeper truth.
Our models and conclusions are like shells in which we hope
to hear an echo of the great ocean beyond.

William Hooke

When you have a father and a mother
who work **all their lives** so you can have
an education and build your body - it's a blessing.

Lou Gehrig

The world breaks everyone..
and afterwards, many are strong at the broken places.

Ernest Hemingway

TABLE OF CONTENTS

| | |
|--|-------|
| Signature Page | iii |
| Dedication | iv |
| Epigraph..... | v |
| Table of Contents | vi |
| List of Symbols and Acronyms | ix |
| List of Supplemental Files | x |
| List of Figures | xi |
| List of Schemes | xiv |
| List of Tables..... | xv |
| Preface | xvi |
| Acknowledgements | xxi |
| Vita..... | xxxii |
| Abstract of the Dissertation | xxxiv |
| Introduction..... | 1 |
| 0.1 Scaling up through the sciences | 4 |
| 0.2 The multiscale nature of biochemistry | 5 |
| 0.3 How mutations affect protein structure: a multiscale perspective | 7 |
| 0.4 Structural methodologies for visualization of biology..... | 9 |
| 0.4.1 X-ray Crystallography | 9 |
| 0.4.2 Hydrogen/Deuterium Exchange Mass Spectrometry | 10 |
| 0.4.3 Small-Angle X-ray Scattering | 12 |
| 0.4.4 Cryo-electron Microscopy..... | 13 |
| 0.4.5 Electron Tomography | 15 |
| 0.5 Biophysical structure modeling methodologies | 16 |
| 0.5.1 Molecular Dynamics | 16 |
| 0.5.2 Brownian Dynamics | 20 |
| 0.6 Subcellular biochemical modeling methodologies | 22 |
| 0.6.1 Modeling biology with differential equations | 23 |
| 0.6.2 Discrete subcellular spatial modeling | 24 |
| 0.7 Conclusion of the Introduction | 25 |

| | | |
|-----------|---|----|
| Chapter 1 | Multiscale Modeling of protein systems | 27 |
| 1.1 | Introduction | 28 |
| 1.2 | Nomenclature | 30 |
| 1.3 | Accessing the Conformational Ensembles of Proteins | 30 |
| 1.3.1 | Molecular Mechanics and Molecular Dynamics Simulations | 30 |
| 1.3.2 | Atomic-scale Markov State Models of a Conformational Ensemble | 30 |
| 1.4 | Investigating Intermolecular Interactions | 32 |
| 1.4.1 | Brownian Dynamics Simulations | 32 |
| 1.4.2 | Considerations for Brownian Dynamics Simulations | 32 |
| 1.4.3 | Unifying MD and BD Simulations through Milestoning | 34 |
| 1.5 | From Atomistic to Protein-scale Models | 35 |
| 1.5.1 | Protein-scale MSM | 35 |
| 1.5.2 | Applying MD and BD modeling to Protein Scale PKA-RI α MSM | 36 |
| 1.6 | Integrating Protein Scale MSM into Whole Cell Models | 37 |
| 1.6.1 | Stochastic and Deterministic MSM in the Whole Cell | 38 |
| 1.6.2 | Advantages in Whole Cell Modeling | 38 |
| 1.7 | Conclusions | 38 |
| 1.8 | Acknowledgments | 43 |
| Chapter 2 | Broad recognition in the interaction of group A Streptococcus M protein hypervariable regions with human C4b-binding protein | 44 |
| 2.1 | Abstract | 45 |
| 2.2 | Introduction | 45 |
| 2.3 | Structural similarity | 47 |
| 2.4 | Uniform reading head | 48 |
| 2.5 | Sequence conservation hidden within hypervariability | 50 |
| 2.6 | Tolerance to hypervariability | 50 |
| 2.7 | Methods | 68 |
| 2.7.1 | DNA manipulation | 68 |
| 2.7.2 | Protein Expression and Purification | 72 |
| 2.7.3 | Crystallization and Data Collection | 73 |
| 2.7.4 | Structure Determination and Refinement | 74 |
| 2.7.5 | Co-Precipitation Assays | 77 |
| 2.7.6 | Molecular Dynamics | 77 |
| 2.8 | Concluding Remarks | 79 |
| 2.9 | Acknowledgment | 80 |
| Chapter 3 | Bridging structural and mechanistic observations with computational modeling techniques | 82 |
| 3.1 | Acknowledgment | 95 |
| Chapter 4 | Subcellular spatial modeling in realistic geometries with stochastic particle methods to understand heart disease | 96 |
| 4.1 | Abstract | 96 |

| | | |
|-------|---|-----|
| 4.2 | Introduction | 97 |
| 4.3 | Methods | 99 |
| 4.3.1 | Building the Geometric Model | 99 |
| 4.3.2 | The MCell model of the CRU | 101 |
| 4.3.3 | Experimental design | 102 |
| 4.3.4 | Cytosolic and SR Ca^{2+} Buffering | 105 |
| 4.3.5 | Sarcolemmal (T-Tubule) Fluxes | 106 |
| 4.3.6 | Sarcoplasmic Reticulum Fluxes | 111 |
| 4.4 | Results and Discussion | 117 |
| 4.4.1 | Buffer and Fluorophore Dynamics | 117 |
| 4.4.2 | Ryanodine Receptor activation profiles | 118 |
| 4.4.3 | Examining disease phenotypes | 118 |
| 4.4.4 | Simulating “heartbeats” with successive action potentials | 128 |
| 4.4.5 | Model caveats and future directions | 130 |
| 4.5 | Conclusion | 131 |
| 4.6 | Acknowledgments | 132 |
| 4.7 | Supplementary Information | 133 |
| | Bibliography | 137 |

LIST OF SYMBOLS AND ACRONYMS

| | |
|---------|---|
| Å | Angstrom |
| AP | Action Potential |
| BD | Brownian Dynamics |
| cAMP | Cyclic Adenosine Monophosphate |
| CBD | Cyclic-nucleotide Binding Domain |
| CMDN | Calmodulin |
| CRU | Calcium Release Unit |
| Cryo-EM | Cryo-Electron Microscopy |
| C4BP | Human C4b-Binding Protein |
| GAS | Group A <i>Streptococcus</i> |
| H/DxMS | Hydrogen/Deuterium Exchange Mass Spectrometry |
| LTCC | L-Type Calcium Channel |
| HVR | Hyper-Variable Region |
| MD | Molecular Dynamics |
| MSM | Markov State Model |
| NCX | Na ⁺ / Ca ²⁺ exchanger |
| ODE | Ordinary Differential Equation |
| PKA | Protein Kinase A |
| PMCA | Plasma Membrane Ca ²⁺ ATPase |
| RyR | Ryanodine Receptor |
| SAXS | Small-Angle X-ray Scattering |
| SERCA | Sarco/Endoplasmic Reticulum Calcium-ATPase |
| SR | Sarcoplasmic Reticulum |
| TT | Transverse Tubule |
| WT | Wild-Type |

LIST OF SUPPLEMENTAL FILES

- Video 2.1 R39 nook in M2-C4BP α 1-2. (Hirakis-SPH-video1.mov)
- Video 2.2 R39 nook in M2 (F75A)-C4BP α 1-2. (Hirakis-SPH-video2.mov)
- Video 2.3 C4BP α 2 contacts in M2-C4BP α 1-2. (Hirakis-SPH-video3.mov)
- Video 2.4 C4BP α 2 contacts in M2 (K65A)-C4BP α 1-2. (Hirakis-SPH-video4.mov)
- Video 2.5 C4BP α 2 contacts in M2 (N66D)-C4BP α 1-2. (Hirakis-SPH-video5.mov)

LIST OF FIGURES

| | | |
|--------------|--|----|
| Figure 0.1. | A drawing of musical frequencies | 3 |
| Figure 0.2. | Levels of protein structure | 5 |
| Figure 0.3. | The interface of the Regulatory and Catalytic subunit of Protein Kinase A (PKA), comparing the mutant to the wild-type | 8 |
| Figure 0.4. | Angstrom-level representations of biomolecules | 14 |
| Figure 0.4. | Brownian Dynamics simulation with BrownDye | 20 |
| Figure 1.1. | Bridging gaps through multiscale modeling | 29 |
| Figure 1.2. | Protein Kinase A cyclic nucleotide binding domain Markov state model | 31 |
| Figure 1.3. | Brownian dynamics simulation method | 33 |
| Figure 1.4. | Milestoning applied to unite MD and BD | 35 |
| Figure 1.5. | The Markov State Model of PKA-RI α R ₂ C ₂ holoenzyme | 37 |
| Figure 2.1. | Structures of M-C4BP complexes | 47 |
| Figure 2.2. | C4BP Binding Mode | 51 |
| Figure 2.3. | C4BP-binding modes of M proteins | 52 |
| Figure 2.4. | M2-C4BP interaction | 54 |
| Figure 2.5. | Schematic of M protein domains | 55 |
| Figure 2.6. | Electron density for the M49 HVR-C4BP α 1-2 complex | 56 |
| Figure 2.7. | Structure of M22-C4BP | 57 |
| Figure 2.8. | Structure of M28-C4BP | 58 |
| Figure 2.9. | Structure of M49-C4BP | 59 |
| Figure 2.10. | Coiled coil parameters of M proteins | 60 |
| Figure 2.11. | Rotation of C4BP α 1-2 | 61 |
| Figure 2.12. | Structure of the M22-C4BP interaction in which C4BP α 1 is tilted rather than rotated | 62 |

| | | |
|--------------|---|-----|
| Figure 2.13. | Sequence alignment of C4BP-binding M protein HVRs of the M2/M49 pattern | 63 |
| Figure 2.14. | Sequence alignment of C4BP-binding M protein HVRs of the M22/M28 pattern | 64 |
| Figure 2.15. | C4BP-binding M protein HVRs that cannot be classified as belonging to either M2/M49 or M22/M28 patterns | 65 |
| Figure 2.16. | C4BP-M2 interaction | 66 |
| Figure 2.17. | Molecular dynamics simulation of the Arg39 hydrophobic nook interaction with wild-type M2 and M2 F75A. | 67 |
| Figure 2.18. | Interactions of M2 and M2 (K65A/N66A) with C4BP α 2 | 68 |
| Figure 2.19. | B-factors of C4BP α 2 bound to M2 or M2 (K65A/ N66A). | 69 |
| Figure 2.20. | Uncropped Gels | 70 |
| Figure 3.1. | Comparison of the novel Flipback heterodimer with resolved PKA RI α Conformations | 85 |
| Figure 3.2. | The stability of the Flipback conformation | 90 |
| Figure 3.3. | Generalized encounter complex of cAMP and the CBD-A/B | 91 |
| Figure 3.4. | Atomic numbering and naming of cAMP molecule in BD simulations | 91 |
| Figure 3.5. | Electrostatic descriptions of the four systems | 92 |
| Figure 4.1. | The Calcium Release Unit geometry | 100 |
| Figure 4.2. | Geometries of Normal and Deformed T-Tubules | 104 |
| Figure 4.3. | Sarcolemmal (T-Tubule) Fluxes | 107 |
| Figure 4.4. | L-Type Calcium Channel Action Potentials | 109 |
| Figure 4.5. | SERCA Models | 111 |
| Figure 4.6. | Sarcoplasmic Reticulum Fluxes | 116 |
| Figure 4.7. | Molecular species in absence of LTCC, RyR and SERCA | 119 |
| Figure 4.8. | Ryanodine Receptor gating modes in absence of L-Type Calcium Channels | 120 |

| | | |
|--------------|---|-----|
| Figure 4.9. | Ryanodine Receptor flux after a single action potential stimulus with a single L-Type Calcium Channel | 121 |
| Figure 4.10. | Ryanodine Receptor flux after a single action potential stimulus compared across different levels L-Type Calcium Channels | 121 |
| Figure 4.11. | Cytosolic Calcium comparing action potential alterations | 123 |
| Figure 4.12. | Ryanodine Receptor gating modes with 1 L-Type Calcium Channel comparing action potential alterations | 123 |
| Figure 4.13. | Ryanodine Receptor gating modes with 4 L-Type Calcium Channels comparing action potential alterations | 124 |
| Figure 4.13. | Cytosolic Calcium Calcium dynamics comparing T-Tubule geometry alterations | 124 |
| Figure 4.14. | Ryanodine Receptor gating modes with 10 L-Type Calcium Channel in deformed T-Tubules comparing action potential alterations | 125 |
| Figure 4.15. | Non-zero numbers of RyR opening with 10 L-Type Calcium Channels comparing RyR loss | 127 |
| Figure 4.17. | Non-zero probability of RyR opening with 10 L-Type Calcium Channels during an action potential train | 129 |

LIST OF SCHEMES

| | | |
|-------------|--|----|
| Scheme 3.1. | Activation Mechanism of PKA RI α with Different R conformations and Relative Association Rates | 86 |
|-------------|--|----|

LIST OF TABLES

| | | |
|------------|---|-----|
| Table 2.1. | Data collection, phasing and refinement statistics for native and SAD (SeMet) structures | 71 |
| Table 2.2. | Ionic Interaction Pair Occupancy in C4BP?2 for Quadrilateral Residues of M2 (%) | 72 |
| Table 2.3. | Ionic Interaction Pair Occupancy in Quadrilateral for residues of M49, M22, and M28 (%) | 72 |
| Table 3.1. | Rates of Association if cAMP with Cyclic Nucleotide Binding Domains of PKA complexes | 85 |
| Table 3.2. | Encounter Complex Description of R subunit and cAMP | 91 |
| Table 4.1. | Parameters of Calcium buffering species | 133 |
| Table 4.2. | Parameters for Sodium-Calcium Exchanger (NCX) and Plasma Membrane Calcium-ATPase (PMCA) pump models | 134 |
| Table 4.3. | Parameters for L-Type Calcium Channel model | 134 |
| Table 4.4. | Parameters for Ryanodine Receptor Markov model | 135 |
| Table 4.5. | Parameters for Sarco/Endoplasmic Reticulum Calcium-ATPase (SERCA) pump | 136 |

PREFACE

I begin this thesis with a feeling of mischievous satisfaction. As a child, I trained as an actor and learned to improvise in order to evoke emotion in my audience. This was my coping mechanism, of sorts, for an awkward and over-zealous personality. But, at my core, I am **an artist**. The truth is that during the course of my life, I fooled *a lot* of people into thinking that I was scientist and not an artist, including myself.

I sang before I could speak, danced before I could walk, and played with anything made sound. Music was not explicitly encouraged by my father. To buy my first guitar, he encouraged me to save my money and I did so, buying it myself at the age of 13 with money that I earned from tutoring. Πέτρος, my father, pushed me towards a career in medicine and I studied hard in order to appease him. From my father, I inherited a thirst for challenges and expanding my boundaries which allowed me to excel, and from my grandfather, Μηνάς an obsession with music, which distracted me fully. My grandfather was not fondest of my educational pursuits, offering marriage as the “logical” solution. But my mother and grandmother pushed me to continue with my education to achieve the dreams that they could never achieve.

In our village, *Óλυμπος Καρπάθου*, more than 80% of women never complete high school. This meant that my grandmother longed for an education that she never had the opportunity to have. Despite this precedent, my grandmother always encouraged me to pursue my education. My mother was married with a child at the age of 21, and she never completed college either. My grandfather never learned to formerly read and write in an academic sense, and my father never completed grade school. I had no precedent, but I *did* have encouragement.

Having gone to school in Greece during my early years and Greek schools throughout my primary education, the Socratic method-dialogue, also known as “conversational learning” or “learning with interruptions” was common. But my tendency towards hyper-stimulation always caused me to ask “too many questions.” Most American teachers found

it disruptive, but in the 9th grade, Ms. Primm asked me to stay after class; what I thought would be some form of rephension to which I was accustomed. Instead, she told me that I “asked the best questions that [she] ever heard from a student.” She encouraged me to become a scientist too, instilling within me the most important thing I’ve carried throughout my educational career: belief in myself. She was the first person that made me see myself as a scientist and helped me begin a journey that began with biology, seeded in chemistry, and flourished with physics.

My love affair with physics began at 4 AM on a June summer night of 2012, while laying on a couch in Baltimore, MD, USA. Exhausted from his shift at Hilltop Carry-out, my *συγχωριανός* *Ελυμπίτης* (fellow villager of Olympos) and dear friend Minas Giorgakis laid in his bed and put on a documentary for me to watch *about the mysteries of the universe* in hopes of occupying me enough so that he could get some sleep.

I fell deeply in love.

The vastness of the universe humbled me. The mere idea of combining chemistry with physics excited me. Just a year before, I was awarded a Bachelors degree in Biochemistry from Manhattan College, and was admitted to the PhD program at the University of California, San Diego, my dream school. After *bearing witness to* the vastness of the universe, I found myself looking for professors to work with in the department that studied cosmic chemistry. I quickly changed my mind when I realized that those beautiful fly-through images of space took decades to systematically study and acquire knowledge about. I was reminded of the advice of my beloved Physical Chemistry professor, Jianwei Fan, PhD: “*Sophia, stick to your strengths.*”

At first, I was offended by Dr. Fan’s suggestion that I stick to the study of organic chemistry, though I admittedly had intuition for synthesis. The Greeks have a word for my response: *πέϊσμα* (*peízma*). I encourage the reader to decide upon a definition for this

poly-semantic word. With *ισχυρογνωμοσύνη* (ishchyrognomosíni), a bull-headed will and stubbornness, I decided I would take the middle road, combining my love for biochemistry with my thirst to understand why we observe what we do, in the study of biophysics.

I packed up my life in my 2007 Spyder Eclipse named “Angie” after the legendary Rolling Stones song and along with my best friend, Yianna (Rodriguez) Zois, took off on a ten day adventure to go to the place I always wanted to be, San Diego, California, USA. The brisk air, the green mountains; those sandy valleys and glorious sunsets of San Diego acted as a mere substitute for the place which I had never been able to call home, my beloved Karpathos.

In retrospect, I sometimes wish I *had* listened to Dr. Fan. I often wish that I took Dr. (Kyriacos Costas) K.C. Nicolaou’s invitation to join his lab at Scripps. I would have likely finished my PhD earlier, since I had a strength for synthesis; but the number of gray hairs in exchange for total synthesis over biophysics might never be known.

But something drew me to biophysics—the ability to “see” what I learned about as a biochemist. Using mathematically-accurate artistic and computational approximations of proteins and cells, I could finally fully understand the concepts that Dr. Chiara Indiani drilled deep into my mind. It felt intuitive to examine proteins and see how their secondary structure was disrupted by molecular motions. The challenges, however, were more than I anticipated.

The learning curve for biophysics was *incredibly* steep. In addition to the complexities of the mathematics I was learning, I suffered endlessly by exchanging “a” with α and “b” with β , a pain that my beloved professor Katja Lindenberg found comical, since Greek is my first language. But at the end of statistical mechanics, I had the pleasure of understanding the elegance with which quantum mechanics and physical chemistry describe molecular motion with each integration of each time step. This, “mathematical magic” coupled with my ability to visualize the biophysical implications of system alterations on the level of the protein convinced me that I was on the right path. I had my heart set on

biophysical chemistry, but still I yearned for more. I yearned to apply my understanding of biochemistry to the complexities of diseases.

Early on in my life, I developed a kind of *obsession* with identifying and understanding illnesses. This started as soon as I could read. My mother had a book by H. Winter Griffith called “A Complete Guide to Pediatric Symptoms, Illnesses and Medications.” I would often find myself correlating my siblings’ symptoms to illnesses suggesting doctors visits, way before WebMD was a thought. My curiosity blossomed in my later years as I learned more about biology and the ways that mutations lead to phenotypical differences that cause diseases; diseases that I would one day witness the real effects of.

In college, I worked for three years in the Department of Surgery Research at Johns Hopkins University Bayview Medical Center. In partnership with the Firefighters Association of Baltimore, MD, we aimed to treat burn wounds on animals and humans with the use of protein cocktails. As a biochemist interested in serious medical conditions, this fascinated me—the concept that small proteins (nm) could heal damaged tissues for the regeneration of skin cells to form collagen and hair follicles to generate α -keratin. I later worked in the Department of Surgical Pathology at Montefiore Medical Center where I dissected human organs. Morbid, I know but I worked with diseased organs daily, trying to understand **why** patients were sick. It was here that I learned the true meaning of multiscale biochemistry.

Like true scientists, after receiving any surgically removed organ during or after surgery, we followed a very strict protocol. The organ would be measured, visually described, and the description dictated to a computer directly into a pathology report. The organ would be fixed in a fixation agent, like formaldehyde, in order to preserve the ultrastructure of the specimen. Successive dissections dependent on the tissue type would be made until an interesting characteristic was identified. This could be, for instance, a calcification of tissue or necrosis of a major artery. My job was to identify the morphology of the disease on the cellular level using a light microscope. In this way, the multiscale

nature of disease was revealed to me.

This experience sparked within me a desire to understand the cause of diseases. In graduate school, my advisor, Dr. Rommie E. Amaro gave me an opportunity to do just this in a way that I could have never imagined. Using a unique set of computational tools, I have learned to tease out the factors that underlie diseases. The most satisfying part of this experience is the use of tools that allow you to visually interpret the results of our experiments. In the following dissertation, I will describe my adventures in using the computational microscope to understand the motion of biological molecules as well as the underlying causes of diseases.

ACKNOWLEDGEMENTS

My doctoral work would not have been possible without the support of my beloved advisor, Dr. Rommie E. Amaro. Thank you, Rommie, for taking a chance on me. Thank you for believing in me when I lost faith in myself. Thank you for pushing me to finish when all I wanted to do was walk away.

I also owe much of my success to my mentor ship by Dr. Terrence J. Sejnowski of the Salk Institute. Thank you, Terry, for challenging me and encouraging me to be the best scientist that I could be.

I would also like to thank my thesis committee for their valuable input and insights: Dr. J. Andrew McCammon, thank you for being a teacher and mentor to me. Thank you for leading me to Rommie and for supporting my research in collaboration with your lab for so many years. I value your advice and the moments I have spent with you.

Dr. Charles L. Perrin, thank you for showing me that Organic Chemists can have a burning passion for the physical forces that underlie chemical function. Thank you for sitting with me and reviewing my mathematical derivations, and for making me question even the smallest details. You have made me a better scientist.

Dr. Elizabeth Villa, thank you for your mentorship and for being a role model to me. Even more, thank you for lifting me up when I was at at my lowest. Knowing and loving your daughter, Isabella, has been so wonderful and healing for me. I can't wait to see the woman that she will become with you as her guide.

Dr. Katja Lindenberg, thank you for challenging me as my professor, leading me as a mentor, and loving me like a friend. You are such an inspiration to me, and I could never repay you for all the love and guidance you've given me throughout the years. Thank you for insisting that I be a mathematician and instilling within me a love for statistical mechanics.

I have also had the pleasure of being mentored by several wonderful scientists, whom I would also like to acknowledge.

Dr. Thomas M. Bartol, thank you for challenging me to be the best version of myself every single day. Without your patience and guidance, I would not be where I am, and for this, I thank you.

Dr. Robert M. Malmstrom, your early mentorship on the complicated history of PKA was invaluable to my success. Thank you for providing me with the simulations wherein-which we discovered the Flipback conformation.

Dr. Alexandr Kornev, for the many hours and discussions on PKA, I thank you. I thank you for pushing me to publish my work and for celebrating along with me when the work was well-received by the community. It was a privilege to be mentored by you throughout my PhD work.

Dr. Susan S. Taylor, for the many discussions we've had trying to uncover the structure of PKA, I thank you. I am grateful for your encouragement and mentorship.

Dr. Elizabeth Komives, my beloved teacher, I thank you for insisting that I be critical of the papers I read and for giving me a framework to upon which I built my understanding of biophysics. Thank you for exposing me to the solution-side of the coin and for encouraging me to publish my discoveries. Thank you utmost for the opportunity to lecture in your courses about my findings.

Dr. Gary S. Huber, my colleague and friend, thank you for fixing BrownDye every time I broke it and for teaching me how to gain insights from Brownian Dynamics simulations.

Dr. Andrew D. McCulloch, my professor, mentor, and co-author. I thank you for your guidance throughout all these years of research. Without your mentorship, my project would not have progressed in the time that it did. I thank you especially for giving me the ability to become an instructor in the graduate course, "Numerical Analysis for Multiscale Biology." It has been my honor to see students grow in a similar way that you watched me grow as a scientist. I wish only that Anushka had the pleasure of seeing the project develop. I hope we made her proud.

Dr. Andy G. Edwards, my colleague and collaborator, thank you for your constant guidance, encouragement, and support throughout this project. Thank you also for the opportunity to be part of the Simula Summer School as a project leader. It has been wonderful to work with you and hope only that our collaboration strengthens.

Dr. Tedd Keating, thank you for giving me my first crack at academic research and for letting me develop the methods for our study. Your belief in me helped me develop confidence in myself.

My wonderful professors chemistry professors from Manhattan College, Dr. Jianwei Fan, Dr. John Capitani, Dr. Gary Kolks, and Dr. Chiara Indiani. Thank you for demanding excellence from me.

I am especially thankful to Dr. James V. McCullagh for his mentorship and for challenging me to beat the odds. Thank you “boss” for your loving guidance and for making me into a true scientist and great presenter.

My beloved physics professor, Dr. Sezar Fesjian, thank you for doubting me initially, and then believing in me fully. You taught me that physics is an art and for that, I cannot thank you enough.

To my dear friends at Manhattan College, especially Mohammed Shoeb Chikhlikar, thank you for standing alongside me as I developed into a biochemist and for making the hard times into great memories.

To my first friend in the Sejnowski lab, Hosam Yousif, I wish you could have seen how far I came. I miss you terribly and thank you for your support in my early days.

To my Amaro lab family, I love you all so much. Special thanks to Dr. Lane W. Votapka for introducing me to the lab and for bringing me closer to God through our time together. To Dr. Adam VanWart, thank you for loving me and encouraging me while you were here in San Diego and after you left. Dr. Jamie M. Schiffer, thank you for the countless hours of encouragement and your friendship during your time in the lab. Dr. Tavina Offut, thank you for your writing advice and for laughing along with me when we

were both at our breaking point with your projects. Dr. Ozlem Demir, I appreciate so much your advice and guidance as my colleague and friend. Dr. Jeffrey Wagner, thank you for making me laugh and giving me advice on how to finish strong. Chris Lee and Brynn Taylor, thank you for making me laugh and encouraging me with high-fives when I needed them. Pek Jeong and Emilia Pecora de Barros, my longest friends in the lab, thank you for all your help and encouragement as we all grew together. My little sisters, Mia Rosenfield and Alexandra Sasha Heyneman, thanks for being there to make me feel like a big sister and for trusting me to impart advice and wisdom on you. I miss you Sash. TO Christian Seitz, I will miss your laugh. Thank you for making it contagious. My big brothers, Dr. Zied Gaib, Dr. Conor Parks, and Dr. Lorenzo Casalino, you are all one of a kind. I am so thankful for your hugs every time I came into the “men’s locker room.” And finally, to my first mentors Dr. Robert Swift and Dr. Jesper Sorensen, you are both my heroes. Thank you for being the first to teach me how to be a computational biophysicist and for making the experience a smooth one.

To the technological and administrative support teams in the Amaro/McCammon and Sejnowski groups, I cannot thank you enough. Dr. Robert Konecny, thank you for all your support of APBS and for making it work for my projects. Bryan Hill and Kevin Smith of the Physics Help Desk, I would not have been nearly as productive or successful as I was without your constant technological support. Thank you for teaching me how NOT to break things. Jorge Aldana, Chris Hiestand, Bryan Nielsen, and Brad Brockmeyer, my work at Salk would not have been possible without you. Jorge, special thanks to you for all your help with the cluster. Additional thanks to Rick Tomas for his responsiveness to issues with anything and everything at UCSD. I also thank Jill Carson for being my biggest supporter for the yearly donations to the Moores Cancer patients. Thank you also to Mary Ellen Perry and Ramona Marchand for all of their support of the group at Salk Institute. Special thanks to Teri Simas for all her loving support and her food pantry that keeps us all going. Your generosity is unparalleled.

I would also like to acknowledge the Chemistry Student Affairs department. Jeanine Sun and Erica Lennard, I thank you both for helping me navigate any challenge in graduate school. Special thanks to my boy Jeff Rances, the first ever person to welcome me to UC San Diego prior to my admission to the program. From playing basketball to grilling for the department, we go way back. I love you Jeff!

My McCammon family, I thank you so much. Special thanks to Dr. Pete Kekeneshuskey for providing a starting point for my simulations with SERCA as well as Dr. Vincent Metzger for being a great friend and for leading me to a fruitful collaboration with Cosmo. Thank you also to Dr. Giulia Palermo for her friendship and adoration.

I would also like to acknowledge Dr. Cosmo Z. Buffalo for his support of the C4BP and M-protein project. You, Cosmo, are the definition of an amazing colleague. I thank you so much for being a friend to me. It was my pleasure to work with you.

My beloved sister and colleague Kimberly McCabe, thank you for making the last year of my PhD a wonderful one. Learning with you, from you, and teaching you what I know has been a wonderful experience. I am SO proud of you and hope we remain colleagues forever.

To my first-ever friend at UC San Diego, Tamika Franklin: thank you for always supporting me from day one all the way to defense. My Jamaican sister in Christ, I love you so much!

To my most consistent friend and colleague, Dr. John (Jack) Berkowitz, thank you for all your help throughout the years. Thank you for being my partner in my first MCell project and for all the times that you helped me to see the light.

My roommate, Osinachi Ajoku, thank you for supporting me at home when I was having the hardest time in school. Thank you for always cheering me on and for making me feel like I was the “MVP.”

My extended family, Robyn Ridley and Jordan Campbell, thanks for laughing with me at the frustrations that we all faced in graduate school, and for giving me a family

feeling when I was far away from home.

To my graduate mentor, Dr. Timia Crisp, thank you. I wouldn't have even known who Rommie was without you. I'll always cherish the memories we had together and am so thankful that we still maintain our friendship despite the distance between us.

My dear loving friend Jacob Thompson, thank you for supporting me with many late night sleep overs and for talking me through some of my darkest times. You're my closest and most consistent friend in San Diego. I love you so much.

My dear friend Dr. Robert Tolley, thank you for all the times you baked me bread and nursed me back to health when I was sick.

My son, Chandler Puritty, thank you for literally saving my life, the many times that I was sick and needed to be in the hospital. Thank you for making me feel like a "Queen" when I felt defeated. Thank you for looking up to me and insisting that I be the best version of myself, always.

Over the course of my graduate career, I spent more hours in meetings with the Associated Residential Housing Community Advisory Committee (ARCHAC) than I did in scientific group meetings split between three labs. Because of this, I forged a bond with a group of very special people. To all the former members of ARCHAC, I thank you for your friendship. It was my greatest honor to serve as a representative for UCSD graduate students, especially those with disabilities and victims of assault.

A special thank you goes out to my dear friends, Ramona Ferreira and Malia Mahi, who sheltered me during many a storm. Additionally, I'd like to thank the most thankless of them all, Mark P. Cunningham, who recognized my passion for activism and embraced it wholeheartedly; allowing me to shine in the darkness.

To my dear friend Dr. Cory S. Stephenson: words can not describe the profound impact you had on my trajectory. You were our leader on ARCHAC, my mentor in the GSA, and one of my dearest friends and role models. Thank you for always giving me the advice I needed to hear. I am forever indebted to you for your love and guidance.

To all my friends and fellow representatives in the GSA Council, thank you for tolerating my antics and for appreciating my tenacity. Thank you to all my political enemies who tested my love for the game. You made a better fighter out of me.

To my peers on the Diversity Advisory Council, thank you for trusting me to collaborate on and lead important projects. Special thanks to Dr. Tara-Lynne Pixley for first trusting me to do the work that needed to be done on our campus and for allowing me to be a part of her legacy as the Vice President of Equity Diversity, and Inclusion.

I would also like to thank the director of the Office for Students with Disabilities, Joanna Boval for her help throughout the years. Joanna, thank you for helping me balance my activism with my work as a scientist. I could never repay you for all that you've done.

My wonderful therapists, Dominique Silvas and Amy Torn, thank you for providing amazing support during the hardest times in my graduate career. Amy, especially, thank you for allowing me to explain my findings to you on many a drawing board, and for encouraging me to get through the darkness.

To the Samatas family, I love you so much. Thank you for making Piraeus a home for me; caring for our small apartment and our properties there while we are away. Maria, Pantelis, Georgia, and Mixalis, I love you all so much.

To my (pseudo)cousin Bill Protopapas, thank you for being the one and only person I could always count on to see when I was back in New York. Our many dinners and good times are something I will cherish forever.

My San Diego Yankees homie, Johnny Antoniadis, thank you for being such a HUGE supporter of me. Special thanks to my San Diego Bronx Bombers group for enjoying the baseball seasons with me. The long seasons would have been lonely without you all.

To my neighbor Sara Solaimani, thank you for being my rock during some of the hardest times. For all the times we nursed each other's sorrows on our balconies and

laughed together, I thank you. Most of all, thank you for letting me be a mentor to your son, Diako, who I love so much. I am so proud of that kid and being in his life was one of the most rewarding things I had the opportunity to do here in San Diego.

My buddy Benjamin Kellman, thanks for being a wonderful person and friend to me. Working together on our service projects with iGem was a wonderful experience and I can't wait for you to join the ranks of Doctor!

My loving friend Jennifer Fauskin, thank you for all the laugh-filled phone calls and your friendship. My darling, Emily Litrell, thank you for giving me an excuse to visit you in Washington and for bringing into this world the most adorable baby, Carter.

My best friend, Maggie Hoban, thank you for being like a therapist to me and for all of our wonderful memories. Thank you especially for being here for my special day. I love you.

To my dear friend, Annamaria "Beba" Diakokomninos, thank you for helping me through the darkest times in the beginning of graduate school and for supporting me from a distance throughout the years.

To the entire Giorgakis family, Antoni, Minas, Yianni, Eleni and George, thank you for welcoming me every time I visited Baltimore and for being so proud of me. I love you all so much.

To my best friend Sarah Young, I can't thank you enough for all your love throughout the years. Thank you to Mary and Victor Young for always welcoming me into their home. And to my special friend, Isaiah Victor Young, I am so thankful God blessed the earth with your birth. I can't wait to see who you become and I hope always to make you smile.

To my first ever friend in America and big brother, Darron "DJ" Allen Jr., thank you for supporting me and telling me I could do this. Thank you also for blessing me with the honor of being a god mother to your son, Trey. Trey, God mommy loves you so much!! Dana Allen, thank you for being the perfect partner to my bestie and for blessing us all with the birth of Jada Allen. I love you sis! And to my family in Christ, the Allen Family,

Mr. Darron and Mrs. Gwen, Lauren, John, and Otanya, thank you for being my family. I love you all to the moon and back.

My darling Rita Tomas, thank you for your unwavering support and helping me focus when I was losing my way. Thank you for making Tomas Fishing and Diving Shop my office when I needed to finish my work. Only God knows how much I love you.

My loving Koumpara Sofia Protopapas, thank you for being the example of strength and humility that I seek to be. Thank you for blessing me with your daughter as her god mother. Kalliopi Rosa, Popi mou, my fairy god daughter, loving you and watching you grow has been an amazing experience. I am so excited to see you blossom as a young lady. I love you baby!! And lastly, my koumpara Kalliopi Protopapas, thank you for loving me like my grandmother, and for making me smile every time I see you and hear your voice. I love you.

I would like to thank my great uncles and aunts for supporting me as the first one in our family to pursue an advanced degree. Specifically, Maria Nicolaidis, Emmanuel and Sofia Mastromanolis, Petros and Evangelia Sotiriou, and Nikos and Mary Prearis. I thank you all for channeling the love of my grandparents, Minas and Kalliope Prearis, and Emmanuel and Sofia Hirakis and encouraging me the entire way through my program.

To my theio Hlia and theia Maria Kanaki, I love you. Thank you for taking care of our home in Greece while we work in America, and always making it a wonderful place to return to each time we come back.

I would also like to thank my loving cousins on both sides of my family. Sofia and Nikos Niotis, thank you for encouraging me and for being like my second parents. My cousins, Billy, Kristen, Manoli, Irine, and Mario, you are the siblings that God forgot to give me. Growing up with you has shaped the person I am today. I love you all so much.

To my older cousin, Suzy Hirakis, I thank you for being my first role model. I know now that I am a role model to your daughters, Gianna and Tina, and I wouldn't have it any other way. I love you and thank you for always being my cheerleader. To my cousin

Manoli, and my theio John and theia Georgia, I thank you for being there for me and for your pride in me.

I would like to thank especially, my cousin Nikos K. Dais who lives in the village of Diafani, Karpathos. Niko mou, thank you for taking all the steps to make sure that I could work at the store with you all summer. You, more than anyone, did everything to support my thesis work in the summers away from San Diego. I can't thank you enough for encouraging me to finish and for demanding the best from me, always.

To my aunts, theia Jenny, Christina, and Anna, uncles, theio John, Spiro, and Paul (Souvlaki) and my cousins, Vicki, Costas, Minas, Louisa, Popi, Fotini, Lia, and Nikita, I love you all so much. Your encouragement throughout the years helped see me through. I love you so much.

The most important people that I can thank and acknowledge are immediate family. My father, Peter; my mother, Marina; my sisters Kally and Maria; and my little brother Manny. Though I had to leave and go far away to accomplish this milestone, I never left you behind. You stood by my side and helped see me through to the other side. I'm so thankful that you are my family. There are no words to describe how much I love and cherish you.

Chapter 1, in full, is a reprint of the material as it appears in *Frontiers in Physiology* 2015. Boras, Britton W.; **Hirakis, Sophia P.**; Votapka, Lane W.; Malmstrom, Robert D.; Amaro, Rommie E.; McCulloch, Andrew D.; "Bridging scales through multiscale modeling: a case study on protein kinase A," *Frontiers in Physiology*, 6: 250, 2015. The dissertation author was the co-primary author of this paper.

Chapter 2, in full, is pre-publication print of an article published in *Nature Microbiology* 2016. Buffalo, Cosmo Z.; Bahn-Suh, Adrian J.; **Hirakis, Sophia P.**; Biswas, Tapan; Amaro, Rommie E.; Nizet, Victor; Ghosh, Partho; "Conserved Patterns hidden within group A *Streptococcus* M protein hypervariability are responsible for recognition of human C4b-binding protein." The dissertation author was the third author of this paper.

Chapter 3, in full, is a reprint of the material as it appears Biochemistry 2017. **Hirakis, Sophia P.**; Malmstrom, Robert D.; Amaro, Rommie E.; “Molecular Simulations Reveal an Unresolved Conformation of the Type IA Protein Kinase A Regulatory Subunit and Suggest its Role in the cAMP Regulatory Mechanism.” The dissertation author was the primary investigator and author of this paper.

Chapter 4, in currently being prepared for submission for publication of the material. **Hirakis, Sophia P.**; Bartol, Thomas M.; Sejnowski, Terrence, J; Amaro, Rommie E; The dissertation author was the primary investigator and author of this material.

VITA

- 2011 Bachelor of Science in Biochemistry, Manhattan College, Riverdale, NY
- 2012-2015 Teaching Assistant, Department of Chemistry and Biochemistry
University of California San Diego
- 2012-2015 Master of Science in Chemistry, University of California San Diego
- 2012-2019 Research Assistant, University of California San Diego
- 2019 Doctor of Philosophy in Chemistry, University of California San Diego

PUBLICATIONS

Hirakis, Sophia P.; Bartol, Thomas M.; Sejnowski, Terrence J.; Amaro, Rommie E.; “Molecules to Membranes: The Calcium Release Unit.” *Biophysical Journal*, 2018, 114(3), pp. 118a.

Hirakis, Sophia P.; Malmstrom, Robert D.; Amaro, Rommie E.; “Molecular Simulations Reveal an Unresolved Conformation of the Type IA Protein Kinase A Regulatory Subunit and Suggest its Role in the cAMP Regulatory Mechanism.” *Biochemistry*, 2017, June 29; 56 (30), pp 3885-3888.

McCullagh, James V. and **Hirakis, Sophia P.**; “Synthesis of the Commercial Fragrance Compound Ethyl 6-Acetoxyhexanoate: A Multistep Ester Experiment for the Second-Year Organic Laboratory.” *Journal of Chemical Education*, 2017, June 29; 94 (9), pp 1347-1351.

Buffalo, Cosmo Z.; Bahn-Suh, Adrian J.; **Hirakis, Sophia P.**; Biswas, Tapan; Amaro, Rommie E.; Nizet, Victor; Ghosh, Partho; “Conserved Patterns hidden within group A Streptococcus M protein hypervariability are responsible for recognition of human C4b-binding protein.” *Nature Microbiology*, 2016, September 5; 1:16155.

Boras, Britton W.; **Hirakis, Sophia P.**; Votapka, Lane W.; Malmstrom, Robert D.; Amaro, Rommie E.; McCulloch, Andrew D.; “Bridging scales through multiscale modeling: a case study on protein kinase A.” *Frontiers in Physiology*, 2015, September 9; 6:250.

Keating, Tedd; **Hirakis, Sophia P.**; McBride, Nicole; “Perceiving the center of Human Figures.” *Proceedings and Abstracts of the Annual Meeting of the Eastern Psychological Association*, March 6, 2012; vol 81, pp s97.

Friedenberg, Jay D.; Keating, Tedd; **Hirakis, Sophia P.**; Toscano, Lisa; “The Effects of a Typical In-season Practice Session on Peak Ground Forces and Select Kinematic

Variables in Women's Division I Basketball Players." Journal of Strength and Conditioning Research, February, 2012; vol 23, pp 87.

FIELDS OF STUDY

Major Field: Chemistry and Biochemistry (Specialization in Multiscale Biology)

Studies in Biophysical Chemistry
Professor Rommie E. Amaro

Studies in Computational Biology
Professor Terrence J. Sejnowski

Major Field: Surgery and Medicine

Studies in Surgical Research
Professors Guy Marti, M.D., John Harmon, M.D.

Studies in Emergency Medicine and Toxicology
Professor Joseph LaMantia, M.D.

Major Field: Synthetic Chemistry and Chemical Education

Studies in Synthetic Organic Chemistry
Professor James V. McCullagh, Ph.D.

Major Field: Kinesiology and Cognitive Perception

Studies in Kinesiology
Professor Tedd Keating, Ph.D.

Studies in Cognitive Perception
Professor Jay D. Friedenberg, Ph.D.

ABSTRACT OF THE DISSERTATION

Multiscale analysis and visualization of
biophysical structure and biochemical function with
computational microscopy

by

Sophia P. Hirakis

Doctor of Philosophy in Chemistry
with a Specialization in
Multiscale Biology

University of California San Diego, 2019

Professor Rommie E. Amaro, Chair
Professor J. Andrew McCammon, Co-Chair

The evasive source and cause of a disease is oftentimes *smaller than you think*. Imagine, though, chasing something that you can't actually *see*. Fortunately for the modern-day biomedical scientist, computational tools harnessing the power of physics using the language of mathematics are able to *see the invisible*. Computational microscopy is a tool developed to visualize the energetic behavior of biological systems. With progressive advancements in computer graphics and the development of mathematical theories to

explain biological behavior, computational microscopy has become a useful tool used by many kinds of scientists over the greater half of the last century to understand the energetic underpinnings of a system's behavior. Unlike most "microscopes," it allows us to visualize extremely small entities like atoms, molecules, proteins, and cells. More importantly, it allows us to spatiotemporally transcend scales to understand the dynamics of our systems. Like a biophysically detailed time-lapse, we are able to *see through time*, to understand chemical "butterfly effects" that transcend the time and space scale at which they operate. In this thesis, the computational microscope is applied to multiple systems to *visualize and analyze* the physicochemical mechanisms that underlie biological function. Specifically, the thesis is centered on the structure of proteins and subcellular mechanisms driving cardiac function and dysfunction. In the first chapter, we address the concept of multiscale biological simulations, integrating information from atomistic scales toward cellular models of Protein Kinase A. The second chapter demonstrates the ways that atomistic simulations can be applied to the study of the structural interactions in protein-protein complexes vital to the infectious mechanisms of Group-A *Streptococcus*. In the third chapter, two scales of biological simulation are used in tandem to understand the structure and the kinetic behavior of Protein Kinase A RI α . The final chapter incorporates the kinetic understanding of relevant species in a realistic subcellular geometry to investigate signaling mechanisms that underlie calcium activation in healthy and diseased hearts. Particular attention is paid to the way that structural alterations on the atomistic, molecular, and membranous level alter the behavior of biological systems. Holistically, this thesis is centered on the use of computational tools and the development of realistic models that can reproduce experimental findings and predict the behavior of systems, driving the creation of new hypotheses.

Introduction

The Greek word, *πολυσήμαντα* (poly-sêm-anta, english: polysemantic) translates to the phrase: a symbol or word that has multiple meanings. Depending on where accents are located in Greek and other accented languages, the meaning of a word can change completely. As such, words can be misinterpreted when translating from one language to another, especially because of the sociological implications of the passage of time. From this phenomenon stems so much of the world's fascination with science: the cycle of observation and theoretical inquiry. Curious minds have learned *simply* that words cannot properly encapsulate the meaning of something fully. Rather, it is *feeling and bearing witness* that satisfies the understanding that the mind craves.

A language exists, which is not subject to misinterpretation of translation, and that is the language of mathematics. From the first known records of mathematics which date back to approximately 3000 B.C.E. until now, what has been common between four sheep and four goats has always been their *value*, equaling to 4. With common units of length, time, volume and the like, we can describe in exponential terms the quantitative *scale* of a value. Using mathematics, humans can describe observable phenomena—from the attractive forces that govern the behavior of the smallest atom, to the revolution of planets in galaxies that are not our own. What is more, mathematics allows us to tease out simple formulas that *predict* the future state of a system. In simple terms, mathematics is used to encode, translate, and communicate *energy*.

To understand the power of mathematics, we must lend some consideration to the ability to *visualize* energetic concepts. This concept is best demonstrated by one of the

most amazing phenomenon known to mankind: the musical genius of Beethoven, a **deaf man**. How could it be possible for a man who could not *hear sound* to produce such melodic masterpieces? The answer is simply, his ability to visualize sound energy. Ludwig van Beethoven once said,

“I always have a *picture in my mind* when composing, and follow its lines.”

Ludwig van Beethoven was able to see and feel the mathematical and energetic relationship of musical notes relative to one another, even though he could not hear them. His words allude to the existence of a phenomenon that so many non-traditional artists and scientists live: a deepened understanding of something is somehow tied to ones ability to *feel it, and visualize it* in their mind’s eye. The famous mathematician, James Joseph Sylvester once said,

“The musician feels mathematics, the mathematician thinks music:
music the dream, mathematics the working life.”

Many of us do not know what it means to “visualize” energy, music, or mathematics. But, if you have the ability to hear sound, you may recognize the note commonly known as “A” or “La.” If you used a pressure reader near the source of the note, the pressure recorder would record 440 energetic detections per second, or 440 Hz (see figure 0.1a). Lower frequencies like the note “D” or “Re” at 240 Hz have longer wavelengths (see figure 0.1b). One can infer that the “lines” that Beethoven referred to are the amplitudes of the sound frequencies.

If you have ever played a musical instrument, you may already know that these notes compliment each other nicely, a phenomenon known in music as a “harmonious interval.” Practically, this means when the two sound waves interact, their waves sum to produce a complex wave (see figure 0.1b). This is but one example of how mathematics and art can be used to *visualize* a physical phenomenon best described by the senses.

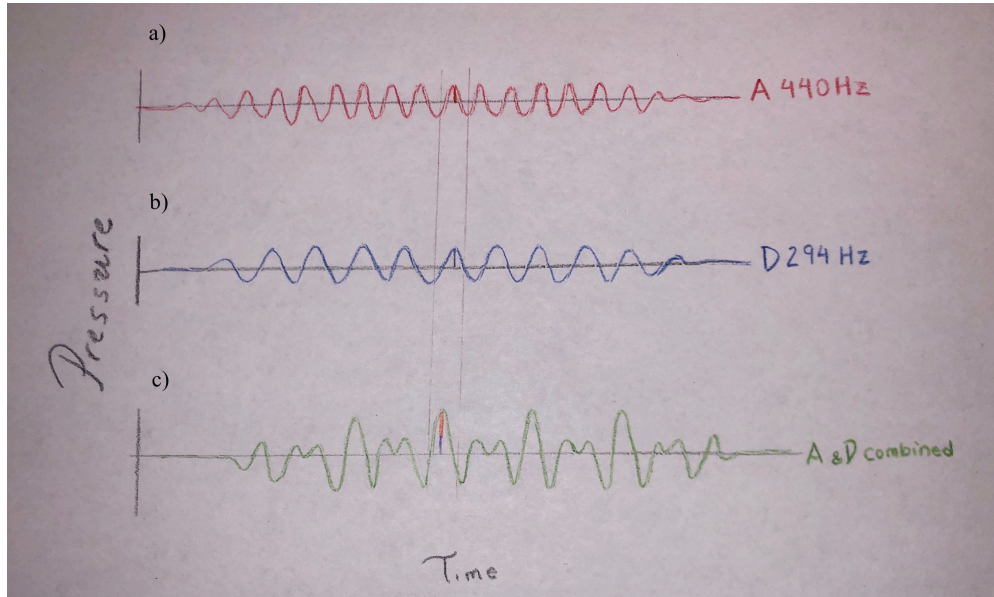


Figure 0.1. A drawing of musical frequencies

(a) A drawing of the high frequency, 440Hz musical note, “A” or “La” (red); (b) A drawing of the lower frequency, 294Hz musical note, “D” or “Re” (blue); (c) A drawing of the additive frequencies of “A” and “D” to produce an complex wave function; Adapted from freely available lessons by 3Blue1Brown.

This begs the question: how is one able to perceive the energies that surround them? The answer is quite complex but can be simplified with the following explanation: biological organisms are *master detectors*. Every organism, from single-celled bacteria to humans, to organisms that have yet to be discovered living in the deepest oceans and the highest peaks, have the ability to detect energetic changes in their environment in some form or another.

With properly functioning sensory faculties, we are able to perceive and interpret the energies around us, allowing us to respond to stimuli, learn, and adapt to our environment. This process, however, only works for phenomena that can be observed—at frequencies and energies that can be detected by our sensory systems. More often than not, interesting phenomena are too small/fast and large/slow to be observed and interpreted without aide of technology. Fortunately, the language of mathematics can be used to describe many observable and non-observable phenomena, allowing for contextualization and

understanding by inquisitive minds.

0.1 Scaling up through the sciences

As noted earlier, our knowledge and understanding of the world around us is timelessly encoded in the language of *mathematics*. The application of mathematics to the laws that govern the physical world is termed, *physics*. The laws and theories of physics applied to the study of matter are collectively, *chemistry*. The use of chemical principles used to describe living systems is *biochemistry*. All of these disciplines are common in their use of mathematics to prove the principles that are upheld by what we are able to observe through our senses and detect with technology.

Oftentimes, biological systems are too complex, and detailed mathematical descriptions of them become difficult to compute by hand. For this, scientists use computational algorithms that describe and predict behaviors. But the more explicit we become in describing a system, the more computational power we use. In 1965, Gordon Moore published a paper entitled, “Cramming more components onto integrated circuits,” [1] where he first theorized about the relationship between the passage of time and the growth of computational power, known today as Moore’s law. Over half a century later, our technology has grown to a point where almost every person carries a small, powerful computer in their pocket that gives them access to volumes of up-to-date information at the mere decision of the user. Though computation is at its most advanced state to date, biological systems operate so fast and at such small time intervals, that computational power is often the limiting factor in obtaining an answer to questions that go beyond the spatiotemporal scale in question.

Nevertheless, using our currently available technology combined with sophisticated mathematical algorithms, we computational biophysicists have collectively created a powerful tool, *the computational microscope*. Like a true microscope, we can choose a

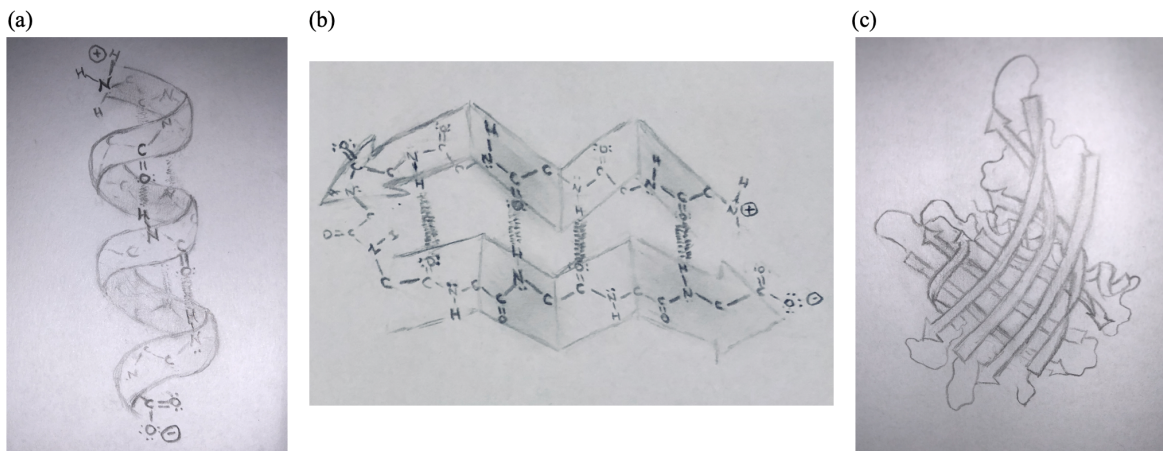


Figure 0.2. Levels of protein structure

(a) α -helix structure; (b) Parallel β -sheet structure; (c) Tertiary structure of a protein comprised of β -barrels and an α -helix

“magnification” or scale to visualize that which we seek to see. Using innovative methods for artistic rendering of atoms, proteins, and cells, the invisible suddenly becomes visible. Most importantly, we can see the biophysical ramifications of small changes that scale upwards, a biophysical butterfly effect.

With humility, I ask the reader: join me in this journey across scales of microscopic biophysical space-time to see the unseen, through the lens of the computational microscope.

0.2 The multiscale nature of biochemistry

The complexity of biological systems stems from the various scales of space and time at which they operate. In terms of the field of biochemistry, there are many levels that operate in unison to produce an observable phenomenon. This section will describe the various mechanisms that underlie structure in biochemical systems.

Firstly, *atoms*, joined by attractive forces called bonds, link together to produce *molecules* called *amino acids* (also known as *residues*). Atoms come in different sizes and shapes but are generally on the order of $1 \times 10^{-10} \text{m}$ (1\AA). There are twenty different amino acids, all differing in their side-chain structure and therefore, their physical properties.

Each amino acid molecule has a common backbone which begins with a reduced nitrogen atom (amino) and ends with an oxidized carbon atom (acid) with a width of about 4Å. Some amino acids are polar and charged, interacting with other charged species and water. Others are hydrophobic and “hide from water,” through water-exclusion forces. These acid interactions dictate the structure or the “fold” of the protein. Figure 0.2 demonstrates the levels of protein structure: from hydrogen bonding in alpha helices (a) and beta sheets (b), to folds that result in tertiary structure (c).

The structure and dynamics of a protein are determined by its sequence. The amino acid sequence is specified by the genetic code of the organism, known as deoxyribose nucleic acid (DNA). Our DNA is *translated* to messenger ribonucleic acid (mRNA) and *transcribed* to the protein sequence. The linkage of amino acids in a linear “code” is known as the *primary structure* of a protein. When the protein chain lengthens, the protein backbone folds onto itself using non-bonding interactions between the backbone hydrogen and oxygen atoms called hydrogen bonds. These interactions lead to the formation of *secondary structure* like α -helices and β -sheets (see Figure 1a,b). Collections of secondary structural moieties folded on top of one another are the protein’s *tertiary structure* (see Figure 1c). Several proteins within the same system are a “complex” and the individual proteins are known as subunits. For example, a protein complex of two regulatory (R) and two catalytic (C) subunits can be written as R₂C₂. Two or more subunits within a complex is termed the *quaternary structure* of the protein. Different quantities and arrangements of the same subunits can have different structures and functions. Thus, quaternary structure yields multiplicity in the number of structures a single protein complex can have.

A slight change in amino acid sequence can alter the *structure* of a protein drastically, and oftentimes, affect its *function*. For example, residues like Glycine (G) confer lots of flexibility in helices, often causing helices to break reversibly. A simple mutation to a more-rigid residue like Proline (P) lead to a physical turn in the structure and this can break a helix without allowing for flexible return the helical moiety. Mutations thus

oftentimes affect the way the protein moves and interacts with other molecules, leading to large-scale changes in signaling cascades on the cellular level. Such is true with changes in expression levels of DNA, RNA, proteins, In this way, atomic-level alterations can have a “*biophysical butterfly effect*”, leading to the presentation of disease phenotypes in the organism harboring the mutation. Protein Kinase A (PKA), one of the systems examined by this thesis, is a prime example of a case in which a single-point mutation can alter structure and chemical interactions, and is discussed in more detail below.

0.3 How mutations affect protein structure: a multiscale perspective

Protein Kinase A is found in every multicellular organism. It is responsible for responding to extracellular signals and eliciting an intracellular response; a signal translator and amplifier of sorts. PKA binds a molecule named cyclic adenosine monophosphate (cAMP) in the cyclic-nucleotide binding domain (CBD) regulatory subunit. Once enough cAMP is bound to PKA, the regulatory subunit relieves its inhibition on the catalytic subunit, which is responsible for phosphorylating protein targets, effectively deactivating or activating them in response to an extracellular signal. This tightly controlled mechanism is vital to the function of the cell. Any dysregulation in the response of PKA has major detrimental effects that are common in cancer [2], and endocrine disorders like Carney Complex [3].

There are several isoforms of the Regulatory subunit, denoted by Types I and II and subtypes α and β . In the RI α isoform, a positively charged residue in PKA, Arginine (R), in position 333 of the regulatory subunit interacts directly with the first molecule of cAMP sensed by PKA. Because of its flexibility, the tetrameric for PKA RI α structure has proven difficult to elucidate [4], but scientists figured out that with a mutation of Arginine 333, the dynamics of PKA are stabilized enough for structural biochemists to be able to

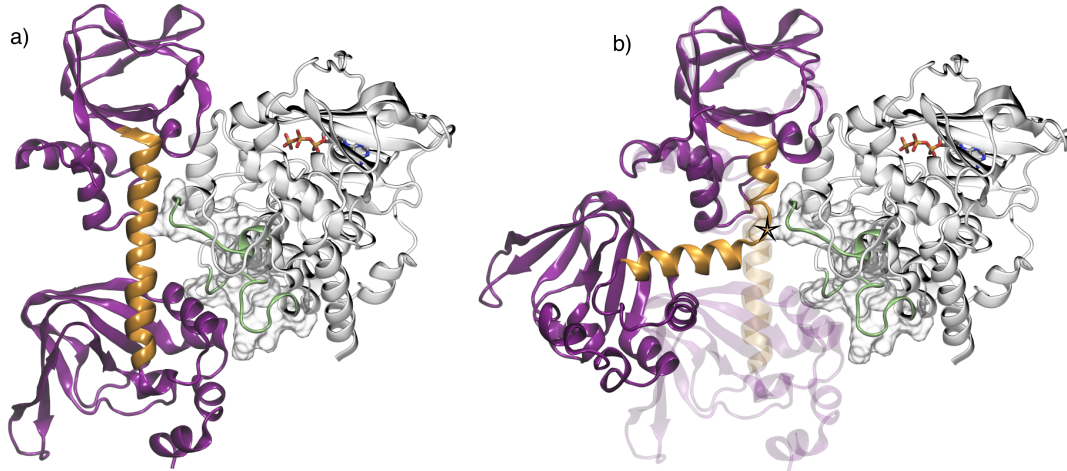


Figure 0.3. The interface of the Regulatory and Catalytic subunit of Protein Kinase A (PKA), comparing the mutant to the wild-type

(a) The mutant Regulatory subunit (R) R333K (purple) in complex with the Catalytic subunit (C). The B/C helix (gold) connecting the two cyclic nucleotide binding domains (CBDs) is extended in the heterodimeric crystal structure; (b) The wild-type R subunit adopts a stable conformation with a break in the B/C helix at Glycine 235 (black star). The structure is juxtaposed with the R333K mutant (transparent purple) to show the consistency of the WT conformation with the solvent-exposed regions in the C subunit (green) determined by solution experiments

visualize it using crystallography [5].

When Arginine (R) 333 is mutated to a Lysine (K), which is *also* positively charged, the shape of the PKA molecule changes significantly (see Figure 0.3). This mutation also affects PKA function, altering the way the protein responds to the cAMP signal. In the wild-type (WT) R333 case, the protein is dynamic, with a swinging domain that creates a “butterfly” motion in the protein. It responds to cAMP by dissociating the regulatory subunit from the catalytic subunit, relieving inhibition and activating PKA. In contrast, the K333 mutation makes the protein more globular, neutralizing the motion of swinging domain [6] and changes the interaction between cAMP and the regulatory subunit, completely altering the way that PKA is activated. This and other mutations are known to be the root cause of diseases like Carney complex [7] and Adrenocortical Cushing’s adenoma [8]. In this way, a single residue mutation effects the dynamics of a

protein, and the way that the protein interacts with its environment.

0.4 Structural methodologies for visualization of biology

As the specialization of this degree is centered on Multiscale Biology, the following chapters will discuss experimental techniques that utilize information gathered from different structure elucidation methods with range in spatial scales from small molecule (\AA), to proteins (nm), to membranous structures on the sub-cellular level (μm). The following subsections review the pertinent methodologies for biological structure determination and alludes to the ways these methods facilitated scientific investigations in the coming chapters.

0.4.1 X-ray Crystallography

Crystal x-ray diffraction also known as X-ray Crystallography is popular and widely used method for determining the atomic positions of small molecules and proteins. First developed in 1912 by Max von Laue and later further developed by William Henry Bragg and Sir William Lawrence Bragg [9], crystallography is based on the principal of X-ray diffraction by the atoms in an ordered crystalline substance.

As the name indicates, the subject of interest is suspended in a crystallographic form and a beam of X-rays at a particular energy, marked by its wavelength, λ , is shot at the sample. Using the angle, θ of the diffracted X-ray beams and a mathematical equation known as the Bragg equation:

$$n\lambda = 2d\sin\theta \quad (0.1)$$

Where n is an integer multiple of the wavelength, λ , d is the distance between the planes of the crystals, and θ is the scattering angle, one can understand the structure of the biological specimen. From the interference and diffraction pattern produced by the X-ray

beam after hitting the crystal lattice structure, one can back-calculate the atomic positions of a molecule.

X-ray crystal structures yield Å-level resolution of structure; a real asset for computational chemists and biophysicists who use the atomic positions as inputs for their models. Section 0.5 of the introduction will elaborate more on how atomistic positions are used to understand dynamics and functions of small molecules and proteins.

It is important to note that an X-ray structure is but a snapshot of a molecule in a single state of its ensemble of conformational states. Although crystal structures are unparalleled in their accuracy of single structure, by their nature, they do not give information about molecular flexibility. Furthermore, the resolution of atomic positions falls off as molecules increase in size. Therefore, crystallography is limited in the size of the molecules that can be investigated with sufficient accuracy. Fortunately, X-ray structures can be complimented with other structural determination and computational methodologies, which will be discussed in the coming sections.

Chapter 2 combines X-ray Crystallography with molecular dynamics (discussed in section 0.5.1) in an investigative study of common structural motifs used by an infectious organism. Chapter 3 demonstrates the ways that molecular dynamics can expand on structural information not offered by crystallographic snapshots of proteins.

0.4.2 Hydrogen/Deuterium Exchange Mass Spectrometry

A remedy for the limited structural states that can be understood with X-ray crystallography is offered by Hydrogen/Deuterium Exchange Mass Spectrometry(H/DxMS). Since the discovery of the “heavy” hydrogen isotope, deuterium in 1932 by Harold Urey et al. [10], deuterium has been used in a myriad of experimental techniques. Early in the 21st century, the application of the deuterium to the study of proteins appeared [11].

As indicated by the name, in H/DxMS, hydrogen and its isotope, deuterium exchange (H/Dx) for one another. This happens specifically with labile hydrogen atoms

on the backbone nitrogen of the amino acid, which exchange nearly instantaneously. In a solution of D₂O (the deuterium version of H₂O), exchange is rapid if backbone nitrogens are surface-exposed or located in an unstructured region of the protein, and slower if the nitrogen is buried in the core of a protein. Exchange can happen at a rate on the order of milliseconds to seconds if amide nitrogens are not hydrogen bonded, but can take minutes to days if hydrogen bonds are stable. Timescales of exchange translate to the “degree of protection” of the amide, and indicate the degree of flexibility of the structural regions that comprise the protein. The exchange reaction of hydrogen for deuterium is then quenched by acidification of the protein (pH 2.5), which “freezes” the deuteration of the protein. Proteolysis then follows—a method that degrades the protein into smaller fragments to be analyzed by Mass Spectrometry (MS). Mass Spectrometry determines the atomic weight of the fragments, yielding the information of which nitrogens have exchanged their hydrogen for deuterium [12].

This technique is used to understand the comparative changes in dynamics induced by protein-protein interactions, ligand binding, as well as signaling modes of a protein [13]. H/DxMS analysis gives a *dynamic* picture of the protein, allowing biophysicists to understand the flexibility of a molecule. Moreover, the technique is less limited in the size of molecules that can be investigated.

This method complements computational methods like Molecular Dynamics which examine the flexibility of proteins. Chapter 3 details how H/DxMS in combination with Molecular Dynamics have brought us closer to a more holistic understanding of the structure of the RI α Protein Kinase A complex, as was briefly discussed earlier.

While X-ray crystallography and H/DxMS give information about protein structure on the angstrom level, a multitude of methods exist that reveal insights about global structural features of proteins on the nanometer scale.

0.4.3 Small-Angle X-ray Scattering

As discussed earlier, X-ray crystallography utilizes the diffraction, or “bending”, of X-rays to resolve the atomic positions comprising molecular structure. But X-rays can also interact with a sample, altering the momentum of the beam in a phenomenon known as *scattering*. In the phenomenon of scattering, the photons in an X-ray beam interact with electrons in a specimen. The scattering pattern indicates the fluctuation of the electron density of the matter being investigated. Put more simply, the scattering profile gives information about the overall shape of a molecule. This is accomplished by measuring the scattering profile in the following way: a photon of wavelength λ scatters off the molecular sample at an angle, 2θ , is related to the scattering vector, q through equation 0.2. The intensity of the scattering vector, $I(q)$ is a function of a multitude of factors, including molecular volume, size, electron density.

$$q = \frac{4\pi \sin(\theta)}{\lambda} \quad (0.2)$$

The elegant mathematical theories used to obtain information SAXS experiments are explained best by experts in the literature [14], but to summarize: experimental scattering profile, information such as the molecular weight, molecular volume, and radius of gyration can be determined. What is important to note is that SAXS provides information about an ensemble of structural states, in stark contrast to X-ray Crystallography, which resolve explicit conformations of a molecule. This means that multiple structural states can be “lumped” together in the same SAXS profile if a molecule is flexible.

Molecular structure is, by nature, affected by its environment. SAXS has been used to understand molecular structure under particular conditions such as temperature, pH [15], and in complex with other molecules and proteins [6]. Thus, SAXS profile of a protein in its apo or “unbound” state will undoubtedly differ from its profile in a “bound” complex with a small molecule, another protein or even another copy of the same protein

(homodimers, homotrimers, etc.)

In the same way, mutations can affect molecular structure resulting in different SAXS profiles when compared to wild-type (WT). Such is the case of a system examined in chapter 3 of this thesis. Using SAXS as a guide, we contextualized the crystallographic structure of PKA combined with the H/DxMS in order to validate the existence of an unresolved conformation of the regulatory subunit obtained by MD, the Flipback conformation.

As is the true with most technologies, crystallography and solution-structure methodologies like H/DxMS and SAXS crystallography have their limitations. That limitation being, molecular size. As the molecular size increases, so often does the difficulty in resolving fine details from the system. As we transcend spatiotemporal scales, we can utilize methods that take advantage of different technologies to gain insight into larger systems.

0.4.4 Cryo-electron Microscopy

Just two years ago, the 2017 Nobel Prize in chemistry was awarded to Jacques Dubochet, Joachim Frank and Richard Henderson for the development of Cryo-electron microscopy (Cryo-EM) in resolving structures of biomolecules in solution. Cryo-EM utilizes an electron beam as an illuminating source to visualize proteins and large molecular species that are flash-frozen in their native structural states.

Cryo-EM remedies the “single crystalline conformation” limitations of X-ray crystallography and combines the strength of solution structural methods that describe ensembles of structural states, as discussed in the earlier introductory sections. The ingenious method behind Cryo-EM completely avoids the cumbersome crystallization of biomolecules by flash-freezing the subject of study in an aqueous solution. This method, also known as vitrification was developed by James Dubochet along with Jean Lepault nearly 40 years ago [16–18]. The original spray-freezing method evolved further with the use of

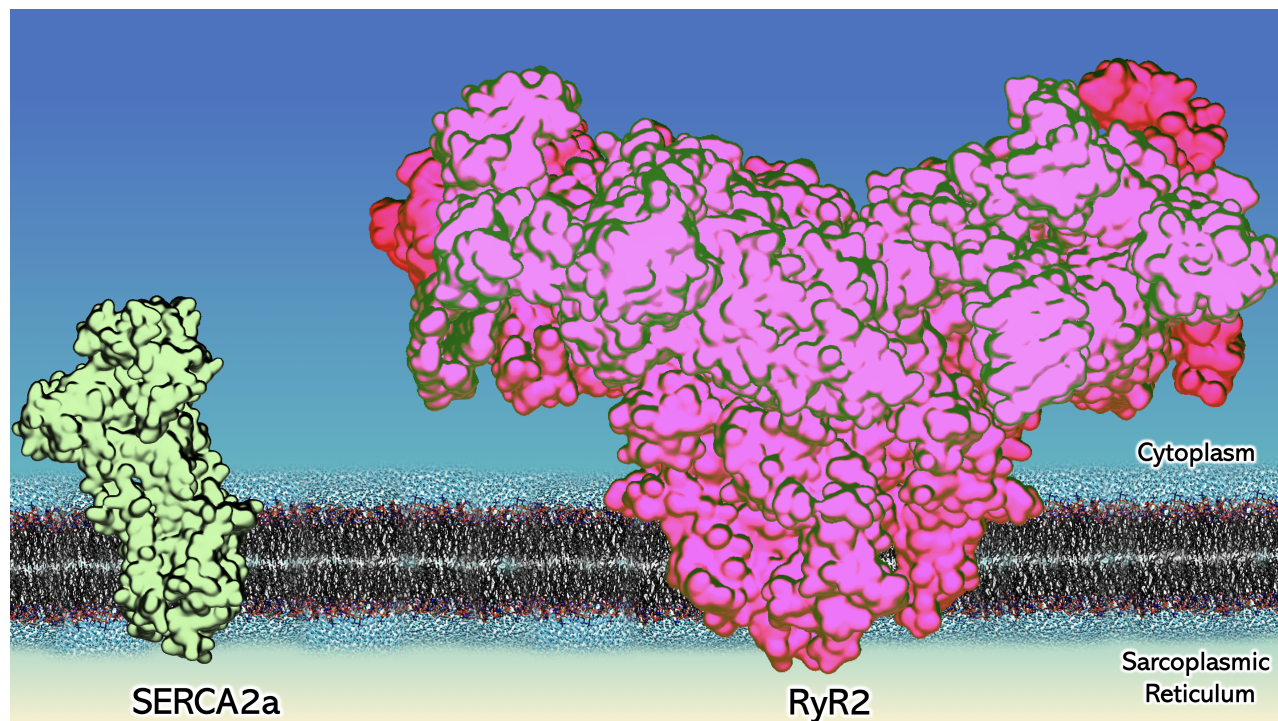


Figure 0.4. Angstrom-level representations of biomolecules
 The Sarco/Endoplasmic Reticulum Calcium-ATPase, SERCA2a, (PDBID:5MPM) crystal-
 lized at 3.3Å resolution (green, left) and the Ryanodine Receptor, RyR2, (PDBID:5GOA)
 resolved at 4.2Å with Cryo-EM.

high-pressure which increases the depth of vitrification [19].

Vitrification involves cooling the sample to very low, cryogenic temperatures thus, embedding it a glass-like, irregularly frozen vitreous state upon treatment with liquid ethane. This is a significant advantage over X-ray crystallography, as many proteins will not crystallize easily due to the inherent dynamic nature of proteins in general. It is also especially advantageous for large membrane proteins which are notoriously hard to crystallize.

Cryo-EM is a popular method for understanding the structure large proteins which can be too “noisy” for solution experiments like H/DxMS and Nuclear Magnetic Resonance [20]. In fact, the larger the system, the more intriguing the results of Cryo-EM can be [21–23]. Figure 0.4 compares the size of SERCA2a, a protein resolved using X-ray crystallography at 3.3Å (PDBID:5MPM, unpublished) and RyR2, a large (>2MDa) protein

at 4.2Å resolved with Cryo-EM [24]. The larger the system, the more possible projections of the molecule that can exist. This is because molecules are not oriented when they are captured in vitreous water. Thus, part of the “magic” of Cryo-EM stems from its use in resolving information from randomly oriented molecules.

Our ability to visualize dynamic, large, and complex Cryo-EM structures is a direct result of Joachim Frank’s work in analyzing and processing the images derived from Cryo-EM [25]. The algorithm, known as cross-correlation, averages ensembles of structural states to resolve what is known as a “single-particle reconstruction” of a biological specimen [26]. Using several thousand 2D images of the same molecule in random orientations, the algorithm generates 3D reconstructions from 2D projections [27]. Cryo-EM gives molecular detail on the order of several angstrom. The size-limitation problem still stands however: the larger the protein, the lower the atomic resolution will be. But the popularity of this method has and will undoubtedly continue to drive the development of both Cryo-EM and the algorithms used to resolve molecular structures. One can only hope that the resolution will improve, yielding atomistic detail. In the context of this thesis, Cryo-EM structures gave us insights into the subcellular context of large structures like the Ryanodine Receptor.

0.4.5 Electron Tomography

As we scale up from atoms to molecules in living systems, the next “step” is towards the subcellular level. This is realm within which we can see the direct effect of protein function. The name, “subcellular,” alludes to the fact that this scale is a part of a cell. At this level, organelles– boundaries that distinguish specialized regions of the cell–become the subject of focus. To see within the cell, we can use a special technique used to determine the 3D structure of a cell known as electron tomography.

As its name indicates, electron tomography uses an electron beam as its source of a penetrating wave of energy. Tomography is an imaging technique that “sections” or penetrates the object being visualized. In the case of electron tomography, a biological

specimen undergoes a tilt-series, where the sample is exposed to the beam at different angles resulting in different projections of the specimen [28,29]. The specimen itself is fixed with chemicals that “lock-in” the physical structure and stained with heavy metals that increase the contrast of lipids and proteins. Gold particles are used as fiduciary markers to align projections. Upon alignment, the boundaries of important structures are traced and used for reconstruction on the sub-volumes [30,31].

Three-dimensional reconstructions of subcellular volumes are used in this thesis to understand the dynamics of calcium signaling in cardiac tissue [32] as first pioneered by Hake et al [33]. In the fourth chapter these realistic geometries are used as a framework within which we reconstitute the biochemical reactions that underlie calcium signaling. Additional details will follow in the final section of the introduction and in chapter 4 of the dissertation.

0.5 Biophysical structure modeling methodologies

Studying the dynamics of biological systems is no simple feat. Proteins are incredibly small and their relevant motions are fast; oftentimes too fast to study with available technology. Fortunately, a number of computational tools that simulate molecular motion exist. This section contains a brief introduction to two such computational tools used to model the dynamics of proteins, Molecular Dynamics and Brownian Dynamics.

0.5.1 Molecular Dynamics

Since the late 70’s, computers have been used in the field of biochemistry, providing visual insights into dynamics of proteins [34,35]. Since the first simulations of bovine pancreatic trypsin inhibitor, the field of biomolecular simulation has expanded both in the systems that have been investigated, along with the algorithms used to compute the forces that govern molecular motion.

Molecular Dynamics (MD) is a computational technique used to compute the

spatial and temporal atomic motions according to Newton's equations of motion, where F is force, m is mass, and a is acceleration (see equation 0.3). The gradient of potential energy, U in space is described by equation 0.4 where \vec{R} are the coordinates of all of the atoms in the structure. These atomic positions oftentimes are often determined by X-ray crystal structures as discussed in earlier sections.

$$F = ma \quad (0.3)$$

$$F(\vec{R}) = -\nabla U(\vec{R}) \quad (0.4)$$

For a large system, the analytic solution of this function is impossible to solve, but a numerical solution is made possible using a collection of equations that describe the forces that influence atomic motions. Time is discretized, or subdivided into small steps. In MD, simulated time is on the order of 1 to 2 femtoseconds (fs), 1×10^{-15} s. After every time step, we solve for the potential energy function by adding the forces acting on every atom in the system. The forces can be split into bonding and non-bonding forces as they are in equation 5, below.

$$\begin{aligned}
 U(\vec{R}) = & \underbrace{\sum_{\text{bonds}} k_i^{\text{bond}} (r_i - r_0)^2 + \sum_{\text{angles}} k_i^{\text{angle}} (\theta_i - \theta_0)^2}_{U_{\text{bonding}}} \\
 & + \underbrace{\sum_{\text{dihedrals}} k_i^{\text{dihe}} [1 + \cos(n_i \phi_i + \delta_i)] + \sum_{\text{improper}} k_i^{\text{impr}} (\varphi - \varphi_0)^2}_{U_{\text{bonding}}} \\
 & + \underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij} \left[\left(\frac{A_{ij}}{r_{ij}} \right)^{12} - \left(\frac{B_{ij}}{r_{ij}} \right)^6 \right] + \sum_i \sum_{j \neq i} \left(\frac{q_i q_j}{\epsilon r_{ij}} \right)}_{U_{\text{non-bonding}}}
 \end{aligned} \quad (0.5)$$

The bonding potential term is defined as the oscillation of the equilibrium bond length. The angular potential is defined as the oscillation of three atoms forming an angle. The dihedral potential is a torsional relation of four atoms oscillating about a central

bonding axis. The improper dihedral potential is defined as the torsional relation of three atoms connected to a single common atom. All bonded forces are described by corresponding spring constants (k_i) with radial distances (r), angular (θ), proper dihedral angle (ϕ), and improper dihedral angle (φ) terms. These terms have an initial value and value after every time step (0 and i , respectively).

The van der Waals potential, which is synonymous with the total intermolecular forces, is expressed as a “6-12” Lennard Jones potential, or the difference between experimentally determined values A_{ij} and B_{ij} over the radial distances, r_{ij} multiplied by the Lennard Jones well depth, ϵ_i . Finally, the electrostatic potential, the effective attraction and repulsion of charged or partially charged atoms, is proportional to the partial atomic charges of the atoms ($q_i q_j$), and inversely proportional to the interatomic distances (R_{ij}).

The summation of the bonding and non-bonding forces over iterative timesteps yields a spatial displacement of each atom in the system. Collectively, it results in the motion of the atoms in the system, or *intramolecular motion*. Iterative solutions of these equations through time result in a trajectory of atomic motion. This is, effectively, a change in structure as a function of time, where the protein moves and responds to its environment. The amount of linear time that can be simulated depends both on the size of the protein (in number of atoms) and amount of computational power available. The larger the system, the more computational power required and the longer it will take to compute the dynamics. Today, it is routine to simulate nanosecond-microseconds of time with MD, though there are methods that can access millisecond timescales and beyond [36, 37].

The energetic terms associated with each atom in a system are collectively called a molecular mechanics “force field.” Depending on what the user is modeling— lipids, carbohydrates, nucleotides, proteins, metals—one can find force fields specialized in the accurate prediction of the dynamics of the chosen system. Some examples of force fields are AMBER [38], CHARMM [39], GROMACS [40], and OPLS [41]. The forcefields are used to design the system in terms of energetics, and the potential energy function is calculated with

a molecular mechanics engine. The engines have been developed with their own force fields such as the GROMOS MD engine [42] that is used with the GROMACS force field. Some popular molecular mechanics engines include AMBER [43], NAMD [44], GROMOS [42], and Desmond [45]. Many force fields and engines can be mixed-and-matched, leaving the decision up to the user on what combination best suits their study.

The protein being investigated using MD can represent the solvent in one of two ways: explicitly or as a continuum. The explicit solvent description models every atom in the solvent including ions and water; sometimes even dummy atoms that are meant to model the behavior of the electrons in the molecules. In the continuum approach, the solvent is modeled inexplicitly as an electric field. The latter approach yields increased computational power as water calculations take up significant amounts of time and computational resources [46]. Nevertheless, water interactions with the system are oftentimes important, so there is an information trade-off for inexplicit water modeling. Additionally, there are a wide variety of explicit water models that can be used to simulate a chosen system [47].

MD has a long, successful track record of uncovering the hidden features of protein motions [48]. In these cases, that which is not revealed by crystallography or other structure determination techniques can be further understood with the use of molecular dynamics. In this thesis, MD will be applied to the intramolecular investigations of several systems. The first chapter highlights ways that molecular dynamics simulations can yield an understanding of the kinetic transitions as well as provide new structural states to be interrogated by intermolecular investigation methods. Molecular dynamics is then used to understand the effects of mutations in a protein-protein system in chapter 2. Finally, in chapter 3, MD is used to elucidate a conformation of PKA that has been elusive to crystallographers.

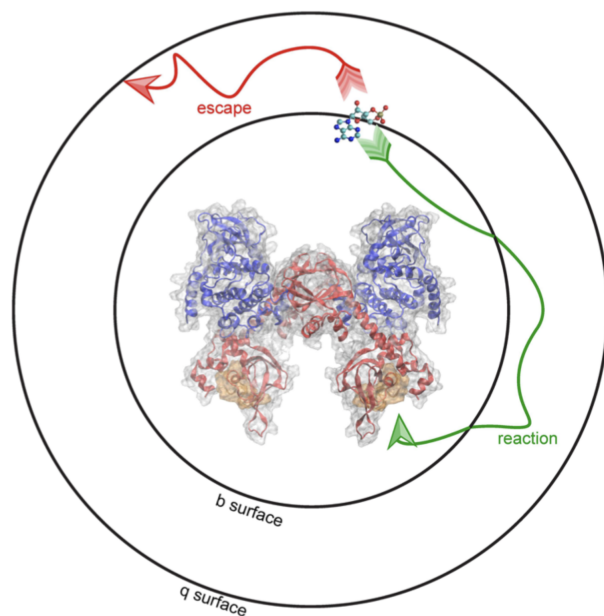


Figure 0.4. Brownian Dynamics simulation with BrownDye
 A molecule begins on the centrosymmetric b-surface and can either escape (red arrow) beyond the q-surface or associate (green arrow) to the specified binding region (gold)

0.5.2 Brownian Dynamics

As we scale upwards through biomolecular space and time, the next level we reach centers on the *intermolecular* interactions *between* molecules. Brownian Dynamics (BD) is a method developed to model bimolecular diffusion that estimates kinetic association events. This is especially useful for the step-wise understanding of complex biochemical mechanisms.

BD is specifically focused on the determination of second-order association kinetics between proteins and their binding partners, be they proteins, small molecules, or ions. BD uses several main assumptions in order to accomplish this. Firstly the solvent is modeled using a dielectric field and ionic continuum. In stark contrast to MD, most often, BD models the atoms in a molecule as rigid bodies. Furthermore, each atom is modeled as a hard-sphere and the atomic forces include mostly electrostatic forces.

The equation of molecular motion (equation 0.6) in a BD is a function of the

position (x), rotation (ϕ), attractive and repulsive forces (F), and the torque (T) of a particle index (i). A random vector (w) simulates the random “kicks” of water on the diffusion of the particle, an effective “Brownian motion” factor. D is the diffusion tensor and s the matrix square root of the diffusion tensor [49].

$$d \begin{pmatrix} x_i \\ \varphi_i \end{pmatrix} = \frac{dt}{k_B T} \mathbf{D} \cdot (\mathbf{T}_i) + \sqrt{2dt} s \cdot w + \nabla \cdot \mathbf{D} dt \quad (0.6)$$

The algorithm used to compute rate constants from the Brownian trajectories was developed by Northrup, Allison and McCammon [50] and is known in the field as the NAM algorithm. Using the NAM algorithm, one can determine the spatial trajectory of diffusing molecules in 3D space. A number of BD simulation programs exist, including SDA [51, 52], ReaDDy [53], Brownmove [54], BD_BOX [55], and BrownDye [49]. In order to set up a simulation, one must parameterize system by assigning the appropriate parameters to each atom. The relevant terms in the case of BD are van der Waals radial (r) and charge (q) values that are assigned using MD force fields like the ones discussed in section 0.5.1, earlier. Using the charge and radial parameters, an electrostatic description of the molecule in the form of a charge field can be calculated using programs like DelPhi [56] and APBS [57]. In this thesis, the program BrownDye is used to perform BD simulations and APBS is used for the electrostatic description of the biomolecules.

In BD simulations with BrownDye, diffusing molecules are placed at a radial distance b from one another. The simulation begins by placing the diffusing molecules at different distances of b distributed over a spherical surface, called the “ b -surface” (see figure 0.4). Molecular diffusion begins at the start of the simulation and ends in one of two ways: 1) diffusion past a certain distance “ q ” or the “ q -surface”; or 2) with association of the two molecules after satisfying the contacts set by the user in the encounter complex description [49]. The timescale of integration is typically on the order of picoseconds ($1 \times 10^{-12}s$). With the speed of BrownDye, millions of trials can be run on a desktop

computer to estimate β , or the probability that a binding partner located on the b-surface will satisfy the encounter complex and “react” rather than escape beyond the q-surface. Figure 0.5 demonstrates the basic principles of the BrownDye simulation package used to determine the rate constant of bimolecular association, k_{on} (see equation 0.7, below).

$$\mathbf{k}_{on} = \mathbf{k}_b \beta \tag{0.7}$$

The rate of diffusion of the ligand to the b-surface, k_b can be calculated using equation 0.8, below. k_b is a function of $U(r)$, the effective potential energy at a distance r from the center of the sphere, the temperature, T , Boltzmann’s constant, k_B , and the diffusion coefficient, D , which varies with radial distance, (r) .

$$\mathbf{k}_b = \frac{4\pi}{\left[\int_{\mathbf{b}}^{\infty} \frac{\exp(-U(r)/k_B T)}{r^2 D(r)} \mathbf{d}r \right]} \tag{0.8}$$

Models of the structural and kinetic properties of biomolecules are useful not only to those experimentalists seeking to further understand and visualize the behavior of biochemical systems, but also to computational biologists seeking to model biochemical phenomena on the cellular level. Using BD simulations, we can tease out the association rates between to molecules which is often difficult to measure with experiments. Chapter 1 will demonstrate the ways in which this kinetic information can be used in higher-order cellular models. In chapter 3, Brownian Dynamics will be applied to the understanding of cAMP association to PKA.

0.6 Subcellular biochemical modeling methodologies

As we continue transcending through the scales of space and time, having already explored intramolecular and intermolecular biophysical methods, we arrive at the scales within which these biochemical properties exert their effects: the subcellular level. Even

the slightest alterations in chemical structure and function can have dramatic effect on the dynamics of cellular systems; what I like to call “a biophysical butterfly effect.” Biological modeling at this level allows for the holistic understanding of the biochemistry within the subcellular space. This is especially useful for complex systems where multiple competing reactions occur. More specifically, it is useful for the comparison and contextualization of reactions in biological systems. A number of methods that model diffusion and reactions on the molecular level have been developed and used extensively to explain and predict biological observables.

0.6.1 Modeling biology with differential equations

Biochemical descriptions of biological systems are almost always complex. In a given cell, anywhere from hundreds to thousands of different molecular species exist. The reactions between these species only intensifies the complexity when you consider that these molecules react with many different partners to create more molecules. Yet again, mathematics comes to the rescue, giving biologists, chemists, and physicists a language that can be used to model biological systems. One can describe the dynamical system in terms of the change in molecular concentration and local density with respect to time.

Modeling the change of a single feature of a molecular system, say a given concentration, with respect to time, is the essence of an Ordinary Differential Equation (ODE). Reactions, too, can be described in this way, where the concentrations of molecular species change as a function of the forward and reverse rate of consumption. Molecular diffusion in single and multiple can also be modeled with ODEs making this a powerful method for understanding the temporal evolution of gradients in cellular systems [58].

One can also model the change in a system as a function of several simultaneously changing parameters. Consider for example, modeling the rate of change as a function of time through a particular three-dimensional region, like an organelle. This is arguably a more complex problem As it involves a *spatial description* of the system. To model systems

like this, one could use spatial ODE solvers [13] which use specialized algorithms to model diffusion through space. Partial differential Equations (PDEs) can also be used for this purpose but are comparatively more computationally intensive to solve as space has to be subdivided into elements within which diffusion is computed. What is common to modeling with differential equations is a very important factor: the continuum assumption.

The fundamental component of all differential equations in biological simulation methods is the use of non-discrete representations of molecular species, or a continuum. That is, all the elements of a species are non-distinguishable from one-another. In ODE and PDE biological simulations, molecular concentrations are modeled as a “field” that evolves in its intensity with the passage of time according to the equations that describe its temporal evolution. This approximation works well for large molecular quantities that are well-mixed. It falls short, however, when we consider situations in which chemical concentrations are extremely low.

0.6.2 Discrete subcellular spatial modeling

As insinuated in the earlier section, there are delicate systems in biology where molecular concentrations are extremely low. For example, a region in the cardiac muscle cell that sits between the plasma membrane and an organelle named the sarcoplasmic reticulum (SR), the dyadic cleft, has a range of zero to three molecules in the subspace at equilibrium [59]. In such cases, it is inappropriate to treat molecules as a continuum. Instead, we can model the molecules discretely using algorithms that track the spatial location of these ions. The more we seek to model realistic phenomena in biology, the more we need to consider the importance of stochasticity and explicit spatial encounters in biological systems. Using the program MCell [60,61], we can accomplish exactly this.

MCell is a Monte-Carlo simulation method that stochastically and explicitly models the 3D diffusion and reaction of molecular species by treating them as particles. The Monte-Carlo technique uses random number generator that “chooses” from an ensemble

of values which correspond to the spatial displacement of molecules in the system. In the context of subcellular modeling, Monte Carlo algorithms are aimed at representing the stochastic nature of a biochemical system. The algorithm describing motion in MCell is a simplified, unbiased random walk. Molecules are modeled as volume-less points that diffuse according to assigned diffusion parameters.

One of MCell’s most impressive attribute is that it is a *master counter*. At every integration timestep, often on the scale of microseconds ($1 \times 10^{-6}s$) molecules have the potential to either react, diffuse, or remain unchanged. MCell explicitly counts not only the number of particles but tracks their spatial location in subcellular geometries. Molecular reactions can be unimolecular or, upon spatial encounter of two species, bimolecular. Reactions can happen with two freely diffusing “volume” molecules, or between membrane-bound “surface” molecules and volume molecules. Surface molecules are also capable of 3D diffusion and changing their orientation in a membrane.

Using mesh generation tools like GaMer [30,31], cellular reconstructions obtained from electron microscopy can be reconstructed into usable meshes for biological simulation. These complex subcellular geometries can be imported into CellBlender [62], the graphical user interface that is used to design the MCell simulation. Kinetic descriptions of the subcellular system, taken from experiments or from simulation techniques like MD and BD can be used to model the reactions between species in the system. The focus of the final chapter is the use of MCell to understand cardiac function and dysfunction on the subcellular level.

0.7 Conclusion of the Introduction

The following body of work examines cardiac function and dysfunction from the perspective of the proteins that comprise the respective systems. This is no trivial feat, as cardiac dysfunction can occur at multiple scales of space and time. How are these

scales examined? How do changes at one spatiotemporal scale translate, resulting in heart disease? How are mathematics and physics applied understand these invisible biological phenomena?

Although the language of mathematics is common, the ability to learn mathematics is often a matter of privilege. The same is true of any systematic study. To devote time to the study of a discipline is likely to come with ones' basic needs being met, though examples to the contrary likely exist. Even more so, exposure to literature and teachers that can aide in the understanding of the concepts underlying higher order mathematics grants even more opportunity to expand the boundaries of mathematical study.

For those of us that have been so privileged to study a discipline systematically, it remains our duty to communicate its power. The goal of this work is to show just how stunningly visual biochemistry in the language of mathematics can be. It aims to highlight the fruit yielded by the collaborative and independent labor of a world full of talented mathematicians, physicists, and chemists alike. This thesis is an ode to the scientists that came before me. It is but an example of the application of mathematical theory and biophysical methods to the chemical understanding of molecular function through the lens of the computational microscope.

Chapter 1

Multiscale Modeling of protein systems

In the following manuscript, we demonstrate ways in which multiscale computational methods can integrate structural and chemical information to better understand how, for example, changes in protein sequence such as point mutations can alter organ-level phenotypes like cardiac contraction. Our case study is focused on Protein Kinase A (PKA) and the manuscript provides examples of multiscale methods that compliment each other, showing how structural and kinetic information can be combined into an integrated understanding of the protein system. Included in the manuscripts are examples of how multiple short molecular dynamics (MD) simulations can be combined into an atomistic Markov State Model (MSM) to provide mechanistic insights and kinetic information about important structural transitions in the cyclic nucleotide binding domain (CBD) of PKA. Moreover, classical and accelerated MD simulations can provide new structures to understand the association kinetics of the second messenger, cAMP using Brownian dynamics (BD) simulations. BD simulations yield kinetic rates of bimolecular association reactions that can be used in protein-scale MSMs. Semi-analogous to MSMs are milestoning methods that have been developed and optimized in the Amaro group at the University of California San Diego that directly integrate MD and BD to understand on and off-rates of ligands to their protein systems. Finally, we discuss how the information provided by the aforementioned methods can be used in higher-order subcellular and whole cell models.



Bridging scales through multiscale modeling: a case study on protein kinase A

Britton W. Boras^{1†}, **Sophia P. Hirakis**^{2†}, **Lane W. Votapka**², **Robert D. Malmstrom**³, **Rommie E. Amaro**^{2,3*} and **Andrew D. McCulloch**^{1,3,4*}

¹ Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA, ² Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA, USA, ³ National Biomedical Computation Resource, University of California, San Diego, La Jolla, CA, USA, ⁴ Department of Medicine, University of California, San Diego, La Jolla, CA, USA

OPEN ACCESS

Edited by:

Mario Nicodemi,
Universita' di "Napoli Federico II," Italy

Reviewed by:

Supriyo Bhattacharya,
City of Hope Medical Center, USA
Andrzej Stasiak,
University of Lausanne, Switzerland

*Correspondence:

Rommie E. Amaro,
Department of Chemistry and
Biochemistry, University of California,
San Diego, 9500 Gilman Drive,
La Jolla, CA 92093-0340, USA
ramaro@ucsd.edu;
Andrew D. McCulloch,
Departments of Bioengineering and
Medicine, University of California,
San Diego, 9500 Gilman Drive,
La Jolla, CA 92093-0412, USA
amcculloch@ucsd.edu

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Biophysics,
a section of the journal
Frontiers in Physiology

Received: 29 June 2015

Accepted: 24 August 2015

Published: 09 September 2015

Citation:

Boras BW, Hirakis SP, Votapka LW,
Malmstrom RD, Amaro RE and
McCulloch AD (2015) Bridging scales
through multiscale modeling: a case
study on protein kinase A.
Front. Physiol. 6:250.
doi: 10.3389/fphys.2015.00250

The goal of multiscale modeling in biology is to use structurally based physico-chemical models to integrate across temporal and spatial scales of biology and thereby improve mechanistic understanding of, for example, how a single mutation can alter organism-scale phenotypes. This approach may also inform therapeutic strategies or identify candidate drug targets that might otherwise have been overlooked. However, in many cases, it remains unclear how best to synthesize information obtained from various scales and analysis approaches, such as atomistic molecular models, Markov state models (MSM), subcellular network models, and whole cell models. In this paper, we use protein kinase A (PKA) activation as a case study to explore how computational methods that model different physical scales can complement each other and integrate into an improved multiscale representation of the biological mechanisms. Using measured crystal structures, we show how molecular dynamics (MD) simulations coupled with atomic-scale MSMs can provide conformations for Brownian dynamics (BD) simulations to feed transitional states and kinetic parameters into protein-scale MSMs. We discuss how milestoning can give reaction probabilities and forward-rate constants of cAMP association events by seamlessly integrating MD and BD simulation scales. These rate constants coupled with MSMs provide a robust representation of the free energy landscape, enabling access to kinetic, and thermodynamic parameters unavailable from current experimental data. These approaches have helped to illuminate the cooperative nature of PKA activation in response to distinct cAMP binding events. Collectively, this approach exemplifies a general strategy for multiscale model development that is applicable to a wide range of biological problems.

Keywords: protein kinase A, multiscale model, molecular dynamics, Brownian dynamics, Markov state model

Introduction

The goal of multiscale modeling is to understand how the hierarchy of biological structures integrates to produce biochemical, cellular and physiological functions. At the single cell scale, signaling networks are analyzed using system analysis methods to provide mechanistic they are

Abbreviations: PKA, Protein Kinase A; MD, Molecular Dynamics; MSM, Markov state model; BD, Brownian Dynamics.

insights into the functional interactions between proteins and second messengers. Network models of cell signaling have recently been developed for neurons (Cowan et al., 2012), myocytes (Bondarenko, 2014), and pancreatic beta cells (Wang et al., 2012), to name a few. These cell-scale network models are helpful to understanding normal cell physiology, pathobiology, and therapeutic mechanisms. Interest in the phenomenological effects of protein mutations (Kirchner et al., 2012; Cong et al., 2013) are driving the development of new methods to incorporate atomic and molecular-scale models and data into whole cell simulations. To this end, advances in atomic-scale modeling, particularly molecular dynamics (MD) and Brownian dynamics (BD) simulations, have provided insights into the effects of mutations on protein folding and protein-protein interactions (Kozack and Subramaniam, 1993; Cregut and Serrano, 1999; De Rienzo et al., 2001; Koukos and Glykos, 2014). However, bridging these scales and disciplines to create models that can predict the effect of a point mutation or post-translational modification on cellular phenotypes remains a daunting task. Frequently, even nomenclature does not readily transcend disciplines, making interdisciplinary collaborations across scales more difficult. Furthermore, understanding the limitations of models and methods at each scale to avoid error propagation is essential to obtaining physiologically meaningful solutions. In this article, we describe atomic and protein-scale Markov state modeling (MSM), as well as milestoning, which allow us to bridge atomic-scale molecular models to cell-scale signaling networks (Figure 1).

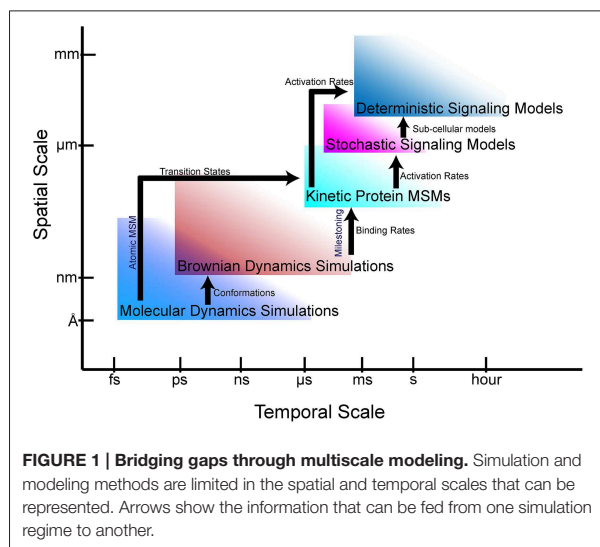
Over the past decade, the availability of high-resolution protein structures and the capabilities of atomistic molecular modeling techniques has improved dramatically. MD and atomic-scale MSMs use atomic-resolution structural data to model the position of atoms in a protein and calculate the forces between them. This is helpful in predicting functional states and rates of conformational change. However, these methods cannot

easily calculate the rates of interactions between molecules, which are needed for higher scale reaction network models.

Advances in BD simulations and milestoning have provided tools that are specialized in calculating diffusion-limited association rate constants. Previously, the data used for parameterization of the transitions in protein-scale MSM came almost exclusively from *in vitro* experiments where conditions are controlled to limit the number of potential states. These data included phosphorylation rates, k_{on}/k_{off} of binding events, and ion channel transitions (Clancy and Rudy, 1999; Campbell et al., 2010; Boras et al., 2014). However, many molecular events occur at time-scales that cannot be easily accessed by experiments (Zhou and Bates, 2013). Fortunately, computational simulations have provided alternative methods for determining parameters for whole-cell models. BD simulations rely on simplifying assumptions that allow simulations of microscopic events that span larger systems and timescales than more detailed methods, such as MD, allow. BD can be used to determine association rate constants (k_{on}) for diffusion-limited protein-protein and protein-small molecule interactions. It specifically examines how electrostatic and steric properties of molecules affect molecular encounter rates. Combining this information with *in vitro* experiments and MD-derived states will enable a new generation of protein-scale MSMs to be developed for incorporation into whole cell models.

As an example problem necessitating the integration of approaches across a broad range of spatial and temporal scales, we focus here on protein kinase A (PKA), which is activated by cAMP and is a key regulator of many cellular processes. In cardiac myocytes, for example, PKA is a critical regulator of intracellular calcium handling cycling, and its dysregulation is well known to be a contributing factor in heart failure (Bers, 2001). The PKA holoenzyme consists of two regulatory (R) subunits and two catalytic (C) subunits. Each R subunit has two cAMP-binding domains (CBD), a DD-docking domain, and a disordered linker region containing the inhibitory sequence that interacts with the C subunit. PKA is activated upon cAMP binding to the CBDs on the R subunit inducing release of the C subunit. Over the last 15 years, several whole-cell models of ventricular myocytes that incorporate calcium release and beta-adrenergic stimulation through a simplified PKA activation mechanism were developed (Saucerman et al., 2003; Bondarenko, 2014). More recently, a mechanistic protein-scale MSM of PKA holoenzyme activation was developed (Boras et al., 2014). Still, incorporating an improved PKA MSM into existing whole cell models will provide a more physiological testing of PKA activation as well as the capability to predict the effects of PKA mutations on the whole cell scale.

In this review, we highlight some of the tools and techniques used to develop integrative models that span scales from the molecule to the cell, including: MD, atomic MSM, BD, milestoning models, protein MSM, and whole cell modeling. We provide the nomenclature necessary to bridge these scales and discuss the limitations of these approaches as well as ways to minimize error propagation. Finally, we show the role of MD and BD simulations have played in the development of a protein scale MSM of PKA $R1\alpha$ and discuss the role this new protein-scale



MSM of PKA will play in existing whole cell models of cardiac function and disease states.

Nomenclature

This paper deals primarily with Markovian models, or models that are only dependent on the current state of the model and not the history of the states it has visited. Both MSM and milestone models operate under a Markovian assumption. Also, for this paper we use “atomistic” or “atomic-scale” to describe any model that treats atoms explicitly. This generally includes MD, MSM, BD, and milestone. These models stand in contrast to “protein-scale” Markov models and cell-systems models which primarily focus on protein and cell function and general protein-protein and small molecule-protein binding events. Even though atomistic MSM and protein MSM are both Markovian models, they serve distinct purposes.

Accessing the Conformational Ensembles of Proteins

A protein’s function is governed by its conformational ensemble, which can be modulated through mutations and intermolecular interactions (Tsai et al., 1999; Henzler-Wildman and Kern, 2007; Boehr et al., 2009; Teilum et al., 2011; Marsh et al., 2012; Motlagh et al., 2014). Therefore, to build multiscale models starting at the atomic scale, one needs to elucidate the key conformational states of a protein and the dynamics of those states from atomistic data associated with those states. This can be achieved through exploration and characterization of the protein’s conformational ensemble. In this section, we review computational methods for modeling the conformational ensembles of proteins important in cell signaling. We begin with an overview of molecular dynamics simulation methods and conclude with a discussion on the use of MSM to determine the conformational ensemble more efficiently.

Molecular Mechanics and Molecular Dynamics Simulations

Atomistic models of conformational ensembles can be computationally generated from molecular mechanics simulations. These simulations require two components: a force field that describes how the atoms interact with each other and a method for exploring the conformational ensemble (Karplus and McCammon, 2002; Adcock and McCammon, 2006).

To simplify the complex quantum mechanical interactions between atoms, molecular mechanics simulations use empirical force fields to describe the interactions between atoms. These force fields are described in terms of classical mechanics (Wang et al., 2001; Ponder and Case, 2003; Adcock and McCammon, 2006). For example, each atom of a system is described as a charged particle in space. Bonding interactions between atoms are described as springs using Hooke’s law. Nonbonding interactions between atoms are described as Columbic and van der Waals interactions. Commonly used force fields include CHARMM (Brooks et al., 2009), AMBER (Cornell et al., 1995),

OPLS (Kaminski et al., 2001), and GROMOS (Oostenbrink et al., 2004). While a discussion on force field selection is beyond the scope of this review, it is important to understand the assumptions and performance bias of a force field used in any simulation (Guvench and Mackerell, 2008; Vitalini et al., 2015).

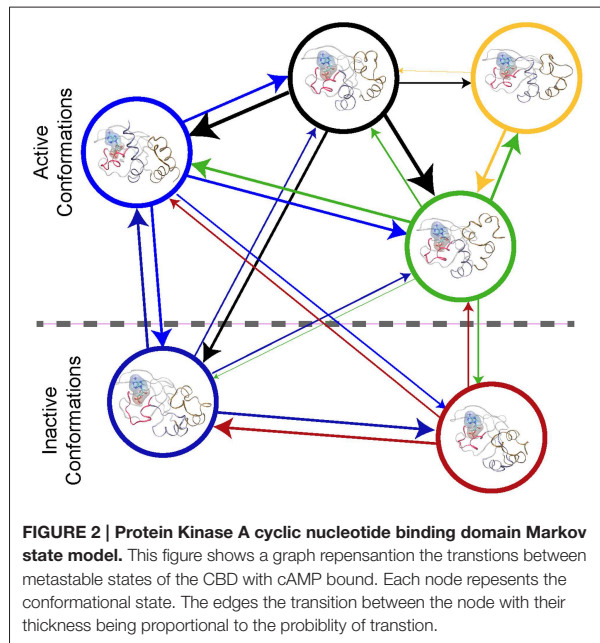
The motion of the atoms resulting from the force field determines the conformational ensemble of the system. The motions of these particles are generally simulated either with Monte Carlo techniques that randomly sample conformational space, or through MD simulations, where Langevin’s or Newton’s laws of motion are solved over time (Karplus and McCammon, 2002; Adcock and McCammon, 2006). While MD is more computationally expensive than MC, it retains the temporal relationship between conformations, which is advantageous when quantification of kinetic parameters is desired. Popular MD programs include AMBER (Pearlman et al., 1995), CHARMM (Brooks et al., 2009), GROMOS (Christen et al., 2005), and NAMD (Phillips et al., 2005).

Theoretically, MD simulations can sample the entire conformational ensemble of a system given infinite simulation time. While certain specialized supercomputers have been built to sample into the millisecond range (Shaw et al., 2010), with current commodity-level resources, MD simulations can only continuously sample a system for a few microseconds at most, which is insufficient to effectively sample most ensembles, including the CBD. However, with the increasing performance of supercomputers, GPU-accelerated MD simulations (Götz et al., 2012; Pierce et al., 2012; Salomon-Ferrer et al., 2013), and the use of highly distributed computing (Pande et al., 2003; Kohlhoff et al., 2014), multiple parallel MD simulations can achieve total non-continuous sampling time approaching the high-microsecond to low-millisecond range. MSMs can subsequently be used to stitch together the many short-timescale simulations into one cohesive framework that allows the extrapolation of longer-timescale data. This MSM framework was used for the CBD system discussed below.

Atomic-scale Markov State Models of a Conformational Ensemble

An atomic-scale MSM describes the conformational ensemble of a protein as the probability of transitioning between discrete collections of conformational states at a fixed time (Pande et al., 2010; Chodera and Noé, 2014). This can be visualized as a bidirectional graph, (see **Figure 2**), where each node represents a cluster of similar conformations. The probability of transition between states is indicated by the thickness of the connecting lines in **Figure 2**. If the conformational states and the transitions can be accurately determined, then the MSM describes the thermodynamics and the kinetics of the system’s conformational ensemble. Thus, one can derive the key parameters required for higher scale models with a MSM (Prinz et al., 2011a).

Atomic-scale MSMs of the conformational ensemble of a protein are built from MD simulations. Each conformation sampled during the simulation is assigned to a discrete conformational state, usually by clustering. Then the transitions between the discrete states are determined from the MD trajectory by counting the transitions. The transition counts



are then used to generate a transition probability matrix, the mathematical representation of the MSM (Pande et al., 2010; Prinz et al., 2011a,b; Chodera and Noé, 2014). The transition probability matrix can be analyzed to determine the equilibrium population of each conformational state, to identify metastable conformational states, to understand the principal motions of the protein, and to study the mechanisms of conformational change (Pande et al., 2010; Prinz et al., 2011a,b; Chodera and Noé, 2014).

Because a MSM depends on the probabilities of transitions between discrete conformational states, the conformational ensemble of the protein can be sampled more efficiently than with traditional MD. To effectively sample the conformational ensemble of a protein at equilibrium using traditional MD simulations requires running the simulation long enough to explore the conformational ensemble multiple times. However, when building a MSM the MD simulations can be focused on the transitions between states avoiding spending time sampling stable conformations and improving the sampling of rare events. For example, a hypothetical transition between active and inactive states can be determined from multiple short simulations that explore the intervening conformations without requiring a single simulation to bridge the two states. Additionally, once a preliminary MSM is built poorly sampled transition can be additionally sampled to improve the quality of overall MSM.

Detailed methods for building MSMs for MD simulations have already been described (Sjoberg et al., 2010; Prinz et al., 2011b). Here we highlight key considerations for building a MSM that will be integrated with higher-scale models with examples from a recently developed MSM of the cyclic-nucleotide binding domain of the R subunit of PKA (Malmstrom et al., 2015). The overall process of building a MSM is as follows: (1) defining the

conformational space; (2) initial molecular dynamics sampling; (3) iterative refinement of the MSM; and finally (4) selection of the final model for analysis.

The goal of our study was to determine the kinetics of the conformational ensemble of the CBD with and without cAMP bound. We defined the conformational space as the atomic coordinates of the alpha carbons in a protein, dividing the conformational states discrete into states using RMSD-based clustering. We started sampling the CBD in either a crystallographic predetermined active or inactive state with and without cAMP bound. Building the final MSMs required over 70 μ s of total sampling time comprised of both long-timescale initial sampling and iterative adaptive sampling to refine the models (Malmstrom et al., 2015).

Throughout the sampling and refinement process, the quality of a MSM is judged using implied timescale plots (Pande et al., 2010; Prinz et al., 2011b). Data points of the plot are constructed with eigenvalues of the transition probability matrix populated at different lag times, or times between events. The plots indicate at what lag times the models are Markovian and if the models are consistently capturing the principal conformational changes of the system. Additionally, a Chapman-Kolmogorov-test is used to validate the consistency of a MSM with molecular dynamics simulations (Prinz et al., 2011b). Using these two metrics, a final model is selected, the statistics of which are sampled at a specific lag time, which represents the fastest transition within the conformational ensemble that is also Markovian. This final model can then be used to derive the parameters for the multiscale model.

As described before, a MSM consists of the equilibrium probabilities for each conformational state. These probabilities are used to derive thermodynamic properties. Spectroscopic analysis can be used to identify metastable states within the conformational ensemble that can be used to build coarse-grained models of the system (Prinz et al., 2011a; Malmstrom et al., 2014). Transition path theory (Vanden-Eijnden and Tal, 2005) can be employed to approximate the kinetics of transitions between states. These rates become the parameters to feed into the multiscale model. For the CBD model we were able to obtain the rates of transitions between the active and inactive states and show how cAMP modulates the conformational ensemble, changing the function of the CBD. These rates have been an important benchmark in understanding the dynamics of the CBD, and form the foundation for examining the total R subunit and its interactions with the C subunit.

While the use of MSMs provides to conformational ensembles, there are still several important considerations and limitations to this method that should be considered in context of integrating them into a multiscale model. First, because conformational space is discretized, all kinetic rates are artificially fast (Prinz et al., 2011a,b), and should be considered an upper bound, especially when applied to high scale models. Second, a recent study indicates that modern force fields used in MD simulations produce varying transition kinetics (Vitalini et al., 2015). Therefore, the same force field should be used for all models of a system, and the limitations of the force fields should be understood. Thirdly, while the MSM is somewhat

robust to errors in clustering, give a sufficiently fine division of conformational space (i.e., a lot of clusters) (Prinz et al., 2011b), the MSM is still dependent on the starting conformation used to initialize the simulations and the limitations of MD. Therefore, it is possible to not have included important conformational states leading to an incomplete model of the conformational landscape and incorrect predictions. However, limitations can be overcome using enhanced sampling methods (Bernardi et al., 2015) and from understanding acquired in the large-scale models. Finally, the MSMs are computationally demanding. This cost limits their usefulness in multiscale models, as a significant amount of time can be required to describe only one state in a higher scale model. Other sampling methods may be sufficient to obtain parameters for larger models. For example, if the opening and closing of a flap on a protein is the only permutation of interest, elastic network models are more computationally efficient in estimating those rates than MSM.

Investigating Intermolecular Interactions

As we extend into larger spatial scales of modeling, the focus of our discussion shifts from intramolecular investigations with MD to the study of intermolecular encounters using BD. BD simulations are used to estimate the rate constants of second-order association events between two molecules. The output of these simulations provides kinetic on-rates used directly in higher levels of modeling. The application of BD simulations has extended beyond bi-molecular encounters in simulations of molecular crowding (McGuffee and Elcock, 2010) in cellular environments. In this section, we discuss the methodology and limitations of BD simulations, what can be gained from their use, and a brief overview of their application to multiscale modeling.

Brownian Dynamics Simulations

In BD, molecular diffusion is modeled using the theory of Brownian motion; where internal dynamics of each molecule are frozen, constraining the molecules into rigid bodies that are free to diffuse and tumble in solution, but may not change shape. Popular programs used to carry out BD simulations include BrownDye (Huber and McCammon, 2010), SDA (Gabdouline and Wade, 1997, 1998), ReaDDy (Geyer, 2011; Schöneberg and Noé, 2013), Brownmove (Geyer, 2011), and BD_BOX (Dlugosz et al., 2011). Similar to MD, one must choose a force field for BD simulations of the molecular system: AMBER (Dickson et al., 2014), CHARMM (Klauda et al., 2010), GROMOS (Oostenbrink et al., 2004), etc. However, the only force field quantities utilized in BD simulations are the partial charges and Van der Waals radii of each of the atoms of the biomolecule. In conjunction, these properties can be used to obtain the electrostatic potential from software that can solve the Poisson–Boltzmann (PB) equation. The electrostatic potentials of the biomolecules determine the long-range forces that the molecules impose on each other. Thus, electrostatics function as one of the most important determinants of the outcomes of BD simulations. Popular software packages that solve the PB equation include APBS (Baker et al., 2001; Holst, 2001) and DelPhi (Honig and Nicholls, 1995; Rocchia et al., 2002). Rigorous derivations and discussions of the form

and proper usage of the PB equation can be found in the literature (Leach, 2001; Fogolari et al., 2002; Gu and Bourne, 2009).

In BD simulations, the solvent is modeled as a continuum; that is, there are no water molecules or dissolved ions modeled in atomic form in the simulation. Instead, the solvent is modeled as a field that surrounds the biomolecules and can have varying degrees of physical realism. This significantly reduces the computational power necessary for BD simulations in comparison to explicit solvent MD. The user typically specifies parameters that control solvent dielectric, hydrodynamics, desolvation, and ion screening, all which affect the realism of the solvent model and the computational cost of the simulation.

In addition to the long-range forces imposed by intermolecular electrostatics, a stochastically determined force is also imposed on the molecules in a BD simulation. This stochastic force is directed randomly with a magnitude sampled from a Gaussian distribution centered at zero whose variance depends on the simulation time-step and the molecule's diffusivity properties. The stochastic force is intended to approximate the random “kicks” that would be caused by the solvent, but are otherwise absent in the continuum model.

Finally, the simulation must ensure that the Van der Waals radii of the atoms of different molecules do not overlap; a phenomenon known as a steric clash. Often, simulation steps that result in a steric clash are discarded and recomputed. Alternatively, many BD programs can impose a Lennard-Jones force at close molecular proximity to prevent a steric clash (Elcock, 2004; Huber and McCammon, 2010). BD simulation and the theory behind it compose a rich and expansive field, and many sources exist to allow the interested reader to improve his or her knowledge and technique (Ermak and McCammon, 1978; Allen and Tildesley, 1987; Gabdouline and Wade, 1996; Elcock et al., 2001; Gabdouline and Wade, 2002; Madura et al., 2002; Elcock, 2004).

Considerations for Brownian Dynamics Simulations

A key starting point for BD simulations is the selection of the encounter complex, which describes the atomic interactions that define a reaction between molecules. Ideally, crystal structures will inform this step. If crystal structures of the encounter complex do not exist, molecular docking programs can serve as a substitute. In the case of PKA, two crystallized structures of the regulatory subunit of protein kinase A RI α show very different conformations when bound to either cAMP (Su et al., 1995) or the catalytic subunit (Kim et al., 2007). To test the effects of structure on cAMP association with BD methods, one can use the crystal structure conformations of the regulatory subunit in separate BD simulations. Alternatively, the two different conformations can be used as starting points of separate MD simulations. A number of structures in the conformational ensemble will be generated and can serve as structures for separate BD simulations.

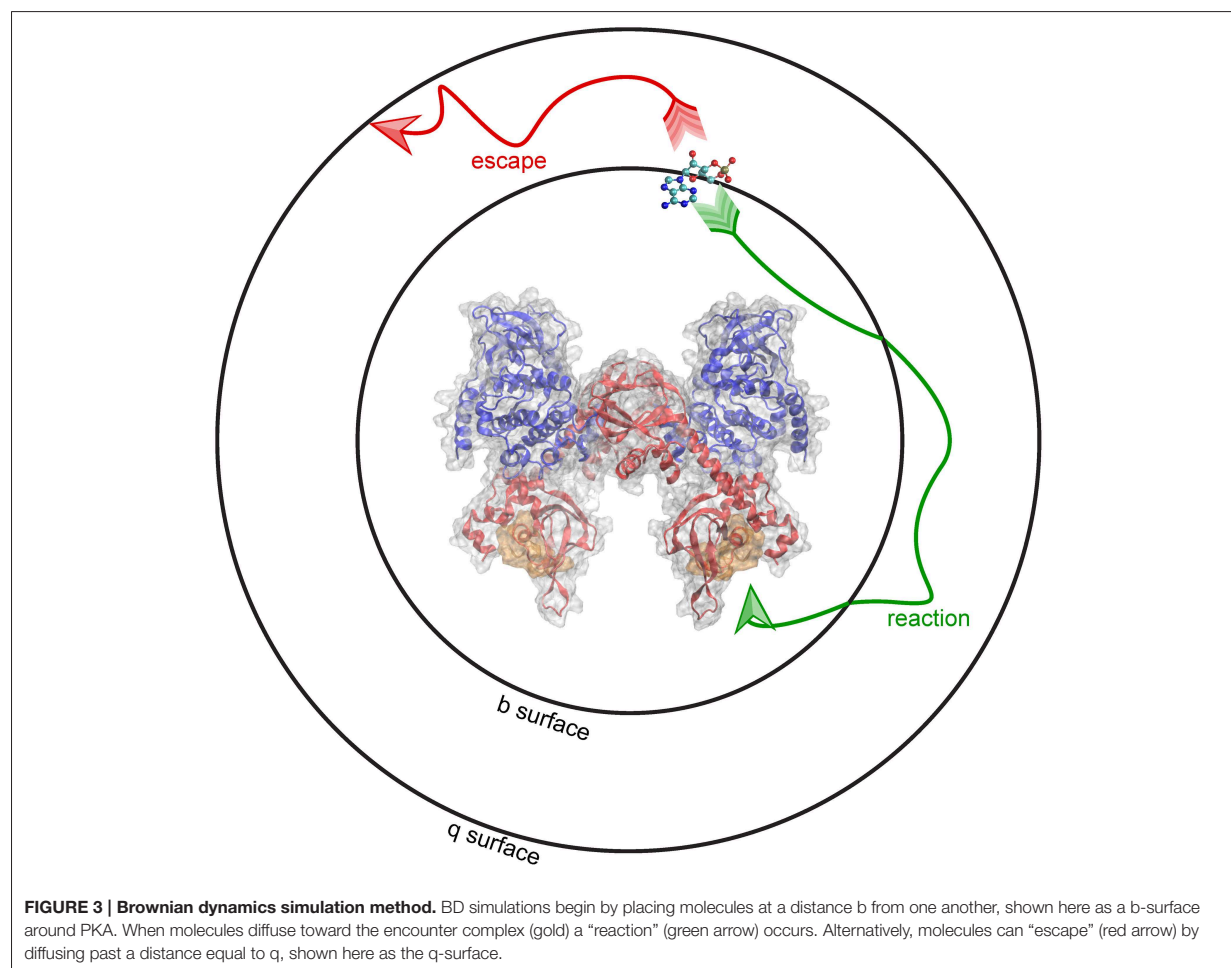
At the start of a simulation, the ligand is placed at a distance b from the receptor, at a location known as the b surface, which is defined as the distance where forces between the two molecules are centrosymmetric. Simulations terminate

either upon the molecules reaching the predefined bimolecular encounter complex (a binding event), or when the molecules separate beyond a greater intermolecular distance q . The distance q , the radius of the q surface, is typically 10–50 nm larger than the distance b (Gabdoulline and Wade, 1998). The probability of association vs. escape is then used to calculate the association rate constant (k_{on}). This schematic, including the surfaces at the b and q distances, are depicted using PKA as the receptor and cAMP as the ligand (Figure 3). BD can be used to model the association of cAMP with PKA, and predict features of the binding event, including the route of approach, the encounter complex, and the rate constant of association.

A second important factor in BD simulations is the structure of the molecules used in the simulations. Recall that BD simulations use a rigid-body approximation of molecules, meaning that the conformation of the molecule will not change throughout the simulation as it does in MD. Typically, crystal structure conformations are used. Another attractive possibility is the use of conformations generated by MD as starting points for BD simulations. Using this method, the user can

select meta-stable or even rare conformations of a protein generated in MD simulations and compare the association rates and probabilities with respect to structural changes in the protein. MD trajectories can also be used to generate ensemble-averaged electrostatics (Votapka et al., 2013) where the simulated molecular motions are combined to form an electrostatic potential that includes the dynamic properties of the molecule. This effectively leads to a more holistic, dynamic representation of the electrostatic potential, effectively mediating some of the limitations of the rigid-body approximation of the simulations.

Solutions to the Poisson-Boltzmann equation include variations in the dipole moment and especially the charge density, with respect to how the solute affects the solvent, but also how the solvent affects *itself*. So while common implementations of the Poisson-Boltzmann equation solvers do not include many features of true aqueous solvents, it at least does assume that certain aspects of the solvent are heterogeneous. In addition, BD simulations themselves often model such things as hydrodynamics and desolvation forces, which are intended



to approximate additional solvent features such as inertia and entropy at a surface, respectively.

Despite their ability to calculate association rate constants with respect to steric and electrostatic properties of molecules, BD simulations have limitations that users should know and recognize. First, the results of BD simulations depend on the encounter complex criteria. Such criteria must usually be tested and optimized in order to reproduce a reasonable association rate constant. Incorrectly chosen encounter criteria can significantly limit the accuracy of the simulation outcomes. Second, the rigid-body approximation of molecules in BD can only represent one part of the binding process: the diffusion—meaning that the rate constant calculated by simulations is that of association and not actual binding. Nevertheless, alternative methods that combine intermolecular investigations of BD with intramolecular dynamics of MD are being developed that promise kinetic rate estimations through simulation (Votapka and Amaro, 2015). These developments represent an approach toward spanning the MD and BD simulation regimes into a unified multiscale framework.

To our knowledge, no systematic method yet exists for estimating the true amount of error propagated by the assumptions inherent in BD. However, general consensus agrees that BD performs relatively well if the rate constants of an event it is estimating can be classified as a diffusion-limited process; that is, a process whose time to completion is primarily limited by particle diffusion. In the case of binding, the range of diffusion-limited rate constants is considered to include values of approximately 10^8 – 10^9 $M^{-1}s^{-1}$ (Bar-Even et al., 2011).

Schemes do exist to approximate the precision of a k_{on} based on the statistical sampling of a binding probability vs. escape. Specifically, the uncertainty of the rate constant of binding is proportional to the inverse square root of the number of trajectories in BD simulations that have completed in a binding event. Since millions or even billions of BD trajectories can usually be completed at relatively little computational cost, it is typically not difficult to obtain relatively high precision of a rate constant using BD. However, while the estimated rate constant may be precise, it still may be inaccurate if the rigid molecules, implicit continuum solvent, or some other approximation assumed by BD do not adequately model the system. Comparison to experimental rate constants of the simulated ligand-receptor system, or perhaps of similar systems, can give an indication of the discrepancy between the “true” value, and the value obtained using BD.

Unifying MD and BD Simulations through Milestoning

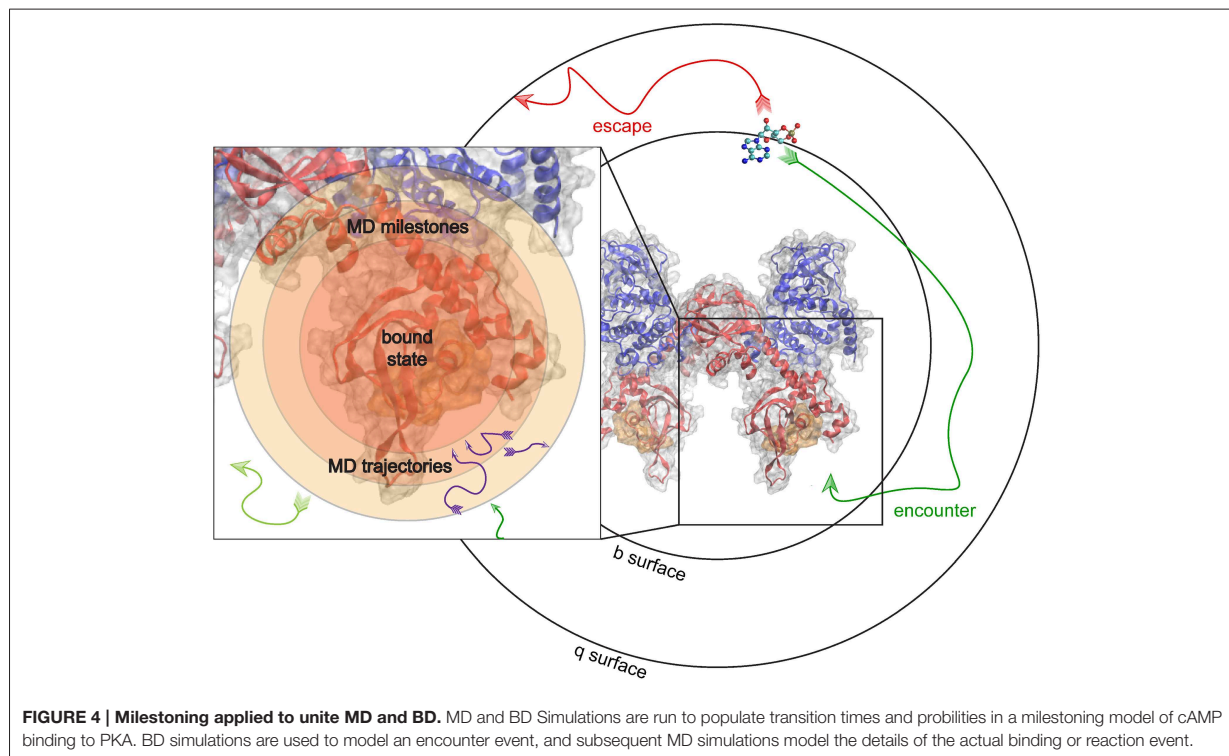
The possibility of combining the speed of rigid-body BD simulations with the flexibility of all-atom MD simulations to predict kinetic and thermodynamic quantities of interest is an attractive option. Ensembles of conformations or trajectories can be sampled from each simulation method, and statistics involving the probability and timescales of transitions between predefined states from the simulations can be combined using MSMs or the theory of milestoning (Faradjian and Elber, 2004) to model the details of intermolecular interactions.

Milestoning is a technique that is similar to the theory used in MSMs and can serve as an alternative approach to investigating biomolecular events, such as conformational sampling (West et al., 2007; Mugnai and Elber, 2015), diffusion (Mugnai and Elber, 2015), and membrane permeation (Cardenas et al., 2012), among others. Milestoning retrieves the kinetics as well as the thermodynamics of chemical processes (Vanden-Eijnden et al., 2008; Májek and Elber, 2010; Kirmizialtin and Elber, 2011), and can make use of extensive parallelization. Although similar to MSMs, milestoning models have a number of key differences, and may or may not be well suited to address a particular biophysical question. Unlike MSM states that are volumes in phase space where the system exists until it crosses into another, milestones are surfaces in phase space that the system traverses, and where the system’s current “state” is the surface that the system has most recently crossed.

To give an example, we examine the hypothetical case where the k_{on} of binding between PKA and cAMP can be predicted. In this milestoning model, we define a set of concentric spheres of different radii, all centered on the binding site of PKA (Figure 4). These concentric spheres define the milestones. MD simulations are started from conformations where cAMP is located on each spherical milestone, and each simulation is similarly terminated once cAMP diffuses to another surface. Thus, to the milestoning model, whichever simulation method is used to populate the transition kernels and incubation time vectors with statistics is of no consequence. The most appropriate simulation method can be chosen when cAMP is started on a particular surface.

Unlike MSMs, milestoning transitions may only occur to states that are adjacent in the positional or conformational space and lag times (or incubation times) can vary between inter-state transitions. Therefore, milestoning may be a desirable technique in situations where the system crossing surfaces would more appropriately represent transitions than the system traversing regions of space. For instance, because current implementations of BD simulations make extensive use of surfaces, such as the surfaces at the b and q distances and the encounter surfaces, milestoning is a natural choice to utilize transition statistics obtained in BD simulations. MD simulations modeling a binding event can make use of either milestoning models or MSMs, but when a combination with BD is desired, milestoning offers a promising framework to combine statistics from the two simulation methods.

Milestoning theory can be used to investigate a wide diversity of biophysical scenarios, and has been applied in a variety of contexts (Faradjian and Elber, 2004; Elber, 2007; West et al., 2007; Elber and West, 2010; Cardenas et al., 2012). In some physical situations, implementations of milestoning outperformed MSMs in resemblance to experimental results (Vanden-Eijnden et al., 2008). The application of milestoning to intermolecular interactions is still a recent development, and many possible improvements may enhance the efficiency and accuracy of the estimation of binding rate constants. Examples of these include further discretizing the system into grid-like milestones, rotational milestones, or milestones that can represent internal degrees of freedom. Extensive derivations and discussions of milestoning theory are discussed elsewhere



(Faradjian and Elber, 2004; Vanden-Eijnden et al., 2008; Májek and Elber, 2010; Kirmizialtin and Elber, 2011).

From Atomistic to Protein-scale Models

Bridging the gap from atomic simulations to whole cell models is challenging. Protein-scale MSMs connect the atomistic scale to cell or tissue phenomena by reducing the complexity of molecular models. This enables simulations on larger time and spatial scales, while maintaining structural details required for protein function. These models simulate biological phenomena relevant to a whole-cell model, including ionic currents, fraction phosphorylated, or percent activation, and the output can be compared to *in vivo* experiments.

Protein-scale MSMs have been used to represent protein interactions since the early-1990's (Edeson et al., 1990). Several papers have been written on the development of protein-scale MSM, particularly of ion channels (Edeson et al., 1990; Giugliano, 2000; Gurkiewicz et al., 2011; Lampert and Korngreen, 2014). Ion channel MSMs have been made possible by the detailed statistical data that comes from single channel patch clamp recordings (Qin et al., 1996). These models have started to replace traditional phenomenological Hodgkin-Huxley style models of ion channel kinetics in whole cell action potential models (Rudy and Silva, 2006). They have been most useful when there is a need to model the effects of specific channel modifications, such as drug binding (Clancy et al., 2007), gene mutation (Rudy and Silva, 2006), or post-translational modifications (Yang and Saucerman, 2012).

But the use of protein-scale MSMs is not limited to those systems where dynamic biophysical recordings are available; instead, these models can be built from BD and MD simulations.

Protein-scale MSM

The first step in model development is to determine the overall structure of the model. Unlike atomic-scale MSMs, protein-scale MSMs do not represent every conformation of atoms as a state; instead, each state represents an ensemble of related atomic conformations that comprise a functional structure. This significantly limits the number of degrees of freedom and decreases the computational power needed, which enables multiple protein-scale MSMs to be combined into system-scale models. However, because these models are a simplification of the total potential states, the choice of which states are relevant becomes essential to making a useful model of a protein.

Functional State Discovery through MD Simulations

Frequently, several different states are captured by molecular-scale experiments, including X-ray crystallography and mass spectrometry. These experimental approaches can provide data on particular stable conformations (e.g., active or inactive states); however, due to the static nature of these tools, significantly less information is known about the transitions between states. For example, there are published structures of the R subunit of PKA bound either to cAMP or to the C subunit, but little is known about the transition between these end states. MD simulations

can suggest intermediate states for incorporation into protein-scale MSMs. Similarly, atomic-scale MSMs provide insights into which states are populated and the rates of transitions between conformations.

Using BD Simulation to Inform Kinetics

For small molecule-protein interactions and protein-protein interactions, BD and experimental data can serve complementary roles in determining kinetics. Dissociation-rates are typically slower than association rates and are therefore easier to measure experimentally using techniques such as surface plasmon resonance (Herberg et al., 1996). Additionally, most dissociation events are limited by conformational changes and not by diffusion—the latter of which BD is designed to model. For these same problems, MD simulations would be required to run for inaccessibly long periods of time (ms to s) to register release events. Association-rates, on the other hand, tend to be orders of magnitude faster and therefore are harder to measure experimentally. BD simulations are ideal for measuring fast interaction rates on the ns to μ s time scales, many of which are limited by diffusion. In combination with equilibrium data, these techniques can be used synergistically to determine rates for small molecule-protein and protein-protein interactions. By basing the ensemble of states of the model on MD simulations, and the kinetic interactions on BD simulations, it is possible to predict the effect of a mutation on protein function and, by extension, on the whole cell.

Testing with Empirical Data

Data from experiments, MD simulations, and BD simulations can be integrated into a simplified protein-scale MSM with states and interactions relevant to protein function. Frequently, these combined methods will suggest several possible functional state ensembles. Competing models are generated, with different states or different relationships between the states. Subsequently, the resulting models are tested to determine their ability to fit relevant experimental data (Boras et al., 2014). For protein-scale MSMs, the data used for fitting most often comes from *in vitro* experiments. Ideally, the data used to differentiate between competing models is collected under conditions that are most relevant to a whole cell. For example, in the PKA-R1 α model developed by Boras et al. (2014) all of the data used for fitting was collected in the presence of excess Mg²⁺ and ATP, both of which have been shown to affect PKA activation (Neitzel et al., 1991). These conditions are similar to what is found in a cell; however, recently published data has also highlighted the role of ADP in PKA activation (Khavrutskii et al., 2009), which could affect the role of PKA in metabolism but is absent in the current MSM.

The accuracy of each theoretical model is determined using an error function based on the weighted sum of squares difference between the model's predictions and the available experimental data. Minimizing this error function optimizes unknown parameters. If the MSM are nested (all possible states in a model with fewer degrees of freedom can be represented in the model with more degrees of freedom) then a statistical *F*-test can be performed to determine if the added degrees of freedom

significantly improve the fit (Anderson and Conder, 2011). This ensures that MSMs do not become needlessly complex without an improvement in the accuracy of the model's predictions.

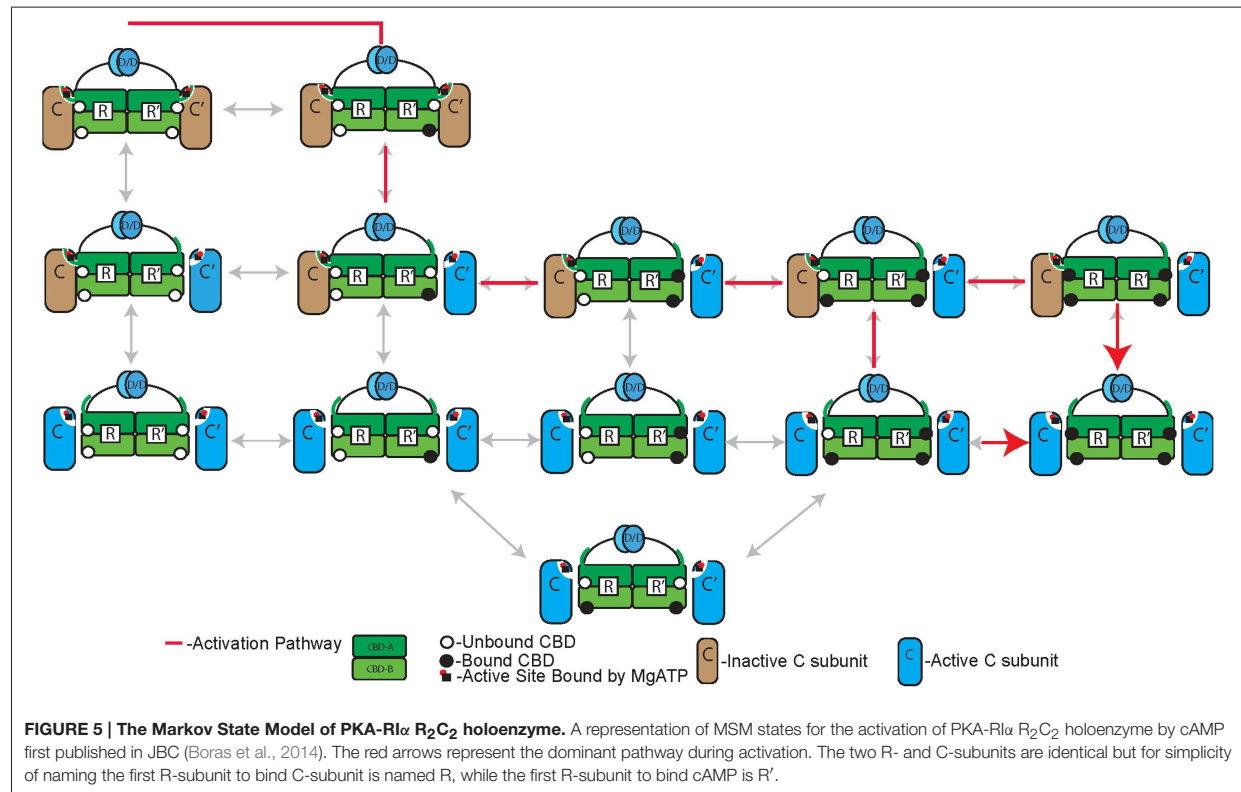
Frequently, data acquired with mutant proteins that cannot reach specific states is used to differentiate competing models. The MSMs are altered slightly by removing those states, without refitting any parameters, and the output is compared against the experimental results (Boras et al., 2014). For example Clancy et al. used MSM of a cardiac sodium channel to show that a mutation in its C terminus can lead to long-QT syndrome, which causes life-threatening arrhythmias (Clancy and Rudy, 2002). This highlights how protein-scale MSM based on atomistic data can predict the effect a mutation will have on the whole cell and eventually on the organ scale as well.

To mitigate error propagation when the protein-scale MSM is added to whole cell models, a sensitivity analysis can be performed to test the robustness of the solution (Campbell et al., 2010). In this process each rate is perturbed to determine its effect on the desired output of the model. States can also be removed to see how essential they are to the final result. The objective is to quantify how much the final result relies on any individual rate or state and compare that to the uncertainty in the experimental measurements. This technique can also highlight which states predicted from atomic scale modeling would have the greatest effect if mutated or pharmaceutically targeted. This is especially useful in quantifying the potential effect of rare conformations. Due to sampling bias they may not be captured in MD simulations but by adding them to the model their potential effect can be determined even if precise kinetic parameters are not known.

Applying to MD and BD Modeling to Protein Scale PKA-R1 α MSM

These techniques have been applied to the development of a novel PKA protein scale MSM (Boras et al., 2014) (Figure 5). First, the effects of cAMP binding on CBD-A of the regulatory subunit of PKA were examined (Malmstrom et al., 2015). Using extensive all atom molecular dynamics simulations integrated with atomic-MSM, the conformation of the CBD with and without cAMP bound was determined. Conformational selection was identified as the general mechanism of allostery within a single CBD, which transitions between an active and an inactive conformation whether or not cAMP is bound. cAMP was found to regulate the function of the CBD by deepening the free energy landscape and selecting conformational states that favor the active conformation. Interestingly, cAMP modifies the transition rate between the active and inactive conformation and not the transition between the inactive and active conformations. Additionally, the roles of each of the signaling motifs in the CBD were elucidated.

Based on these findings and crystal structure data, five nested protein scale MSM were considered. Each model was structured to test competing theories of PKA R₂C₂ activation based on MD simulations. The crystal structures suggested that a model that treated each R-C heterodimer as independent would be insufficient to fit the data, due to the compactness of the R₂C₂ holoenzyme. Atomistic MSM predicted that a conformational



selection mechanism would most accurately fit the data for isolated CBDs. The models were developed in the Virtual Cell computational environment (Moraru et al., 2008) before being translated into MATLAB to take advantage of optimization programs (Marsden et al., 2008). The models were fitted to kinase activity and cAMP binding data under physiological conditions (Christensen et al., 2003). Additionally, the various models were then compared to mutant PKA experimental data with either an inhibited CBD-A or CBD-B binding site. One model was shown to fit the wild type and predict the experimental results better than any other. This model validated the atomistic MSM by showing that CBD-B binding leads to release of the C-subunit prior to CBD-A binding similar to a conformational selection mechanism and created a thermodynamic protein-scale MSM of PKA activation.

However, since the fitting data, as well as the mutant data, were collected at long enough intervals that an assumption of thermodynamic equilibrium was valid, the resulting MSM could only reproduce equilibrium behavior. From a cellular perspective, PKA's response to a stimulus over time is essential to understanding PKA function. The single turnover rate in response to a stimulus has been implicated in activation due to A kinase anchoring proteins that bind PKA near one target (Scott and Santana, 2010). Therefore, the addition of kinetic rates would significantly increase the utility of the MSM a whole cell model.

Owing to the fast rate of activation of PKA in the presence of cAMP, the amount of experimental kinetic data on PKA activation, particularly on-rates, is limited. This is a problem ideally suited to solving using atomic simulations. The atomic scale MSM suggested that cAMP binding only affects the rate of transition from active to inactive states but not the reverse. BD simulations can be combined with experimental data to suggest binding and release rates for R-C and R-cAMP interactions. In conjunction with *in vitro* experimental data this type of data will allow the thermodynamic MSM to become a kinetic MSM better suited to whole cell scale analysis of signaling network properties.

Integrating Protein Scale MSM into Whole Cell Models

The potential of molecular and protein-scale models culminate in whole-cell and tissue-scale models that can predict phenotypes and mechanistically explain disease states. These models combine several MSMs to predict cellular responses to either internal or external stimuli by tracing behavior down to molecular interactions. When developing these protein-scale MSMs, it is best to keep in mind what broader biological function will be modeled at a larger scale since this will determine not only what states are relevant but also what type of model is best for a given phenomena.

Stochastic and Deterministic MSM in the Whole Cell

Some cell functions are best simulated using a continuum of species concentrations and smooth probability distributions, while statistically rare events are better modeled when individual molecules are tracked and the stochasticity of interactions is accounted for. Correspondingly, protein-scale MSM can be either stochastic or deterministic in nature. The stochastic models, like the atomic-scale MSM described earlier, are based on Monte Carlo simulations where the probability of transitioning between states is dependent on the kinetics of the binding and/or the conformational shift that each transition represents. This is the most accurate representation, where each event is dependent on the chance that two molecules will interact or that a conformation will be sampled based on random motion.

Many biological processes, such as calcium sparks in cardiac myocytes, can be explained with stochastic simulations. Calcium sparks occur when calcium is released from the sarcoplasmic reticulum via an isolated cluster of ryanodine receptor calcium release channels in the absence of a depolarizing event. In other words, a single cleft or a cluster of clefts acts differently than the rest of the cell. Whole-cell deterministic models require that every channel of a given type are identical and therefore every channel could be fractionally open but no one channel could be fully open while the others were fully closed without changing the conditions. Therefore, to model phenomena like this, a stochastic model is necessary. When translating these models up to the whole cell, the stochastic models are ideal for agent based spatial modeling tools, such as MCell (Kerr et al., 2008), where each molecule is tracked and diffusion is represented by a random walk; although, it is worth noting that a whole-cell model can consist of a continuum diffusion approximation but still contain stochastic protein-scale MSMs. Agent-based models are ideal for small numbers of molecules or short time and spatial scales, where tracking each molecule is computationally reasonable or average approximations may be invalid.

Over a long enough time-scale or a large enough population of molecules, the Monte Carlo simulation will approach the deterministic solution. The deterministic solution is represented by a system of ordinary differential equations, instead of being represented by a transition matrix of probabilities. In these models, the states of the MSM are frequently populated by concentrations instead of a specific number of molecules.

Many biological processes can be represented deterministically, most often when the system has a large number of molecules, or covers a long time and spatial scales. Models of the calcium concentration in a cell, for example, would require a deterministic model because computationally there are too many molecules to follow and the simulation becomes intractable. However, even on a small scale a deterministic approximation can be valid. For example, Hake et al. (2014) showed that for a single dyadic cleft in a cardiac myocyte, the random walk and the deterministic continuum approximation gives the same result for a calcium induced calcium release event, even though a continuum approximation of the calcium in the cleft is unrealistic due to the scarcity of calcium ions. By treating the continuum as deterministic but the protein-scale MSMs as

stochastic we can reproduce the stochastic sparks while limiting the required computational power.

Advantages in Whole Cell Modeling

The potential of molecular and protein-scale models culminate in whole-cell and tissue-scale models that can predict phenotypes and mechanistically explain disease states. These models combine several MSMs to predict cellular responses to either internal or external stimuli by tracing behavior down to molecular interactions. The power of building atomic-scale and protein-scale MSMs for wild type and disease mutants comes from their integration into whole cell models. At the whole cell scale, differences in sub-cellular dynamics of protein mutations can be studied comparatively with their wild-type counterparts. Several disease states come from known protein mutations. For example, in the case of PKA-R1 α , 117 polymorphisms and mutations have already been discovered (Horvath et al., 2010). Owing to the complexity of signaling pathways, how these mutations affect cell function is frequently unclear but by creating a whole cell model from molecular mechanisms it is possible to predict how a given mutation will lead to a particular cellular phenotype.

Whole cell models based on atomic resolution information have opened entirely new avenues of research into drug discovery. In addition to suggesting which protein is a viable target, mechanistic whole cell models can suggest which protein conformation is most favorable and even the chemical shape of a small molecule necessary to inhibit/promote activation. This allows a scale of specificity that could decrease toxicity and limit side effects.

Cardiac arrhythmias are a prime example of the potential relevance of whole cell models. Currently, one of the most commonly prescribed classes of drugs to treat arrhythmias are beta-blockers. Beta blockers bind the beta adrenergic receptors to inhibit epinephrine and norepinephrine binding to reduce the chance of a second heart attack (Cruickshank, 2010). However, this inhibits the entire beta-adrenergic pathway. By combining this new PKA protein-scale MSM with previously published adrenergic signaling models of the heart (Saucerman et al., 2003; Bondarenko, 2014), it is possible to suggest drug targets and even specific binding pockets to inhibit parts of the pathway while limiting their effect on the rest of the cell.

Conclusions

For years, atomic-resolution protein structures have aided our understanding of protein function not only through static structures provided by NMR and crystallographic experiments, but also through the prediction of dynamic properties with MD simulations. MD has revealed ensembles of structures that comprise the conformational landscape of a protein. Due to computational limitations, classical MD simulations are only able to generate microseconds (or less) of simulated time. This significantly limits the extent of the protein conformational ensemble sampled. However, information generated by MD simulations can be integrated into atomic scale MSMs, which are used to link states generated in a conformational ensemble through a kinetic scheme. The outputs of structures from MD

simulations are analyzed according to a chosen conformational state description and are discretized into microstates. The MD trajectory informs which transitory states are most favorable and calculates the transition rates between these states to be used in different scales of modeling.

MD simulations subsequently inform both the atomic scale MSMs and BD simulations. BD simulations typically use rigid-body representations; therefore selected conformations are important for understanding the effect of different structures on association probability. MD simulations provide relevant conformations for BD by generating stable conformations. In addition, ensemble-averaged electrostatics can be generated from the MD trajectories, reflecting the dynamic properties of a molecule in a static electrostatic potential map. Finally, MD and BD simulations can be directly integrated through milestoning to derive association rate constants (k_{on}) of diffusion-limited processes; a process that combines the two distinct simulation regimes—utilizing the advantages and minimizing the disadvantages of each. Such a scheme can vastly expand the time and length scales accessible in the simulation of multimolecular interactions between proteins and small molecules and/or other proteins to be combined with experimental data in protein-scale MSMs.

Protein-scale MSMs draw on every facet of the atomistic models to bridge the atomic and cellular scales. MD simulations and atomistic-scale MSMs suggest which ensemble of states will reproduce a molecular function. BD simulations combined with milestoning predict association rate constants that would be difficult to experimentally reproduce. This information, when combined with *in vitro* experimental data and statistical analysis tools, leads to the development of protein-scale MSMs for incorporation into whole cell models. Whole cell models based on atomic level details provide a new scale of specificity. The ability to scale up the effects of a protein mutation on a cellular level function is the ideal goal of a robust MSM of this kind.

As discussed throughout this paper, during the process of multiscale modeling it is essential to consider error propagation, or the effect of inaccuracies in small-scale models and the

translation of this error into higher levels. For example, conditions such as molecular crowding in the cell likely affect the energetic landscape of a protein, a phenomenon not explicitly represented in an MD simulation. The limited sampling time of MD simulations can bias the conformational landscape of the protein, affecting the kinetic rates determined by the MSM. Structures and kinetics abstracted from biased simulations can further limit the accuracy of BD simulations and protein-scale MSM, respectively. Furthermore, coarser-grained simulations such as BD and protein-scale MSMs are not free of their own inaccuracies. The sources of errors in many modeling methods may or may not be easy to recognized and the best practices for quantifying the errors is still an active area of research. Iterating through the multiscale modeling process is extremely time consuming, since frequently MSMs must be fully recreated when new constraints are added. With ample computational resources, a multiscale modeler can incorporate recursive feedback loops from multiple scales to converge to a steady solution of the represented system.

Building a multiscale model, despite inaccuracies, is extremely useful. The findings of larger scale models can be used to inform the finer scales, like the identification of unknown conformational states in protein-small molecule energetic landscape. Carefully considering the whole-cell constraints of a given disease state or drug target before creating the initial model can allow these multiscale models to be a powerful and efficient tool for understanding the mechanisms behind some of the most intriguing biological questions.

Funding

This work was funded in part by the National Biomedical Computation Resource, NIH P41 GM103426, NIH Director's New Innovator Award Program DP2-OD007237 (to REA), through NSF XSEDE Supercomputer resources grant RAC CHE060073N (to REA), through NIH grants 5 P01 HL098053, 1 R01 HL105242, and 1 R03 EB014593 (to ADM) and by NSF graduate fellowship DGE-1144086 (to LWV).

References

- Adcock, S. A., and McCammon, J. A. (2006). Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* 106, 1589–1615. doi: 10.1021/cr040426m
- Allen, M. P., and Tildesley, D. J. (1987). *Computer Simulation of Liquids*. Oxford, UK; New York, NY: Clarendon Press; Oxford University Press.
- Anderson, K. B., and Conder, J. A. (2011). Discussion of multicyclic hubbert modeling as a method for forecasting future petroleum production. *Energy Fuels* 25, 1578–1584. doi: 10.1021/ef1012648
- 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. doi: 10.1073/pnas.181342398
- Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., et al. (2011). The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50, 4402–4410. doi: 10.1021/bi2002289
- Bernardi, R. C., Melo, M. C. R., and Schulten, K. (2015). Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta* 1850, 872–877. doi: 10.1016/j.bbagen.2014.10.019
- Bers, D. M. (2001). *Excitation-contraction Coupling and Cardiac Contractile Force*. Dordrecht; Boston, MA: Kluwer Academic Publishers. doi: 10.1007/978-94-010-0658-3
- Boehr, D. D., Nussinov, R., and Wright, P. E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* 5, 954–954. doi: 10.1038/nchembio1209-954d
- Bondarenko, V. E. (2014). A compartmentalized mathematical model of the beta(1)-adrenergic signaling system in mouse ventricular myocytes. *PLoS ONE* 9:e89113. doi: 10.1371/journal.pone.0089113
- Boras, B. W., Kornev, A., Taylor, S. S., and McCulloch, A. D. (2014). Using Markov state models to develop a mechanistic understanding of protein kinase A regulatory subunit RIalpha activation in response to cAMP binding. *J. Biol. Chem.* 289, 30040–30051. doi: 10.1074/jbc.M114.568907
- Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., et al. (2009). CHARMM: the biomolecular simulation program. *J. Comput. Chem.* 30, 1545–1614. doi: 10.1002/jcc.21287

- Campbell, S. G., Lionetti, F. V., Campbell, K. S., and McCulloch, A. D. (2010). Coupling of adjacent tropomyosins enhances cross-bridge-mediated cooperative activation in a markov model of the cardiac thin filament. *Biophys. J.* 98, 2254–2264. doi: 10.1016/j.bpj.2010.02.010
- Cardenas, A. E., Jas, G. S., Deleon, K. Y., Hegefeld, W. A., Kuczera, K., and Elber, R. (2012). Unassisted transport of N-acetyl-L-tryptophanamide through membrane: experiment and simulation of kinetics. *J. Phys. Chem. B* 116, 2739–2750. doi: 10.1021/jp2102447
- Chodera, J. D., and Noé, F. (2014). Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* 25, 135–144. doi: 10.1016/j.sbi.2014.04.002
- Christen, M., Hünenberger, P. H., Bakowies, D., Baron, R., Bürgi, R., Geerke, D. P., et al. (2005). The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* 26, 1719–1751. doi: 10.1002/jcc.20303
- Christensen, A. E., Selheim, F., de Rooij, J., Dremier, S., Schwede, F., Dao, K. K., et al. (2003). cAMP analog mapping of Epac1 and cAMP kinase - Discriminating analogs demonstrate that Epac and cAMP kinase act synergistically to promote PC-12 cell neurite extension. *J. Biol. Chem.* 278, 35394–35402. doi: 10.1074/jbc.M302179200
- Clancy, C. E., and Rudy, Y. (1999). Linking a genetic defect to its cellular phenotype in a cardiac arrhythmia. *Nature* 400, 566–569. doi: 10.1038/23034
- Clancy, C. E., and Rudy, Y. (2002). Na(+) channel mutation that causes both Brugada and long-QT syndrome phenotypes: a simulation study of mechanism. *Circulation* 105, 1208–1213. doi: 10.1161/hc1002.105183
- Clancy, C. E., Zhu, Z. I., and Rudy, Y. (2007). Pharmacogenetics and anti-arrhythmic drug therapy: a theoretical investigation. *Am. J. Physiol. Heart Circ. Physiol.* 292, H1641–H1642. doi: 10.1152/ajpheart.00312.2006
- Cong, S., Ma, X. T., Li, Y. X., and Wang, J. F. (2013). Structural basis for the mutation-induced dysfunction of human CYP2J2: a computational study. *J. Chem. Inf. Model.* 53, 1350–1357. doi: 10.1021/ci400003p
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., et al. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 118, 2309–2309. doi: 10.1021/ja955032e
- Cowan, A. E., Moraru, I. I., Schaff, J. C., Slepchenko, B. M., and Loew, L. M. (2012). Spatial modeling of cell signaling networks. *Methods Cell Biol.* 110, 195–221. doi: 10.1016/B978-0-12-388403-9.00008-4
- Cregut, D., and Serrano, L. (1999). Molecular dynamics as a tool to detect protein foldability. A mutant of domain B1 of protein G with non-native secondary structure propensities. *Protein Sci.* 8, 271–282. doi: 10.1110/ps.8.2.271
- Cruckshank, J. M. (2010). Beta blockers in hypertension. *Lancet* 376, 415–415. doi: 10.1016/S0140-6736(10)61217-2
- De Rienzo, F., Gabbouline, R. R., Menziani, M. C., De Benedetti, P. G., and Wade, R. C. (2001). Electrostatic analysis and Brownian dynamics simulation of the association of plastocyanin and cytochrome F. *Biophys. J.* 81, 3090–3104. doi: 10.1016/S0006-3495(01)75947-4
- Dickson, C. J., Madej, B. D., Skjevik, A. A., Betz, R. M., Teigen, K., Gould, I. R., et al. (2014). Lipid14: the amber lipid force field. *J. Chem. Theory Comput.* 10, 865–879. doi: 10.1021/ct4010307
- Dlugosz, M., Zielinski, P., and Trylska, J. (2011). Software news and updates brownian dynamics simulations on CPU and GPU with BD_BOX. *J. Comput. Chem.* 32, 2734–2744. doi: 10.1002/jcc.21847
- Edeson, R. O., Yeo, G. F., Milne, R. K., and Madsen, B. W. (1990). Graphs, random sums, and sojourn time distributions, with application to ion-channel modeling. *Math. Biosci.* 102, 75–104. doi: 10.1016/0025-5564(90)90056-5
- Elber, R. (2007). A milestone study of the kinetics of an allosteric transition: atomically detailed simulations of deoxy Scaapharca hemoglobin. *Biophys. J.* 92, L85–L87. doi: 10.1529/biophysj.106.101899
- Elber, R., and West, A. (2010). Atomically detailed simulation of the recovery stroke in myosin by Milestoning. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5001–5005. doi: 10.1073/pnas.0909636107
- Elcock, A. H. (2004). Molecular simulations of diffusion and association in multimacromolecular systems. *Numerical Computer Methods* 383(Pt D), 166–198. doi: 10.1016/S0076-6879(04)83008-8
- Elcock, A. H., Sept, D., and McCammon, J. A. (2001). Computer simulation of protein-protein interactions. *J. Phys. Chem. B* 105, 1504–1518. doi: 10.1021/jp003602d
- Ermak, D. L., and McCammon, J. A. (1978). Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys.* 69, 1352–1360. doi: 10.1063/1.436761
- Faradjian, A. K., and Elber, R. (2004). Computing time scales from reaction coordinates by milestoneing. *J. Chem. Phys.* 120, 10880–10889. doi: 10.1063/1.1738640
- Fogolari, F., Brigo, A., and Molinari, H. (2002). The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J. Mol. Recognit.* 15, 377–392. doi: 10.1002/jmr.577
- Gabbouline, R. R., and Wade, R. C. (1996). Effective charges for macromolecules in solvent. *J. Phys. Chem.* 100, 3868–3878. doi: 10.1021/jp953109f
- Gabbouline, R. R., and Wade, R. C. (1997). Simulation of the diffusional association of Barnase and Barstar. *Biophys. J.* 72, 1917–1929. doi: 10.1016/S0006-3495(97)78838-6
- Gabbouline, R. R., and Wade, R. C. (1998). Brownian dynamics simulation of protein-protein diffusional encounter. *Methods* 14, 329–341. doi: 10.1006/meth.1998.0588
- Gabbouline, R. R., and Wade, R. C. (2002). Biomolecular diffusional association. *Curr. Opin. Struct. Biol.* 12, 204–213. doi: 10.1016/S0959-440X(02)00311-1
- Geyer, T. (2011). Many-particle brownian and langevin dynamics simulations with the brownmove package. *BMC Biophys.* 4:7. doi: 10.1186/2046-1682-4-7
- Giugliano, M. (2000). Synthesis of generalized algorithms for the fast computation of synaptic conductances with Markov kinetic models in large network simulations. *Neural Comput.* 12, 903–931. doi: 10.1162/08997660030015646
- Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* 8, 1542–1555. doi: 10.1021/ct200909j
- Gu, J., and Bourne, P. E. (2009). *Structural Bioinformatics*. Hoboken, NJ: Wiley-Blackwell.
- Gurkiewicz, M., Korngreen, A., Waxman, S. G., and Lampert, A. (2011). Kinetic modeling of Nav1.7 provides insight into erythromelalgia-associated F1449V mutation. *J. Neurophysiol.* 105, 1546–1557. doi: 10.1152/jn.00703.2010
- Guvenc, O., and Mackerell, A. D. Jr. (2008). Comparison of protein force fields for molecular dynamics simulations. *Methods Mol. Biol.* 443, 63–88. doi: 10.1007/978-1-59745-177-2_4
- Hake, J., Kekenus-Huskey, P. M., and McCulloch, A. D. (2014). Computational modeling of subcellular transport and signaling. *Curr. Opin. Struct. Biol.* 25, 92–97. doi: 10.1016/j.sbi.2014.01.006
- Henzler-Wildman, K., and Kern, D. (2007). Dynamic personalities of proteins. *Nature* 450, 964–972. doi: 10.1038/nature06522
- Herberg, F. W., Taylor, S. S., and Dostmann, W. R. (1996). Active site mutations define the pathway for the cooperative activation of cAMP-dependent protein kinase. *Biochemistry* 35, 2934–2942. doi: 10.1021/bi951647c
- Holst, M. (2001). Adaptive numerical treatment of elliptic systems on manifolds. *Adv. Comput. Math.* 15, 139–191. doi: 10.1023/A:1014246117321
- Honig, B., and Nicholls, A. (1995). Classical electrostatics in biology and chemistry. *Science* 268, 1144–1149. doi: 10.1126/science.7761829
- Horvath, A., Bertherat, J., Groussin, L., Guillaud-Bataille, M., Tsang, K., Cazabat, L., et al. (2010). Mutations and polymorphisms in the gene encoding regulatory subunit Type 1-Alpha of Protein Kinase A (PRKAR1A): an update. *Hum. Mutat.* 31, 369–379. doi: 10.1002/humu.21178
- Huber, G. A., and McCammon, J. A. (2010). Browndye: a software package for Brownian dynamics. *Comput. Phys. Commun.* 181, 1896–1905. doi: 10.1016/j.cpc.2010.07.022
- Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001). Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* 105, 6474–6487. doi: 10.1021/jp003919d
- Karplus, M., and McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9, 646–652. doi: 10.1038/nsb0902-646
- Kerr, R. A., Bartol, T. M., Kaminsky, B., Dittrich, M., Chang, J. C. J., Baden, S. B., et al. (2008). Fast monte carlo simulation methods for biological reaction-diffusion systems in solution and on surfaces. *SIAM J. Sci. Comput.* 30, 3126–3149. doi: 10.1137/070692017
- Khavrutskii, I. V., Grant, B., Taylor, S. S., and McCammon, J. A. (2009). A transition path ensemble study reveals a linchpin role for Mg(2+) during rate-limiting

- ADP release from protein kinase A. *Biochemistry* 48, 11532–11545. doi: 10.1021/bi901475g
- Kim, C., Cheng, C. Y., Saldanha, S. A., and Taylor, S. S. (2007). PKA-I holoenzyme structure reveals a mechanism for cAMP-dependent activation. *Cell* 130, 1032–1043. doi: 10.1016/j.cell.2007.07.018
- Kirchner, F., Schuetz, A., Boldt, L. H., Martens, K., Dittmar, G., Haverkamp, W., et al. (2012). Molecular insights into arrhythmogenic right ventricular cardiomyopathy caused by plakophilin-2 missense mutations. *Circ. Cardiovasc. Genet.* 5, 400–411. doi: 10.1161/CIRCGENETICS.111.961854
- Kirmizialtin, S., and Elber, R. (2011). Revisiting and computing reaction coordinates with Directional Milestoning. *J. Phys. Chem. A* 115, 6137–6148. doi: 10.1021/jp111093c
- Klauda, J. B., Venable, R. M., Freites, J. A., O'Connor, J. W., Tobias, D. J., Mondragon-Ramirez, C., et al. (2010). Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* 114, 7830–7843. doi: 10.1021/jp101759q
- Kohlhoff, K. J., Shukla, D., Lawrenz, M., Bowman, G. R., Konerding, D. E., Belov, D., et al. (2014). Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nat. Chem.* 6, 15–21. doi: 10.1038/nchem.1821
- Koukos, P. I., and Glykos, N. M. (2014). Folding molecular dynamics simulations accurately predict the effect of mutations on the stability and structure of a vaminin-derived peptide. *J. Phys. Chem. B* 118, 10076–10084. doi: 10.1021/jp5046113
- Kozack, R. E., and Subramaniam, S. (1993). Brownian dynamics simulations of molecular recognition in an antibody antigen system. *Protein Sci.* 2, 915–926. doi: 10.1002/pro.5560020605
- Lampert, A., and Korngreen, A. (2014). Markov modeling of ion channels: implications for understanding disease. *Prog. Mol. Biol. Transl. Sci.* 123, 1–21. doi: 10.1016/B978-0-12-397897-4.00009-7
- Leach, A. R. (2001). *Molecular Modelling: Principles and Applications*. Harlow, New York, NY: Prentice Hall.
- Madura, J. D., Briggs, J. M., Wade, R. C., and Gabbouline, R. R. (2002). Brownian dynamics. *Encyclopedia Comput. Chem.* doi: 10.1002/0470845015.cba003. [Epub ahead of print].
- Májek, P., and Elber, R. (2010). Milestoning without a reaction coordinate. *J. Chem. Theory Comput.* 6, 1805–1817. doi: 10.1021/ct100114j
- Malmstrom, R. D., Kornev, A. P., Taylor, S. S., and Amaro, R. E. (2015). Allosteric through the computational microscope: cAMP activation of a canonical signaling domain. *Nat. Commun.* 6:7588. doi: 10.1038/ncomms8588
- Malmstrom, R. D., Lee, C. T., Van Wart, A. T., and Amaro, R. E. (2014). Application of molecular-dynamics based Markov state models to functional proteins. *J. Chem. Theory Comput.* 10, 2648–2657. doi: 10.1021/ct5002363
- Marsden, A. L., Feinstein, J. A., and Taylor, C. A. (2008). A computational framework for derivative-free optimization of cardiovascular geometries. *Comput. Methods Appl. Mech. Eng.* 197, 1890–1905. doi: 10.1016/j.cma.2007.12.009
- Marsh, J. A., Teichmann, S. A., and Forman-Kay, J. D. (2012). Probing the diverse landscape of protein flexibility and binding. *Curr. Opin. Struct. Biol.* 22, 643–650. doi: 10.1016/j.sbi.2012.08.008
- McGuffee, S. R., and Elcock, A. H. (2010). Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput. Biol.* 6:e1000694. doi: 10.1371/journal.pcbi.1000694
- Moraru, I. I., Schaff, J. C., Slepchenko, B. M., Blinov, M. L., Morgan, F., Lakshminarayana, A., et al. (2008). Virtual Cell modelling and simulation software environment. *IET Syst. Biol.* 2, 352–362. doi: 10.1049/iet-syb:20080102
- Motlagh, H. N., Wrabl, J. O., Li, J., and Hilser, V. J. (2014). The ensemble nature of allostery. *Nature* 508, 331–339. doi: 10.1038/nature13001
- Mugnai, M. L., and Elber, R. (2015). Extracting the diffusion tensor from molecular dynamics simulation with Milestoning. *J. Chem. Phys.* 142, 014105. doi: 10.1063/1.4904882
- Neitzel, J. J., Dostmann, W. R., and Taylor, S. S. (1991). Role of MgATP in the activation and reassociation of cAMP-dependent protein kinase I: consequences of replacing the essential arginine in cAMP binding site A. *Biochemistry* 30, 733–739. doi: 10.1021/bi00217a023
- Oostenbrink, C., Villa, A., Mark, A. E., and Van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* 25, 1656–1676. doi: 10.1002/jcc.20090
- Pande, V. S., Baker, I., Chapman, J., Elmer, S. P., Khaliq, S., Larson, S. M., et al. (2003). Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers* 68, 91–109. doi: 10.1002/bip.10219
- Pande, V. S., Beauchamp, K., and Bowman, G. R. (2010). Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* 52, 99–105. doi: 10.1016/j.ymeth.2010.06.002
- Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Debolt, S., et al. (1995). Amber, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* 91, 1–41. doi: 10.1016/0010-4655(95)00041-D
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802. doi: 10.1002/jcc.20289
- Pierce, L. C. T., Salomon-Ferrer, R., de Oliveira, C. A. F., Mccammon, J. A., and Walker, R. C. (2012). Routine access to millisecond time scale events with accelerated molecular dynamics. *J. Chem. Theory Comput.* 8, 2997–3002. doi: 10.1021/ct300284c
- Ponder, J. W., and Case, D. A. (2003). Force fields for protein simulations. *Protein Simul.* 66, 27. doi: 10.1016/S0065-3233(03)66002-X
- Prinz, J. H., Keller, B., and Noe, F. (2011a). Probing molecular kinetics with Markov models: metastable states, transition pathways and spectroscopic observables. *Phys. Chem. Chem. Phys.* 13, 16912–16927. doi: 10.1039/c1cp21258c
- Prinz, J. H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., et al. (2011b). Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.* 134:174105. doi: 10.1063/1.3565032
- Qin, F., Auerbach, A., and Sachs, F. (1996). Estimating single-channel kinetic parameters from idealized patch-clamp data containing missed events. *Biophys. J.* 70, 264–280. doi: 10.1016/S0006-3495(96)79568-1
- Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A., and Honig, B. (2002). Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J. Comput. Chem.* 23, 128–137. doi: 10.1002/jcc.1161
- Rudy, Y., and Silva, J. R. (2006). Computational biology in the study of cardiac ion channels and cell electrophysiology. *Q. Rev. Biophys.* 39, 57–116. doi: 10.1017/S0033583506004227
- Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. *J. Chem. Theory Comput.* 9, 3878–3888. doi: 10.1021/ct400314y
- Saucerman, J. J., Brunton, L. L., Michailova, A. P., and Mcculloch, A. D. (2003). Modeling beta-adrenergic control of cardiac myocyte contractility *in silico*. *J. Biol. Chem.* 278, 47997–48003. doi: 10.1074/jbc.M308362200
- Schöneberg, J., and Noé, F. (2013). ReaDDy - a software for particle-based reaction-diffusion dynamics in crowded cellular environments. *PLoS ONE* 8:e74261. doi: 10.1371/journal.pone.0074261
- Scott, J. D., and Santana, L. F. (2010). A-Kinase anchoring proteins getting to the heart of the matter. *Circulation* 121, 1264–1271. doi: 10.1161/CIRCULATIONAHA.109.896357
- Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., et al. (2010). Atomic-level characterization of the structural dynamics of proteins. *Science* 330, 341–346. doi: 10.1126/science.1187409
- Sjoberg, T. J., Kornev, A. P., and Taylor, S. S. (2010). Dissecting the cAMP-inducible allosteric switch in protein kinase A R1alpha. *Protein Sci.* 19, 1213–1221. doi: 10.1002/pro.400
- Su, Y., Dostmann, W. R. G., Herberg, F. W., Durick, K., Xuong, N. H., Teneyck, L., et al. (1995). Regulatory Subunit of Protein-Kinase-a - structure of deletion mutant with camp Binding Domains. *Science* 269, 807–813. doi: 10.1126/science.7638597
- Teilmann, K., Olsen, J. G., and Kragelund, B. B. (2011). Protein stability, flexibility and function. *Biochim. Biophys. Acta* 1814, 969–976. doi: 10.1016/j.bbapap.2010.11.005

- Tsai, C. J., Kumar, S., Ma, B., and Nussinov, R. (1999). Folding funnels, binding funnels, and protein function. *Protein Sci.* 8, 1181–1190. doi: 10.1110/ps.8.6.1181
- Vanden-Eijnden, E., and Tal, F. A. (2005). Transition state theory: variational formulation, dynamical corrections, and error estimates. *J. Chem. Phys.* 123:184103. doi: 10.1063/1.2102898
- Vanden-Eijnden, E., Venturoli, M., Ciccotti, G., and Elber, R. (2008). On the assumptions underlying milestoning. *J. Chem. Phys.* 129:174102. doi: 10.1063/1.2996509
- Vitalini, F., Mey, A. S. J. S., Noe, F., and Keller, B. G. (2015). Dynamic properties of force fields. *J. Chem. Phys.* 142:084101. doi: 10.1063/1.4909549
- Votapka, L. W., and Amaro, R. E. (2015). Multiscale estimation of binding kinetics using molecular dynamics, brownian dynamics, and milestoning. *J. Biomol. Struct. Dyn.* 33, 26–27. doi: 10.1080/07391102.2015.1032587
- Votapka, L. W., Czaplá, L., Zhenirovskyy, M., and Amaro, R. E. (2013). DelEnsembleElec: computing ensemble-averaged electrostatics using DelPhi. *Commun. Comput. Phys.* 13, 256–268. doi: 10.4208/Cicp.170711.111111s
- Wang, W., Donini, O., Reyes, C. M., and Kollman, P. A. (2001). Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* 30, 211–243. doi: 10.1146/annurev.biophys.30.1.211
- Wang, Y. F., Khan, M., and van den berg, H. A. (2012). Interaction of fast and slow dynamics in endocrine control systems with an application to beta-cell dynamics. *Math. Biosci.* 235, 8–18. doi: 10.1016/j.mbs.2011.10.003
- West, A. M., Elber, R., and Shalloway, D. (2007). Extending molecular dynamics time scales with milestoning: example of complex kinetics in a solvated peptide. *J. Chem. Phys.* 126, 145104. doi: 10.1063/1.2716389
- Yang, J. H., and Saucerman, J. J. (2012). Phospholemman is a negative feed-forward regulator of Ca²⁺ in beta-adrenergic signaling, accelerating beta-adrenergic inotropy. *J. Mol. Cell. Cardiol.* 52, 1048–1055. doi: 10.1016/j.yjmcc.2011.12.015
- Zhou, H. X., and Bates, P. A. (2013). Modeling protein association mechanisms and kinetics. *Curr. Opin. Struct. Biol.* 23, 887–893. doi: 10.1016/j.sbi.2013.06.014

Conflict of Interest Statement: Andrew D. McCulloch is required to disclose that he is a co-founder and equity holder of Insilicomed, Inc., a licensee of software developed by his laboratory. This relationship has been disclosed to and is managed by the University of California, San Diego. The software was not used in this research, and the company has had no involvement in this research or any interest in its outcome. The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Boras, Hirakis, Votapka, Malmstrom, Amaro and McCulloch. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

1.8 Acknowledgments

Chapter 1, in full, is a reprint of the material as it appears in *Frontiers in Physiology* 2015. Boras, Britton W.; **Hirakis, Sophia P.**; Votapka, Lane W.; Malmstrom, Robert D.; Amaro, Rommie E.; McCulloch, Andrew D.;“Bridging scales through multiscale modeling: a case study on protein kinase A,” *Frontiers in Physiology*, 6: 250, 2015. The dissertation author was the co-primary author of this paper.

This work was the result of a direct collaboration with wonderful co-authors whom I would like to formally acknowledge. Dr. Britton W. Boras, is the co-first author who wrote a portion of the paper and created figures used in the manuscript. Dr. Lane W. Votapka and Dr. Robert D. Malmstrom also wrote sections of the paper and helped to construct figures. Dr. Rommie E. Amaro and Dr. Andrew D. McCulloch provided support for the manuscript through editing and revision of the text. I extend my gratitude to the above-mentioned scientists for their creative and scientific contributions to the manuscript.

Chapter 2

Broad recognition in the interaction of group A *Streptococcus* M protein hypervariable regions with human C4b-binding protein

The following manuscript demonstrates the ways that protein-protein contacts involved in bacterial infectious mechanisms can be elucidated by protein crystallography and further understood using MD simulations. The subject of the manuscript, Group A *Streptococcus* (GAS), is incredibly infectious with a wide range of symptoms ranging from sore throat to flesh-eating bacteria that affects the skin and internal organs such as the heart. Development of a vaccine treatment for GAS has been slow due to the large variance in sequence of its surface antigen, the M protein. Over 200 different types of M proteins have been identified and antibodies recognize a hypervariable region (HVR) of the M protein, offering narrow specificity due to sequence variability. The infectious mechanism of the protein is quite remarkable, as the Human C4b-binding protein (C4BP) interacts with around 90% of M proteins, making it a great candidate for understanding the broad specificity of the HVR. In this study, crystal structures of C4BP in complex with four different M-proteins reveal the uniform and sequence-variant “tolerant” reading head of the protein-protein complex.

In context of the thesis, this manuscript demonstrates how observations made by crystallographic methods are enhanced by molecular simulation. The MD study in this

manuscript allowed us to visualize and understanding of the atomistic contacts which could not be explained by the static structure itself. A number of mutants of the WT, which strengthened and weakened the contacts of the M protein-C4BP complex. Most notably, the MD showed how water is able to penetrate the protein complex due to a point mutation of a hydrophobic residue and how charged residue mutations counter-intuitively strengthen the interactions between C4BP and M protein complexes.

2.1 Abstract

Group A *Streptococcus* (GAS) is a leading cause of worldwide morbidity and mortality. A severe hurdle for the development of a GAS vaccine is the hypervariability of its major antigen, the M protein, with over 200 different M types in existence. Neutralizing antibodies typically recognize the hypervariable region (HVR) of M proteins, and thus confer narrow specificity directed against a single M type. In stark contrast, human C4b-binding protein (C4BP) interacts with a remarkably large number of M protein HVRs (~89%). Broad recognition such as this is rare, and we investigated it through structure determination of four sequence-diverse M proteins in complex with C4BP. The structures revealed a uniform and tolerant reading head in C4BP, which detected conserved sequence patterns hidden within hypervariability. Our work explains not only how sequence-divergent M proteins capture C4BP, but also has implications for the design of a unique broadly neutralizing antibody targeting GAS.

2.2 Introduction

Group A *Streptococcus* (GAS, *S. pyogenes*) is a major cause of worldwide morbidity and mortality [63]. This bacterial pathogen is responsible for mucosal infections (e.g. pharyngitis), acute invasive diseases (e.g. necrotizing fasciitis) and autoimmune sequelae

(e.g. rheumatic heart disease) [64]. Currently, no vaccine against GAS exists [65,66]. A major impediment to immunization is the hypervariability of the antigenic M protein, a surface-anchored virulence factor [67,68] that is also the target of neutralizing antibodies. These antibodies typically recognize the hypervariable region (HVR, N-terminal ~50 amino acids) [69–71] of M proteins, which are dimeric α -helical coiled coils, and thus confer M type-specific immunity. One approach to overcoming hypervariability is to include multiple M protein HVRs in a vaccine, and indeed a vaccine candidate that includes 30 HVRs [72] has advanced into early clinical testing. However, with >200 distinct M protein HVRs [73] and the complexity of global GAS epidemiology [63], even the most extensive multivalent vaccine is unlikely to offer universal protection. Here, we offer structural details that inform an alternative approach to overcoming M protein hypervariability for vaccine design. This approach is based on the finding that human C4b-binding protein (C4BP) binds M protein HVRs with broad specificity [74], in stark contrast to the narrow type-specificity displayed by antibodies. In one study, a remarkable ~89% of GAS strains of differing M types bound C4BP [75]. C4BP [76] is a negative regulator of the complement system and binds the complement protein C4b, and thereby disables the C3 convertase of the classical and lectin pathways. GAS recruits C4BP to its surface, like a number of others pathogens [77,78], to evade opsonophagocytic killing [79,80].

Broad specificity in recognition is a rare phenomenon, having been observed only in a few cases. A prominent one is the interaction between major histocompatibility complex (MHC) glycoproteins and peptides [81,82]. The breadth of this interaction is explained by the fact that MHC glycoproteins contact mainly the main chain of peptides. To understand the basis for broad specificity in the case of M protein and C4BP, co-crystal structures of four M protein HVRs (M2, M22, M28, and M49) bound to the first two domains of the C4BP α chain were determined (Figs. 1a and 5, Table 1). C4BP consists of 7 α chains disulfide-bonded to a single β chain, with each of these chains being composed of multiple ~60-residue complement control protein (CCP) domains [83]. The first two CCP domains

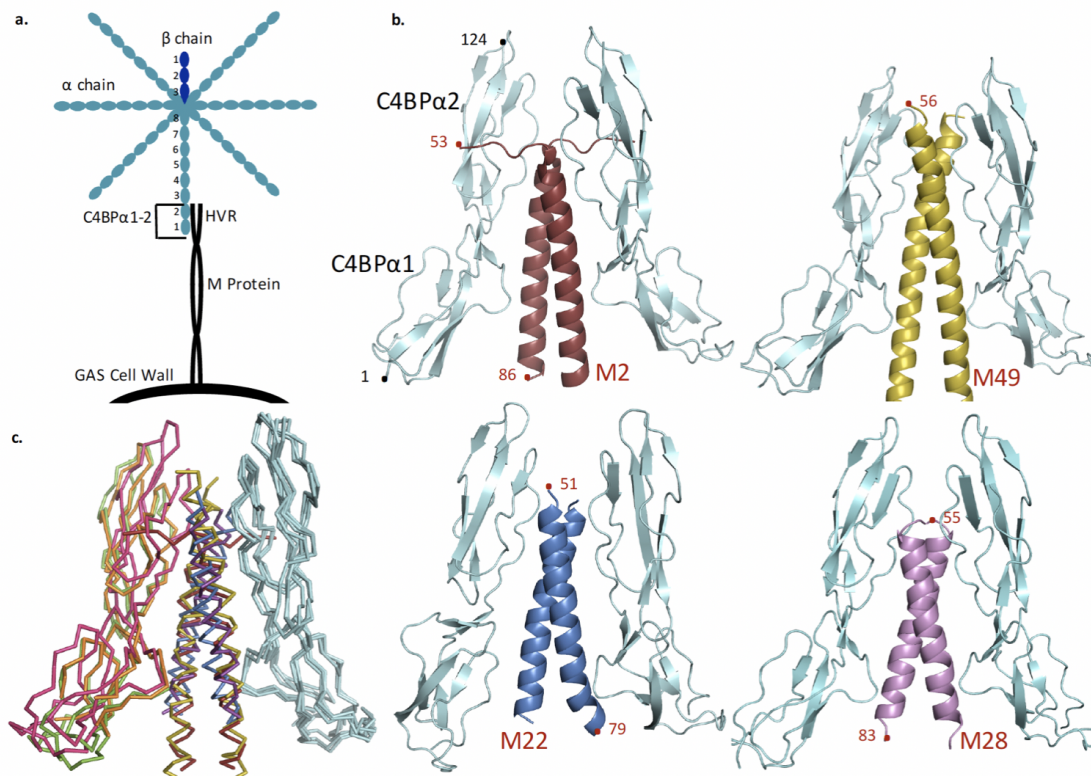


Figure 2.1. Structures of M-C4BP complexes

(a) Schematic of C4BP (blue) bound to surface-associated M protein (black), highlighting the M HVR-C4BP α 1-2 interaction; (b) C4BP α 1-2 (cyan) in complex with M2 (red), M49 (yellow), M22 (blue), and M28 (magenta). Terminal residues are numbered; (c) Superposition of M-C4BP complexes, based on the C4BP α 1-2 molecule shown at right in cyan. M2 is red and its second bound C4BP α 1-2 molecule green; M49 is yellow and its second bound C4BP α 1-2 orange; M22 is blue and its second bound C4BP α 1-2 is omitted (because a crystal contact restricts its orientation into an artifactual conformation); and M28 is magenta and its second bound C4BP α 1-2 pink.

of the α chain (C4BP α 1-2) are sufficient to bind M protein HVRs [84] and C4b [84, 85] (Fig. 2.1a). Overlapping but non-identical sites on C4BP are engaged by M protein HVRs and C4b [85].

2.3 Structural similarity

The structures of the four M protein-C4BP complexes (determined between 2.54-3.02 angstrom resolution limits) were astonishingly similar, given the lack of sequence relationship among the M proteins (fig. 2.5). The M protein HVRs form parallel, dimeric

α -helical coiled coils, with two C4BP α 1-2 molecules bound to each M protein dimer, as prior reports suggested [83, 86] (Figs. 1b-c; detailed view of M2 shown later, and detailed views of M22, M28, and M49 are in Figs. 6-8). The portions of the M proteins that contact C4BP α 1-2 are in canonical coiled-coil conformation, except for M2, which is underwound (fig. 2.19). C4BP α 1 is proximal to the C-terminal portion of the M protein HVR and C4BP α 2 to the N-terminal portion, in agreement with the approach of intact C4BP to the streptococcal surface (Fig 1a). The C4BP α 1 and α 2 domains are relatively unchanged from their unbound NMR structures [83] (average RMSD \sim 1.5 and \sim 1.0 angstrom for domains 1 and 2, respectively), except that domain 1 is rotated 180 $^\circ$ with respect to domain 2 (Figs. 10 and 11). This rotation is consistent with evidence from mutagenic [85] and structural [83] studies, and is discussed further below. The M-C4BP interface is extensive, with a total of \sim 1450-1690 angstrom² of surface area being buried (in the 2:2 complex). Most of this surface area is polar, and the fit is far from hand-in-glove (surface complementarities 0.56-0.66) [87], except for M22 which has a better fit (0.72). These observations suggest a modest binding affinity, consistent with the 0.5 μ M K_d [83] for the interaction between C4BP α 1-2 and the M4 HVR. A much tighter association of picomolar K_d [88] results from avidity between multi-armed C4BP and surface-localized M protein.

2.4 Uniform reading head

Most significantly, the four structures revealed a uniform set of amino acids in C4BP that act as a reading head for recognizing M protein HVRs. Most of this reading head resides in C4BP α 2 (fig. 2.2a) and takes the form of a quadrilateral that is composed of: (1) a hydrophobic pocket that contains C4BP H67, I78, and L82; (2) a hydrogen bonding group in the form of the main chain nitrogen of C4BP H67; and two positively charged residues, C4BP (3) R64 and (4) R66. The segment that holds this quadrilateral is structurally invariant, being stabilized by a disulfide bond at C65 and limited in conformation by

P68 (not depicted). The M proteins supply amino acid side chains that interact with these C4BP residues and form complementary quadrilaterals (fig. 2.2b). In all four M-C4BP structures, a hydrophobic M protein residue (usually an aromatic) fits into the (1) hydrophobic pocket, and a polar M protein residue immediately following in sequence hydrogen bonds to the (2) main chain nitrogen of H67. The contacts to C4BP (3) R64 and (4) R66 are predominantly electrostatic (usually salt bridges), but in the case of M49, a polar residue is absent and instead R64 makes hydrophobic contacts, extending its alkyl chains across several M49 residues. Substitution of C4BP residues R64, R66, or H67 with Gln affects binding to M4 and M22 [85]. Decreased affinity results in the case of R64Q and H67Q, but increased affinity occurs for R66Q (likely a gain-of-function). Uniform reading head contacts from C4BP α 1 were far fewer. The key C4BP α 1 residue was R39, which forms electrostatic contacts through its guanidinium group as well as hydrophobic contacts through its alkyl chain, creating a hydrophobic nook in conjunction with main chain atoms of C4BP α 1 (Fig 2c). Thus, out of the six C4BP residues that form uniform contacts, three are arginines. This high proportion is likely significant, as the combination of polar and apolar atoms in Arg along with its chain length increase the possibilities for interactions with variable residues. Substitution of C4BP R39 with Gln results in decreased binding to M4 but increased binding to M22 [85] (again, likely a gain-of-function). All four M proteins have hydrophobic residues that insert into the C4BP α 1 hydrophobic nook. M2 and M49 also have negatively charged residues that interact with C4BP R39, whereas neither M22 nor M28 do. The importance of C4BP R39 provided an explanation for the aforementioned 180° rotation of C4BP α 1 (around a hinge at K63, fig. 2.10). In free C4BP, the C4BP α 1 R39 nook and the C4BP α 2 quadrilateral are on opposite sides, and require a 180° reorientation to interact simultaneously with M protein. This 180° rotation was seen in all four structures. However, in one of the two C4BP α 1-2 molecules bound to M22, the 180° rotation was prevented due to a crystal contact (Figs. 10c, d and 11). A similar 180° rotation appears to be necessary for the interaction of C4BP with C4b [85]. The

purpose of C4BP having different free and bound conformations is unclear.

2.5 Sequence conservation hidden within hypervariability

The evidence gathered from these structures proved powerful in bringing to light weak sequence conservation in M protein HVRs. Comparison of the heptad position of M protein residues that interacted with C4BP made it clear that there were two binding patterns observed in these structures (Fig. 2.3a): M2 and M49 belonged to one, and M22 and M28 to a second one. These two patterns were also evident in the way the coiled coils interacted with C4BP. In the case of M2 and M49, the coiled coils ran roughly parallel to C4BP α 1-2, such that each α -helix contacted a single C4BP α 1-2 molecule (Figs. 1b and 2b). But in the case of M22 and M28, the coiled coils lay crosswise across C4BP α 1-2 such that both α -helices contacted a single C4BP α 1-2 molecule. Remarkably, these same two patterns were evident in a larger number of M proteins (Fig. 2.3b, 12, and 13). We were able to assign 13 M proteins to the M2/M49 pattern and 32 to the M22/M28 pattern, suggesting that these two patterns explained the interaction of nearly half of the M strains that were studied for C4BP binding [74]. A further 46 M proteins from this study [74] could not be assigned to either pattern (Fig. 2.14), making it likely that there are still other arrangements by which M proteins interact with C4BP.

2.6 Tolerance to hypervariability

We next sought to understand tolerance in C4BP to sequence variation in M protein, as single amino acid changes have been shown to alter recognition by antibodies but not C4BP [74]. Alanine substitutions were created in the M2 residues mentioned above that make contact with the uniform reading head. In addition, substitutions were made in two M2 residues, K65 and E83, that make contacts observed only for M2 (Fig. 2.4). The

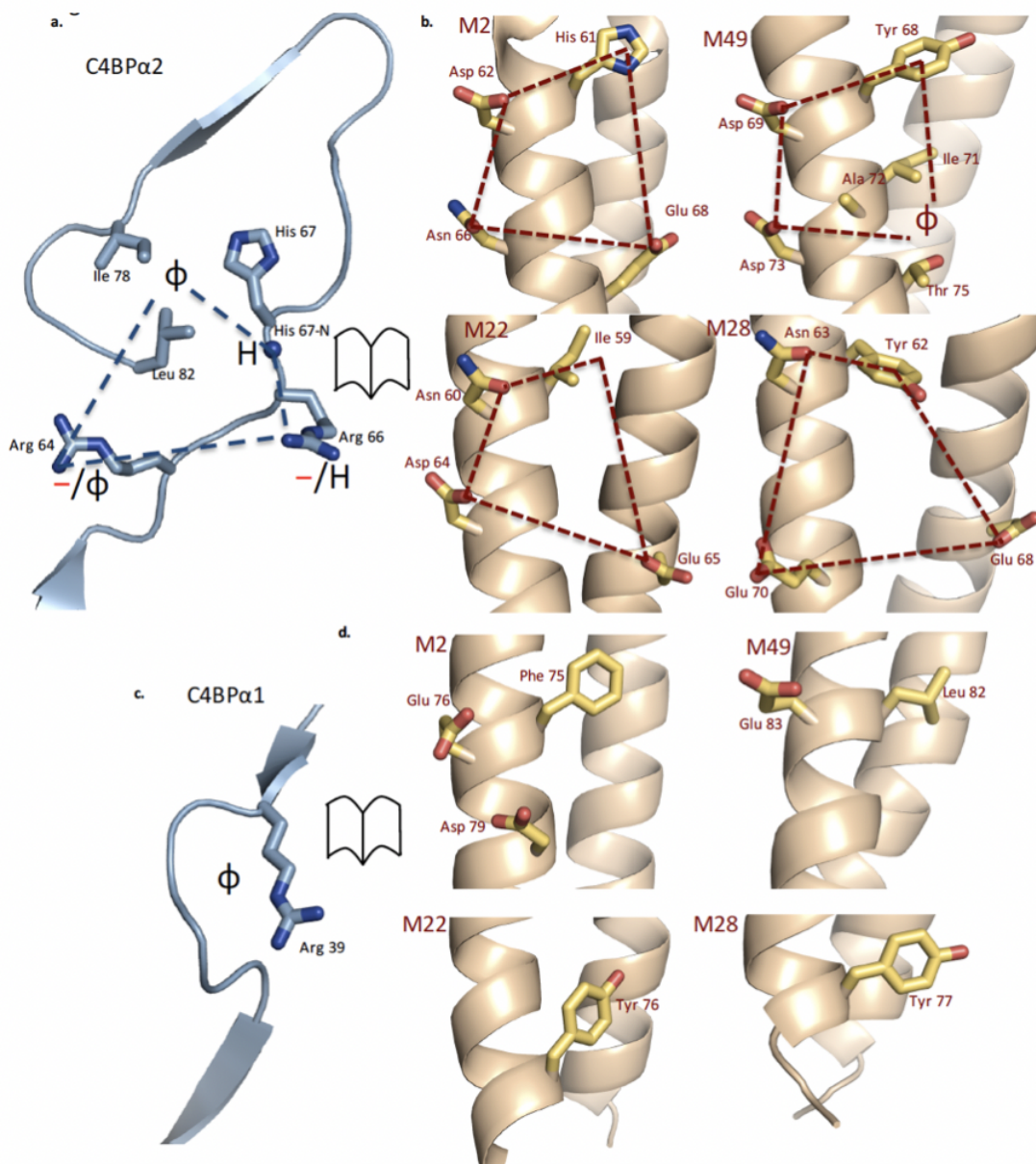


Figure 2.2. C4BP Binding Mode

(a) The C4BP α 2 quadrilateral (blue dashed lines), with the C4BP α 2 backbone shown in ribbon representation and key side chains shown as bonds (here and in following panels). The chemical character of M protein residues that interact with the quadrilateral is depicted: ϕ , hydrophobic; -, negative; H, hydrogen bond forming; (b) M2, M49, M22, and M28 residues that interact with the C4BP α 2 quadrilateral, shown in open-book representation with respect to C4BP α 2. The M protein residues form the complementary quadrilateral (red dashed lines); (c) The C4BP α 1 Arg39 nook. Symbols are as described for the first panel. (d) M2, M49, M22, and M28 residues that interact with the C4BP α 1 Arg39 nook shown in open-book representation.

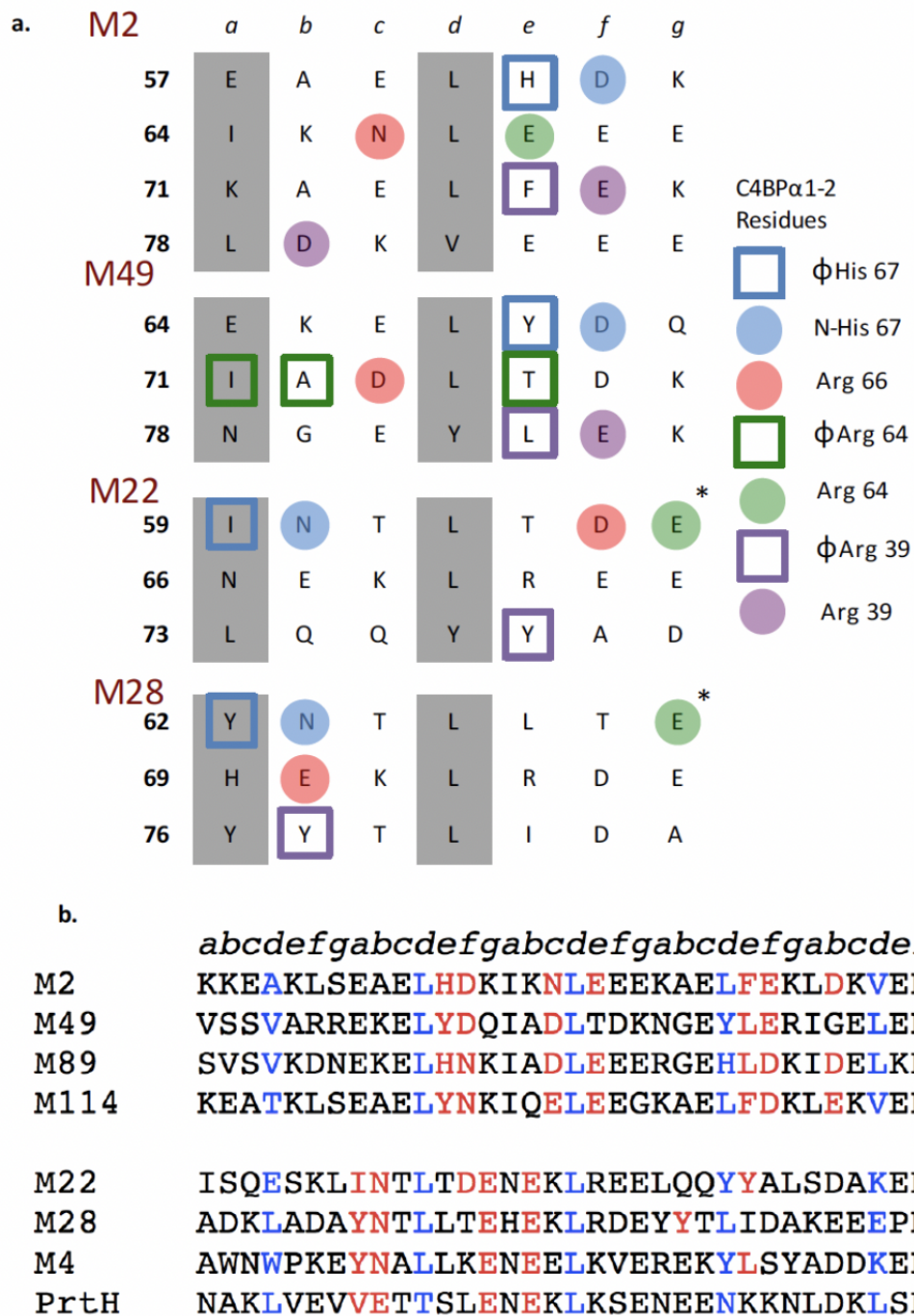


Figure 2.3. C4BP-binding modes of M proteins

(a) Heptad register of M2, M49 and M28 HVRs (a and d residues in grey). M protein residues interacting with C4BP α 1-2 residues are highlighted according to their corresponding C4BP α 1-2 interaction ;(b) Sequence alignment of M protein HVRs that belong to the M2/M49 group (top) or the M28/M22 group (bottom). Residues that contact or are predicted to contact C4BP α 1-2 are in red. Protein H (Pr \dagger H) is an M-like protein expressed by certain M1 strains

M2-C4BP interaction was evaluated by a Ni²⁺-NTA agarose coprecipitation assay using His-tagged C4BP α 1-2 (Figs. 4a and 15). Of the single-site substitutions, only F75A, which binds in the C4BP α 1 nook, showed decreased binding. Strikingly, all other residues could be mutated to Ala without obvious loss in binding, and indeed in two cases, increased binding was observed (see below). Providing verification for the structural observations, two double substitutions resulted in loss of binding: D62A/E68A, which removed two of the polar contacts to the C4BP α 2 quadrilateral, and E76A/D79A, which removed the two salt bridges to C4BP α 1 R39.

Surprisingly, two of the single-site mutations, M2 K65A and N66A, increased binding, as did the K65A/N66A double mutant (Fig 4a). The structure of M2 (K65A/N66A) in complex with C4BP α 1-2 was determined to 2.29 angstrom resolution limit, and no re-ordering of the binding site was evident (RMSD 0.15 angstrom) (Fig. 2.16). This result suggested that M2 K65 and N66 were tolerated in the binding site but not optimal. While M2 K65 formed a hydrogen bond to the main chain oxygen of C4BP R64, it was sandwiched between two positively charged side chains (i.e., C4BP R64 and R66), providing an explanation for why Ala substitution of K65 led to enhanced binding. Molecular dynamics (MD) simulations reinforced this interpretation, as Ala substitution of K65 led to better contacts between M2 and C4BP, especially evident in the increased frequency of hydrogen bonding between C4BP R66 and M2 N66 (Table 2.2). The simulations suggested that the hydrogen bond between these two residues was otherwise infrequent (Videos 2.1 and 2.2, Table 2.2), and indeed these two residues had the highest B-factors in the binding site (Fig. 2.17). In other M proteins belonging to the M2/M49 pattern, the equivalent of N66 is almost always Asp or Glu (Figs. 12 and 13). Consistent with this notion, substitution of M2 N66 with Asp resulted in increased binding (Fig 4a), and MD simulations provided evidence of the favorable interactions between C4BP R66 and negatively charged M protein amino acids (Video 2.3 and Table 2.3). Puzzlingly, substitution of M2 N66 with Ala (and thus loss of hydrogen bonding to C4BP R66) also

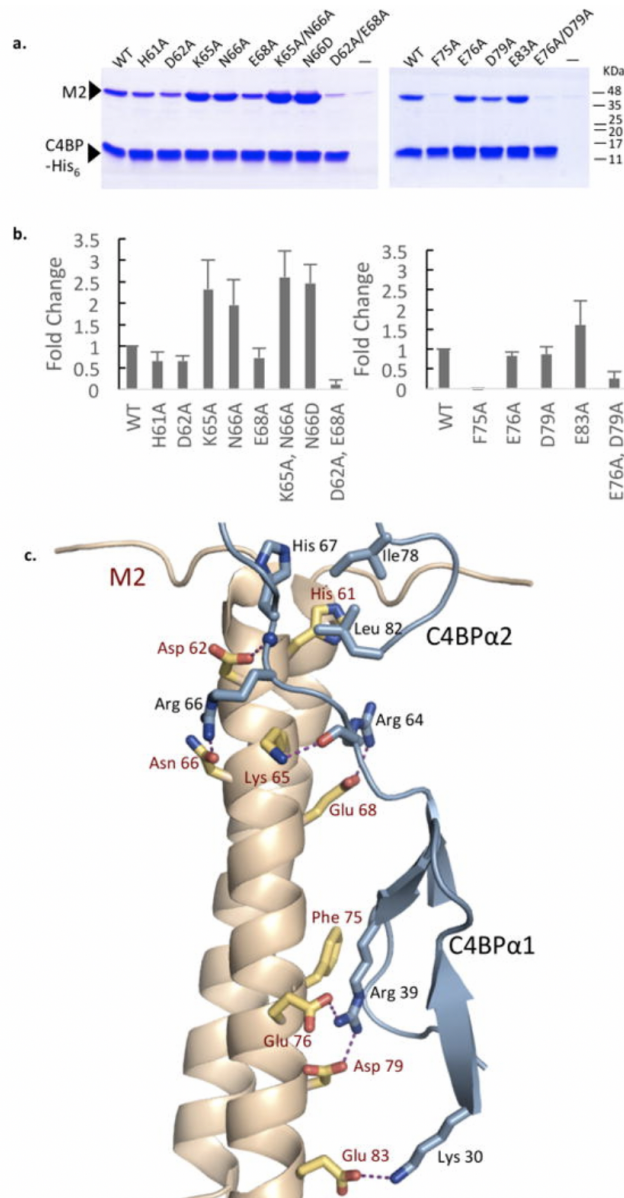


Figure 2.4. M2-C4BP interaction

(a) Association of His-tagged C4BP α 1-2 with wild-type and mutant M2 HVR at 37°C, as assessed by a Ni²⁺-NTA agarose coprecipitation assay and visualized by non-reducing, Coomassie-stained SDS-PAGE. Only bound fractions are shown here. Input samples shown in Fig. 2.16. This gel is representative of four experimental replicates. Molecular mass markers were not run on these particular gels; their positions are based on measurements from equivalent gels; (b) Quantification of the interaction between C4BP α 1-2 and wild-type M2 or M2 mutants proteins. The values shown are averages of four experimental replicates, corrected for the level of background binding (i.e., no C4BP α 1-2) and normalized to wild-type M2. Standard deviations are depicted; (c) Structure of M2 (gray ribbon representation with key side chains in bonds representation, in which carbons are yellow, oxygens red, and nitrogens blue) bound to C4BP α 1-2 (cyan ribbon representation, with key side chains in bonds representation, in which carbons are cyan and nitrogens blue). Hydrogen bonds and salt bridges depicted by dashed magenta lines.

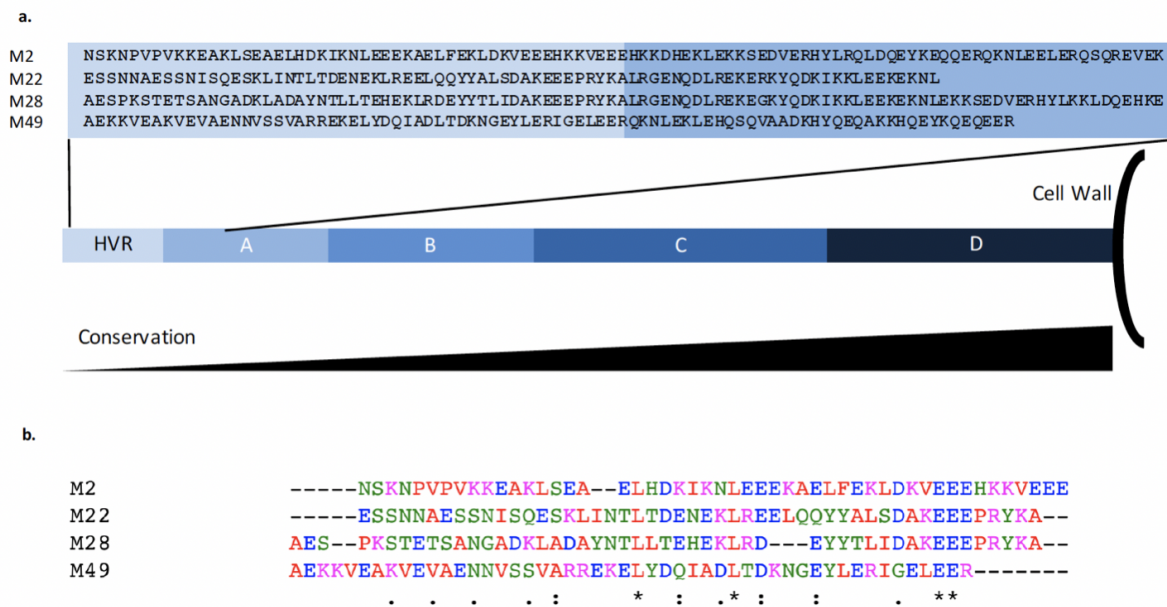


Figure 2.5. Schematic of M protein domains

The sequences of M2, M22, M28, and M49 co-crystallized with C4BP α 1-2 are depicted, with the HVR in light blue and other regions of the protein in darker blue or black. M proteins are hypervariable at their N-termini, with conservation increasing towards their C-termini; (b) Multiple sequence alignment of the M2, M22, M28, and M49 HVRs is shown, as carried out using MUSCLE. The asterisks indicate identical amino acids, the colons indicate amino acids with strongly similar properties, and the periods indicate amino acids with weakly similar properties.

resulted in better binding. It is worth noting that C4BP R66 had an even higher relative B-factor when contacting M2 (K65A/N66A) as compared to wild-type M2 (Fig. 2.17). Thus, it appears that C4BP R66 prefers a salt bridge (e.g., N66D) or no interaction (e.g., N66A) to a hydrogen bond, because the salt bridge provides sufficient binding energy to relieve the entropic cost of ordering the Arg, whereas the hydrogen bond does not. In short, the mutagenesis experiments reinforced the notion that the reading head in C4BP is highly tolerant to variation in the M protein.

Video Legends

Video 2.1 R39 nook in M2-C4BP α 1-2.

Portion of the molecular dynamics simulation of the M2-C4BP α 1-2 complex focusing on the R39 hydrophobic nook. M2 is in gold ribbon representation, and C4BP in cyan ribbon representation.

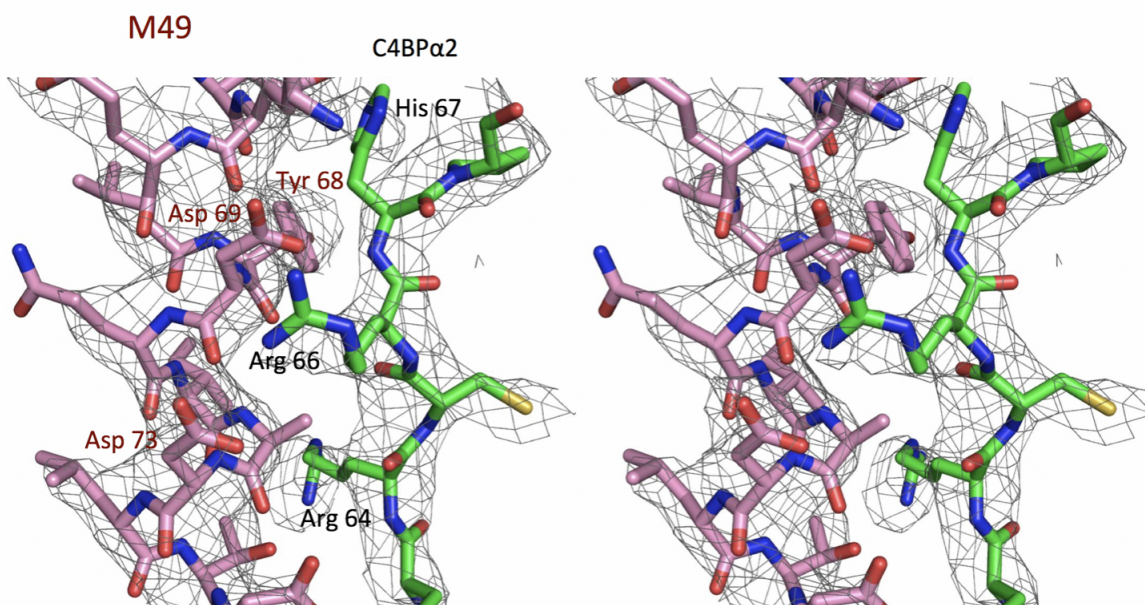


Figure 2.6. Electron density for the M49 HVR-C4BP α 1-2 complex

(a) Electron density from a simulated annealing $2F_o-F_c$ omit map (contoured at 1σ) for the M49 HVR-C4BP α 1-2 structure. Residues M49 68-75 and C4BP α 1-2 64-67 and 77-82 were excluded along with a 3.5 angstrom sphere around these residues, prior to performing simulated annealing (2500 K) and phase calculation. The final model is overlaid in bonds representation. M49 is depicted with pink carbons, and C4BP with green carbons; nitrogens and oxygens are blue and red, respectively, for both. The numbering of M proteins is such that the initiator Met is residue 1

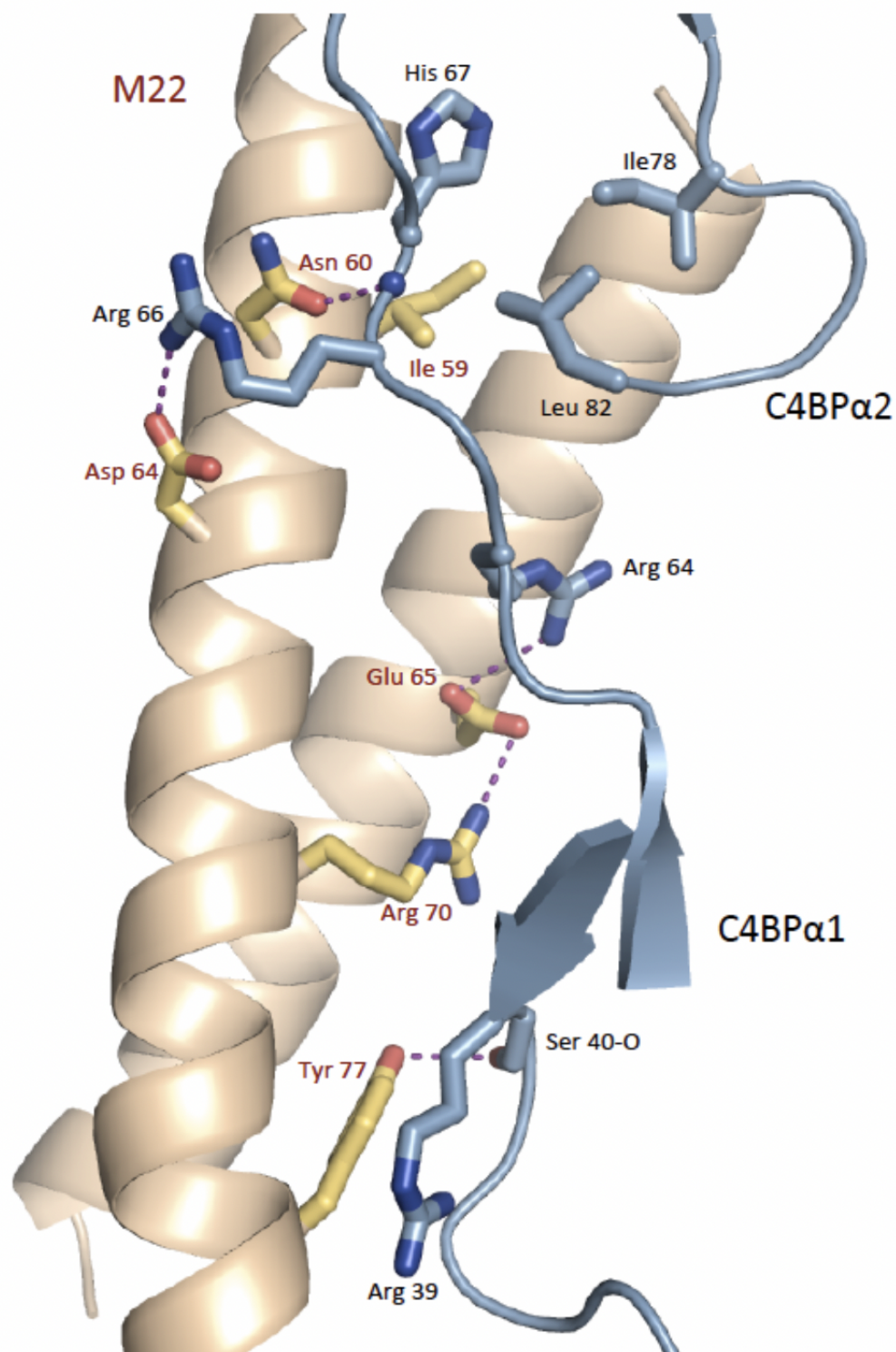


Figure 2.7. Structure of M22-C4BP

M22 is in gold ribbon representation with key side chains in bonds representation, in which carbon are yellow, oxygen red, and nitrogen blue. C4BP α 1-2 is in cyan ribbon representation, with key side chains in bonds representation, in which carbon are cyan, oxygen red, and nitrogen are blue. Hydrogen bonds and salt bridges depicted by dashed magenta lines.

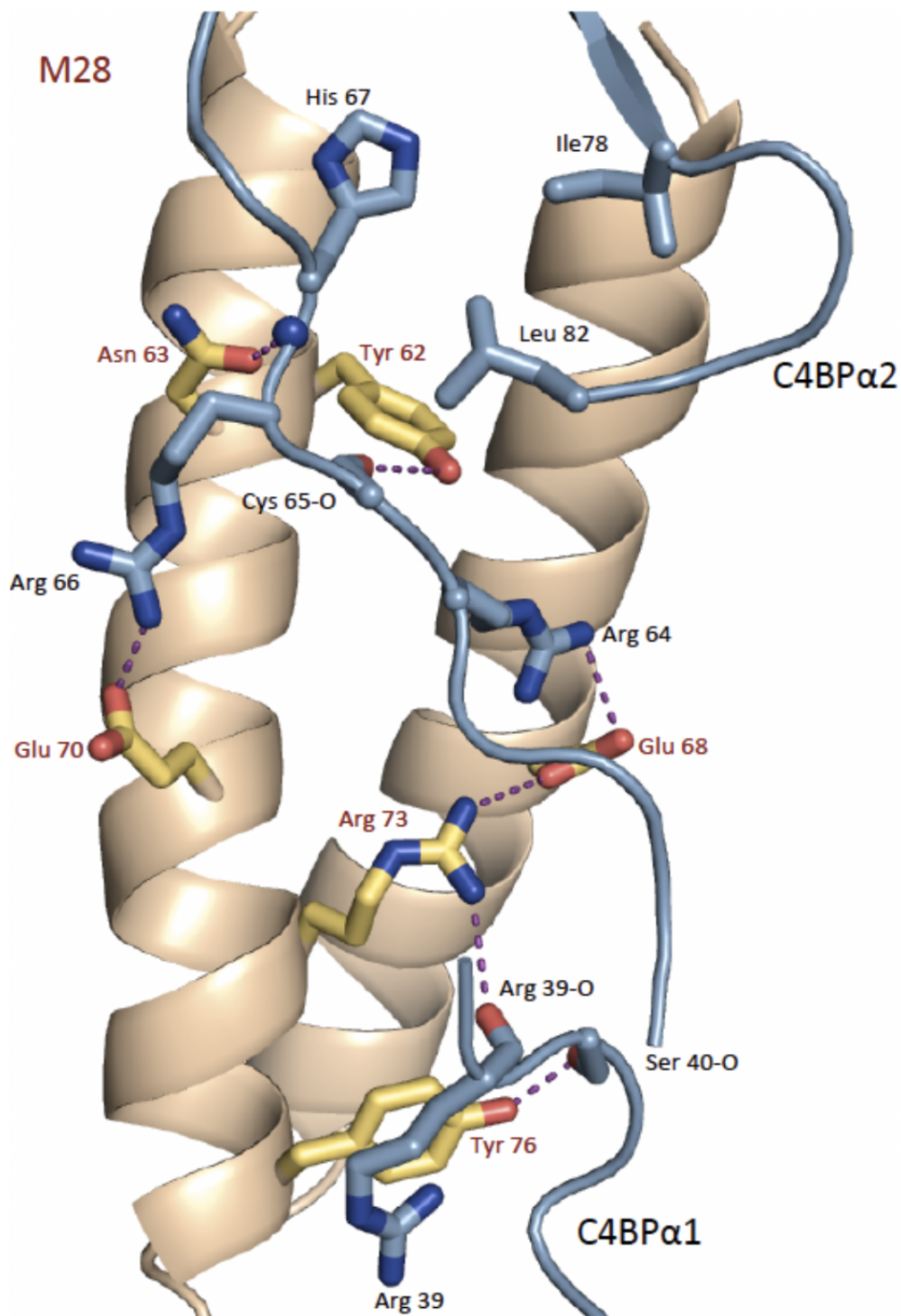


Figure 2.8. Structure of M28-C4BP

M28 is in gold ribbon representation with key side chains in bonds representation, in which carbon are yellow, oxygen red, and nitrogen blue. C4BP α 1-2 is in cyan ribbon representation, with key side chains in bonds representation, in which carbon are cyan, oxygen red, and nitrogen are blue. Hydrogen bonds and salt bridges depicted by dashed magenta lines.

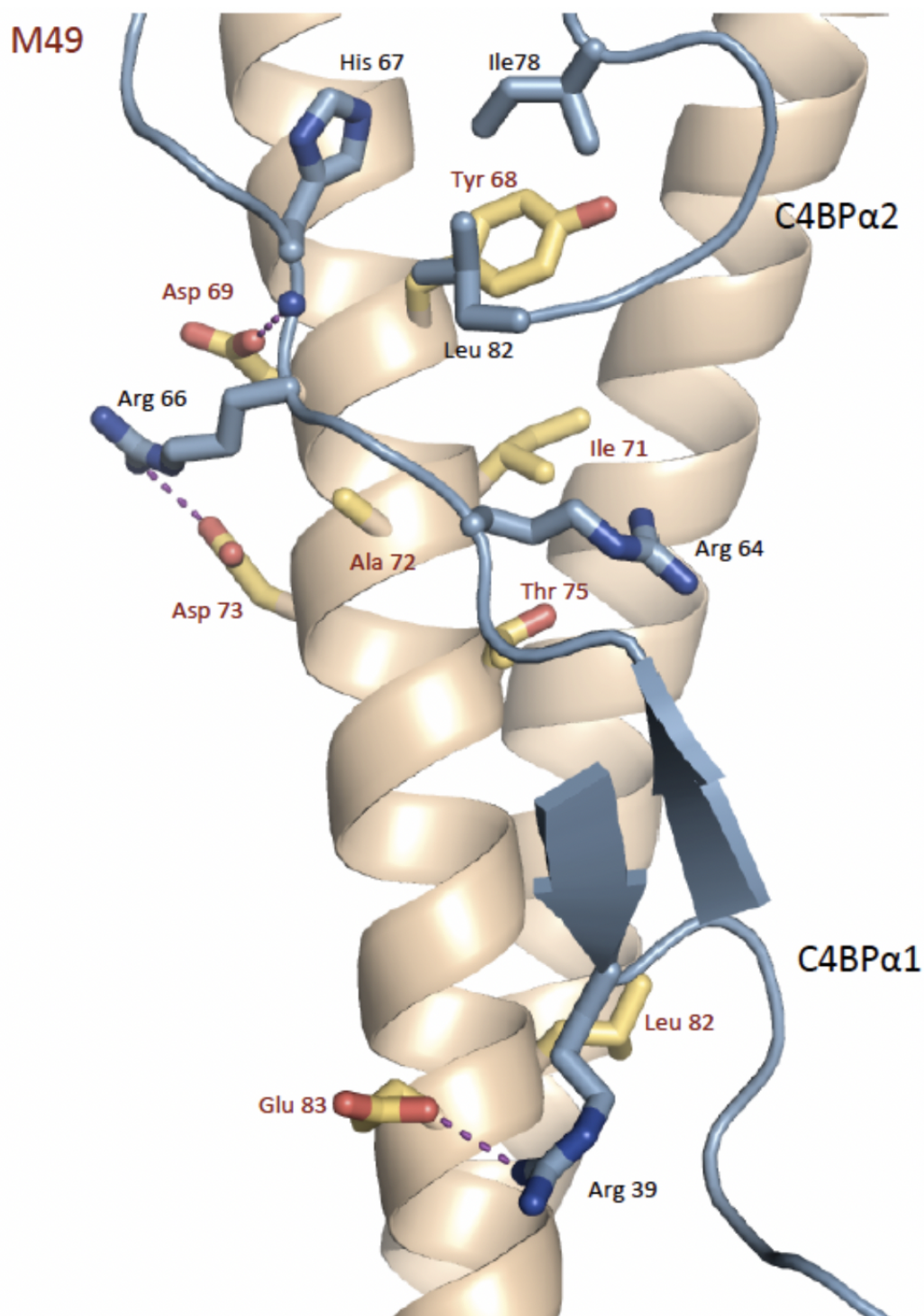


Figure 2.9. Structure of M49-C4BP

M49 is in gold ribbon representation with key side chains in bonds representation, in which carbon are yellow, oxygen red, and nitrogen blue. C4BP α 1-2 is in cyan ribbon representation, with key side chains in bonds representation, in which carbon are cyan, oxygen red, and nitrogen are blue. Hydrogen bonds and salt bridges depicted by dashed magenta lines.

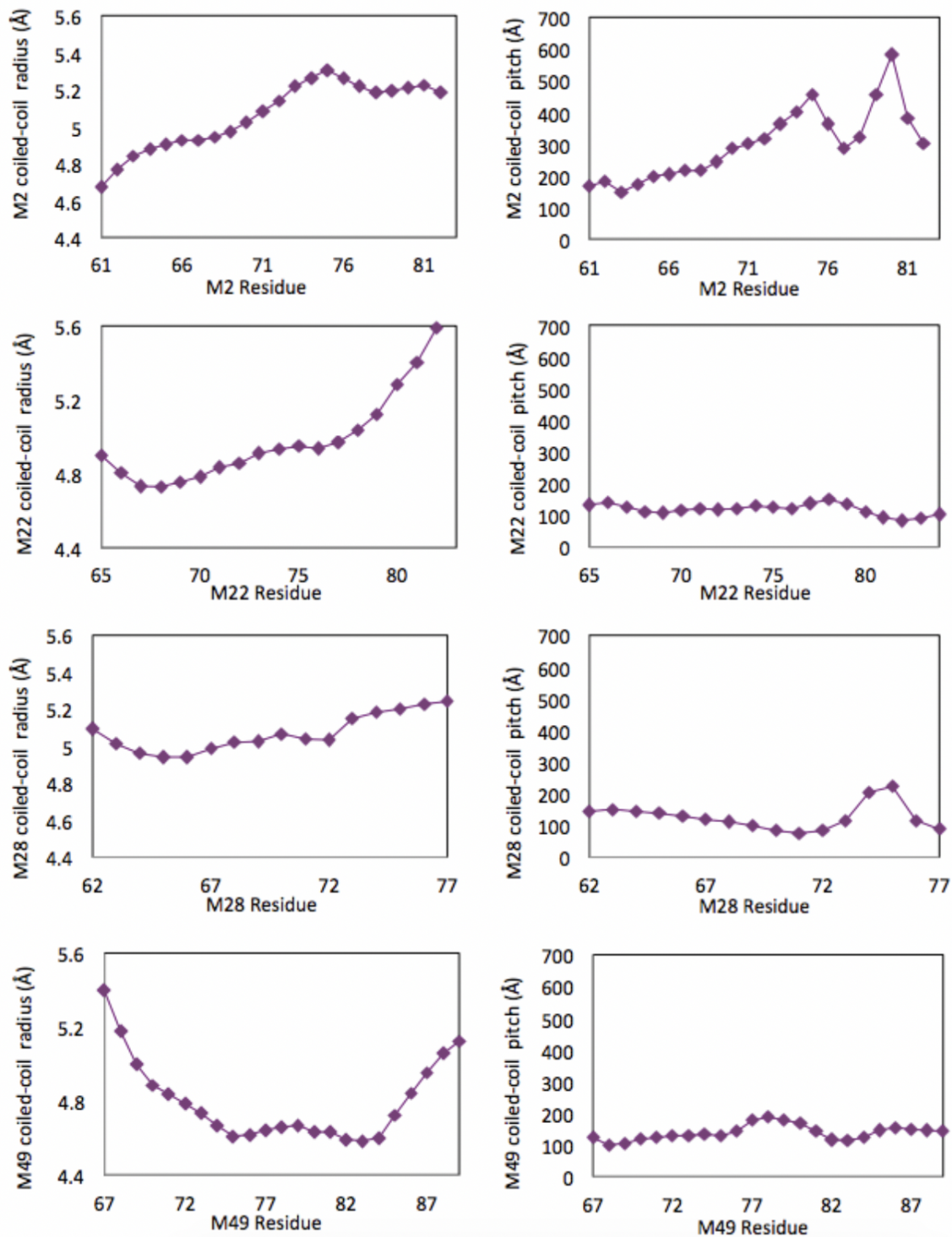


Figure 2.10. Coiled coil parameters of M proteins
 Radius and pitch of the M protein HVR α -helical coiled coils, reporting on the residues that are at the interface with C4BP α 1-2.

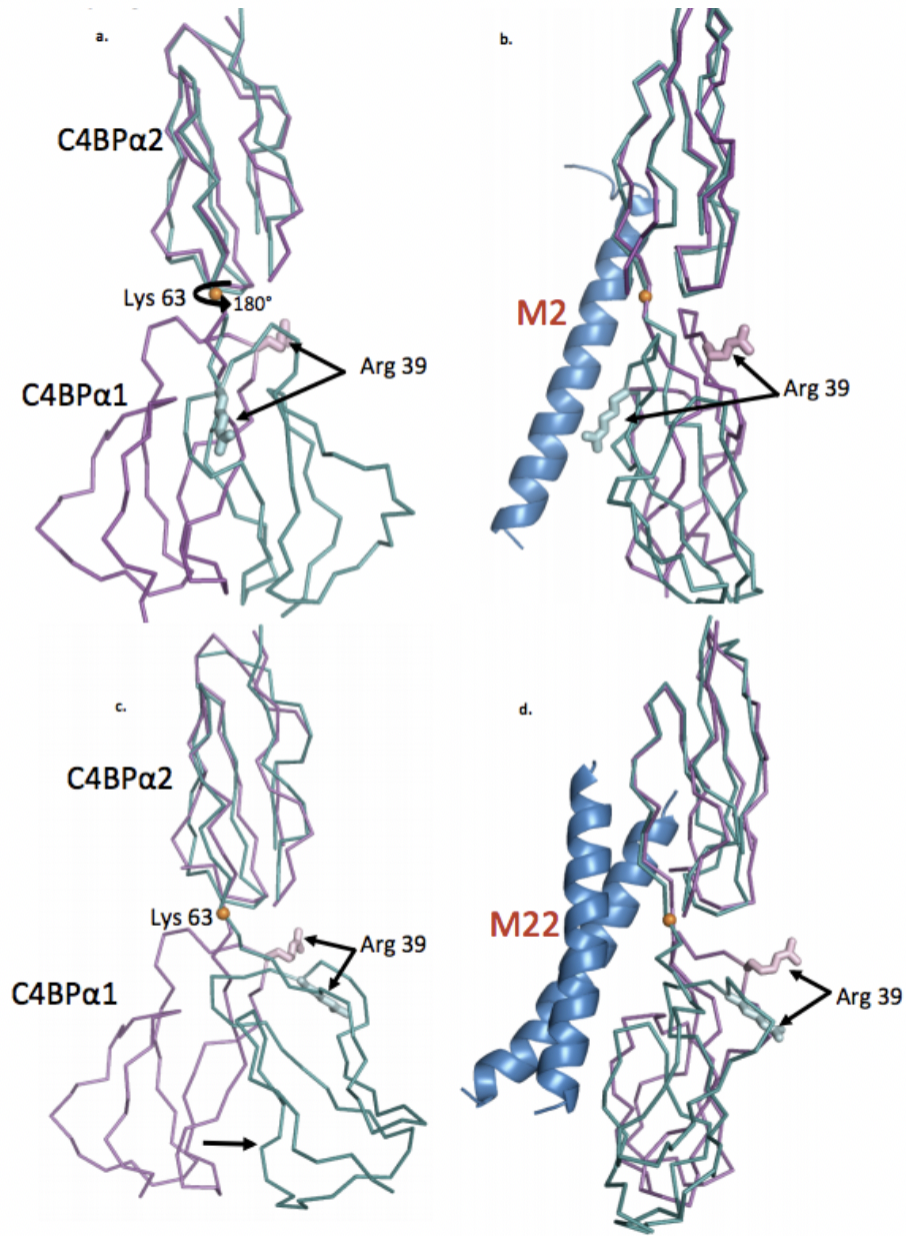


Figure 2.11. Rotation of C4BP α 1-2

(a) Superposition of free (magenta) and M protein-bound C4BP α 1-2 (cyan) based on the C4BP α 2 domain, showing the 180° rotation of C4BP α 1 around Lys 63 (indicated by golden ball). The molecules are depicted as C α chain traces. The position of C4BP Arg 39 is shown in bonds representation. The same depiction is used in the following panels; (b) A 90° rotated view of the superposition shown in panel a, with one α -helix of the M2 coiled coil shown as a blue ribbon; (c) One of the C4BP α 1-2 molecules bound to M22 that is restricted from undergoing a 180° rotation by a crystal contact is depicted. Superposition of free (magenta) and the M22 protein-bound C4BP α 1-2 (cyan) is based on the C4BP α 2 domain. The bound C4BP α 1 domain is related to free C4BP α 1-2 by a tilt rather than a rotation; (d) A 90° rotated view of the superposition shown in panel c, with the M22 coiled coil shown as a blue ribbon.

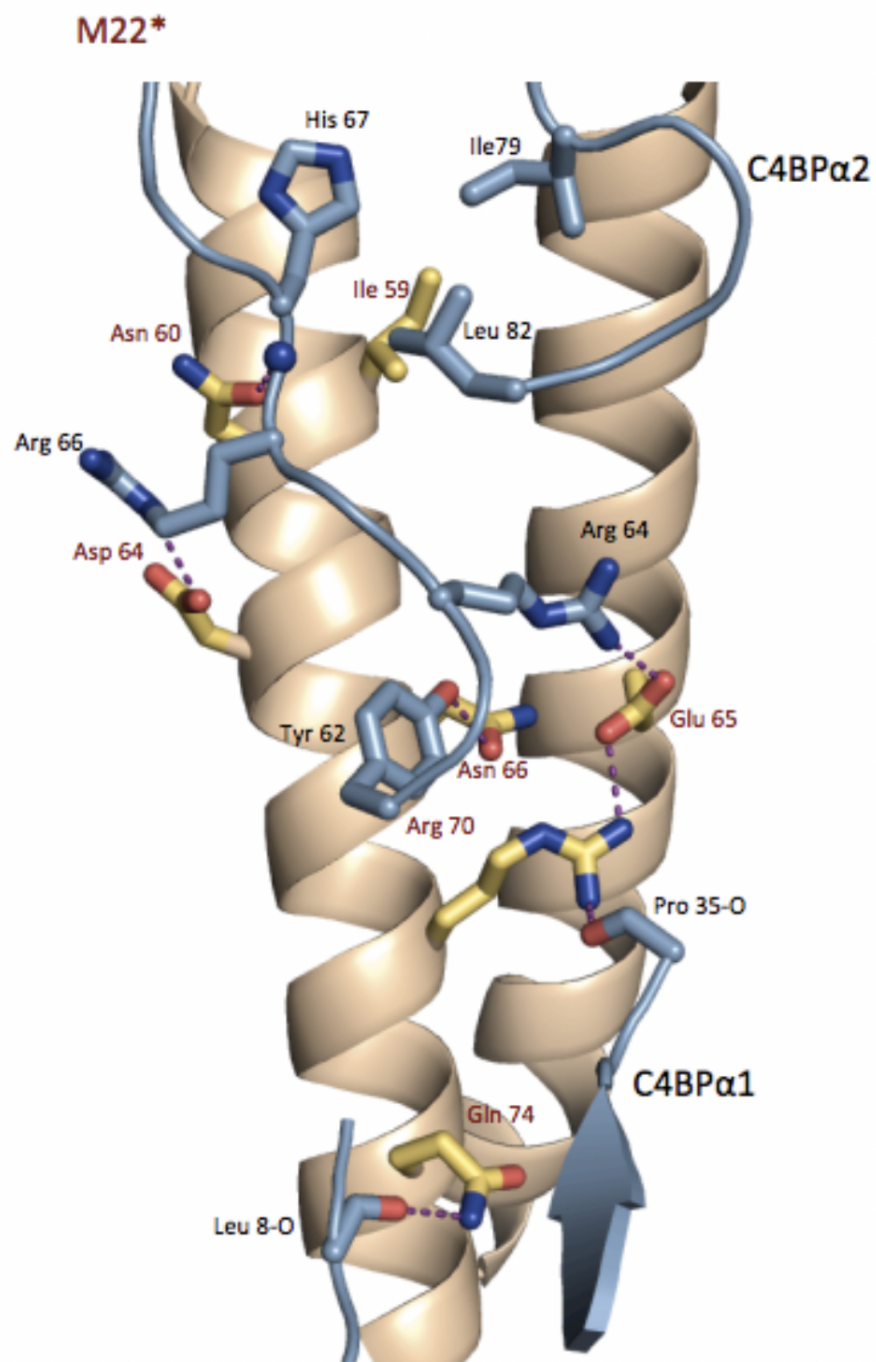


Figure 2.12. Structure of the M22-C4BP interaction in which C4BP α 1 is tilted rather than rotated
 The depiction is as in Figure 7. M22 is denoted with an asterisk to indicate that in this complex, the C4BP α 1 domain is tilted rather than rotated.

| | <i>abcde</i> | <i>fgabcde</i> | <i>fgabcde</i> | <i>fgabcde</i> | <i>fgabcde</i> | <i>fg</i> |
|--------|--------------|----------------|----------------|----------------|----------------|-----------|
| M2 | KKE | AKLSEAE | LHDKIKN | LEEEKAEL | FEKLDK | VVEE |
| M49 | VSS | VARREKEL | YDQIADL | TDKNGEY | LERIGE | LEER |
| M114 | KEA | TKLSEAE | LYNKIQE | LEEGKAEL | FDKLEK | VVEE |
| M73 | KEA | KKLNEAE | LYNKIQE | LEEGKAEL | FDKLEK | VVEE |
| M77 | EGV | SVGSDAS | LHNRI | TDLEERE | KLLNKLD | KVVEE |
| M84 | ASV | KKNNEEEL | HNKIADL | LDQNEEY | LNKIDEL | KEG |
| M89 | SVS | VKDNEKEL | HNKIADL | EEERGEH | LDKIDEL | KEE |
| M97 | GPV | PRSLWL | REYDKNQ | ELTKKL | TEFEK | LLQN |
| M102.1 | SSV | PVKKAAE | LYDKIKE | LEEGREEL | LNDL | DKVKED |
| M106 | QKQ | NVSSNGR | IYEIYDE | LQTKYDE | LQTKHE | ELLGE |
| M112 | SSV | SVKNEVK | LHNEIAA | LQEEKEK | LLNELD | VKEEH |
| M118 | ADS | NASSVAK | LYNQIADL | TDKNGEY | LERIEE | LEER |
| M124 | ATK | SKLSEAE | LHDKIKN | LEEEKAEL | FEKLDK | VVEE |

Figure 2.13. Sequence alignment of C4BP-binding M protein HVRs of the M2/M49 pattern. Residues that contact (first two lines) or are predicted to contact C4BP are in red, and the heptad register is indicated above. Residues observed (first two lines) or predicted to be at core d positions of the heptad register are highlighted in blue for visual reference.

Individual side chains are shown in bonds representation, with carbon cyan, nitrogen blue, and oxygen red. Water molecules are shown as purple spheres. Of particular note is the lack of water molecules at the M2-C4BP α 1 interface, and the stability of the loop on which C4BP R39 is located.

Video 2.2 R39 nook in M2 (F75A)-C4BP α 1-2.

Portion of the molecular dynamics simulation of the M2 (F75A)-C4BP α 1-2 complex. The depiction is the same as in Video 2.1. Of particular note is the increased number of water molecules at the M2 (F75A)-C4BP α 1 interface, and the instability of the loop on which C4BP R39 is located.

Video 2.3 C4BP α 2 contacts in M2-C4BP α 1-2.

Portion of the molecular dynamics simulation of the M2-C4BP α 1-2 complex. The depiction is the same as in Video 2.1. Of particular note is the infrequent interaction between C4BP α 2 R66 and M2 N66.

Video 2.4 C4BP α 2 contacts in M2 (K65A)-C4BP α 1-2.

Portion of the molecular dynamics simulation of the M2 (K65A)-C4BP α 1-2 complex. The depiction is the same as in Video 2.1. Of particular note is the increased frequency of interaction

```

                Abcdefgabcdefgabcdefgabcdefgabcdefg
M22      ISQESKLINTLTDENEKLREELQQYYALSDAKEEE
M28      ADKLADAYNTLLTEHEKLRDEYYTLIDAKEEEEPRY
M4       AWNWPKEYNALLKENEEELKVEREKYLSYADDKEKD
M4.1    AWNWPKEYNALLKENEEFKVEREKYLSYADDKEKD
PrtH    NAKLVEVVETTSLENEEKLKSENEENKNNLDKLSKD
M8       NEQLINELNNLIEENNDLKDKLARNDLLDDNTREK
M9.2    LSVPKTEYDKLYDDYDKLQEKSAEYLERIGELEER
M11     TNVSADLYNSLWDENKTLREKQEEYITKIQNEETK
M15.1   WKLTIEEYNKLLDENEEKLKEKNEEYLEKIGEQEER
M25     AKAAEAKVDKLEQLEGYKKLEEDYFNLEKR
M42     KVKLEVLYNSLWEENKTLREEQEEYIAKIDKLDEK
M44     GSVSLELYDKLSDENDILREKQDEYLTKIDGLDKE
M48.1   ELPPEARYKAWKSENDELRENHRKILDKFNAEQNK
M55     LYQERQRLQDLKSKFQDLKNRSEGYIQQYYDEEKN
M59     ELTLQQKYDALTNENKSLRRERDNYLNYLYEKEEL
M60     ISKERELINTLVDENNKLMEERARHLDLIDNI
M61     GSVSLELYDKLSDENDILREKQDEYLTKIDGLDKE
M63.4   KLTLEHKYNALTNENKSLIRREKDKYLYEKEELEK
M66     QNTWEKRYQKLSDDHTLLQDAIEEISSENEKLKSE
M67     GGVRLDLYDKLSKENDILREKQDEYLTKIDELGEK
M69.1   GGVSLDLYSKLLNENDILRDKQDDYLTKIDELTEK
M76     SNVSINLYNELQAEHDKLQTKHEELLAEHDALKEK
M78     SITNEQLIDKLVEENNDLKEERAKYLLDLDNREKD
M81     ENVPKQQYNALWEENEDLRGRERKYIAKLEKEEIQ
M85     TSVSADLYNSLWDENKTLREKQGEYITKIQNEETK
M88.1   ISNNERLINELTDENNELKDKLARSLLDLDNTREK
M92     SGSVSTPYNNLLNEYDDLLAKHGELLSEYDALKEK
M96     LDQFGRDYDELQKKYDKLDKENKEYASQLGK
M109    ADNLAKEYNTLLTENEKLREELQQYYALIDAKEEE
M110    HEELWKEYDILKEKLDKDQEEREKIELNYLK
M111.1  VTAPAHFWENQRREIEKLKGEIDQLKLLGKS
M117    GSVSIDRYNELSGEYNKLLDQNGNLLDENEILREK

```

Figure 2.14. Sequence alignment of C4BP-binding M protein HVRs of the M22/M28 pattern. Residues that contact (first two lines) or are predicted to contact C4BP are in red, and the heptad register is indicated above. Residues observed (first two lines) or predicted to be at core d positions of the heptad register are highlighted in blue for visual reference.

M14.5 DRVSRMSRDDLLNRAQNLEAKNHGLEHQNTKLSTENKTLQEQAEARQKE
M29.2 RVYITRRMTKEDVEKIANDLDTENHGLKQONEQLSTEQGLEEQNKQLST
M38 EGEPREVSEELVNSNPVLLNKKIAKLEELANKQESKESKEAIDALNNI
M54 EVLTRRQSQDPKYVTQRISDLEVKNHDLNKNKLTSENQNLKNKTELE
M62 EEAGASRTITSSENISKLYDENSKLIEERADLLGKLEEKEDKLESVERQY
M90 EGKAAAVSRSNSEQNNSEQNNLEKRYRKLSDDYTVLQEAIEGISSENEKL
M94 EEASNNGQLTLQHKNNAITSENESLRREKEELEKKNKELDSQVAGLIGVV
M99 DGERVPPKNNRLSKKYSELSEKYGALSEKYGALLDKQGALLDKQEELEKEN
M105 EVNTRSRAQDAGYQKGRADKLE TENHGLKQNEKLNQNNNDLKTQTATLT
M18.6 APLTRATADNKDELIKRANDYEIQNHQLTVENRKLKTDKEQLTKENDDLK
M32 KAVTRGTVSDPETARQTIKDYDIKNHQLTQENEKLTKEKEELTQENEKLT
M36 KALTRSTASNSEARQTIKDYDIKNHQLTQENEKLTQENEKLTSEKELT
M46 AAVTRHMSTEQLKQVREYDIENHKLKTDKARLEAEKQLETKKNELEAK
M71 RAITRATSDPAKQKQVVEGYELENHTLKNDEKLTENSALTTEKNRLT
M100.1 RVTTRSQAQDAAGLKEKADQYEVNRHELEHNNEKLTENSCLKTENSCLT
M115 KAVTRSTASDPEKARQTIKDYDIKNHQLTQENEKLTQENEKLTSEKELT
M58 DSSREVTNLTASMWKAQADS AKAKAKELEKQVEEYKKNYETLEKGYDDL
M79.1 DSRDITGTLPATMWKQKAEAAKAKASNLKQLEEARQDYSQIEEKLEQFG
M87 ESPREVTNELAASVWKKKVEEAKEKASKLEKQLEEAQKDYSEIEGKLEQF
M103 DSPRDVTSDLTSMWKKKAEAAKAKASKFEKQLEDYKKAQKDYIEEKL
M104 EGVNRHNSQNTWEKRYRELS EDHALLEATIDDI SLENEKLSSENKKNLE
M33 EEHEKVTQAREAVIREMQRGTNFGPLLASTMRDNHNLKETLTKKEID
M41.2 EGNARLAQAQEEALRDVLLNTPHNQLRDYAGAFRRNNELEKIQEKERE
M43.2 EEHPDVVAARESVLNNVRVPGTLWLRLQKEENDKLKSEKKGLETELQEKEQ
M52 DQPVDDHRYTEANDAVLQGRVTSARALLHEINKNGQLRSENEELKADLQK
M64 DRLHPGYTAANNAARNEFLVPAGAVLHEREKNDLRLKNEELKADLQKKE
M68.4 EEVKKAEVVKKAESESKSAKMWENMYKELDRDYSLEKTVENMSLENM
M70 EEHESVTRAREAAIEMMRQGRGDFAPLLANAIRDNNNL TETLTKKEI
M72 NRADDARREVLRGQFVEAELWHHQIQENDQLKLEKEELKSDLQKQELK
M74 FTVTRSMTRDYLAQVVDFTKNHELET HNSELSATNQLQGGVBAEQK
M75 EEERTFTELPYEARYKAWKSENDELRENYRRTLDKFNTEQGGKTRLEEON
M80 HQLADAAARREVLKGETVPAHLWYYQKEENDKLSANELETTLQKKEQEL
M82 DSSSRDI TEAGVSKFWKSKFDAEQNRANELEKLSGYSYKDYKLEQYEN
M83.1 DNPRYTDHNAVTOGRVPLQNLLEHMDKNGKLRSENEELKADLQKQEQ
M86 DNVGRVDVDKIREALHQAIGGMTNVQLRNTLAGSFRMNDLKKAIQEK
M91 ADDHPGAVAARNDVLSGFSVPGNVWYRQHQEIGKLSKSEKEELETTELQEK
M93 ENNTRYNEAYEQALREVLGGMNNVQQLRGALAGSFHRNNELOKTIQEKER
M98 DRYTDHNAVTOGRVPLRNLLLEMDKNSKLRSENEELQAGLQEKERELE
M101 ADHPSYTAAKDEVLSKFSVPGHVWAHEREKNDKLSSENEGLKAGLQEKEQ
M107 AEAQAQAQAEAKAEAKAPAPAKAPAKAQTREKQLLLLEEYRKLLEEGYFNL
M108 KEHESVTRAREAAIROMMQGGRDFAPLLADTIRDNNNLRETLDLTKKEI
M116.2 DHPLYTAANNAVRNGLSPSDRAVLAIEDKNDKLRLENKELKAGLQEKEQ
M119.2 DQPNHPRYTDANNAVVRNGLSPDRVLAIEDKNDKLRLENKELKAGLEEL
M120 DDNPRYTAQDEVLRLELPGQAQAFSRAFLYERQKNGELRLENEGLKLTALQ
M121 DQPNHPGYTEANNAVLRNGYSVPLRYWAHEREKNDKLSSENEELKAGLQEK
M123 AENHPLAESARQVLGESTVPASAWYYQKEENDKLSSENEGLKTDLQKKE

Figure 2.15. C4BP-binding M protein HVRs that cannot be classified as belonging to either M2/M49 or M22/M28 patterns
Residues predicted to be at core d positions of the heptad register are highlighted in blue.

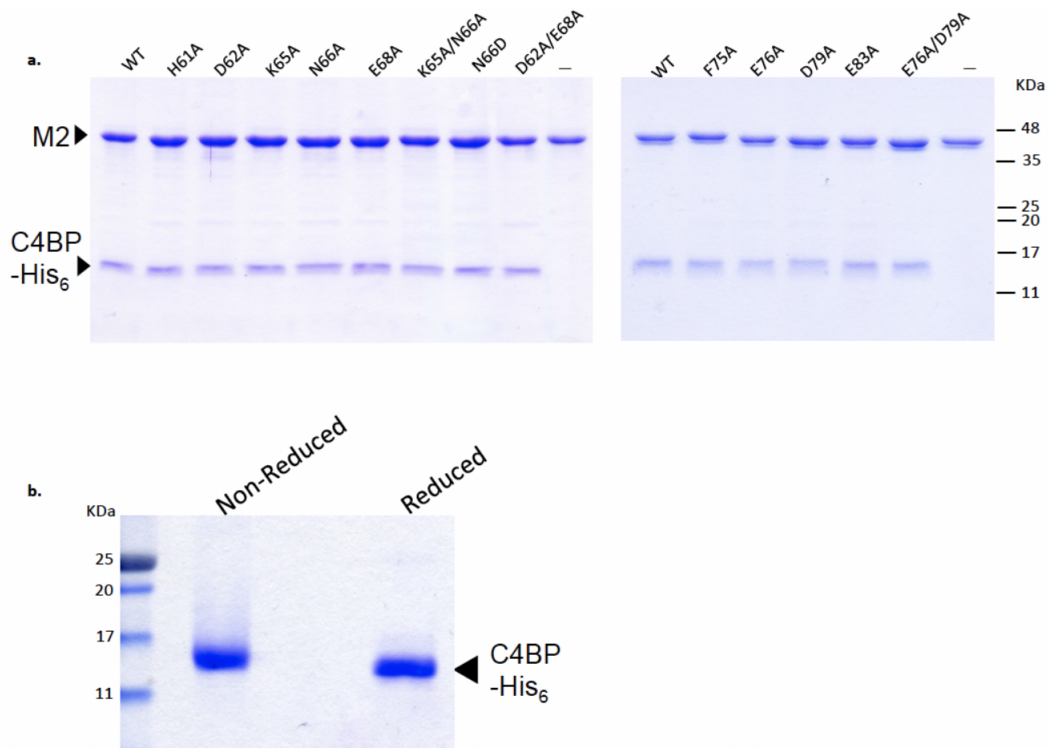
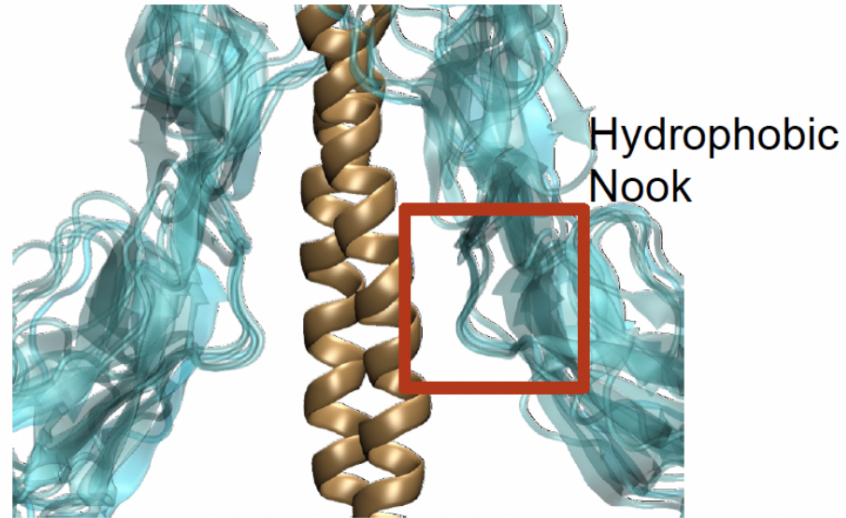


Figure 2.16. C4BP-M2 interaction

(a) Input samples from the experiment shown in Figure 4a visualized by non-reducing, Coomassie-stained SDS-PAGE. Molecular mass markers were not run on these particular gels; their positions are based on measurements from equivalent gels. ;(b) C4BP α 1-2 visualized on Coomassie-stained SDS-PAGE under non-reducing and reducing conditions, showing that C4BP α 1-2 used in these experiments is intact and not nicked. Molecular mass markers are in the leftmost lane.

M2-C4BP α 1



M2 (F75A)-C4BP α 1

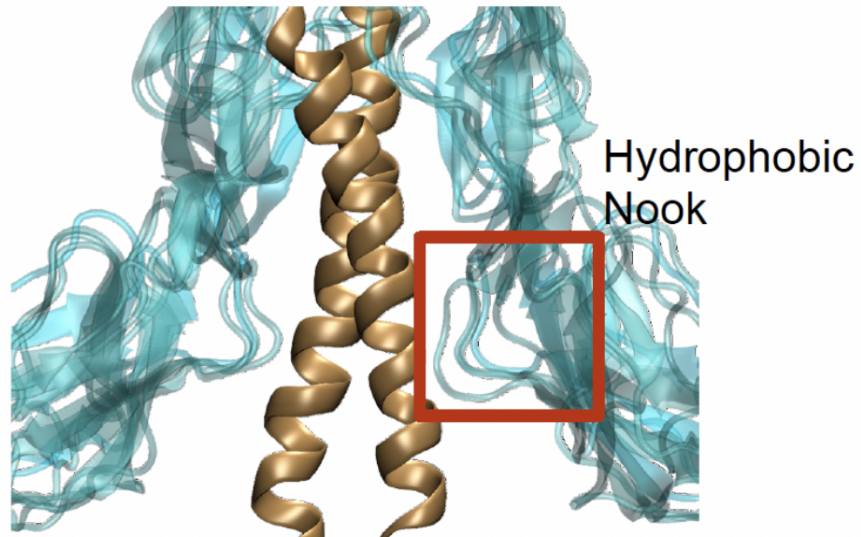


Figure 2.17. Molecular dynamics simulation of the Arg39 hydrophobic nook interaction with wild-type M2 and M2 F75A.

Five overlaid replicates of simulated interactions between C4BP α 1-2 (cyan, ribbon representation) and wild-type M2 (upper panel) or M2 F75A (lower panel); C4BP α 1-2 is in cyan ribbon representation, and M2 in gold ribbon representation. The simulated coordinates of the five replicates were aligned on M2 protein. The images highlight the hydrophobic C4BP α 1 Arg39 hydrophobic nook.

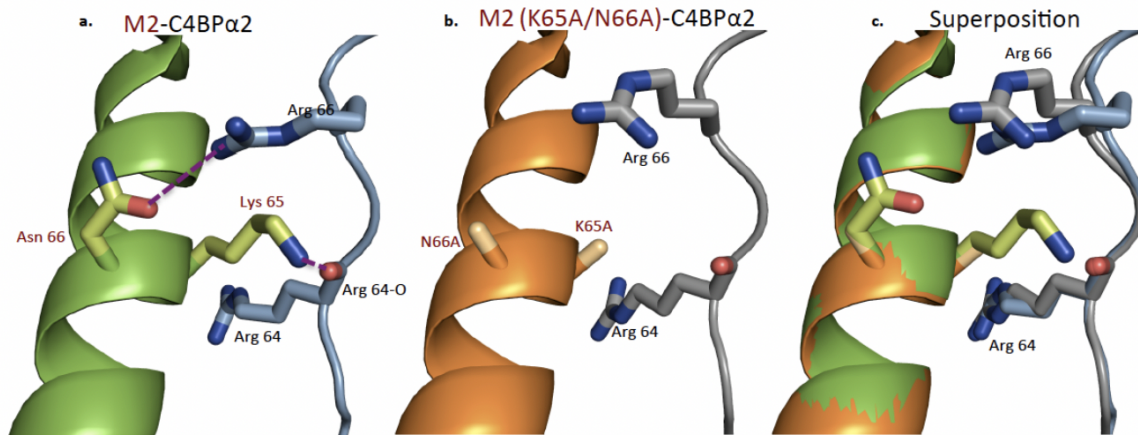


Figure 2.18. Interactions of M2 and M2 (K65A/N66A) with C4BP α 2

(a) M2 is depicted as a green ribbon, with Lys 65 and Asn 66 in bonds representation (carbons are green, oxygens red, and nitrogens blue). C4BP α 2 is in cyan ribbon representation, with Arg 64 and its main chain carbonyl and Arg 66 in bonds representation (carbons are cyan, oxygens red, and nitrogens blue). Hydrogen bonds are depicted as red dashed lines; (b) M2 (K65A/N66A) is depicted as a gold ribbon, with Ala 65 and Ala 66 in bonds representation (carbons are light gold). C4BP α 2 is in gray ribbon representation, with the same groups as in panel a shown; (c) Superposition of the structures shown in panels a and b.

between C4BP α 2 R66 and M2 N66.

Video 2.5 C4BP α 2 contacts in M2 (N66D)-C4BP α 1-2.

Portion of the molecular dynamics simulation of the M2 (N66D)-C4BP α 1-2 complex. The depiction is the same as in Video 2.1. Of particular note is the increased frequency of interaction between C4BP α 2 R66 and M2 N66D.

2.7 Methods

2.7.1 DNA manipulation

The coding sequences of mature M2 (amino acids 42-367), M22 (42-335), M28 (42-363), and M49 (42-359) proteins were cloned from GAS strains M2 (AP2), M22 (Sir22), M28 (strain 4039-05), and M49 (NZ131), respectively, into a modified version of the pET28a vector (Novagen) modified such that it encoded an N-terminal His₆-tag followed by a PreScissionTM protease (GE Healthcare) cleavage site. Constructs that encoded truncated versions of these proteins, which consisted of only the N-terminal 79, 86 or 100 amino acids, were generated through the insertion of

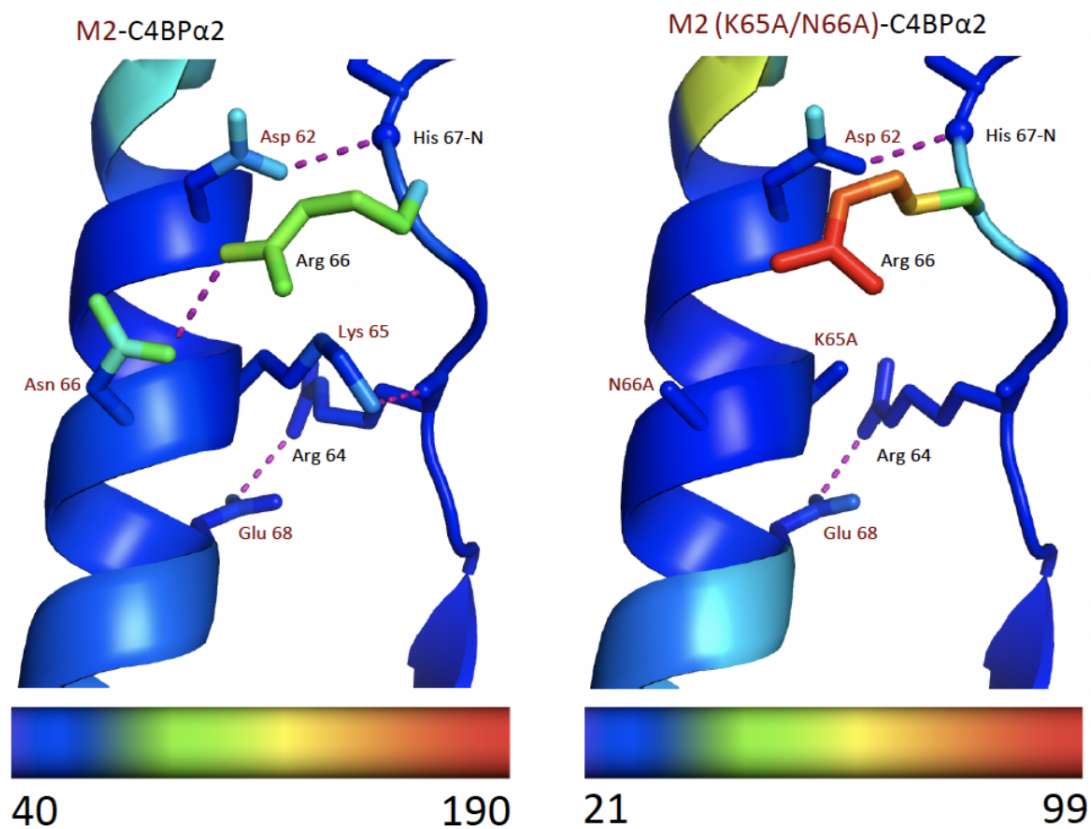


Figure 2.19. B-factors of C4BP α 2 bound to M2 or M2 (K65A/ N66A). B-factors in the M2-C4BP α 1-2 (left) and M2 (K65A/N66A)-C4BP α 1-2 (right) structures represented by a color spectrum. The coloring of the M2-C4BP α 1-2 complex is from the lowest to highest B-factor in the structure. The scale for the M2 (K65A/N66A)-C4BP α 1-2 complex was set equivalent to that of the M2-C4BP α 1-2 complex, but adjusted to account for the difference in Wilson B-factors between these two structures. This was done by dividing the B-factors of the M2-C4BP α 1-2 structure by the \sim 2-fold lower Wilson B-factor of the M2 (K65A, N66A)-C4BP α 1-2 structure (30.9 vs 59.7 angstrom²).

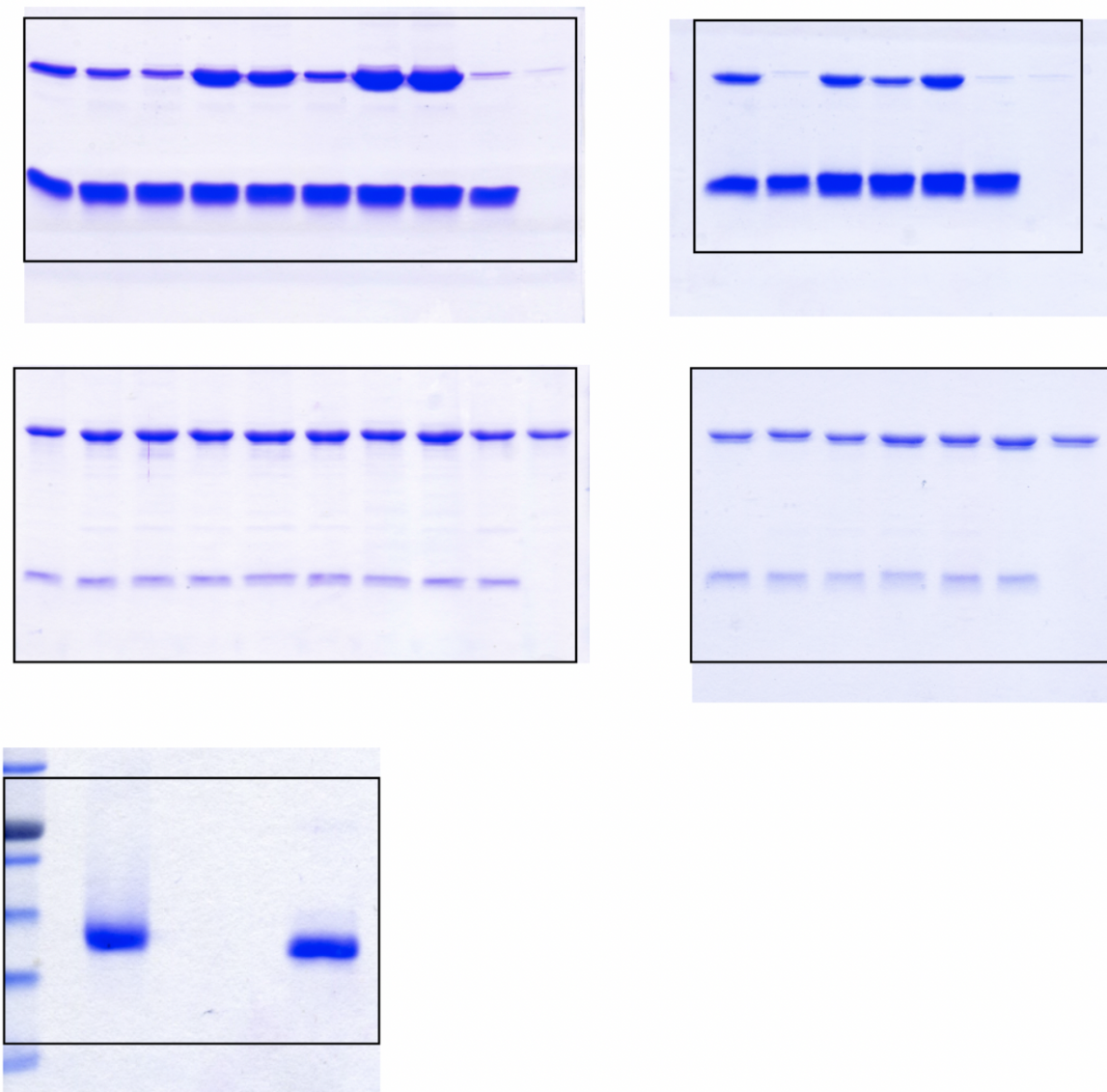


Figure 2.20. Uncropped Gels
Uncropped gels from Figures 4a (top two) and Figure 16 (bottom three)

Table 2.1. Data collection, phasing and refinement statistics for native and SAD (SeMet) structures

| PDB ID | M2-C4BP 5HYU | M2-C4BP(SeMet L46M/ L71M) | M2-C4BP(SeMet L29M/ L46M) | M28-C4BP 5HYP | M22-C4BP 5HYT | M49-C4BP(SeMet L29M/ L46M) 5HZP | M2 (K65A/ N66A)- C4BP 5IQQ |
|---|----------------------|---------------------------|---------------------------|----------------------|------------------------|------------------------------------|-------------------------------|
| Data collection | | | | | | | |
| Space group | P 4 ₃ 3 2 | P 4 ₃ 3 2 | P 4 ₃ 3 2 | P 4 ₃ 3 2 | P 2 ₁ 21 21 | P 4 ₃ 21 2 | P 4 ₃ 3 2 |
| Cell dimensions | | | | | | | |
| <i>a</i> , <i>b</i> , <i>c</i> (angstrom) | 148.3 | 148.7 | 148.3 | 133.7 | 68.08 | 78.1 | 148.6 |
| | 148.3 | 148.7 | 148.3 | 133.7 | 80.35 | 78.1 | 148.6 |
| | 148.3 | 148.7 | 148.3 | 133.7 | 152.9 | 345.3 | 148.6 |
| α , β , γ (°) | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 | 90, 90, 90 |
| Wavelength | 0.984 angstrom | 0.979 angstrom | 0.979 angstrom | 0.979 angstrom | 0.979 angstrom | 0.979 angstrom | 0.979 angstrom |
| Resolution (angstrom) | 50.00- | 50.00- | 50.00- | 50.00- | 80.40- | 86.40- | 50.00- |
| | 2.56(2.60- | 2.90(2.95- | 3.00(3.05- | 3.02(3.07- | 2.54(2.68- | 2.74(2.89- | 2.29(2.37- |
| | 2.56)a | 2.90)a | 3.00)a | 3.02)a | 2.54)a | 2.74)a | 2.29)a |
| <i>R</i> merge | 0.18(1.00) | 0.18(1.00) | 0.15(1.00) | 0.14(1.00) | 0.14(1.00) | 0.23(1.00) | 0.13(1.00) |
| <i>I</i> / σ <i>I</i> | 67.3(4.17) | 21.3(2.17) | 22.4(3.44) | 13.2(0.81) | 10.7(1.2) | 11.3(1.00) | 46.4(3.22) |
| Completeness (%) | 100(100) | 99.9(100) | 100(100) | 99.8(100) | 99.8(99.9) | 99.8(99.1) | 100(100) |
| Redundancy | 42.8(43.1) | 40.6(40.7) | 22.9(22.6) | 9.5(9.7) | 6.6(6.7) | 9.4(9.6) | 40.5(32.6) |
| cc1/2 | 1.00(0.86) | 0.99(0.86) | 0.99(0.86) | 0.99(0.48) | 1.00(0.60) | 1.00(0.61) | 0.99(0.86) |
| Refinement | | | | | | | |
| Resolution (angstrom) | 35.97- | | | 44.40- | 76.50- | 71.16- | 49.54- |
| | 2.56(2.65- | | | 3.02(3.21- | 2.54(2.60- | 2.74(2.84- | 2.29(2.37- |
| | 2.56) | | | 3.02) | 2.54) | 2.74) | 2.29) |
| No. reflections | 18514 | | | 7741 | 28328 | 28700 | 47375 |
| | (1808) | | | (1084) | (1975) | (2449) | (4655) |
| <i>R</i> work / <i>R</i> free | 0.21(0.27)/ | | | 0.25(0.33)/ | 0.21(0.34)/ | 0.25(0.42)/ | 0.20(0.26)/ |
| | 0.22(0.28) | | | 0.29(0.39) | 0.27(0.42) | 0.31(0.43) | 0.22(0.29) |
| No. atoms | | | | | | | |
| Protein | 1259 | | | 1197 | 4844 | 3146 | 1252 |
| Ligand/ion | 0 | | | 0 | 0 | 25 | 0 |
| Water | 76 | | | 0 | 70 | 20 | 132 |
| <i>B</i> -factors | | | | | | | |
| Protein | 76.7 | | | 110.9 | 70.6 | 106.7 | 53.4 |
| Ligand/ion | | | | | | 145.2 | |
| Water | 103.8 | | | | 54.6 | 68.2 | 67.0 |
| R.m.s deviations | | | | | | | |
| Bond lengths (Å) | 0.01 | | | 0.01 | 0.01 | 0.01 | 0.01 |
| Bond angles (°) | 1.35 | | | 1.31 | 1.48 | 1.21 | 1.35 |
| MolProbity score | 3.03[44th]b | | | 3.33[52nd] b | 2.49[78th] b | 3.03[53rd]b | 2.80[37th]b |
| Ramachandran | | | | | | | |
| % preferred | 90.3 | | | 86.7 | 92.6 | 90.4 | 92.9 |
| % allowed | 8.4 | | | 11.3 | 4.2 | 7.8 | 6.4 |
| % disallowed | 1.3 | | | 2.0 | 3.2 | 1.8 | 0.7 |
| Clashscore | 18.23[81st] | | | 16.4[96th] | 9.81[97th] | 11.69[97th] | 9.78[94th] |

(a) Highest resolution bin in parentheses here and other rows. (b) Percentile in brackets here and other rows.

an amber stop codon at an appropriate site by site-directed mutagenesis. Site-specific mutations were also introduced into the M2 coding sequence by site-directed mutagenesis. All site-directed mutagenesis was performed according to the Agilent QuickChangeTM manual, except that 50 μ L reactions were set up for polymerase chain reactions (PCR) instead of 12.5 μ L reactions. The coding sequence of the CCP1-2 domains of human C4BP α chain (C4BP α 1-2) [83] (a kind gift from G. Lindahl) was cloned into the modified pET28a vector described above, and also into a pET28b vector that encoded a non-cleavable C-terminal His₆-tag. The cleavable N-terminal His₆-tag version of C4BP α 1-2 was used for crystallographic studies, and the non-cleavable C-terminal His₆-tagged version for co-precipitation binding studies. To obtain selenomethionine (SeMet)-substituted protein to be used in phase determination, methionines were introduced in the coding sequence of C4BP α 1-2 at amino acid positions 29, 46, and/or 71 by site-directed mutagenesis.

Table 2.2. Ionic Interaction Pair Occupancy in C4BP α 2 for Quadrilateral Residues of M2 (%)

| C4BP-M2 | M2 Wild-type | M2 (K65A) | M2 (N66D) | M2 (N66A) | M2 (K65A/N66A) |
|--------------------|--------------|-----------|-----------|-----------|----------------|
| H67-D62 | 64.3/94.4a | 96.5/100 | 93.4/88.6 | 75.8/77.5 | 99.0/98.4 |
| R64-E68 | 73.6/44.2 | 33.8/41.8 | 50.1/73.0 | 61.2/35.8 | 56.2/28.0 |
| R66-N/D/A66 | 5.70/4.70 | 59.0/44.4 | 45.7/26.0 | - | - |

(a) The two percentages reflect the two binding sites in the 2:2 C4BP-M protein complex

Table 2.3. Ionic Interaction Pair Occupancy in Quadrilateral for residues of M49, M22, and M28 (%)

| C4BP-M49 | | C4BP-M22 | | C4BP-M28 | |
|----------------|------------|----------------|-----------|----------------|-----------|
| H67-D69 | 75.3/90.4a | H67-N60 | 100/100 | H67-N63 | 100/100 |
| | | R64-E65 | 71.4/55.7 | R64-E68 | 52.4/30.2 |
| R66-D73 | 54.1/43.3 | R66-D64 | 95.6/96.0 | R66-E70 | 21.1/28.8 |

(a) The two percentages reflect the two binding sites in the 2:2 C4BP-M protein complex.

2.7.2 Protein Expression and Purification

M proteins were expressed in Escherichia coli BL21 (DE3) and purified as described [89] with minor modifications to the procedure. Specifically, bacteria were lysed with a C-5 Emulsiflex (Avestin Inc., Ottawa, Canada) and ion exchange chromatography was omitted, and in the case of purification of M2 (wild-type and variants), imidazole was not included in the lysis and wash buffers. C4BP α 1-2 was expressed in E. coli Rosetta 2 (Novagen) cells. The protein was purified and refolded as described [86], except for the use of a C 5 Emulsiflex for lysis. Where needed, the N-terminal His₆-tags of M proteins and C4BP α 1-2 were removed by PreScissionTM protease cleavage according to manufacturers instructions, and the cleaved protein was purified by reverse Ni²⁺-NTA chromatography. M proteins and C4BP α 1-2 were lastly purified by size-exclusion chromatography (Superdex 200) in a buffer composed of 150 mM NaCl, 50 mM Tris, pH 8.5. Proteins were then concentrated to ~20 mg/mL by ultrafiltration; protein concentrations were determined by absorbance at 280 nm using calculated molar extinction coefficients. Aliquots of concentrated protein were flash-frozen in liquid N₂ and stored at -80 . SeMet was incorporated into C4BP α 1-2 (L29M/L46M), C4BP α 1-2 (L29M/L71M), and C4BP α 1-2 (L46M/L71M) using methionine pathway inhibition as described⁴. SeMet-labeled C4BP α 1-2 was purified as described above.

2.7.3 Crystallization and Data Collection

For preparation of complexes, M2 (amino acids 42-141), M2 (K65A/N66A) (42-141), M22 (42-120), M28 (42-141), or M49 (42-127) protein was mixed with C4BP α 1-2 (wild-type or SeMet-substituted mutant) at a 1:1 molar ratio (final concentration of complex \sim 5 mg/mL), and dialyzed overnight at 4 in 10 mM Tris, pH 8. The samples were then concentrated by ultrafiltration to \sim 20 mg/mL. Crystallization was performed by the hanging drop vapor-diffusion method. The M2-C4BP α 1-2, M2 (K65A/N66A)-C4BP α 1-2, and M28-C4BP α 1-2 complexes and the SeMet-labeled M2-C4BP α 1-2 (L29M/L46M) and M2-C4BP α 1-2 (L46M/L71M) complexes were co-crystallized at 20 by mixing 1 μ L of complex with 1 μ L of the reservoir solution, which was 1.5 M (NH₄)₂SO₄, 0.1 M Bis-Tris Propane, pH 7.0. These crystals were transferred to the reservoir solution supplemented with 20% ethylene glycol for cryopreservation, mounted in fiber loops, and flash-cooled in liquid N₂. Crystals containing SeMet-labeled protein were treated similarly, except the reservoir solution was supplemented with freshly prepared 1 mM TCEP. The M22-C4BP α 1-2 complex was co-crystallized similarly, except the reservoir solution was 2 M (NH₄)₂SO₄, 2% PEG 400, and HEPES pH 7.5. The SeMet-labeled M49-C4BP α 1-2 L29M/L46M complex was also co-crystallized similarly, except the reservoir solution was 1.6 M Na/K PO₄, pH 6.9. These two co-crystals were transferred to their respective reservoir solutions supplemented with 20% glycerol before being flash-cooled in liquid N₂. Diffraction data were collected from crystals under cryogenic conditions. Diffraction data for M2-C4BP α 1-2 were collected at the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 9-2, for M22-C4BP α 1-2 at the Advanced Photon Source (APS) beamline 24-ID-C, and for M2 (K65A/N66A)-C4BP α 1-2 and M28-C4BP α 1-2 at the Advanced Lightsource (ALS) beamline 8.2.1. Single-wavelength anomalous dispersion (SAD) data were collected from SeMet-labeled M2-C4BP α 1-2 (L29M/L46M) and M2-C4BP α 1-2 (L46M/L71M) at the APS beamline 19-ID, and from SeMet-labeled M49-C4BP α 1-2 (L29M/L46M) at the APS beamline 24-ID-E. Diffraction data from crystals of M22-C4BP α 1-2 and M49-C4BP α 1-2 (L29M/L46M) were indexed, integrated, and scaled using XDS [90], while HKL2000 [91] was used for data from all other crystals.

2.7.4 Structure Determination and Refinement

M2-C4BP α 1-2

For structure determination of M2-C4BP α 1-2, Se sites were located from SAD data of SeMet-labeled M2-C4BP α 1-2 (L29M/L46M) and M2-C4BP α 1-2 (L46M/L71M), and phases calculated for each data set using Autosol (within Phenix [92]). The two sets of phases were combined using the Reflection File Editor program (within Phenix). From the combined phase set, four Se sites, three at substituted methionines and one at the native Met 14, were identified per asymmetric unit, which contained one M2 α -helix and one C4BP α 1-2 molecule.

Here and in all cases below, model building was carried out with Coot [93] as guided by inspection of SAD-phased maps or σ_A -weighted $2mF_o-DF_c$ and mF_o-DF_c maps, and refinement was carried out with Refine (within Phenix) using default parameters. Between 15 and 75 iterative cycles of building and refinement, with each refinement step consisting of 1-10 rounds, were performed in each case. In later stages of refinement, TLS parameterization was used in Refine. Individual B-factors were refined isotropically. Water molecules were added in the final stages of refinement using Phenix with default parameters (3σ peak height in σ_A -weighted mF_o-DF_c maps).

In order to model M2-C4BP α 1-2 (L29M/L46M/L71M), the NMR structure of C4BP α 1-2 was manually fit into SAD-phased density, with the two domains of C4BP α 1-2 being treated as individual rigid bodies. The M2 molecule was then built into density, with the register of the coiled coil being assigned from well-defined density corresponding to large side chains (i.e., His 20, Phe 75, and His 85). The SeMet residues in the model were changed to leucines, and the model was then refined against the higher resolution (2.56 angstrom resolution limit) data collected from crystals of M2-C4BP α 1-2. TLS parameterization involved the following groups: For M2, 53-57 and 58-86; for C4BP α 1-2, 0-59 and 60-124. Continuous electron density was evident for the entire main chain of C4BP α 1-2 and for residues 53-86 of the M2 protein. Here and in all cases below, side chains for the entire M protein were visible in electron density except certain solvent exposed side chains with longer alkyl chains (i.e. Lys, Arg, Glu). Most side chains of C4BP α 1-2 were also visible, except for some in long loops. All sidechains not visible were distant from the

interface with M protein with the exception of C4BP Arg 66, which showed partial side chain electron density. Here and in most cases, Ramachandran outliers were primarily confined to the larger loop structures of C4BP where electron density was not as well defined. The only residue with a Ramachandran outlier in the M2 protein was at Ala 58. This is probably due to multiple possible conformations of the N-terminal residues of M2 within the crystal structure and such flexibility results in this outlier during refinement. The structure of M2 (K65A/N66A)-C4BP α 1-2 was determined by difference Fourier synthesis using the refined structure of M2-C4BP α 1-2. The set of reflections used for Rfree calculations for the refinement of M2-C4BP α 1-2 were maintained. TLS parameterization was equivalent to that for M2-C4BP α 1-2.

M28-C4BP α 1-2

The structure of M28-C4BP α 1-2 was determined by molecular replacement using the program Phaser (within Phenix [92]). The C4BP α 1-2 molecule from the structure of the M2-C4BP α 1-2 complex served as the search model. The molecular replacement solution had a log-likelihood gain score of 379. The asymmetric unit contained one C4BP α 1-2 molecule and one M28 α -helix, whose register was determined by well-defined density corresponding to large side chains (i.e. Tyr 62, Tyr 76, Tyr 77). The model was first subjected to cycles of rigid body refinement, followed by the refinement protocol described above. TLS parameterization involved the following groups: For M28, 55-83; for C4BP α 1-2, 0-59, 60-86, and 87-124. Continuous electron density was evident for the entire main chain of C4BP α 1-2, except for breaks in some of the longer loops, and amino acids 53-83 of M28.

M22-C4BP α 1-2

The structure of the M22-C4BP α 1-2 complex was determined by molecular replacement using the program Phaser. The search model consisted of an M28 α -helical, dimeric coiled-coil in complex with a single C4BP α 1-2 molecule. The solution, which had a log-likelihood gain score of 166, resulted in two copies of the search model in the asymmetric unit, while the solvent content suggested that the asymmetric unit was composed of two M22 α -helical, dimeric coiled-coils and four C4BP α 1-2 molecules; this latter composition was found to be accurate. After refinement of

the initial molecular replacement model, two additional C4BP α 1-2 molecules became evident in electron density maps, and were placed stepwise into density, with the two domains of C4BP α 1-2 being treated as individual rigid bodies, between rounds of iterative refinement. Both these additional copies had similar conformations to one another, and had a tilted orientation of the C4BP α 1 and C4BP α 2 domains relative to these domains in unbound C4BP α 1-2. This tilted orientation differs from the 180° rotation observed in the two other copies of C4BP α 1-2 bound to M22, as well as in copies of C4BP α 1-2 bound to M2, M28, and M49. Side chains for M22 were subsequently built into density, with the register being assigned based on well-defined density corresponding to large side chains (i.e., Tyr 66 and Tyr 67). The model was then subjected to cycles of rigid body refinement followed by the refinement procedures described above. TLS parameterization involved the following groups: For M22 chain A, 52-80; for M22 chain C, 52-79; for M22 chain E, 52-79; for M22 chain G, 52-80; for C4BP α 1-2 chain B, 1-13, 14-27, 28-59, 60-73, 74-86, 87-102, 103-109, 110-115, and 116-124; for C4BP α 1-2 chain D, 0-59 and 60-124; for C4BP α 1-2 chain F, 1-59 and 60-124; for C4BP α 1-2 chain H, 0-13, 14-33, 34-47, 48-59, 60-74, 75-86, 87-109, and 110-124. Continuous electron density was evident for the entire main chain of C4BP α 1-2, except for breaks in some of the longer loops, and for residues 52-79 (or 80, depending on the chain) of M22.

M49-C4BP α 1-2

For structure determination of M49-C4BP α 1-2, Se sites were located from SAD data collected for SeMet-labeled M49-C4BP α 1-2 (L29M/L46M), and phases calculated using the program Autosol. Six Se sites were identified per asymmetric unit, which was found to contain an M49 α -helical, coiled-coil dimer and two C4BP α 1-2 molecules. This is consistent with the total of two SeMet substitutions introduced into C4BP α 1-2. The crystal structure of C4BP α 1-2 from the M2-C4BP α 1-2 co-crystal structure was manually fit into SAD-phased density, with the two domains of C4BP α 1-2 being treated as individual rigid bodies. A model of the M49 protein was then built into density, with the amino acid register for the coiled coil being assigned based on well-defined density corresponding to large side chains (i.e., His 20, Phe 75, and His 85). TLS parameterization involved the following groups: For M49 chain A, 56-60 and 61-126; for M49

chain C, 56-126; for C4BP α 1-2 chain B, 0-10, 11-62 and 63-124; for C4BP α 1-2 chain D, 0-13, 14-27, 28-33, 34-44, 45-53, 54-62, 63-73, 74-86, 87-102, and 103-124. Continuous electron density was evident for most of the main chain of C4BP α 1-2, except for some of the longer loops of the C4BP α 1 domain, and for amino acids 56-124 (or 126, depending on the chain) of M49. The only Ramachandran outlier in the M49 protein dimer was Ala 106 of chain A. This residue was distant from the interface with C4BP.

Validation

Structural models were validated with MolProbity [94] (Table 2.1). Molecular figures were made with PyMol (<http://pymol.sourceforge.net>). Coordinates and structure factors have been deposited in the RSCB PDB.

2.7.5 Co-Precipitation Assays

Forty μg of C4BP α 1-2-His₆ protein was mixed with 120 μg of intact M2 protein (wild-type or mutant) in 50 μL of phosphate buffered saline (PBS) at 37°C for 30min. Fifty μL of Ni²⁺-NTA agarose beads were equilibrated in PBS, then added to the protein mix in a 1:1 beads:PBS (100 μL) slurry and incubated for 30 min at 37°C under agitation. The beads were washed three times with 0.5 mL of PBS supplemented with 15 mM imidazole, and eluted with 40 μL PBS supplemented with 500 mM imidazole. Proteins in the input and eluted fractions were resolved by non-reducing SDS-PAGE and visualized by Coomassie-staining.

2.7.6 Molecular Dynamics

System Preparation

Heavy atom coordinates were taken from the co-crystal structures of M protein-C4BP α 1-2 complexes. Structures of complexes containing M2 substitution mutants were created by computational point mutations at the desired amino acid(s). Due to the varying resolutions of crystal structures, crystallographic waters were removed prior to solvating the system. Each structure was prepared for simulation using the Amber14SB force field [38,43,95]. The ionization states of titratable residues at pH 7 were predicted using PROPKA 3.1 [96,97] and visually

inspected to ensure the accuracy of assigned states. Free cysteine residues were converted to disulfide-bonded pairs manually and built using tLeap, a system preparation program from the Amber Tools 2015 package [?]. The C-termini of proteins were capped to remove charges. The solvent was modeled explicitly using the TIP4P water model [98] and a 0.15 M NaCl concentration was applied after neutralizing the overall charge of the protein complexes. The Particle Mesh Ewald electrostatic summation method [99,100] was employed to evaluate electrostatics during simulation. In total, eight different M proteins in complex with C4BP α 1-2 were simulated: 1) M2 (amino acids 53-86), 2) M2 K65A (53-86), 3) M2 N66D (53-86), 4) M2 N66A (53-86), 5) M2 K65A/N66A (53-86), 6) M22 (52-79), 7) M28 (55-80), and 8) M49 (56-126). All systems contained residues 1 to 124 of C4BP α .

Minimization, Equilibration, and Production Molecular Dynamics

The NAMD simulation package [44,101] was used to minimize, heat, equilibrate, and simulate each system using a 2 fs time-step. Every system underwent a series of separate minimization, heating, and equilibration stages in preparation for production runs. The minimization spanned five stages in 10 ps intervals using the NVT ensemble: 1) 5,000 steps of hydrogen-only minimization, 2) 5,000 steps of solvent minimization, 3) 5,000 steps of side-chain minimization, 4) 5,000 steps of protein-backbone minimization, and 5) 5,000 steps of full-system minimization. Following minimization, the Langevin thermostat [102,103] was used to slowly heat the system to 310 K using the NVT ensemble over 250,000 steps (500 ps). The system was then subjected to three sequential equilibration stages using the NPT ensemble for 125,000 steps/stage (250 ps/stage). The pressure was set to 1 atm and maintained using the Berendsen barostat [104]. In the first MD production run, atoms were assigned a random starting velocity, and sequential steps carried over the velocities from the previous step. Five replicates of each system were performed to enhance sampling of the conformational landscape [105] and the total simulation time for each system was 25 ns/replicate. Therefore, the total aggregate simulation time for each system was 125 ns.

Percent Occupancy (Footprinting) Analysis

The five replicates comprising each system (125 ns total) were combined using cpptraj [106], a simulation processing software in the AmberTools package [107]. Trajectories were aligned against the first frame and an average structure was calculated using all atoms in the appropriate protein complex. The average conformation was used to realign the trajectories with respect to $C\alpha$ atoms. The average conformation was then used to calculate the root mean squared fluctuation (angstrom (RMSF)) of individual residues in the protein complex. A single concatenated 125 ns trajectory consisting of the five replicates was written by cpptraj and used for the following analysis. Using VMD [108], the radial distribution function (RDF) of pairwise interactions for a number of protein-protein contacts was calculated over the duration of the concatenated trajectory [85]. Distances in the RDF analysis were explicitly calculated for the following heavy atoms of residues: backbone nitrogen of histidine; $C\beta$ of alanine and valine; $C\gamma$ of aspartate, leucine, and isoleucine; $C\delta$ of glutamate; and $C\zeta$ of arginine. A 5 angstrom cutoff was applied to all pairwise interactions to include salt bridges and hydrogen bonds between hydrogen atoms and heavy atoms that were not explicitly analyzed. This was done to capture interactions between equivalent atoms, e.g. $O\delta$ and $O\delta$ of aspartate interacting with $H\omega$ and $H\omega$ of arginine.

2.8 Concluding Remarks

Our work shows that broad recognition between M proteins and C4BP is not due to contacts to the main chain, as it is for MHC-peptide complexes. Instead, the breadth of recognition for the M-C4BP complexes is explained by two attributes. First, the C4BP binding site is tolerant, notably due to the presence of arginines, and thus restrictions on the nature of interacting M protein side chains are not stringent. Second, the M protein coiled coil can align with C4BP in multiple ways, with at least two different arrangements being possible, which enables M protein side chains to reside at different positions of the heptad repeat. Our work importantly also has implications for vaccine design. Broadly neutralizing antibodies (bNAbs) have been identified for several highly antigenically variable microbial pathogens, including HIV and influenza virus [109,110]. These antibodies ignore hypervariable regions and instead focus on

invariant but hidden structural regions, for example, hidden by glycosylation in HIV or sterically occluded in influenza. The lack of ready accessibility to these structurally invariant regions explains why such bNAbs are rare. A different approach to creating bNAbs is suggested by our finding that C4BP, rather than ignoring hypervariability, detects conserved patterns that are hidden within hypervariability. The structures described here provide the initial details for the rational transplantation of the uniform and tolerant reading head of C4BP to the antigen-combining site of an antibody. We predict such an antibody would have broad specificity against M proteins as seen in C4BP, and would provide neutralization due to targeting of the M protein HVR. A potential challenge is that such an antibody may recognize C4b, but differences in C4BP binding modes between M protein HVRs and C4b suggest that selectivity would be possible [85]. A second challenge may be escape through further M protein variation. However, M protein HVRs vary from strain to strain but are stable within the type [111], suggesting that their overall sequence variation is limited by selection. Binding to C4BP appears to be a major selective pressure [76], and escape from such a broadly neutralizing antibody through further sequence variation may be limited by the pressure to maintain C4BP interaction.

2.9 Acknowledgment

Chapter 2, in full, is pre-publication print of an article published in Nature Microbiology 2016. Buffalo, Cosmo Z.; Bahn-Suh, Adrian J.; **Hirakis, Sophia P.**; Biswas, Tapan; Amaro, Rommie E.; Nizet, Victor; Ghosh, Partho; “Conserved Patterns hidden within group A Streptococcus M protein hypervariability are responsible for recognition of human C4b-binding protein.” The dissertation author was the third author of this paper.

This work was the result of a direct and indirect collaboration with a collection of scientists whom I would like to acknowledge. My dear friend and colleague, Dr. Cosmo Z. Buffalo and I began our computational investigations of the M-proteins in complex with Human C4BP, albeit without the blessing of our advisors. He is responsible for the crystallographic characterization of the four complexes as well and wrote the manuscript. Adrian J. Bahn-Suh is Cosmo’s student who also assisted with the experimental components of the work and figures. Dr. Tapan Biswas,

a scientist on the project, assisted Cosmo in the refinement of the structures. My beloved advisor, Dr. Rommie E. Amaro, advised me on the data analysis and helped to write and revise the manuscript. Though not an author on the paper, Pek Jeong, a former scientist in the Amaro lab, provided invaluable advice for the analysis of the binding regions through fingerprinting algorithms that she developed using Kepler software. Dr. Victor Nizet and Dr. Patho Ghosh helped write the manuscript and advised on scientific studies to further refine the work. I thank them all humbly for welcoming my computational insights and for allowing me to collaborate on this wonderful project.

This work was supported by NIH grant T32 GM007240 (CZB), an AHA Predoctoral Fellowship (CZB), NIH R01 AI096837 (PG and VN), and NIH R01 AI077780 (VN). The work was also funded in part by the National Biomedical Computation Resource, NIH P41 GM103426, NIH Director's New Innovator Award Program DP2-OD007237 and through the NSF XSEDE Supercomputer resources grant RAC CHE060073N to REA. SPH was supported by the Interfaces multi-scale analysis of Biological Structure and function training grant NIH T32 EB009380.

Chapter 3

Bridging structural and mechanistic observations with computational modeling techniques


In the following manuscript, we demonstrate how two distinct scales of molecular simulations can be used to understand the structure of a protein complex and suggest paths for the activation mechanism by a small molecule regulator. The subject of the manuscript is the RI α heterotetrameric complex of Protein Kinase A (PKA). PKA is a ubiquitous eukaryotic protein responsible for turning proteins on and off in response to an extracellular stimulus. PKA is comprised of two catalytic (C) subunits that phosphorylate protein targets and regulatory (R) subunits (R₂C₂). Each R subunit has two Cyclic-nucleotide binding Domains (CBD), which bind a total of four equivalents of cAMP for each R dimer. Upon binding cAMP, the inhibition of the catalytic subunits is relieved through dissociation, and the C subunits are free to phosphorylate protein targets. The holoenzyme structure has been the subject of some controversy. Specifically, the interface between the Type 1A R and C subunit has been debated extensively by those in the PKA community. Solution-structure methodologies such as small-angle X-ray scattering (SAXS) and Hydrogen/Deuterium exchange Mass Spectrometry (H/DxMS) suggest a protein-protein interface that involves mostly CBD-A. Crystallography of the full-length R in the holoenzyme complex proved challenging for the wild-type PKA, but a CBD-B mutant of R described a more-extensive interface that involved both CBD-A and CBD-B.

In this manuscript, MD simulations helped bridge the gap between information obtained from crystallographic and solution structures, revealing a stable conformation of the R subunit that was elucidated after only a few nanoseconds of simulations starting from the extended


conformation of the R subunit in the crystal structure. In the wild-type form, the R subunit relaxes into a conformation which we named the Flipback conformation. With the new R subunit structure, a model of the heterotetramer was constructed and proved extremely consistent with the solution structure description. More specifically, the Flipback structure is consistent with solvent-exposure of the C subunit determined by H/DxMS and the shape of the molecule determined by SAXS.

Using the flipback model of the heterotetramer, we used Brownian Dynamics (BD) to investigate the differential association kinetics of the small-molecule regulator cAMP to the two proposed structures: 1) the (Holo) holoenzyme model derived from two crystal structures and 2) the Flipback model derived from our simulations and based on a crystal structure model of the heterotetramer. Earlier studies determined that cAMP preferentially binds to CBD-B. This observation was confirmed by our BD simulations. A novel insight from the simulations was the association to the A-domain. In the heterotetrameric Flipback conformation, the association of cAMP to CBD-A was two orders of magnitude higher than to CBD-A of the crystallographic “Holo” model. In the heterodimeric forms, the association difference was more pronounced, with four orders of magnitude higher association rates to the Flipback conformation over the Holo conformation. The difference in association kinetics is due to a re-distribution of the electrostatic charges as a result of the conformational differences between Holo and Flipback states. The electrostatic potential on the surface of the protein is responsible for the long-range attractive forces guide cAMP to the CBD target. These findings suggest that the R subunit may adopt the Flipback conformation in order to bind cAMP in the A domain of PKA. With these new insights from computational simulations, we have a better understanding of the heterotetrameric structure of PKA, bridging the gap between solution and crystallographic observations.

Molecular Simulations Reveal an Unresolved Conformation of the Type IA Protein Kinase A Regulatory Subunit and Suggest Its Role in the cAMP Regulatory Mechanism

Sophia P. Hirakis, Robert D. Malmstrom, and Rommie E. Amaro*

Department of Chemistry and Biochemistry and National Biomedical Computational Resource, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0304, United States

 Supporting Information

ABSTRACT: We identify a previously unresolved, unrecognized, and highly stable conformation of the protein kinase A (PKA) regulatory subunit $RI\alpha$. This conformation, which we term the “Flipback” structure, bridges conflicting characteristics in crystallographic structures and solution experiments of the PKA $RI\alpha$ heterotetramer. Our simulations reveal a hinge residue, G235, in the B/C helix that is conserved through all isoforms of RI. Brownian dynamics simulations suggest that the Flipback conformation plays a role in cAMP association to the A domain of the R subunit.

Protein kinase A (PKA) is a ubiquitous eukaryotic kinase that modulates the function of proteins through targeted phosphorylation. An ancient cellular second messenger, cAMP, modulates the activity of PKA in a diverse set of biological processes, from synaptic plasticity¹ to cardiac signaling.² Inactive PKA exists as a heterotetramer (R_2C_2), where two catalytic (C) subunits are maintained in the inactive state by binding to the regulatory (R) subunit homodimer.³ In response to extracellular signals like adrenaline,⁴ adenylyl cyclase activity leads to an increase in the level of cAMP. When four molecules of cAMP bind to the cyclic nucleotide binding domains (CBD) of the R dimer, a global conformational change in R occurs, releasing two active C subunits, free to phosphorylate protein targets. This activation cycle is the generalized mechanism for all nonredundant forms of R ($RI\alpha$, $RI\beta$, $RII\alpha$, and $RII\beta$) and C ($C\alpha$, $C\beta$, and $C\gamma$). PKA subtypes are expressed in every cell and encoded by separate genes, differing in structure, activity, and cellular localization.^{5,6} Thus, elucidating the structural organization of PKA complexes and their relationship to activation by cAMP is important for the development of novel therapeutics and the understanding of PKA's fundamental biochemistry.

Structural biologists have elucidated PKA isoforms in various stages of the activation cycle. Although the heterotetrameric (R_2C_2) structure of the type IA protein kinase, $RI\alpha$, has never been fully resolved, it has been the subject of several studies and models. Su et al. crystallized $RI\alpha$ in complex with two molecules of cAMP.⁷ This structure of the PKA R subunit at the end of the activation cycle, known as the “Bound” conformation or “B form” (R^{Bound}), revealed the amino acid residues that are important for coordinating cAMP. Each PKA R subunit has two cyclic nucleotide binding domains (CBD-A and CBD-B) joined by a helical moiety known as the B/C helix.

The CBD is a conserved sensor of cAMP and composed of noncontiguous α helices and β barrel subdomains. At the N-terminus, a 3_{10} helix–loop region (N3A motif⁸) is followed by a β sandwich containing cAMP binding residues, and a terminal helical region (B and C-terminal helices). The cAMP-bound structure has served as an invaluable resource for understanding cAMP activation through molecular simulations⁹ and experiments.

The first structure of the RC “holoenzyme” heterodimer featured the C subunit in complex with $RI\alpha$ CBD-A ($R_A C$).¹⁰ Point mutations made at a key cAMP-interacting residue, Arg333^R in CBD-B, led to the crystallization of the dual-domain R subunit with the C subunit,¹¹ RC^{Holo} . The R subunit “Holoenzyme” conformation or “H form”, R^{Holo} , is different from the cAMP-bound structure, “Bound” or “B form”, R^{Bound} (Figure 1). The Bound conformation is globular, with CBD contacts resulting from a bent B/C/C' helix. In B form, the B/C helix breaks at L233 and Y244 (Figure 1A), bringing CBD-A into contact with N3A^B. In the RC^{Holo} structure, the R subunit wraps around the C subunit (Figure 1B), and the CBDs are separated by an extended B/C helix. In the absence of the C subunit, the H form is stabilized by cAMP analogues.¹² The cAMP-bound, “B” conformation cannot physically accommodate the C subunit because N3A^B interacts with C in RC^{Holo} while N3A^B interacts with CBD-A in R^{Bound} (Figure 1).

Despite extensive efforts, the structure of the full-length (R_{AB}), wild-type type IA ($RI\alpha$) PKA heterotetramer remains elusive. Structural models of the tetramer are available¹³ but not fully consistent with the structure and dynamics of PKA in solution. Specifically, the heterodimeric mutant R333K crystal structure has an R/C interface larger than described by hydrogen/deuterium exchange mass spectrometry (HDXMS) (Figure 1B).^{14–16}

Here, we use molecular simulation techniques to make sense of discordant experimental findings from X-ray crystallography, scattering, and HDXMS experiments. We present a novel conformation of the regulatory subunit that resolves these disparities, the “Flipback” or “F form”. Finally, we use electrostatic descriptions of the biomolecules to understand the effects of structural changes in cAMP association; offering a role for Flipback in the regulatory mechanism of PKA $RI\alpha$.

Received: May 14, 2017

Revised: June 19, 2017

Published: June 29, 2017

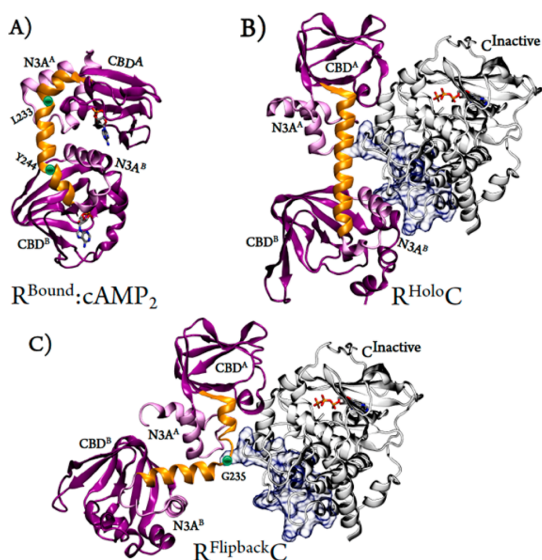


Figure 1. Comparison of the novel Flipback heterodimer with resolved PKAR1 α conformations. (A) (PDB ID 1RGS) R^{Bound}:cAMP₂, the Bound or B conformation, RI α (purple ribbon) with two molecules of cAMP (licorice). (B) (PDB ID 2QCS) R^{Holo}:C, the Holo or H form, RI α (purple ribbon) in complex with the C subunit (white ribbon). (C) R^{Flipback}:C, the Flipback conformation or F form, an MD-derived metastable state, aligned with the C α subunit (white ribbon). B/C helices are shown as gold ribbons. The N3A motifs of A and B (N3A^A and N3A^B) are colored light purple. In white, the C subunit is shown with ATP (licorice). A blue surface representation highlights the H/D exchanging regions of the C subunit measured by HDXMS.

To understand the flexibility of the apo WT R subunit, we performed all-atom molecular dynamics (MD) simulations starting from the H conformation in the absence of the C subunit and cAMP. We simulated WT H in five 200 ns replicates using the AMBER¹⁷ force field in the NTP ensemble at 310 K. MD of WT-R^{Holo} reveals a very flexible B/C helix as observed in other simulations.^{18,19} Mutagenesis of B/C helix residues recently showed pronounced effects on PKA activation.¹⁴

Our MD simulations reveal a unique, stable conformation of the R subunit (Figure 1C and Figure S1), which we call the “Flipback” conformation or “F form”, R^{Flipback}. Like R^{Bound}, R^{Flipback} features interdomain (CBD-A/B) interactions and a break in the B/C helix. However, R^{Flipback} breaks in the opposite direction, with interactions between alternate CBDs and N3A motifs. R^B uses contacts between N3A^B and CBD-A, while R^F uses N3A^A to interact with CBD-B (Figure 1B,C). In R^{Flipback}, the B/C helix breaks at Gly235^R early in the trajectory (\sim 20 ns), bringing the CBDs in contact for the rest of the simulation in a stable conformation (Figure S1). Mutations limiting the flexibility of the B/C helix (G235P) result in poor C subunit binding,²⁰ promoting activation. Residues 230–238 of the B/C helix exhibit nearly equal hydrogen bond propensities in the holoenzyme, cAMP-bound, and cAMP-free forms,¹⁵ suggesting the B/C helix is equally flexible in all structures. Gly235 is conserved among all forms of RI.²¹ Thus, a helical break at this position may be important in the activation of other type I PKA R subunits.

When R^{Flipback} is aligned with CBD-A of RC^{Holo}, it is apparent that the Flipback conformation can accommodate the C subunit, unlike R^{Bound}. We created a PKA R₂C₂ model using the F form (R₂C₂^{Flipback}). As F and H are the only known conformations that can accommodate the C subunit, we were curious to understand how conformational changes in R affect cAMP association. The diffusion of cAMP, a polar molecule, is likely influenced by long-range electrostatic forces that are estimated computationally by Brownian dynamics (BD) simulation methods.^{22–24}

Using the existing¹³ and newly constructed models of the PKA heterotetramers and heterodimers (R₂C₂^{Holo}, R₂C₂^{Flipback}, RC^{Holo}, and RC^{Flipback}, respectively), we examine the relative rate of cAMP encounter to individual CBDs via BD simulations to determine association rates ($k_{\text{association}}$). An “encounter complex” is formed when a specified distance between a set of atoms is reached. We chose three conserved residues to define encounter complexes in CBD-A/B: Val184/300, Glu200/324, and Arg209/333 (Figure S2 and Table S1). Using BrownDye,²⁵ we compare the effect of Holo and Flipback on cAMP association to tetrameric and heterodimeric conformations.

Predicted BD rates are remarkably consistent with apparent k_{on} values from experiments: 4.52×10^6 and 1.00×10^5 (CBD-B and CBD-A, respectively).²⁶ The fastest rate of cAMP encounter with PKA tetramers is that for R₂C₂^{Holo} CBD-B ($\sim 10^7$ M⁻¹ s⁻¹), while CBD-A association is slowest [$\sim 10^5$ M⁻¹ s⁻¹ (see Table 1)]. Domain B preference in H validates the

Table 1. Rates of Association of cAMP with Cyclic Nucleotide Binding Domains of PKA Complexes

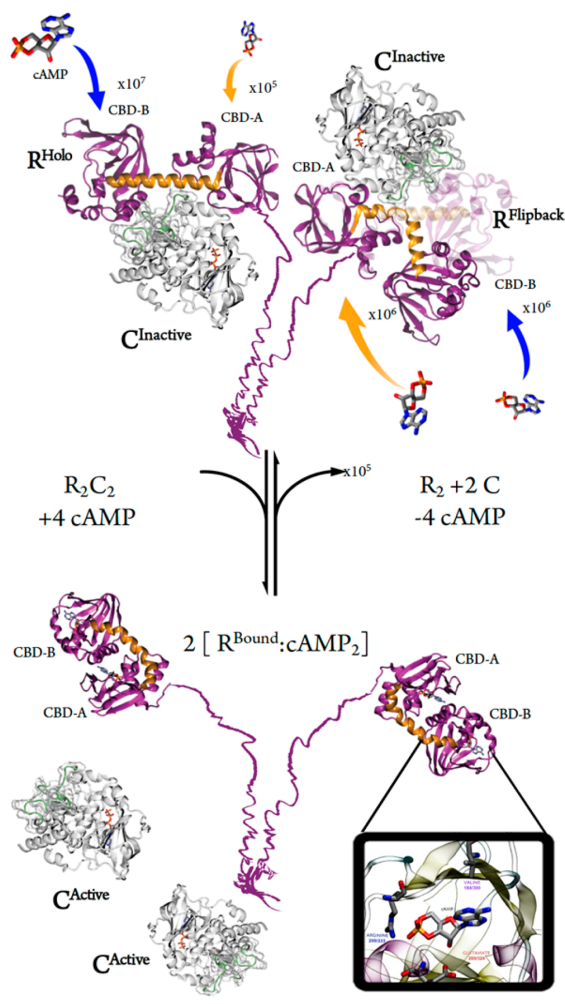
| cAMP binding domain (CBD) | PKA conformation | $k_{\text{association}}$ (M ⁻¹ s ⁻¹) |
|---------------------------|---|---|
| A | R ₂ C ₂ ^{Holo} | 3.07×10^5 |
| B | R ₂ C ₂ ^{Holo} | 2.57×10^7 |
| A | R ₂ C ₂ ^{Flipback} | 4.29×10^6 |
| B | R ₂ C ₂ ^{Flipback} | 4.16×10^6 |
| A | RC ^{Holo} | 2.50×10^4 |
| B | RC ^{Holo} | 2.72×10^4 |
| A | RC ^{Flipback} | 1.38×10^8 |
| B | RC ^{Flipback} | 1.98×10^6 |

standing “gatekeeper” theory of PKA RI α , which holds that cAMP binds to CBD-B first.²⁷ Domain B preference is neutralized in R₂C₂^{Flipback}, where both CBD-A and CBD-B bind on the order of $\sim 10^6$ M⁻¹ s⁻¹ (Table 1 and Scheme 1).

Electrostatically, the systems differ in the distribution of charge on the surfaces of PKA complexes. Flipback has a more electropositive CBD-A than Holo (Figure S4), yielding a higher cAMP association rate. The phenomenological preference for CBD-A in Flipback is most pronounced in heterodimers, with association rates being 2–4 orders of magnitude higher in F. This difference in association rates is due to the very different electrostatic potential surfaces of the heterodimer versus heterotetramer forms in H and F conformations (see Figure S4). From these results, we hypothesize that the Flipback conformation is important for association of cAMP with CBD-A. It is not known if RC heterodimers are important players in the activation mechanism of PKA, but if they are, we predict that the Flipback conformation plays a role in CBD-A association.

The only resolved conformation of full-length R in complex with C necessitated the R333K mutation to stabilize CBD-B.¹¹

Scheme 1. Activation Mechanism of PKA RI α with Different R Conformations and Relative Association Rates



However, it has been shown that the R333K mutant ($R^{C^{\text{Holo}}}$) has a solution structure different from that of WT. Small-angle X-ray scattering (SAXS) of the full-length WT heterodimer and heterotetramer²⁸ exhibit a shouldering region, one not observed in symmetric SAXS $p(r)$ distributions of mutant RC (R333K) and $R_A C$, a heterodimer lacking the B domain.²⁹

The Flipback heterotetramer structure corroborates observations from multiple solution experiments. First, the C subunit interface of the WT full-length holoenzyme measured by HDXMS is consistent with the R/C interface of the $R_2C_2^{\text{Flipback}}$ structure, where the amides of residues 212–221^C and 278–289^C were reported to be unprotected.¹⁴ Second, structural models from scattering experiments suggest that conformational changes in R cause the release of one set of R/C contacts,³⁰ consistent with the Flipback conformation (Scheme 1). It has been shown that binding of cAMP to CBD-B leads to an increase in the rate of H/D exchange at the B/C helix.³¹ It is likely that once binding to CBD-B has occurred, the Flipback conformation is formed, resulting in CBD-A association. Attempts to elucidate the heterotetrameric structure of PKA

RI α with SAXS have proposed a Flipback-like conformation.²⁸ Finally, the majority of R/C on RI α contact have been traced to interactions in CBD-A, consistent with Flipback.^{10,14–16,21,29,32,33}

Flipback dynamics are consistent with other molecular simulations and models.^{27,34} Guo and Zhou recently observed a flexible B/C helix when simulating the apo-B form,⁹ the opposite of the starting point of our H form simulations. A backward-bending B/C helix is observed in these simulations, in a conformation resembling Flipback. This suggests that F is a state that can be accessed from both B and H forms.

An interesting role for Flipback emerges when we consider the termination phase of PKA regulation. The phosphodiesterase enzyme (PDE) hydrolyzes cAMP to 5'AMP, regulating concentration of the second messenger. Computational docking and HDXMS determined that for PDE to bind RI α subunit, the B/C helix requires a complete reorganization.³⁵ It is possible that Flipback is a binding partner of PDE, though further examination of this hypothesis is necessary.

Our simulations, coupled with experimental data, make the case for a viable and stable Flipback conformation of PKA RI α that may play important roles in the cAMP regulatory mechanism. Our work reveals a new structure of the WT PKA R subunit, which supports observations from ensemble-averaged solution structures and experiments. BD suggests a role for the R^{Flipback} conformation in the mechanism of PKA activation. We hope our findings will lead to a re-examination PKA, especially with regard to differences between the conformations of WT and R333K mutants and the role of structural ensembles in ligand binding and, ultimately, signal transduction.

■ ASSOCIATED CONTENT

§ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.7b00461.

Detailed experimental methods, MD of the Flipback structure (Figure S1), the encounter complex (Figures S2 and S3 and Table S2), and electrostatic profiles of Holo and Flipback (Figure S4) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: ramaro@ucsd.edu.

ORCID

Rommie E. Amaro: 0000-0002-9275-9553

Author Contributions

S.P.H. wrote the manuscript, analyzed MD, and performed and designed BD experiments. R.D.M. performed and designed MD experiments. R.E.A. directed the research.

Funding

This research was funded by a fellowship from the Interfaces Graduate Training Program (NBIB T32 EB009380), by the National Biomedical Computational Resource (NIH P41-GM103426), and by National Institutes of Health Grant DP2 OD007237.

Notes

The authors declare the following competing financial interest(s): R.E.A. is a co-founder of, on the scientific advisory board of, and equity holder in Actavalon, Inc.

ACKNOWLEDGMENTS

We especially acknowledge Dr. Alexandr Kornev for frequent and helpful discussions. We acknowledge Drs. Jamie M. Schiffer, Susan S. Taylor, and Elizabeth Komives for their insights. Finally, we acknowledge Dr. Gary Huber and Dr. Lane W. Votapka for help and advice with BD simulations.

ABBREVIATIONS

PKA, protein kinase A; cAMP, cyclic adenosine monophosphate; CBD, cyclic nucleotide binding domain; B, cAMP-Bound; H, Holo; F, Flipback.

REFERENCES

- (1) Esteban, J. A., Shi, S. H., Wilson, C., Nuriya, M., Huganir, R. L., and Malinow, R. (2003) *Nat. Neurosci.* 6, 136–143.
- (2) Kamp, T. J., and Hell, J. W. (2000) *Circ. Res.* 87, 1095–1102.
- (3) Johnson, L. N., Noble, M. E. M., and Owen, D. J. (1996) *Cell* 85, 149–158.
- (4) Chasiotis, D. (1985) *Acta Physiol. Scand.* 125, 537–540.
- (5) Taylor, S. S., Zhang, P., Steichen, J. M., Keshwani, M. M., and Kornev, A. P. (2013) *Biochim. Biophys. Acta, Proteins Proteomics* 1834, 1271–1278.
- (6) Taylor, S. S., Ilouz, R., Zhang, P., and Kornev, A. P. (2012) *Nat. Rev. Mol. Cell Biol.* 13, 646–658.
- (7) Su, Y., Dostmann, W. R. G., Herberg, F. W., Durick, K., Xuong, N. H., Teneyck, L., Taylor, S. S., and Varughese, K. I. (1995) *Science* 269, 807–813.
- (8) Kornev, A. P., Taylor, S. S., and Ten Eyck, L. F. (2008) *Proc. Natl. Acad. Sci. U. S. A.* 105, 14377–14382.
- (9) Guo, C., and Zhou, H. X. (2016) *Proc. Natl. Acad. Sci. U. S. A.* 113, E6776–E6785.
- (10) Kim, C., Xuong, N. H., and Taylor, S. S. (2005) *Science* 307, 690–696.
- (11) Kim, C., Cheng, C. Y., Saldanha, S. A., and Taylor, S. S. (2007) *Cell* 130, 1032–1043.
- (12) Badireddy, S., Gao, Y. F., Ritchie, M., Akamine, P., Wu, J., Kim, C. W., Taylor, S. S., Lin, Q. S., Swaminathan, K., and Anand, G. S. (2011) *Mol. Cell. Proteomics* 10, M110.004390.
- (13) Boettcher, A. J., Wu, J. A., Kim, C., Yang, J., Bruystens, J., Cheung, N., Pennypacker, J. K., Blumenthal, D. A., Kornev, A. P., and Taylor, S. S. (2011) *Structure* 19, 265–276.
- (14) Anand, G. S., Hotchkro, M., Brown, S. H., Ten Eyck, L. F., Komives, E. A., and Taylor, S. S. (2007) *J. Mol. Biol.* 374, 487–499.
- (15) Anand, G. S., Hughes, C. A., Jones, J. M., Taylor, S. S., and Komives, E. A. (2002) *J. Mol. Biol.* 323, 377–386.
- (16) Anand, G. S., Law, D., Mandell, J. G., Sneed, A. N., Tsigelny, I., Taylor, S. S., Ten Eyck, L. F., and Komives, E. A. (2003) *Proc. Natl. Acad. Sci. U. S. A.* 100, 13264–13269.
- (17) Salomon-Ferrer, R., Gotz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013) *J. Chem. Theory Comput.* 9, 3878–3888.
- (18) Pecora de Barros, E., Malmstrom, R. D., Nourbakhsh, K., Del Rio, J. C., Kornev, A. P., Taylor, S. S., and Amaro, R. E. (2017) *Biochemistry* 56, 1536–1545.
- (19) Malmstrom, R. D., Kornev, A. P., Taylor, S. S., and Amaro, R. E. (2015) *Nat. Commun.* 6, 7588.
- (20) Cheng, C. Y. (2009) Dissecting the allosteric regulation of PKA-I alpha activation. Ph.D. Thesis, University of California, San Diego, La Jolla, CA.
- (21) Sjoberg, T. J., Kornev, A. P., and Taylor, S. S. (2010) *Protein Sci.* 19, 1213–1221.
- (22) Northrup, S. H., Allison, S. A., and Mccammon, J. A. (1984) *J. Chem. Phys.* 80, 1517–1526.
- (23) Ermak, D. L., and Mccammon, J. A. (1978) *J. Chem. Phys.* 69, 1352–1360.
- (24) Votapka, L. W., and Amaro, R. E. (2015) *PLoS Comput. Biol.* 11, e1004381.
- (25) Huber, G. A., and McCammon, J. A. (2010) *Comput. Phys. Commun.* 181, 1896–1905.
- (26) Herberg, F. W., Taylor, S. S., and Dostmann, W. R. G. (1996) *Biochemistry* 35, 2934–2942.
- (27) Boras, B. W., Kornev, A., Taylor, S. S., and McCulloch, A. D. (2014) *J. Biol. Chem.* 289, 30040–30051.
- (28) Vigil, D., Blumenthal, D. K., Heller, W. T., Brown, S., Canaves, J. M., Taylor, S. S., and Trehwella, J. (2004) *J. Mol. Biol.* 337, 1183–1194.
- (29) Cheng, C. Y., Yang, J., Taylor, S. S., and Blumenthal, D. K. (2009) *J. Biol. Chem.* 284, 35916–35925.
- (30) Heller, W. T., Vigil, D., Brown, S., Blumenthal, D. K., Taylor, S. S., and Trehwella, J. (2004) *J. Biol. Chem.* 279, 19084–19090.
- (31) Moorthy, B. S., Badireddy, S., and Anand, G. S. (2011) *Int. J. Mass Spectrom.* 302, 157–166.
- (32) Hamuro, Y., Anand, G. S., Kim, J. S., Juliano, C., Stranz, D. D., Taylor, S. S., and Woods, V. L. (2004) *J. Mol. Biol.* 340, 1185–1196.
- (33) Huang, L. J., and Taylor, S. S. (1998) *J. Biol. Chem.* 273, 26739–26746.
- (34) Boras, B. W., Hirakis, S. P., Votapka, L. W., Malmstrom, R. D., Amaro, R. E., and McCulloch, A. D. (2015) *Front. Physiol.* 6, 250.
- (35) Krishnamurthy, S., Moorthy, B. S., Xiang, L. X., Shan, L. X., Bharatham, K., Tulsian, N. K., Mihalek, I., and Anand, G. S. (2014) *Biophys. J.* 107, 1426–1440.

Supplementary information

Molecular simulations reveal an unresolved conformation of the Type-IA Protein Kinase A regulatory subunit and suggest its role in the cAMP regulatory mechanism

Sophia P. Hirakis, Robert D. Malmstrom, Rommie E. Amaro*

Department of Chemistry and Biochemistry and National Biomedical Computational Resource, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0304

*corresponding author

Methods

Molecular Dynamics using GPU accelerated AMBER 12

We identified R^{Flipback} as part of a series of molecular dynamics (MD) simulations studying the stability of the WT R^{Holo} conformation (residues 113-379) in the absence of the catalytic subunit. The atomic coordinates for the starting conformation of the regulatory subunit (R^{Holo}) were taken from the heterodimeric structure of the R333K mutant, RC^{Holo} . As a starting conformation for R^{Holo} , we used the Protein Databank structure, 2QCS.¹ Residue K333 was changed back to R333 to undo the mutation and simulate wild-type (WT) R subunit. The system was prepared using Schrodinger's Maestro (Suite 2012: Maestro, version 9.3, Schrödinger, LLC, New York, NY, 2012) PDB prep and Desmond system preparer. Terminal residues were capped to remove charges. Titratable residues were protonated at pH 7.0 using Maestro integrated PROPKA.²⁻⁵ The system was solvated in cubic water box with 10Å buffer using TIP3 waters⁶ and 0.15M NaCl. Using xleap from Ambergtools⁷ the systems parameterized using AMBER ff99SB force field⁸ and periodic boundaries were applied. The system was minimized in the following four stages: 1) 2,000 steps of hydrogen only minimization, 2) 4,000 steps of solvent minimization, 3) 20,000 steps with constrained backbone, and 4) 40,000 steps full minimization. MD simulations were performed as an NTP ensemble, at 310K and 1 bar, with a 2 fs time step and a 10 Å non-bonding interaction cutoff with a partial mesh Ewald approximation for long-range electrostatics interactions. The system was heated and then equilibrated over 1ps with harmonic constraints on the backbone. Production was performed as five parallel runs with new initial starting vectors for 200 ns each. The Flipback structure (R^{Flipback}) was observed in one of the five runs and was identified through visual inspection with VMD.⁹

Preparation of Structures for Brownian Dynamics

The atomic structure of R subunit from the cAMP-bound structure (R^{Bound} ; PDBID:1RGS)¹¹ was used to determine the encounter complex. The RC^{Holo} heterodimer (PDBID:2QCS)¹ was used for BD simulations of and RC^{Holo} . The R^{Flipback} structure was a configuration taken from one of five molecular dynamics simulations starting from the R^{Holo} conformation. The Susan Taylor lab at the University of California, San Diego, provided the structural model of heterotetrameric $R_2C_2^{\text{Holo}}$ to us. This structure was validated through mutational experiments and published in 2011¹². We created the $R_2C_2^{\text{Flipback}}$ tetrameric structure by aligning the A-domain of the R^{Flipback} conformation with the CBD-A with of the regulatory subunit of the $R_2C_2^{\text{Holo}}$ structure using PyMol¹³. Residues 113 to 376 of the regulatory subunit were used for all simulations. Residues 13-350 were used for the C_α subunit. Titratable residues were assigned a protonation state using PROPKA 3.1 pKa prediction software^{3-5, 14} on the PDB2PQR server²⁻⁵ and manually inspected for accurate state assignment at a pH of 7.0.

To facilitate the crystallization of the heterotetramer trapped in the inactive conformation, RC^{Holo} was crystallized with phosphoaminophosphonic acid-adenylate ester (ANP) and Manganese ions (Mn^{2+}) in the active site of C subunit. To ensure the accurate electrostatic description of C subunit, we replaced ANP with adenosine triphosphate (ATP) and Mn^{2+} with Magnesium ions (Mg^{2+}). Parameters for ATP and Mg^{2+} were obtained from the Bryce Group AMBER parameters database.^{15, 16} The edited PDB files were fed through tleap¹⁷ to create parameter topology and coordinate files consistent with the Amber ff14SB forcefield⁸. PQR files consistent with the addition of Mg^{2+} and ATP in the structures were exported and formatted using ambpdb.¹⁸ The atomic coordinates of cAMP from the R^{Bound} crystal structure and parameterized using Antechamber.¹⁹

To eliminate the effects of simulated vs. crystal structure conformations of R, all structures were solvated and relaxed into a stable conformation. The NAMD simulation package^{20, 21} was used to minimize, heat, equilibrate, and simulate each system using a 2fs time-step. Every system underwent a series of separate minimization, heating, and equilibration stages in preparation for production runs in the following order. Minimization spanned five stages in 10ps intervals using the NVT ensemble: 1) 5,000 steps of hydrogen-only minimization, 2) 5,000 steps of solvent minimization, 3) 5,000 steps of side-chain minimization, 4) 5,000 steps of protein-backbone minimization, and 5) 5,000 steps of full-system minimization. The Langevin thermostat^{22, 23} was used to slowly heat the system to 310K using the NVT ensemble over 250,000 steps

(500ps) following minimization. The system was then subject to three sequential equilibration stages using the NPT ensemble for 125,000 steps/stage (250ps/stage). The pressure was set to 1 atm and maintained using the Berendsen barostat.²⁴ The solvent was modeled explicitly using the TIP4P water model²⁵ and a 0.15M sodium chloride concentration was applied after neutralizing the overall charge of the protein complexes. The Particle Mesh Ewald electrostatic summation method^{26,27} was employed to evaluate electrostatics during simulation.

Brownian Dynamics Simulations using BrownDye

Brownian Dynamics simulations with BrownDye²⁸ require PQR files as input, which list the position, charge, and radius of each atom in the respective molecules to be simulated. All necessary PQR files were generated using *ambpdb*—a program from the AMBER suite.⁷ PQR files were converted to XML format through *pqr2xml*, an accessory program of BrownDye. APBS²⁹ was used to generate the electrostatic field of all molecules immersed in a 0.15M NaCl implicit solvent. *Bd_top*, an accessory program of BrownDye, generated all necessary input files for BD simulations. 5,000,000 single trajectory simulations were performed using *nam_simulation* from the BrownDye suite. Association rate constants and reaction probabilities were calculated using the *compute_rate_constant* function of BrownDye.

The encounter complex is described by the distance between interacting atoms in a “bound” complex. The criteria for cAMP association to R subunit was based on cAMP interactions in CBD-A of R^{Bound} crystal structure.¹¹ To control for the effect of sidechain orientation differences, CBD-A of R^{Bound}, R^{Holo}, and R^{Flipback} were aligned and atoms with less than 1.0 Å deviation were selected for the binding criteria. The atomic interactions chosen for the encounter complex are shown in Figure S1 and listed in Table S1. A distance of 7 Å was chosen for interacting atoms, consistent with the crystal structure conformation of R^{Bound}.

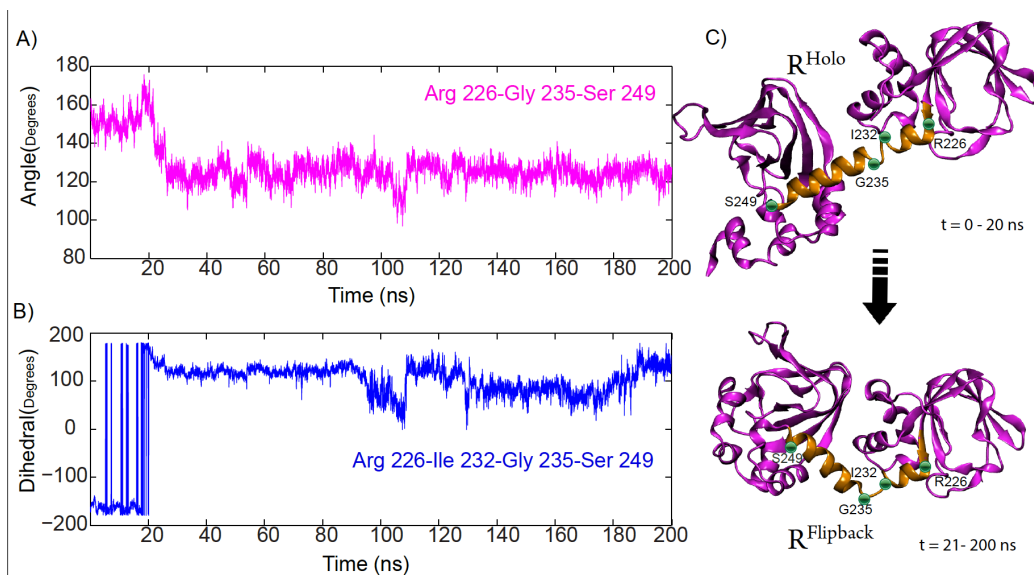


FIGURE S1. The stability of the Flipback conformation. (A) Change in B/C helix angle over the course of the MD simulation. The angle was measured between Arg226-Gly235-Ser249; (B) Change in B/C helix dihedral angle over the course of the MD simulation. The angle was measured between Arg226-Ile233-Gly235-Ser249; (C) Change in conformation of R throughout the MD simulation. CBD-A and CBD-B are shown in pink. The B/C helix is shown in gold. The position of B/C helix residues used in Figure S1A and Figure S1B are shown in green dots.

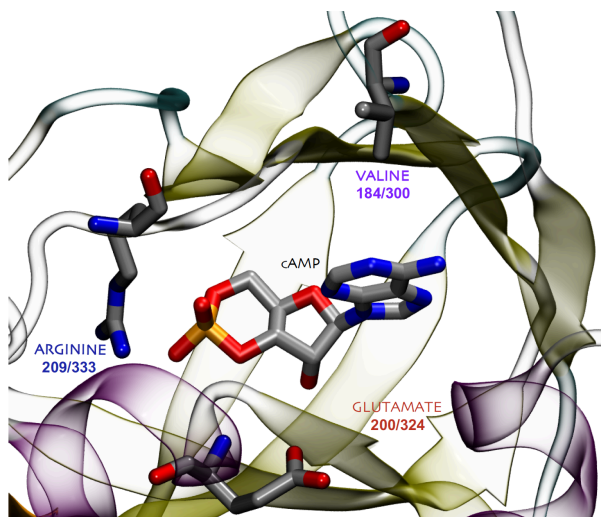


Figure S2. Generalized Encounter complex of cAMP and the CBD-A/B. cAMP is shown in the cyclic-nucleotide binding domain (CBD) of the regulatory subunit of Protein Kinase A RI α . Residue names are included next to the licorice representation of the sidechains and numbering of residues are numbered in CBD-A/CBD-B format. Hydrogen atoms are excluded. Interacting residues of the R subunit were chosen to be Valine, Arginine, and Glutamate for both CBD-A and CBD-B.

Table S1. Encounter Complex Description of R subunit and cAMP

| CBD | R subunit residue number | R subunit atom name | cAMP Atom name and number |
|-----|--------------------------|---------------------|--|
| A | VAL 184 | CB | C8 (imidazole ring) |
| A | GLU 200 | CA, CB | O5 (ribose ring) |
| A | ARG 209 | CB | O1, O2 (α , β Phosphate) |
| B | VAL 300 | CB | C8 (imidazole ring) |
| B | GLU 324 | CA, CB | O5 (ribose ring) |
| B | ARG 333 | CB | O1, O2 (α , β Phosphate) |

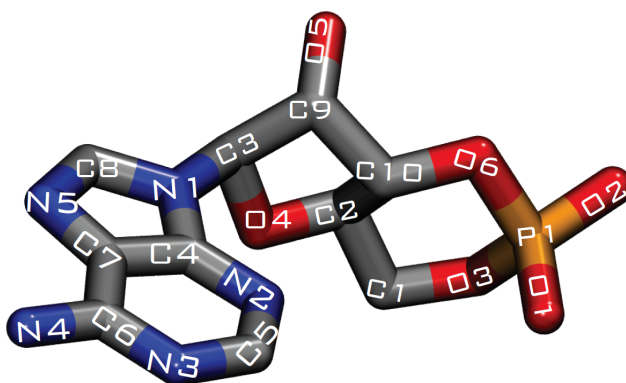


Figure S3. Atomic numbering and naming of cAMP molecule in BD simulations. cAMP atoms are named and numbered according to PDB format. Carbon (gray), Nitrogen (blue), Oxygen (red), and Phosphorous (Orange) atoms are labeled with the corresponding atom name overlaid in white

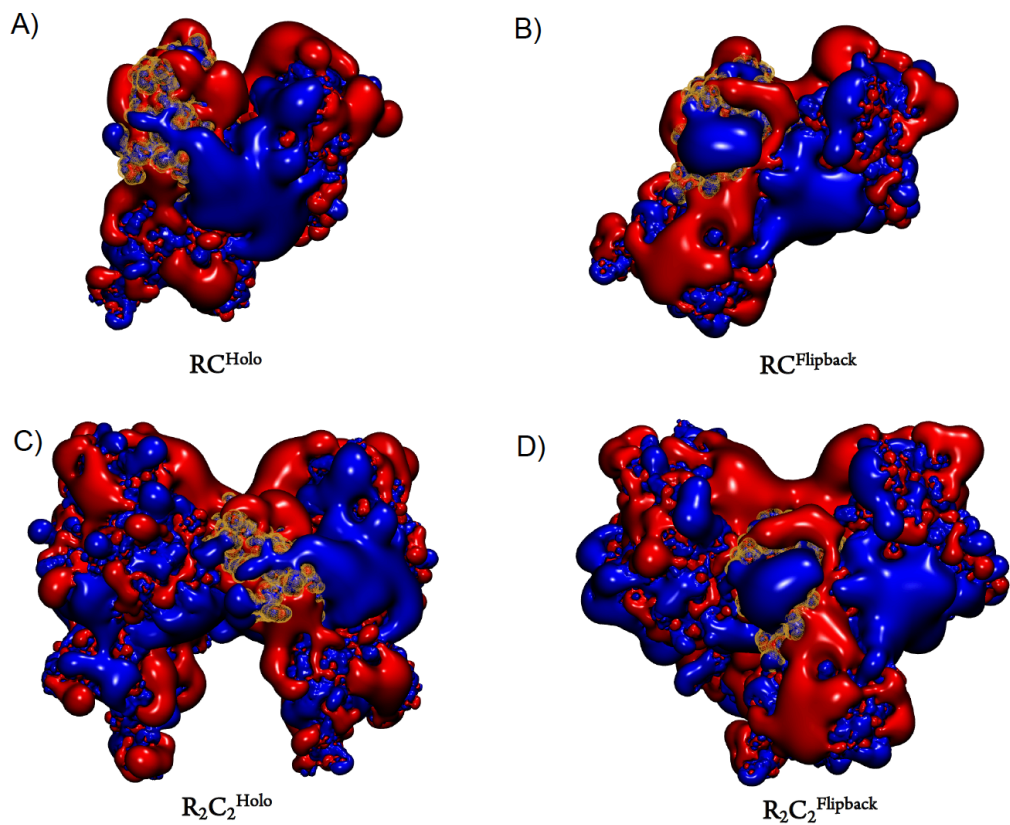


Figure S4. Electrostatic descriptions of the four systems. (A) Electrostatic description of RC^{Holo} ; (B) Electrostatic description of $RC^{Flipback}$; (C) Electrostatic description of $R_2C_2^{Holo}$; (D) Electrostatic description of $R_2C_2^{Flipback}$; Domain A is highlighted in gold wire mesh. An isosurface value of 0.75 was used to create the positive and negative potentials. Electropositive potentials are shown in blue and electronegative potentials are shown in red.

References

- [1] Kim, C., Cheng, C. Y., Saldanha, S. A., and Taylor, S. S. (2007) *Cell* 130, 1032-1043.
- [2] Li, H., Robertson, A. D., and Jensen, J. H. (2005) *Proteins* 61, 704-721.
- [3] Bas, D. C., Rogers, D. M., and Jensen, J. H. (2008) *Proteins* 73, 765-783.
- [4] Olsson, M. H. M., Sondergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011) *J Chem Theory Comput* 7, 525-537.
- [5] Sondergaard, C. R., Olsson, M. H. M., Rostkowski, M., and Jensen, J. H. (2011) *J Chem Theory Comput* 7, 2284-2295.
- [6] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) *J Chem Phys* 79, 926-935.
- [7] D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Götz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman. (2012) AMBER 12, University of California, San Francisco.
- [8] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) *Proteins* 65, 712-725.
- [9] Humphrey, W., Dalke, A., and Schulten, K. (1996) *J Mol Graph Model* 14, 33-38.
- [10] Levine, B. G., Stone, J. E., and Kohlmeyer, A. (2011) *J Comput Phys* 230, 3556-3569.
- [11] Su, Y., Dostmann, W. R. G., Herberg, F. W., Durick, K., Xuong, N. H., Teneyck, L., Taylor, S. S., and Varughese, K. I. (1995) *Science* 269, 807-813.
- [12] Boettcher, A. J., Wu, J., Kim, C., Yang, J., Bruystens, J., Cheung, N., Pennypacker, J. K., Blumenthal, D. A., Kornev, A. P., and Taylor, S. S. (2011) *Structure* 19, 265-276.
- [13] The PyMOL Molecular Graphics System, Version 1.2r3pre ed., Schrödinger, LLC.
- [14] Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., and Baker, N. A. (2007) *Nucleic Acids Res* 35, W522-W525.
- [15] Meagher, K. L., Redman, L. T., and Carlson, H. A. (2003) *J Comput Chem* 24, 1016-1025.
- [16] Allner, O., Nilsson, L., and Villa, A. (2012) *J Chem Theory Comput* 8, 1493-1502.
- [17] Christian E. A. F. Schafmeister, W. S. R. a. V. R. (1995) Leap, University of California, San Francisco.
- [18] Case D.A., D. T. A., Cheatham T.E, III, Simmerling C.L., Wang J. , Duke R.E., Luo R., Walker R.C., Zhang W., Merz K.M., Roberts B., Hayik S., Roitberg A., Seabra G., Swails J., Götz A.W., Kolossváry I., Wong K.F., Paesani F., Vanicek J., Wolf R.M., Liu J., Wu X., Brozell S.R., Steinbrecher T., Gohlke H., Cai Q., Ye X., Wang J., Hsieh M.-J., Cui G., Roe D.R., Mathews D.H., Seetin M.G., Salomon-Ferrer R., Sagui C., Babin V., Luchko T., Gusarov S., Kovalenko A., and Kollman P.A. . (2012) AMBER 12, University of California, San Francisco.
- [19] Case D.A., B. J. T., Betz R.M., Cerutti D.S., Cheatham T.E., III, Darden T.A., Duke R.E., Giese T.J., Gohlke H., Goetz A.W., Homeyer N., Izadi S., Janowski P., Kaus

- J., Kovalenko A., Lee T.S., LeGrand S., Li P., Luchko T., Luo R., Madej B., Merz K.M., Monard G., Needham P., Nguyen H., Nguyen H.T., Omelyan I., Onufriev A., Roe D.R., Roitberg A., Salomon-Ferrer R., Simmerling C.L., Smith W., Swails J., Walker R.C., Wang J., Wolf R.M., Wu X., York D.M., and Kollman P.A. (2015) AMBER 2015, University of California, San Francisco.
- [20] Nelson, M. T., Humphrey, W., Gursoy, A., Dalke, A., Kale, L. V., Skeel, R. D., and Schulten, K. (1996) *Int J Supercomput Ap* 10, 251-268.
- [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) *J Comput Chem* 26, 1781-1802.
- [22] Izaguirre, J. A., Catarella, D. P., Wozniak, J. M., and Skeel, R. D. (2001) *J Chem Phys* 114, 2090-2098.
- [23] Jiang, W., Hardy, D. J., Phillips, J. C., MacKerell, A. D., Schulten, K., and Roux, B. (2011) *J Phys Chem Lett* 2, 87-92.
- [24] Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984) *J Chem Phys* 81, 3684-3690.
- [25] Li, P. F., Roberts, B. P., Chakravorty, D. K., and Merz, K. M. (2013) *J Chem Theory Comput* 9, 2733-2748.
- [26] Darden, T., York, D., and Pedersen, L. (1993) *J Chem Phys* 98, 10089-10092.
- [27] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) *J Chem Phys* 103, 8577-8593.
- [28] Huber, G. A., and McCammon, J. A. (2010) *Comput Phys Commun* 181, 1896-1905.
- [29] Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) *P Natl Acad Sci USA* 98, 10037-10041.

3.1 Acknowledgment

Chapter 3, in full, is a reprint of the material as it appears Biochemistry 2017. Hirakis, Sophia P.; Malmstrom, Robert D.; Amaro, Rommie E.; “Molecular Simulations Reveal an Unresolved Conformation of the Type IA Protein Kinase A Regulatory Subunit and Suggest its Role in the cAMP Regulatory Mechanism.” The dissertation author was the primary investigator and author of this paper.

This discovery would not have been possible without the original simulations by Dr. Robert D. Malmstrom. Dr. Malmstrom allowed me to analyze his simulations and through this, I discovered the Flipback conformation. Dr. Malmstrom assisted in the manuscript review. Though not listed as an author, special thanks goes out to Dr. Alexandr Kornev who assisted in the understanding of the Flipback conformation in the context of the existing literature on PKA heterotetramer. He provided critical edits to the document which helped project it towards success and acceptance in the field. Dr. Susan S. Taylor provided insights and I thank her for the many conversations about this. Dr. Lane W. Votapka and Dr. Gary Huber assisted with the construction of the BrownDye simulations and for this, I thank them. Dr. Jamie M. Schiffer helped understand the SAXS profiles of PKA tetramer. Dr. Elizabeth Komives provided support for the discovery of the structure and much welcomed enthusiasm for the publication of the manuscript. Finally, I thank my beloved advisor, Dr. Rommie E. Amaro for pushing me to publish this work and for helping me to see the value in my findings, which the world may never have known of if not for her invaluable encouragement. Thus, cardiac function spans multiple scales of space and time, from that which can be seen with the naked eye, to that which only a computational microscope can observe.

Chapter 4

Subcellular spatial modeling in realistic geometries with stochastic particle methods to understand heart disease

4.1 Abstract

The evasive source and cause of diseases is oftentimes *smaller than you think*. Imagine, though, chasing something that you can't actually *see*. This is the plight of almost all scientists whose study does not involve the use of a microscope. Fortunately for the modern-day biomedical scientist, computational advancements harnessing the near-limitless potential of mathematics in the language of physics are able to *see the unseen*. Computational microscopy is a tool developed by thousands of talented scientists that yields the power of visualization to the scientist seeking to visually understand biological systems. With progressive advancements in the power of computer graphics and the development of mathematical theories to explain biological behavior, computational microscopy is a term given to a collection of methods developed by hundreds of scientists throughout the greater half of the last century. Using this powerful tool, the artist within the scientist is able to visualize what they know in their mind's eye to be true through their experiments. It also allows scientific discoveries to be translated without the explicit use of language. With respect to diseases, the culprit is oftentimes an atomic-level alteration that has a butterfly effect throughout the cell, scaling up through space and time to affect organs and

eventually, the whole organism. In accordance with its name, the computational microscope not only allows us to visualize extremely small entities like atoms, molecules, proteins, cells, and the like. But it is especially unique, as it allows us to move through time in order to understand the dynamics of the system we seek to “see”. Like a biophysically detailed time-lapse, we are able to *see through time*, the development of a microscopic system and compare how small changes can have large-scale effects.

4.2 Introduction

Over 135 years ago, Dr. Singer demonstrated that a human heart would not beat in the absence of Ca^{2+} ions [112]. Yet, human heart is the size of a fist. Thus the process of the heartbeat is an inherently multiscale process. The atria and ventricles of the heart make up the “working chambers” which pump oxygenated blood through the body, and deliver deoxygenated blood to the lungs. The thickness of these walls are several millimeters depending on the region of the heart. Each cell is a complicated but regular structure of membranous invaginations that range from 20-450 nm in diameter. Cardiomyocytes, or muscle cells, contain bundles of muscle fibers that are responsible for the contraction of the heart. The thick filaments are made out of myosin protein and are about that are 160-170 Å in diameter while thinner actin filaments are 6 to 10 Å in diameter. Finally, large proteins in close contact with each other (roughly 20 nm) sense calcium ions (roughly 231 pm radius), triggering processes the contraction and relaxation of muscle fibers, resulting in the beating of the heart. Thus, cardiac function spans scales that can be observed with the naked eye, to invisible phenomena that only a computational microscope can see.

A healthy heart contracts in response to a synchronous electrical stimulation of the plasma membrane, also known as the sarcolemma, initiated by the sinoatrial (SA) node. The membrane action potential travels from the SA node and is propagated from cell to cell through gap junctions [113]. This electrical stimulation results in membrane depolarization, to which the cell responds with a muscular contraction, a phenomenon known as excitation-contraction coupling (ECC) [114].

Cardiomyocytes have specialized structures within which this process occurs. Invaginations in the sarcolemma (cell membrane), known as axial and transverse-tubules (TT) are positioned directly adjacent to the sarcoplasmic reticulum (SR), the intracellular calcium store. Depolarization of the sarcolemma activates and opens L-Type Calcium Channels (LTCC), through which calcium enters the cell. The LTCC are in close proximity large ($> 2\text{MDa}$) SR membrane proteins called Ryanodine Receptors (RyR) [115]. The space between the extracellular membrane and SR membrane is known as the dyadic junction and *this is where the magic happens*.

High calcium concentrations in the extracellular matrix cause influx of Ca^{2+} through LTCC, into the cardiac muscle cells which maintain a very low Ca^{2+} concentration. Directly adjacent to LTCC are RyR, which are regulated by more than 30 proteins [116]. When RyR sense changes in cytosolic calcium, a *signal amplification* results, whereby RyR release thousands of Ca^{2+} ions from the SR into the cell, an event known as a triggered calcium spark or Calcium Induced Calcium Release (CICR). In response to the change in cytosolic Ca^{2+} , the inhibition of action and myosin is relieved by Troponin C (TnC) which directly bind Ca^{2+} . This causes the muscle fiber proteins to slide across each other leading to muscle fiber shortening, or muscular contraction that squeezes blood out of the atrial chambers of the heart into the ventricles and out towards the rest of the body. An important SR protein, the Sarco/Endoplasmic Reticulum Ca^{2+} -ATPase (SERCA) pump, clears the cytosol of the high Ca^{2+} levels, allowing for muscle relaxation. This cycle is repeated every time the heart beats.

Early on, LTCC (also known as Dihydropyridine Receptors, and Voltage-Dependent Calcium Channels) [117] and RyR [118] were implicated in the generation of calcium sparks. Though theories about ECC existed [119], the first Ca^{2+} spark event was visualized and confirmed 25 years ago using laser-scanning confocal microscopy [120] and quickly confirmed in subsequent studies [121,122]. Within a few years, computer simulations were applied to this system to model the elementary events responsible to elucidate the subcellular mechanisms responsible for what was visualized with the early fluorescence measurements [123].

Microscopic fluorescence imaging techniques have limitations in terms of their ability to resolve spatial changes in calcium levels; especially at short space and timescales. Therefore,

computational models have played an important role to tease out important features of CICR and ECC. The field of computational cardiology has an impressive track-record of over 30 years of computational modeling of cardiac excitation phenomena that have explained and predicted mechanisms that underlie calcium signaling [124]. Almost all of the existing models use idealized geometries of myocardium that do not encapsulate the structural complexities of cardiac cells. Moreover, the field largely uses continuum approaches to model the dynamics of cardiac systems. At low molecular concentrations such as those exhibited in the cardiomyocyte.

The first use of stochastic, explicit-particle simulations with MCell to investigate cardiac Ca^{2+} signaling mechanisms was performed by Koh et al. in 2006 [125]. Their study showed the effects of altering dyadic distances had a pronounced effect on Ca^{2+} SR fluxes. It featured the use of simplistic planar geometries to model the SR and plasma membrane, but paved the way for future discrete-modeling methods in cardiac systems. In 2012, Hake et al. debuted the first subcellular model of a Ca^{2+} spark [33] using realistic geometries of a cardiac calcium release unit [32]. The methods used in the study were continuum-based and deterministic, and modeled phenomenological calcium activation.

In the present study, we combine the realistic geometries used by Hake et al. with stochastic models approaches using particle-based, spatial reaction-diffusion modeling methods [125] by Koh to debut the first-ever model of discrete calcium dynamics in realistic geometries of mouse myocardium. We use our model to investigate the effects of disease phenotypes, such as T-Tubule deformation [126], RyR dispersion [127], and alterations of the mouse action potential [128] on calcium signaling.

4.3 Methods

4.3.1 Building the Geometric Model

The our model builds upon and extends the model built by Hake et al. [33], which was the first to use electron tomography-derived geometries for calcium spark simulations. We used the same realistic geometry of the calcium release unit (CRU) imaged by Hayashi et al. [32] which were segmented by IMOD [129] and meshed with GAMer [30,31]. The geometry files

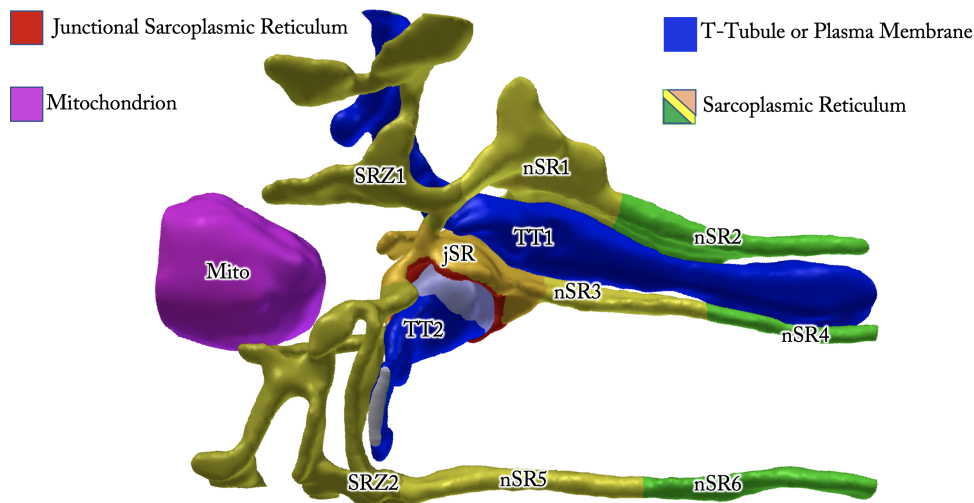


Figure 4.1. The Calcium Release Unit geometry

The Sarcoplasmic Reticulum (SR) (green, yellow, orange) is subdivided into network SR (nSR1, nSR2, nSR3, nSR4, nSR5, nSR6) Z-line SR (SRZ1, SRZ2) and junctional SR (jSR in red). The Axial T-Tubule (TT1) and Transverse T-Tubule (TT2) (blue) are in close proximity to the SR.

A single mitochondrion (Mito in pink) is shown in the geometry.

were graciously provided to us through correspondence with Dr. Johan E. Hake, the primary author of the original realistic CRU paper [33]. We were also very fortunate to have access to the original EM images through the local National Computational Microscopy Imaging Resource (NCMIR) at UCSD. Although the original segmented images did not include the explicit locations of the Ryanodine Receptors (RyR) in the junctional SR, images of the locations were graciously provided to us by Masahiko Hoshijima of NCMIR. In total, 96 RyR were observed in the original CRU tomograms and this number of RyR were used in our simulations and manually placed in the CRU.

The geometry features a contiguous sarcoplasmic reticulum (SR), two mitochondria, and one axial and one transverse tubule (TT1 and TT2 respectively). In order for the geometry mesh to be usable by our simulation interface and simulation engine, (CellBlender and MCell) it was necessary to further refine the mesh. For this, we used an improved version of GAMer developed locally at the National Biomedical Computational Resource (NBCR) at UC San Diego (UCSD) [30, 31].

To divide the mesh into “model objects” usable by MCell and Cellblender, we used the

colors attributed to the original meshes as a way to classify the objects into separate regions. The model objects were named as follows: 1) T-Tubule 1 (TT1, axial); 2) T-Tubule 2 (TT2, transverse); 3) Mitochondrion 1 (Mito1); 4) Mitochondrion 2 (Mito2); 5) Sarcoplasmic Reticulum (SR). The SR was further subdivided into network SR, Z-line SR, and junctional SR subdivided in the following ways: a) nSR1, nSR2, nSR3, nSR4, nSR5, nSR6; b) SRZ1, SRZ2; c) jSR release site, jSR rim, jSR back, preserving the original nomenclature used by Hake et al. [33] (see figure 4.1 for reference).

4.3.2 The MCell model of the CRU

Our model uses a stochastic modeling engine named MCell [60, 61, 130] to track the positions of each molecule in the system individually. This approach treats molecules as point-particles in space, able to diffuse in three dimensions. Reactions between species happen only when two molecules spatially encounter each other. Each molecule is accounted for explicitly in contrast to the continuum methods used by Hake et al. [33].

The same cytosolic and SR Ca^{2+} buffers were used in our MCell model; namely, ATP [33, 131, 132], Calmodulin (CMDN) [132–135], Troponin C (TRPN) [136], and Fluo-4 [132] in the cytosol and Calsequestrin (CSQN) [131, 132, 137] and Fluo-5 in the SR [132]. The models of the sarcolemmal (T-Tubule) pumps and SR pumps used by Hake et al. did not use elementary reactions, and so we used analogous models well-suited to model the behavior of the Plasma Membrane Calcium-ATPase (PMCA) [138–140], Sodium-Calcium Exchanger (NCX) [140, 141], Sarco/Endoplasmic Reticulum Calcium-ATPase (SERCA) pump [140, 142], and (RyR) [143]. The most important feature of these models is the ability to model the kinetics of individual calcium ions. Considerable adaptation of the models was necessary to achieve steady-state behavior in the absence of stimulus, which is detailed in the corresponding molecule descriptions below.

The CRU model of a Ca^{2+} spark by Hake et al. [33] was not designed to model the phenomenon known as Calcium Induced Calcium Release (CICR); an action potential-mediated excitation mechanism of L-Type Calcium Channel (LTCC) fluxes triggering Ca^{2+} efflux from the SR through RyR channels. Instead, Hake’s original model was used to understand spark

termination. It utilized a phenomenological model of the Ryanodine Receptor that occupied one of two states, on or off. Ryanodine receptors were not activated in response to a particular signal, and the number of open receptors was set to a constant number. Moreover, the Ca^{2+} signal terminated in response an SR Ca^{2+} concentration pre-set by the model builders. We sought out to simulate CICR stochastically in response to sarcolemmal excitation. To do this, we implemented a Markov Model of RyR [143] as well as a model of the LTCC action [144] that features both voltage dependent and calcium dependent activation and inactivation properties. Effectively, our RyR respond dynamically to an calcium sparklet generated by LTCC openings in response to a membrane voltage change. Most importantly, our RyR are activated only upon the spatial interactions of Ca^{2+} with RyR in the junctional SR.

4.3.3 Experimental design

One major difference between our model and the earlier model that used the same geometry [33] is our use of stochastic simulations to understand Ca^{2+} spark dynamics in the CRU. The discretization of the system is an important advancement in the modeling of CICR because the resting level of Ca^{2+} in the CRU is extremely low ($140\mu M$). In our particular volume, the number of Ca^{2+} ions are on average, 35. Owing to the low number of ions at rest, the spatial location of the ions is more representative of biological reality than a continuum representation of the concentration which is traditionally used [59, 124].

Utilizing the power of MCell, we can count the exact positions and numbers of the molecules that comprise our system. The molecules in our systems are modeled as point particles that diffuse according to a specified molecular diffusion rate according to the equations of Brownian motion [60]. The reactions in our systems can be unimolecular (state transitions) or bimolecular. Bimolecular reactions occur only upon spatial encounter, and can occur between two cytosolically diffusing molecules (volume molecules) or with membrane-bound species (surface molecules). The Monte Carlo algorithm used by MCell ensures that each simulation gives an independent result that is non-deterministic.

Another major difference between our model and Hake’s CRU model is the use of an

electrical stimulus to induce a Ca^{2+} spark. As noted earlier, the model by Hake et al. used a phenomenological description of the RyR in the dyadic junction, whereby the number of RyR open was set to a constant. When set to open, RyR would “release” Ca^{2+} from the SR that is equivalent to a concentration gradient. Our model requires the use of an electrical stimulus to activate LTCC in the T-Tubules. Upon binding Ca^{2+} on the cytosolic side, the of the Ryanodine Receptor releases discrete Ca^{2+} ions from the sarcoplasmic reticulum into the cytosol. It is important to note that a calcium spark can be triggered by an LTCC “sparklet” or the spark can be spontaneous. With these improvements, we are able to investigate cases that were not possible to test using the earlier model design.

The use of stimulus-induced LTCC opening allowed for interrogation of the effects of action potential alteration on calcium spark genesis. We accomplished this by using action potentials derived from healthy and diseased left ventricle myocytes. More specifically, we used a model of over-expressed Calmodulin-dependent Kinase II (CaMKII-OE) [128,145]. Our goal in comparing the action potentials is to see whether or not the action potential alone can alter the activation of the RyR in the single CRU.

The use of the realistic geometry also allowed us to interrogate the effects of morphological changes in membranous structures. Disrupted T-tubule networks have been observed in diseased cardiomyocytes [126, 146, 147] This disruption oftentimes results in increased dyadic junction distances [148] which effectively decouple LTCC and RyR spatially. Using the open-source 3D graphics software, Blender, we deformed the junctional T-Tubule (TT2) to recreate the deformations observed in diseased tissues. The T-Tubule was altered in such a way that minimally affected the volume and surface area of the mesh but doubled the dyadic volume. This permutation was intended to model the phenomenon known as “detubulation” where the T-tubule network becomes disrupted in diseased forms of cardiac tissue. The original, WT or “normal” T-Tubule and the “deformed” T-Tubule simulations were then compared against one-another to understand the effects of the deformations in cardiac signaling.

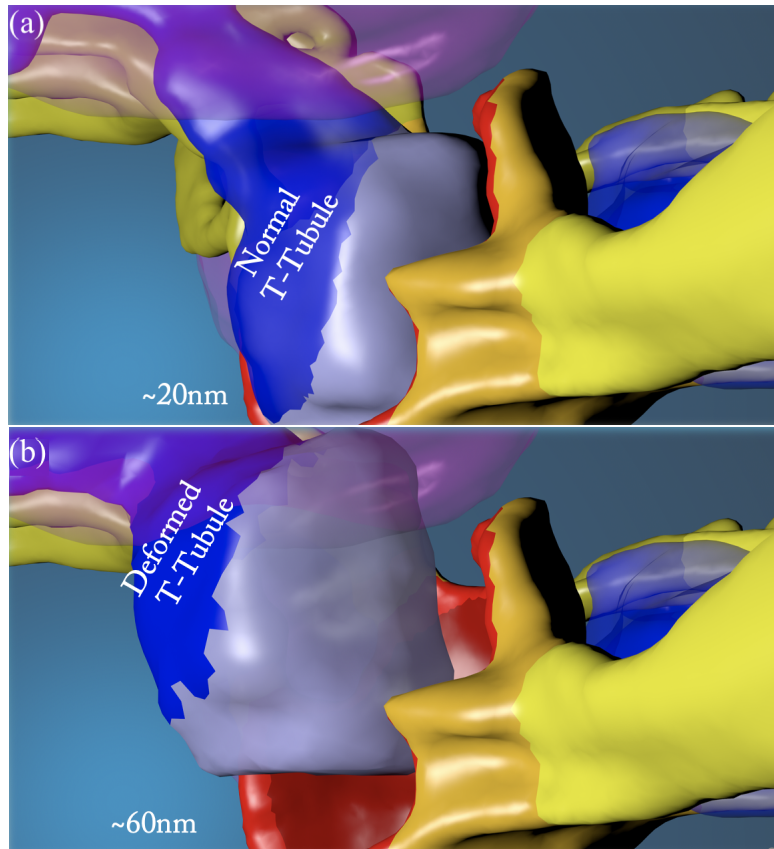


Figure 4.2. Geometries of Normal and Deformed T-Tubules

(a) The original, Wild-Type, “Normal” T-Tubule (blue) has an approximate distance of 20nm from the Sarcoplasmic Reticulum; (b) The deformed T-Tubule has a distance of around 60 nm with total displacement of about 40nm from the original position of the T-Tubule

4.3.4 Cytosolic and SR Ca^{2+} Buffering

The kinetics for our cytosolic and SR Ca^{2+} buffers were modeled similarly to Hake et al (2012) according to equation 4.1. The cytosolic buffers in our simulations are ATP, Calmodulin (CMDN), Troponin-C (TRPN), and Fluo-4 and our SR buffers are Calsequestrin (CSQN) and Fluo-5. The buffer reaction mechanism is a simple two-step reaction, where the buffer can exist in either apo or Ca^{2+} -bound states.



At equilibrium, the product of the forward rate constant, k_f and the concentration of Ca^{2+} and the buffer is equal to the product of the concentration of the Ca^{2+} -bound buffer and the reverse rate constant, k_r , according to equation 4.2.

$$k_f[B][Ca^{2+}] = k_r[B \cdot Ca^{2+}] \quad (4.2)$$

At equilibrium, the concentration of Ca^{2+} is assumed to be constant. In this way, we can assume a pseudo-first order rate constant, k , equaling to the product of the forward rate constant k_f multiplied by the concentration of Calcium ($140\mu\text{M}$), according to equation 4.3.

$$k = k_f \cdot [Ca^{2+}] \quad (4.3)$$

The pseudo-first order relationship in equation 4.3 can be substituted into equation 4.2, yielding equation 4.4.

$$k[B] = k_r[B \cdot Ca^{2+}] \quad (4.4)$$

The above equation can be rearranged to give a ratio of the concentrations of the apo and Calcium-bound buffer state equaling to the ratio of the reverse and pseudo-first order rate constants.

$$\frac{[B]}{[B \cdot Ca^{2+}]} = \frac{k_r}{k} \quad (4.5)$$

Since the buffer exists in either apo or calcium-bound states, according to equation 4.6, the relationship can be substituted into equation 4.5 solving for the concentration of the apo buffer species in 4.7.

$$[Total_B] = [B] + [B \cdot Ca^{2+}] \quad (4.6)$$

$$[B] = \frac{\left(\frac{k_r}{k}\right) [Total_B]}{1 + \left(\frac{k_r}{k}\right)} \quad (4.7)$$

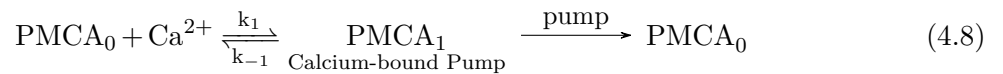
Using this relationship, we are able to solve for the concentration of the buffer species in either apo or Ca^{2+} -bound states according to their total concentrations used by Hake et al. The values for the initial concentrations can be found in Table 4.1.

4.3.5 Sarcolemmal (T-Tubule) Fluxes

PMCA and NCX fluxes

There are two major sarcolemmal pumps known to maintain homeostasis in cardiomyocytes, the Plasma Membrane Calcium-ATPase (PMCA) pump and the Sodium-Calcium exchanger (NCX) pump [59]. Hake et al. originally modeled the sarcolemmal pumps using three separate fluxes, PMCA (termed pCa in Hake et al.), NCX, and a background calcium flux, (termed Cab in Hake et al.). In our model, we capture the background calcium flux utilizing “leak” reactions in both our PMCA and NCX models (see figure 4.3a/b).

The PMCA pump is modeled as a two-state reaction where one ion of Ca^{2+} can reversibly bind to the first state, and in a separate, irreversible reaction, calcium is pumped out of the cytosol (see figure 4.3a and equation 4.8).



The concentration of either state is defined, then, by the total concentration minus the concentration of the other state, according to equation 4.8.

$$[PMCA_1] = [PMCA_{Total}] - [PMCA_0] \quad (4.9)$$

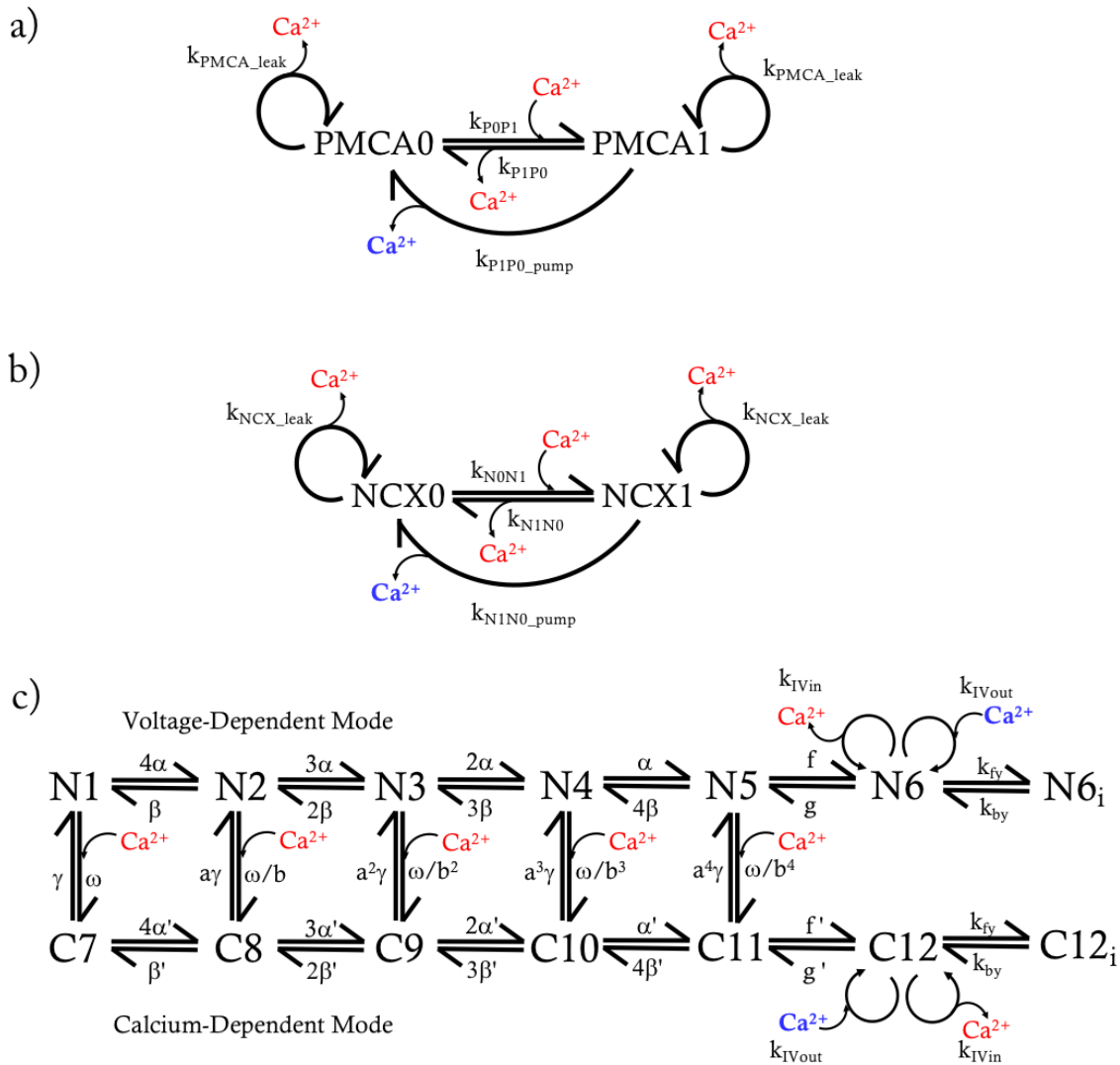


Figure 4.3. Sarcolemmal (T-Tubule) Fluxes

(a) Plasma Membrane Calcium-ATPase Model. Adapted from the model by Bartol et al. (2015) and based on measurements by Brini and Carafoli (2009) and Penheiter et al (2003). Both states are capable of leaking calcium into the cytoplasm (red), corresponding to a background calcium level of 140mM. The transition from state PMCA1 to state PMCA0 pumps one Calcium ion per reaction to the extracellular space (blue). Reaction rate constants can be found in Table 4.2; **(b)** Sodium/Calcium Exchanger Model. Adapted from the model by Bartol et al. (2015) and based on measurements by Hilgemann (1991). Both states are capable of leaking calcium into the cytoplasm (red), corresponding to a background calcium level of 140mM. The transition from state NCX1 to state NCX0 pumps one Calcium ion per reaction to the extracellular space (blue). Reaction rate constants can be found in Table 4.2; **(c)** Sodium/Calcium Exchanger Model. Adapted from the model by Bartol et al. (2015) and based on measurements by Hilgemann (1991). Both states are capable of leaking calcium into the cytoplasm (red), corresponding to a background calcium level of 140mM. The transition from state NCX1 to state NCX0 pumps one Calcium ion per reaction to the extracellular space (blue). Reaction rate constants can be found in Table 4.2

At equilibrium the concentration of Ca^{2+} is constant, and thus, a pseudo-first order rate constant, k_f , can be assumed, as in equation 4.10.

$$k_f = [Ca^{2+}] \cdot k_1 \quad (4.10)$$

The rate of change of a state, for example, $PMCA_0$, can be written as the equation in 4.11, which is simply the sum of the $PMCA_0$ producing reactions minus the $PMCA_0$ consuming reactions.

$$\frac{d[PMCA_0]}{dt} = [PMCA_1]k_{-1} + [PMCA_1]k_{pump} - [PMCA_0]k_f \quad (4.11)$$

At equilibrium, the rate of change is zero, and can be used to solve for the concentration of one of the two states as in equations 4.12-4.15.

$$0 = \frac{d[PMCA_0]}{dt} \quad (4.12)$$

$$0 = (k_{-1} - k_{pump})([PMCA_{Total}] - [PMCA_0]) - [PMCA_0]k_f \quad (4.13)$$

$$0 = [PMCA_{Total}](k_{-1} + k_{pump}) - [PMCA_0](k_{-1} + k_{pump}) - [PMCA_0]k_f \quad (4.14)$$

$$[PMCA_0] = \frac{[PMCA_{Total}](k_{-1} + k_{pump}) - [PMCA_0](k_{-1} + k_{pump})}{k_f} \quad (4.15)$$

To account for the background Ca^{2+} level that maintains an equilibrium of $140\mu M$, each state of the pump was assigned to a leak rate which is defined as the ratio of the product of the forward pump rate constants over all the produce of the total rate constants.

$$k_{leak} = \frac{k_{pump} \cdot k_f}{k_f + k_{-1} + k_{pump}} \quad (4.16)$$

In this way, we can solve for the steady state concentrations of PMCA in either state. The same relationships were used to model the NCX pump, but are not shown in the interest of brevity.

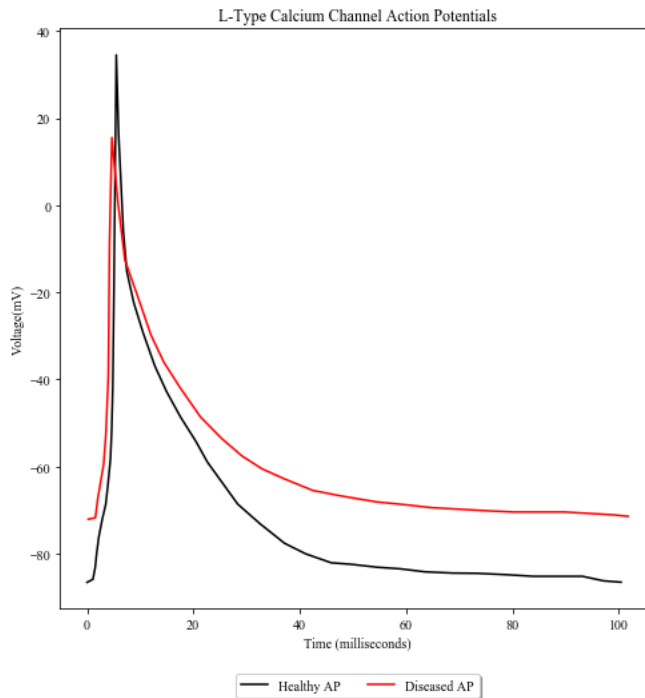


Figure 4.4. L-Type Calcium Channel Action Potentials
 Healthy (black) and Diseased (red) action potentials used to stimulate the LTCC, based on the model by Morotto et al (2014).

L-Type Calcium Channel flux

In an effort to further improve upon the the original Calcium Release Unit mode, we aimed to model a *triggered* Ca^{2+} spark, or the process known as Calcium-Induced Calcium Release. In order to do this, we included a model of the voltage-dependent Calcium Channel, known as the L-Type Calcium Channel (LTCC). We adapted the model of LTCC dynamics developed by Greenstein and Winslow [144] to be used in our system. We used two different action potentials to stimulate our LTCC 1) “healthy” or WT and 2) “diseased” or Calmodulin-Dependent Kinase II over-expression (CaMKII-OE) action potentials as modelled by Morotti et al. [128] (see figure 4.4). The LTCC model has both voltage-dependent modes and Calcium-dependent modes. Using this LTCC model we were able to model the rate constants of LTCC state transitions as a function of membrane calcium reactions, voltage and time, according to the relationships described by Greenstein and Winslow (see figure 4.3c), described below.

The forward rate constants of the voltage-dependent modes, α are a function of the

membrane voltage, according to equation 4.17.

$$\alpha = 2.0e^{0.012(V_m - 35)} \quad (4.17)$$

The reverse rate constants of the voltage-dependent modes, β are a function of the membrane voltage, according to equation 4.18.

$$\beta = 0.0882e^{-0.05(V_m - 35)} \quad (4.18)$$

The forward rate constants of the calcium-dependent modes, α' are a scalar multiple of α given by equation 4.19.

$$\alpha' = a\alpha \quad (4.19)$$

The reverse rate constants of the calcium-dependent modes, β' are a scalar multiple of β given by equation 4.20.

$$\beta' = b\beta \quad (4.20)$$

The parameters defining the forward rate constant, k_{fy} , and reverse rate constant, k_{ry} , of the voltage and calcium-dependent inactivation are defined by equations 4.21 through 4.24.

$$y_\infty = 0.4 / \left(1 + e^{(V_m + 12.5)/5}\right) + 0.6 \quad (4.21)$$

$$\tau_y = 340 / \left(1 + e^{(V_m + 30)/12}\right) + 60 \quad (4.22)$$

$$k_{f,y} = y_\infty / \tau_y \quad (4.23)$$

$$k_{b,y} = (1 - y_\infty) / \tau_y \quad (4.24)$$

In order to calculate the inward flux of calcium through the LTCC k_{IVin} and the outward flux k_{IVout} , we converted the permeability of Ca^{2+} P_{Ca} to units ions/sec using equation 4.25 and 4.26. We assume 2mM extracellular calcium concentration. We use the following two equations where N_A is Avogadro's number, z is charge/ion, V_m is the membrane voltage and F is Faraday's

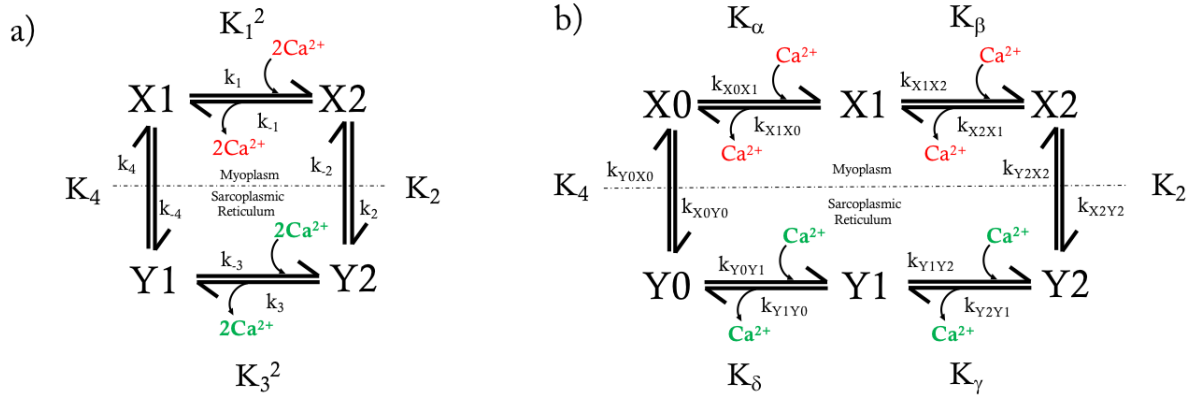


Figure 4.5. SERCA Models

(a) Sarco/Endoplasmic Reticulum Calcium-ATPase (SERCA) Model by Higgins et al. Two cytoplasmic-facing states (X) and two sarcoplasmic reticulum-facing states (Y) translocate cytosolic Calcium (red) to the sarcoplasmic reticulum (green).

(b) Sarco/Endoplasmic Reticulum Model. Adapted from the model by Higgins et al 2006 by separating the event of two Ca^{2+} binding/unbinding into individual Ca^{2+} binding/unbinding reactions. Three cytoplasmic-facing states (X) and three sarcoplasmic reticulum-facing states (Y) translocate cytosolic Calcium (red) to the sarcoplasmic reticulum (green). Reaction rate constants can be found in Table 4.5

constant, R is the universal gas constant, and T is temperature (see table 4.3).

$$k_{IVin} = \frac{[Ca^{2+}]^{\frac{N_A}{z}} P_{Ca} 4V_m F 0.341}{RT e^{\frac{2V_m F}{RT} - 1}} \quad (4.25)$$

$$k_{IVout} = \frac{\frac{N_A}{z} P_{Ca} 4V_m F e^{\frac{2V_m F}{RT}}}{RT e^{\frac{2V_m F}{RT} - 1}} \quad (4.26)$$

4.3.6 Sarcoplasmic Reticulum Fluxes

Sarco/Endoplasmic Reticulum Calcium-ATPase (SERCA) pump kinetics

Of utmost importance in the dynamics of Ca^{2+} during the phenomenon of CICR is the action of the SERCA pump, which is estimated to be responsible for the clearing of 70-90% of cardiomyocyte Ca^{2+} after a release event event [59]. In a single forward cycle of the SERCA pump, one molecule of ATP is consumed to translocate two ions of Ca^{2+} from the cytoplasm into the SR, the intracellular Ca^{2+} store. In order to explicitly model the dynamics of Ca^{2+} in the myoplasm and inside the SR, it is necessary to use models that treat calcium as individual species.

In order to achieve this, we adapted a model of SERCA by Higgins et al. [142] as modeled in our previous work [140] to reproduce the dynamics of SERCA2a, maintaining 140nM cytoplasmic Ca^{2+} while satisfying microscopic reversibility.

The four-state model by Higgins et al. [142] (Figure 4.5) groups both Ca^{2+} binding and unbinding events into a single step (see Figure 4.5,a) K_1 and K_3 , respectively). In our adapted version of the Higgins model, we model SERCA as a six-state pump that binds (see Figure 4.5,b) K_α and K_β) and unbinds (see figure 4.3,b) K_γ and K_δ) Ca^{2+} ions discretely. In order to accomplish this, the equilibrium constants for binding and unbinding Ca^{2+} (K_1 and K_3 , respectively) needed to be split up into two equilibrium constants for the individual binding (K_α and K_β) and unbinding (K_γ and K_δ) events. This was accomplished using the theory of ligand occupancy developed by Sine and Taylor [149–151].

Firstly, we assume that the Ca^{2+} equilibrium constant for the binding of 2 Ca^{2+} ions, K_1^2 is the product of the individual equilibrium constants, K_α and K_β , according to equations 4.27-4.30, below.

$$K_1^2 = K_\alpha \times K_\beta \quad (4.27)$$

$$K_\alpha = \frac{k_{X1X0}}{k_{X0X1}} \quad (4.28)$$

$$K_\beta = \frac{k_{X2X1}}{k_{X1X2}} \quad (4.29)$$

$$K_1^2 = \frac{k_{X1X0}}{k_{X0X1}} \times \frac{k_{X2X1}}{k_{X1X2}} \quad (4.30)$$

Because there are two sites to which Ca^{2+} can bind in state X0 and only one state to which Ca^{2+} can bind to in state X1, the relationship of the Ca^{2+} binding rate constants in equation 4.31 is assumed.

$$k_{X0X1} = 2 \times k_{X1X2} \quad (4.31)$$

Likewise and in the same way, there are two sites from which Ca^{2+} can unbind in state X2 and only one site to which Ca^{2+} can unbind to in state X1. And so, the relationship of the

Ca^{2+} unbinding rate constants in equation 4.32 is assumed.

$$k_{X_2X_1} = 2 \times k_{X_1X_0} \quad (4.32)$$

Substituting into equation 4.30 for the rate constant relationships 4.31 and 4.32, yields equation 4.34.

$$K_1^2 = \frac{k_{X_1X_0}}{2 \times k_{X_1X_2}} \times \frac{2 \times k_{X_1X_0}}{k_{X_1X_2}} = \left(\frac{k_{X_1X_0}}{k_{X_1X_2}} \right)^2 \quad (4.33)$$

According to the relationship of the equilibrium constants in equations 4.31-32, the relationship between the equilibrium constants K_α and K_β is the one given by equation 4.34.

$$K_\alpha = \frac{K_\beta}{4} \quad (4.34)$$

We solve for the individual rate constants by substituting the relationship given in 4.34 into equation 4.35 yielding equations 4.36-38.

$$K_1 = \sqrt{K_\alpha \times K_\beta} \quad (4.35)$$

$$K_1 = \sqrt{\frac{K_\beta^2}{4}} \quad (4.36)$$

$$K_\alpha = \frac{K_1}{2} \quad (4.37)$$

$$K_\beta = K_1 \times 2 \quad (4.38)$$

In the same way, we can derive the individual rate constants, K_γ and K_δ from the equilibrium constant for the unbinding of 2 Ca^{2+} ions, K_3^2 .

$$K_3^2 = K_\gamma \times K_\delta \quad (4.39)$$

$$K_\gamma = \frac{k_{Y_1Y_2}}{k_{Y_2Y_1}} \quad (4.40)$$

$$K_\delta = \frac{k_{Y_0Y_1}}{k_{Y_1Y_0}} \quad (4.41)$$

$$K_3^2 = \frac{k_{Y1Y2}}{k_{Y2Y1}} \times \frac{k_{Y0Y1}}{k_{Y1Y0}} \quad (4.42)$$

There are two sites to which Ca^{2+} can unbind from in state Y2 and only one state from which Ca^{2+} can unbind state Y1. Thus, the relationship between Ca^{2+} unbinding rate constants in equation 4.43 is assumed.

$$k_{Y1Y0} = 2 \times k_{Y2Y1} \quad (4.43)$$

There are two sites to which Ca^{2+} can bind in state Y0 and only one site to which Ca^{2+} can bind to in state Y1. As such, the relationship of the Ca^{2+} binding rate constants in equation 4.44 is assumed.

$$k_{Y0Y1} = 2 \times k_{Y1Y2} \quad (4.44)$$

Substituting into equation 4.42 for the rate constant relationships 4.43 and 4.44, yields equation 4.45.

$$K_3^2 = \frac{k_{Y1Y2}}{2 \times k_{Y1Y0}} \times \frac{2 \times k_{Y1Y2}}{k_{Y1Y0}} = \left(\frac{k_{Y1Y2}}{k_{Y1Y0}} \right)^2 \quad (4.45)$$

According to the relationship of the equilibrium constants in equations 4.43 and 4.44, the relationship between the rate constants K_γ and K_δ is the one given by equation 4.46.

$$K_\delta = \frac{K_\gamma}{4} \quad (4.46)$$

We solve for the individual rate constants by substituting the relationship given in 4.46 into equation 4.47 yielding equations 4.48-4.50.

$$K_3 = \sqrt{K_\gamma \times K_\delta} \quad (4.47)$$

$$K_3 = \sqrt{\frac{K_\gamma^2}{4}} \quad (4.48)$$

$$K_\gamma = K_3 \times 2 \quad (4.49)$$

$$K_\delta = \frac{K_3}{2} \quad (4.50)$$

According to Higgins et al., the relationship between the Gibbs Free energy of 1 molecule of ATP and the kinetic cycle is given by equation 4.51.

$$K_1^2 K_2 K_3^2 K_4 = e^{\Delta G_{\text{ATP}}^0 / RT} \quad (4.51)$$

Using the rate constants described in Table 4.5 and the relationships described above, yields a Gibbs free energy of -47.0943 kJ/mol. The product of the forward rates constants equals the product of the reverse rate constants. At steady state, this is also true for the product of the forward reaction rates and the product of the reverse reaction rates, satisfying detailed balance. Lastly, the pump reaches steady state at a cytosolic concentration, $Ca_{\text{cyt}_{ss}}^{2+}$, of 140nM and a SR concentration, $Ca_{\text{sr}_{ss}}^{2+}$, of 1.3mM.

$$Ca_{\text{sr}_{ss}}^{2+} = \frac{Ca_{\text{cyt}_{ss}}^{2+}}{K_1 K_3 \sqrt{K_2 K_4}} \quad (4.52)$$

Ryanodine Receptor (RyR) kinetics

In order to simulate a triggered Ca^{2+} spark, it was necessary to model the complexities of the Sarcoplasmic Reticulum Ca^{2+} release channel, the Ryanodine Receptor. Hake's original simulations used a simplistic, binary model of RyR that could exist in either an open or a closed state. Once a certain voltage was sensed, the RyR would close and never re-open. Most notably the RyR in these simulations did not activate in response to an action potential stimulus or change in dyadic Ca^{2+} levels, but instead were set to "open" to initiate the spark. RyR is known to interact with 30+ binding partners [152], and thus, can exist in a multitude of states. In an attempt model this complexity, we incorporated a Markov model of RyR that captures its low and high gating modes as well as its ability to bind Ca^{2+} as a basis for its activation.

We adapted the Markov model of RyR dynamics by Saftenku et al. [143] for use in MCell simulations (see Figure 4.6a). Coincidentally, this was the same model of RyR used by Koh et al in their MCell simulations using simplistic geometries [125].

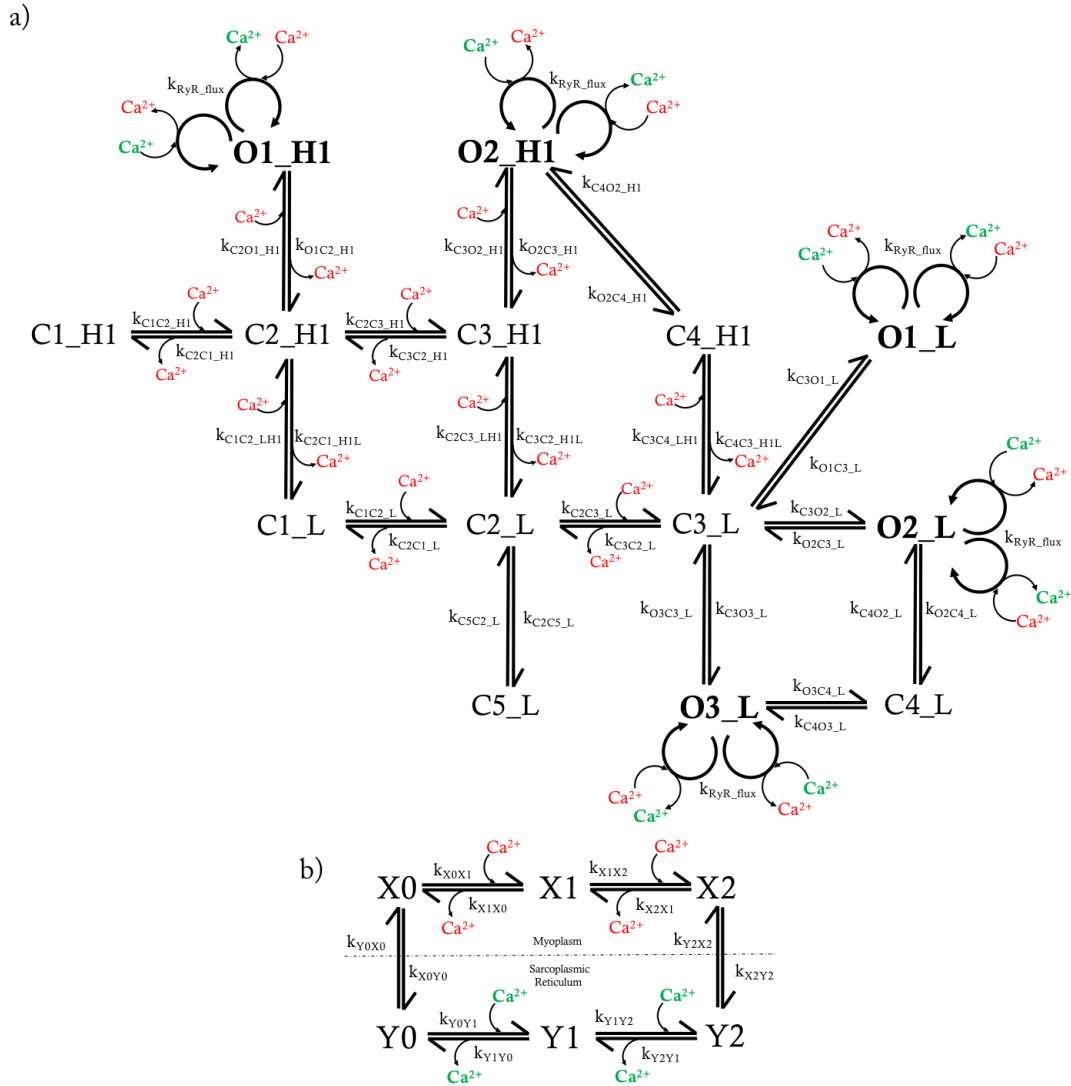


Figure 4.6. Sarcoplasmic Reticulum Fluxes

(a) Ryanodine Receptor Markov Model. Adapted from the Markov Model by Saftenu et al. (2001); Nine closed (C) states and five open (O) states describe the kinetic scheme of the Ryanodine Receptor. Cytosolic Calcium (Red) binds to closed states in High (H1) and Low (L) gating modes. In open states, Sarcoplasmic Reticulum Calcium (Green) and Cytosolic Calcium are translocated through the receptors. Reaction rates can be found in Table 4.4.

(b) Sarco/Endoplasmic Reticulum Model. Adapted from the model by Higgins et al 2006. Three cytoplasmic-facing states (X) and three sarcoplasmic reticulum-facing states (Y) translocate cytosolic Calcium (red) to Calcium the sarcoplasmic reticulum (green). Reaction rate constants can be found in Table 4.5

The parameters for the RyR model (see Table 4.4) were used with little adaptation with the exception of the RyR flux. For the purpose of our simulations, the RyR flux had to be converted to units of Ca^{2+} ions/second. This was accomplished using the single-channel current, measured at 0.35 pA at 1mM by Guo et al [153] and the relationship described in equation 4.53, below.

$$k_{RyRflux} = \frac{I^{RyR} 6.242 \times 10^{18} \frac{\text{Charge}}{\text{Coulomb}}}{[\text{Ca}^{2+}] z} = \frac{9.01 \times 10^{-9} \text{Ca}^{2+} \text{ions}}{s} \quad (4.53)$$

Amps are defined as units of $\frac{\text{Coulomb}}{\text{second}}$. Thus, the RyR current, I^{RyR} in pA can be converted to charges by multiplying the conversion factor, $6.242 \times 10^{18} \frac{\text{Charge}}{\text{Coulomb}}$. Using the 1mM concentration of Calcium and a charge or, z , value of 2, equals to a rate constant of $9.01 \times 10^{-9} \frac{\text{Ca}^{2+} \text{ions}}{s}$.

4.4 Results and Discussion

4.4.1 Buffer and Fluorophore Dynamics

We began our investigations by building the model from a baseline, using cytosolic buffers, sarcolemmal pumps, and fluorophores as the subjects of our simulations. To make sure equilibrium is maintained in our simulations, we investigated the dynamics of Calcium and its buffers in simulations lacking the sarcolemmal stimulus, L-Type Calcium Channel. Because Ryanodine receptors are known to spontaneously activate in the presence of Calcium [120, 122], and SERCA itself is known to act as a buffer [142] we sought to eliminate these effects to establish that our “baseline system,” comprised of cytosolic and SR buffers and fluorophores, maintains equilibrium. A question addressed by Hake et al. in their original investigation centered on the effect of fluorophores in the simulations. We also decided to test this effect in our initial simulations.

The system dynamics comparing the effects of the fluorophores on the baseline system (containing only buffers and sarcolemmal pumps) are nearly identical (see Figure 4.7). The most pronounced effect is on the cytosolic buffer, Troponin C (See Figure 4.7e) which differs, on average, only 14 molecules between simulations with and without fluorophores. Due to a high

concentration of TRPN and TRPN-Ca²⁺ in the simulations (70 μ M total, 13.2 μ M bound to Calcium, 56.8 μ M free), this corresponds to a difference of 0.036% on average between the two systems, which is arguably negligible.

4.4.2 Ryanodine Receptor activation profiles

Spontaneous calcium sparks in absence of LTCC

In order to validate our model, we sought confirmation that that our system could generate spontaneous Ca²⁺ sparks without an LTCC stimulus as observed in experiments [120,122,154]. In order to test this, we simulated our CRU system in absence of an action potential stimulus but in the presence of RyR with and without SERCA (see Figure 4.8). Though the average of total number of RyR openings (n=128) are low, there is a non-zero probability that RyR in be triggered spontaneously.

Triggered calcium sparks: RyR response to LTCC alterations

Having established that RyR are capable of firing spontaneously in the absence of an AP stimulus, we sought to confirm that a single LTCC could activate RyRs as previously demonstrated by Sobie et al [155]. Figure 4.9 demonstrates that one LTCC with a single action potential is sufficient to activate RyR.

Following our simulations with 1 LTCC, we sought to understand the effect that increasing the number of LTCC would have on triggered Calcium release. We comparatively investigated CICR with 1, 2, 4, 6, 8, and 10 LTCC. We predicted that increasing LTCC would result in increased gain, or SR Ca²⁺ efflux through RyR. Consistent with experimental findings, [122] increasing the LTCC results in positive gain as demonstrated in figure 4.10.

4.4.3 Examining disease phenotypes

Alterations in L-Type Calcium Channel Action Potential stimuli

With the use of the LTCC model developed by Greenstein and Winslow [144], we were able to interrogate the effect of diseases sarcolemmal action potential stimuli on Ca²⁺ signaling (see figure 4.4). A number of recent studies have examined disease electrical stimulus

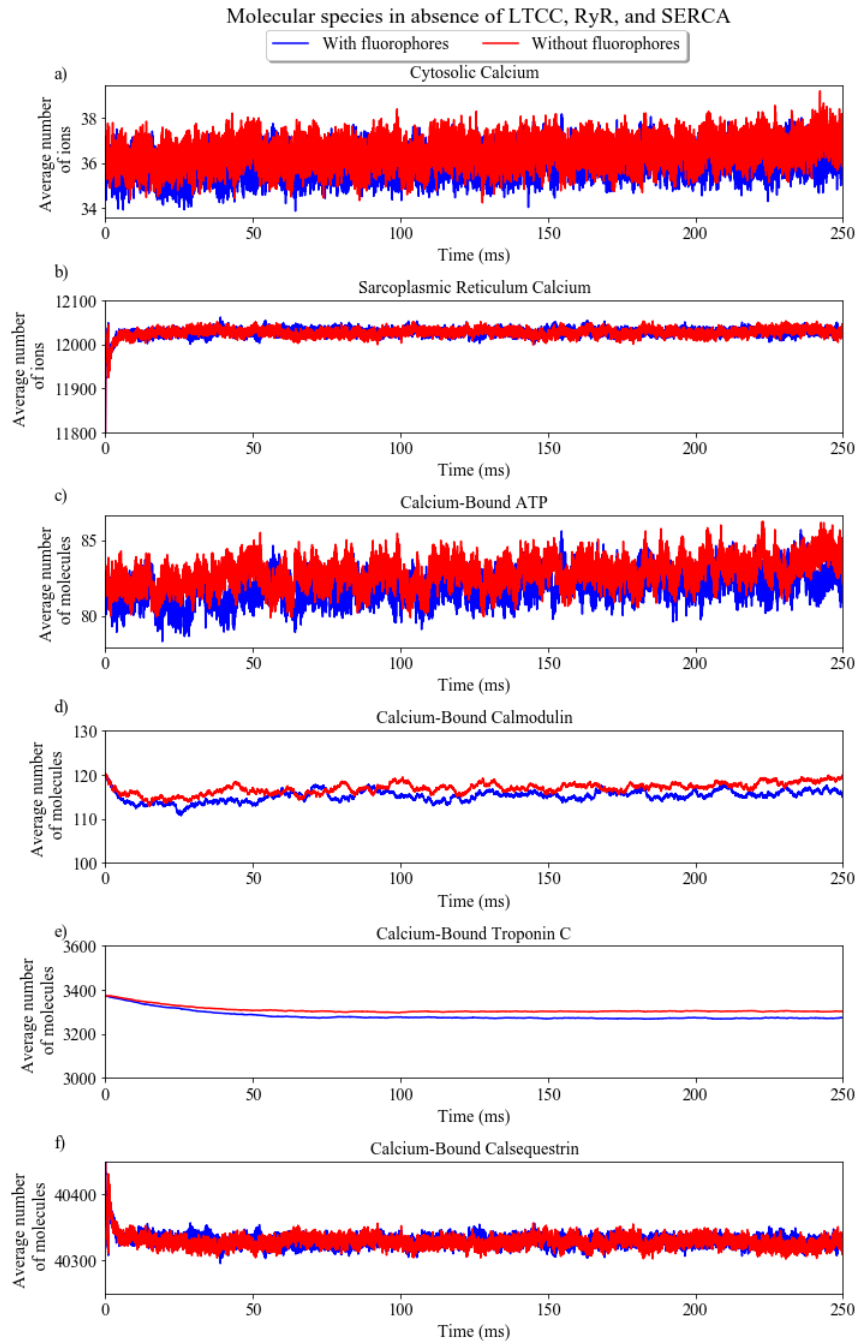


Figure 4.7. Molecular species in absence of LTCC, RyR and SERCA

(a) Cytosolic Calcium dynamics; (b) Sarcoplasmic Reticulum Calcium dynamics; (c) Calcium-bound ATP; (d) Calcium-bound Calmodulin; (e) Calcium-bound Troponin-C; (f) Calcium-bound Calsequestrin; Comparing simulations with only RyR (blue) and RyR and SERCA (red)

Ryanodine Receptor gating modes in absence of L-Type Calcium Channels

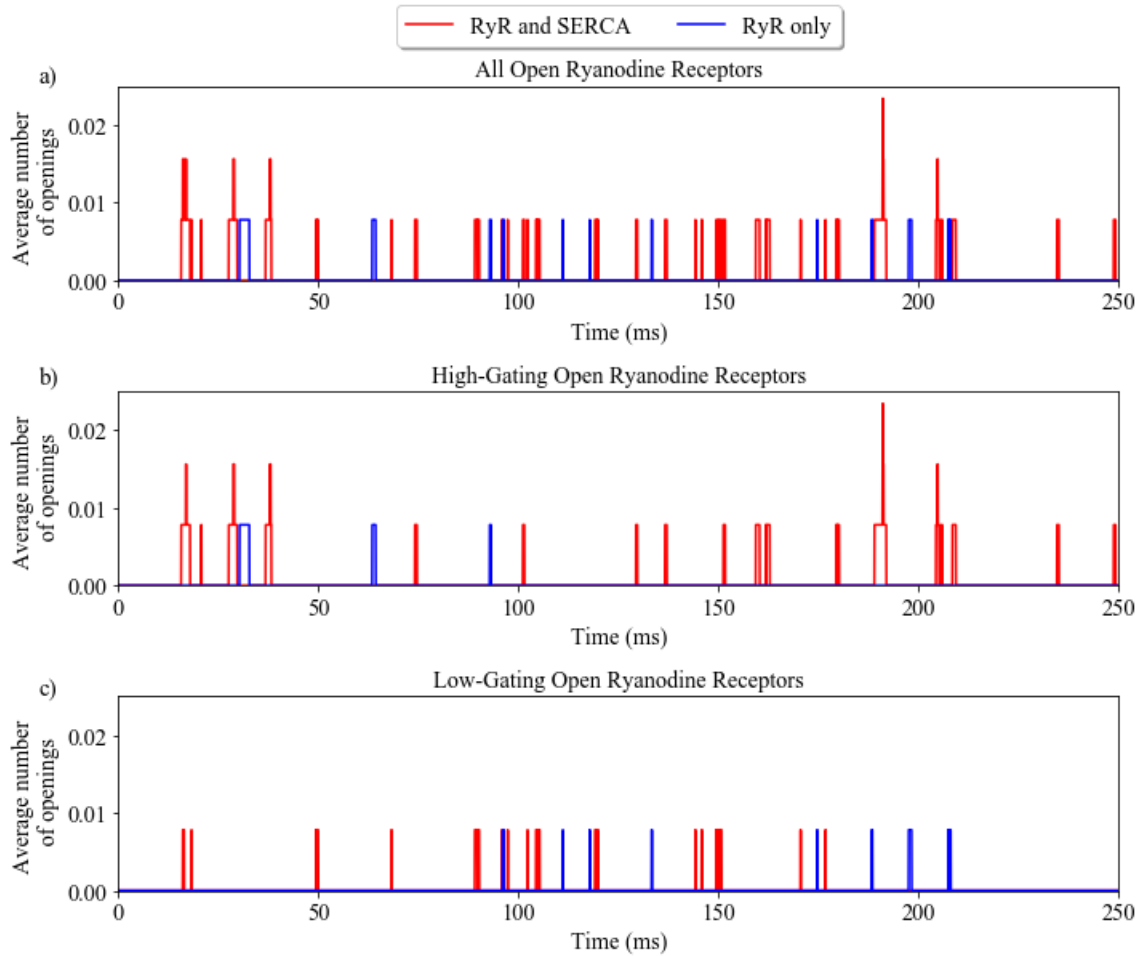


Figure 4.8. Ryanodine Receptor gating modes in absence of L-Type Calcium Channels

(a) All open Ryanodine Receptors; (b) High-gating open Ryanodine Receptors; (c) Low-gating open Ryanodine Receptors; RyR and SERCA (red) and only RyR (blue)

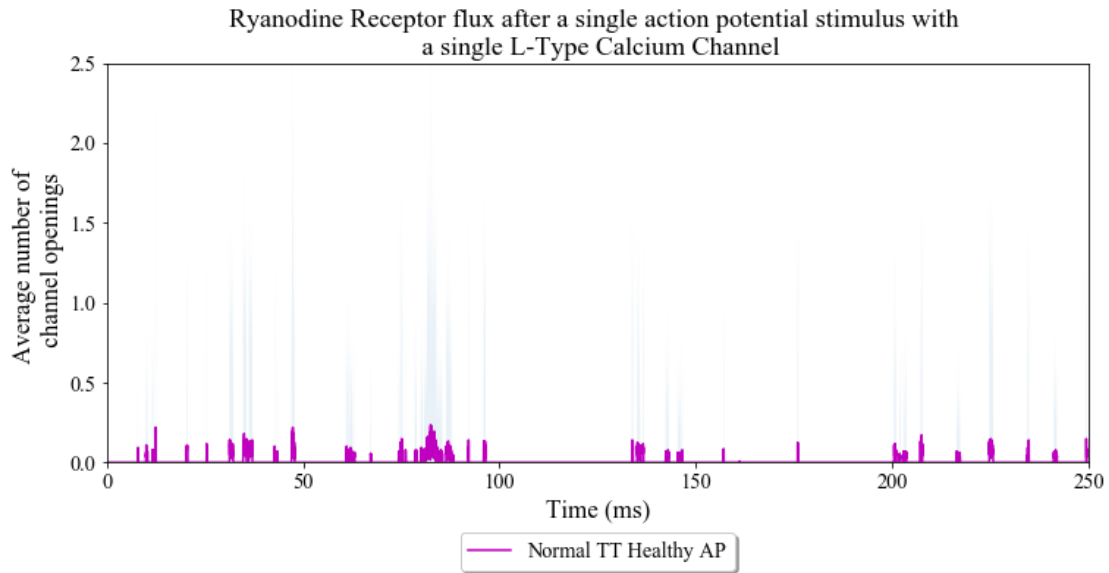


Figure 4.9. Ryanodine Receptor flux after a single action potential stimulus with a single L-Type Calcium Channel
 Average number of openings of RyR in normal T-Tubules simulated with a single L-Type Calcium Channel and the healthy action potential (purple). Standard deviations are shown in light blue.

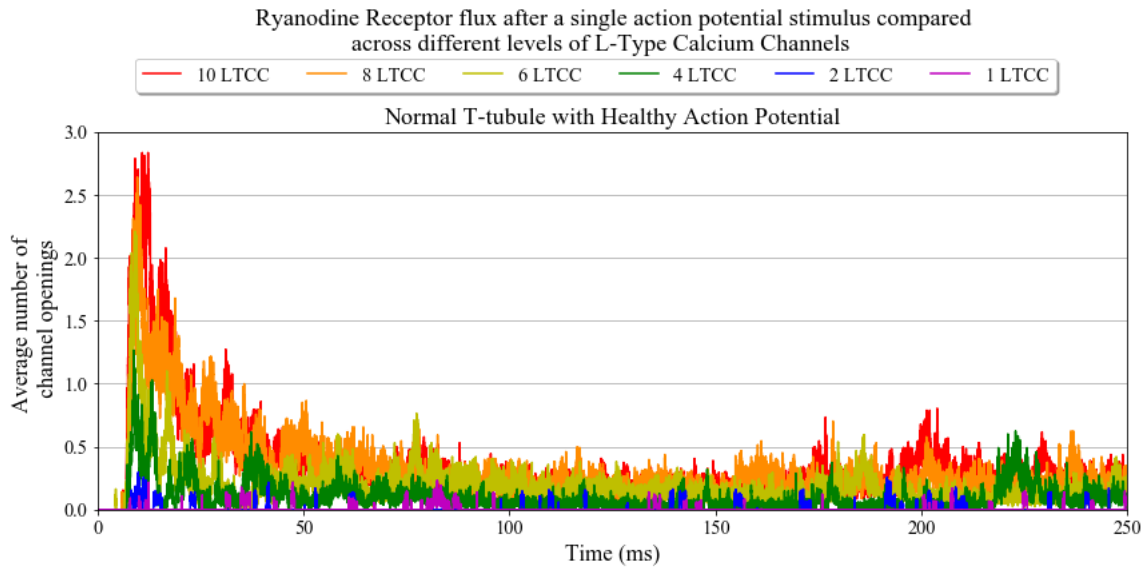


Figure 4.10. Ryanodine Receptor flux after a single action potential stimulus compared across different levels L-Type Calcium Channels
 Average number of openings of RyR in normal T-Tubules with a healthy action potential compared across a series of LTCC numbers: 10 LTCC (red), 8 LTCC (orange), 6 LTCC (yellow), 4 LTCC (green), 2 LTCC (blue), 1 LTCC (purple)

phenotypes [128,156], demonstrating that indeed, diseased left-ventricle mouse action potentials have pronounced differences from that of healthy myocytes. We chose to model the effect of Calcium-Calmodulin Kinase II over-expression (CaMKII-OE) which leads to heart-failure in mice.

The effect of the diseased action potential on Calcium signaling is the same across all L-Type Calcium channels levels. The diseased AP leads to an overall increase in calcium levels in the cytosol (see figure 4.11) as a result of frequent activations of Ryanodine Receptors in disease AP cases (see figure 4.12, 4.13). This is an expected result, as depolarization of the membrane is relatively higher in the later half of the AP in the CAMKII-OE disease state. Most interesting is the later-stage RyR activation seen in the disease states. Significantly more RyR fire at later stages of the AP in diseased states compared to those with a healthy AP. This will likely lead to asynchronous firing of the calcium release units. In the case of CAMKII-OE, the disease phenotype known as Late Ca^{2+} Sparks (LCS) [157]. It should be noted that this effect is seen in all variations of LTCC numbers but only the cases of 1 LTCC and 4 LTCC are shown here in the interest of brevity. Our model result is consistent with the Ca^{2+} signaling profile CAMKII-OE models in mouse models, which indicate more frequent sparks as well as greater time to peak calcium and larger maximal rate of rise [158]. This effect that is most apparent in figure 4.11, where levels of Ca^{2+} release are higher overall in diseased AP cases and time to achieve maximum calcium efflux is increased.

Geometric alterations

In an additional effort to understand disease phenotypes in heart failure, we investigated the effects of T-Tubule deformations on cardiac calcium release units. In the literature, this effect is termed “de-tubulation” of myocyte tissue [126]. Because of the close juxtaposition of T-Tubule membranes and SR membranes, we predicted an increase in the dyadic space would result in significantly less RyR activation and overall low rates SR calcium efflux, resulting in low cytosolic calcium.

Our results validated our hypothesis, showing little to no calcium spark activity in deformed cells when compared to healthy cells (see figure 4.13). A similar distance (40 nm

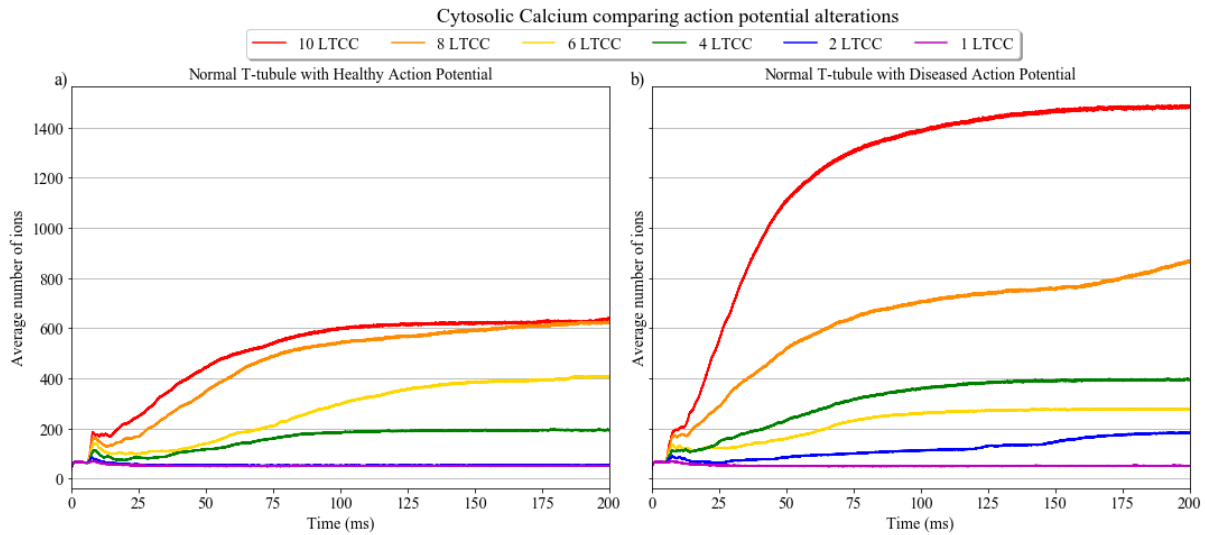


Figure 4.11. Cytosolic Calcium comparing action potential alterations

(a) Cytosolic calcium in Normal T-Tubules with healthy action potential across range of L-Type Calcium Channels; (b) Cytosolic calcium in Normal T-Tubules with diseased action potential across range of L-Type Calcium Channels;

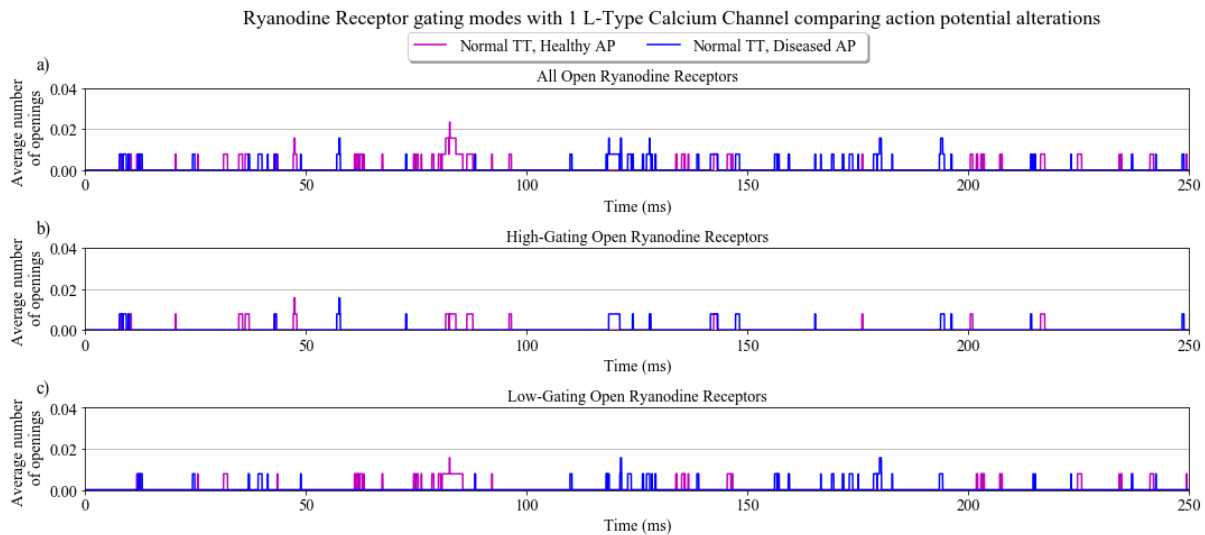


Figure 4.12. Ryanodine Receptor gating modes with 1 L-Type Calcium Channel comparing action potential alterations

(a) All open Ryanodine Receptors; (b) High-gating open Ryanodine Receptors; (c) Low-gating open Ryanodine Receptors: Healthy action potential (purple), Diseased action potential (blue)

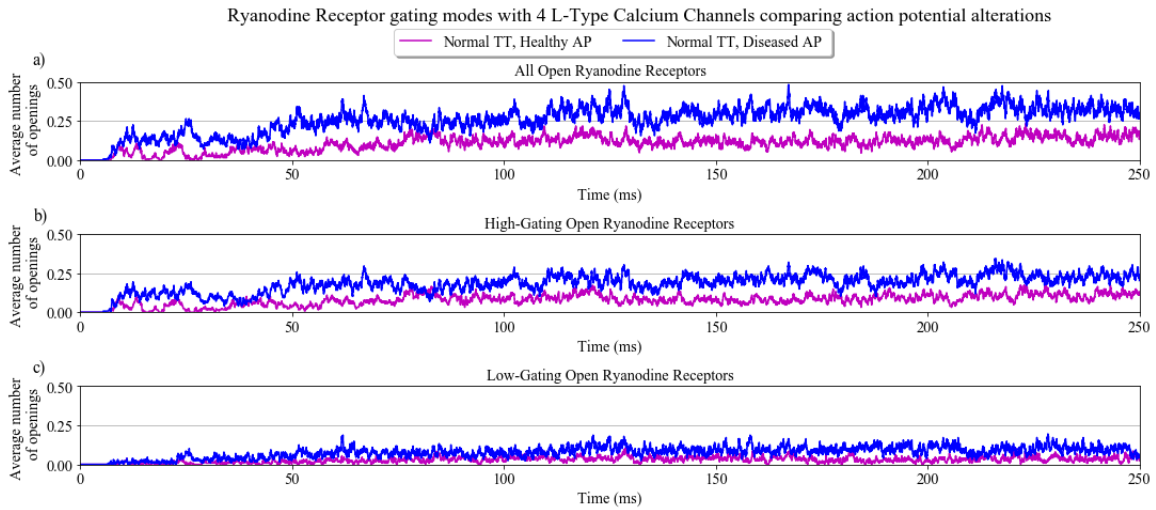


Figure 4.13. Ryanodine Receptor gating modes with 4 L-Type Calcium Channels comparing action potential alterations

(a) All open Ryanodine Receptors; (b) High-gating open Ryanodine Receptors; (c) Low-gating open Ryanodine Receptors; Healthy action potential (purple), Diseased action potential (blue)

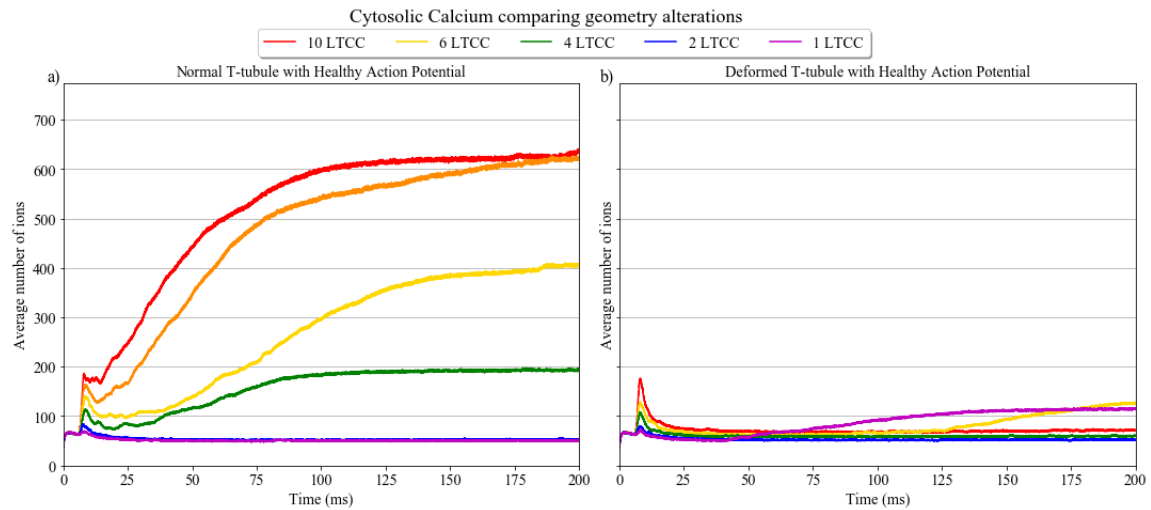


Figure 4.13. Cytosolic Calcium dynamics comparing T-Tubule geometry alterations

(a) Normal T-Tubule with Healthy action potential showing Calcium dynamics in the cytosol; (b) Deformed T-Tubule with Healthy action potential showing Calcium dynamics in the cytosol; 10 LTCC (red), 8 LTCC (orange), 6 LTCC (yellow), 4 LTCC (green), 2 LTCC (blue), 1 LTCC (purple)

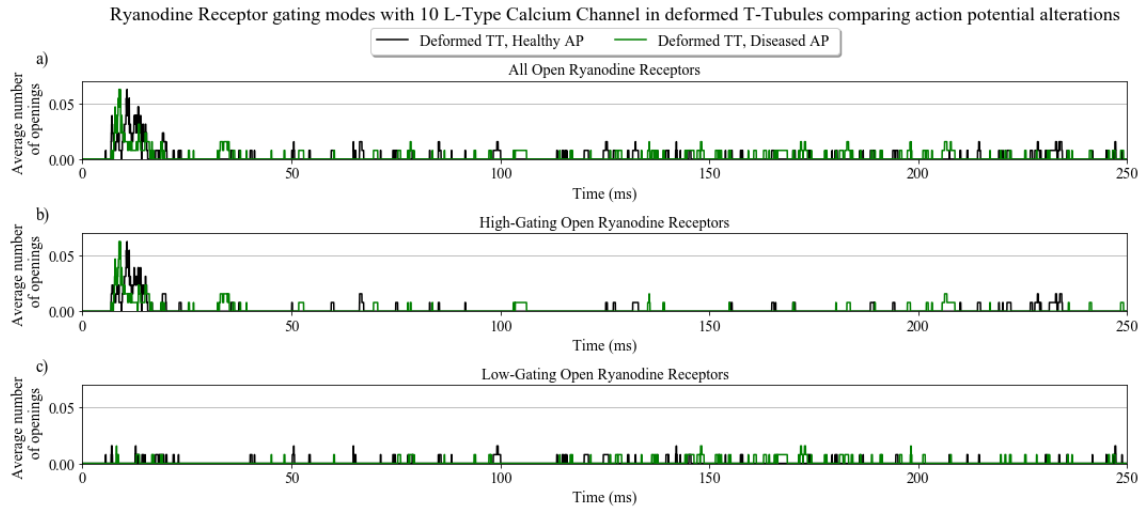


Figure 4.14. Ryanodine Receptor gating modes with 10 L-Type Calcium Channel in deformed T-Tubules comparing action potential alterations

(a) All open Ryanodine Receptors; (b) High-gating open Ryanodine Receptors; (c) Low-gating open Ryanodine Receptors; Healthy action potential (black), Diseased action potential (green)

displacement of TT2) was used by Koh et al. [125], who reported similar results. In future simulations, we will successively increase the dyadic volume in order to further analyze the effect of disease progression, systematically.

An interesting result of our simulations was the compensation effect exhibited by disease action potentials in simulations of diseased geometries. We compared the activation profiles of healthy and diseased action potentials in diseased geometries in our experiments. Under these conditions, the disease action potentials lead to more RyR flux events even in the diseased morphologies (see figure 14). Even with a dramatic detubulation, this effect is still visible.

We hypothesize that at less severe levels of T-Tubule deformations, the disease AP compensation effect will be pronounced enough to compensate for the deformity. Ideally, we will apply our MCell model to new tomographic CRU geometries from both healthy and diseased myocytes to gain more insights into the potential for compensation by affected action potentials.

Modeling Ryanodine Receptor dispersion

The use of a stochastic spatial technique combined Markov Model of RyR marks a significant advancement in the realistic modeling of calcium sparks, in stark contrast to phe-

nomenclological, deterministic, continuum models of RyR that are traditionally used to describe Ca^{2+} sparks [124]. With the use of Markov Model descriptions of RyR, we are able to directly investigate the phenomenon of Calcium-Induced Calcium Release. The spatial representations exhibited in our model of the CRU importantly allow for investigation of the effects of disease phenotypes that implicate RyR.

It has been shown by multiple investigators that RyR function and localization are affected by heart failure [159]. More specifically, regulators of RyR including PKA [160] and CaMKII [145] have been implicated in altered calcium signaling. In heartfailure, RyR dispersion, or loss of RyR in junctional SR has been observed and shown to have pronounced effects on Ca^{2+} signaling [127]. Earlier reports have determined that 50% RyR loss results in significant alterations in firing propensity at the level that can be detected by cellular imaging and biochemical measurements [161]. In order to visualize these effects on the molecular level, we performed simulations within our CRU and successively removed clusters of RyR in the junctional space. Starting from a baseline of 96 RyR as imaged in the original CRU [32], we successively deleted clusters of RyR for a total of 80, 60, 40, 20, and 10 Ryanodine Receptors juxtaposed against 10 L-Type Calcium Channels, the results of which are shown in figure 15.

As we decrease the number of RyR from 96 to 80 (figure 15 a-d) we see a dramatic decrease in the disease AP case. This is likely because of the phenomenon of CICR, where RyR efflux activates neighboring RyR. The decrease also affects the peak levels of RyR which in the former case are 7-10 RyR open at a time, and in the latter case 5-7. Decreasing the number of RyR to 60 (figure 15 e,f) significantly affects the number of RyR that can be activated as well as the total range of RyR that are open. As we decrease the number of RyR by more than half, to 40 (figure 15 g-i), it becomes nearly impossible to see activation of RyR. These results are consistent with previous reports that demonstrate a 50% decrease in junctional SR RyR results in significant alterations in activation [161].

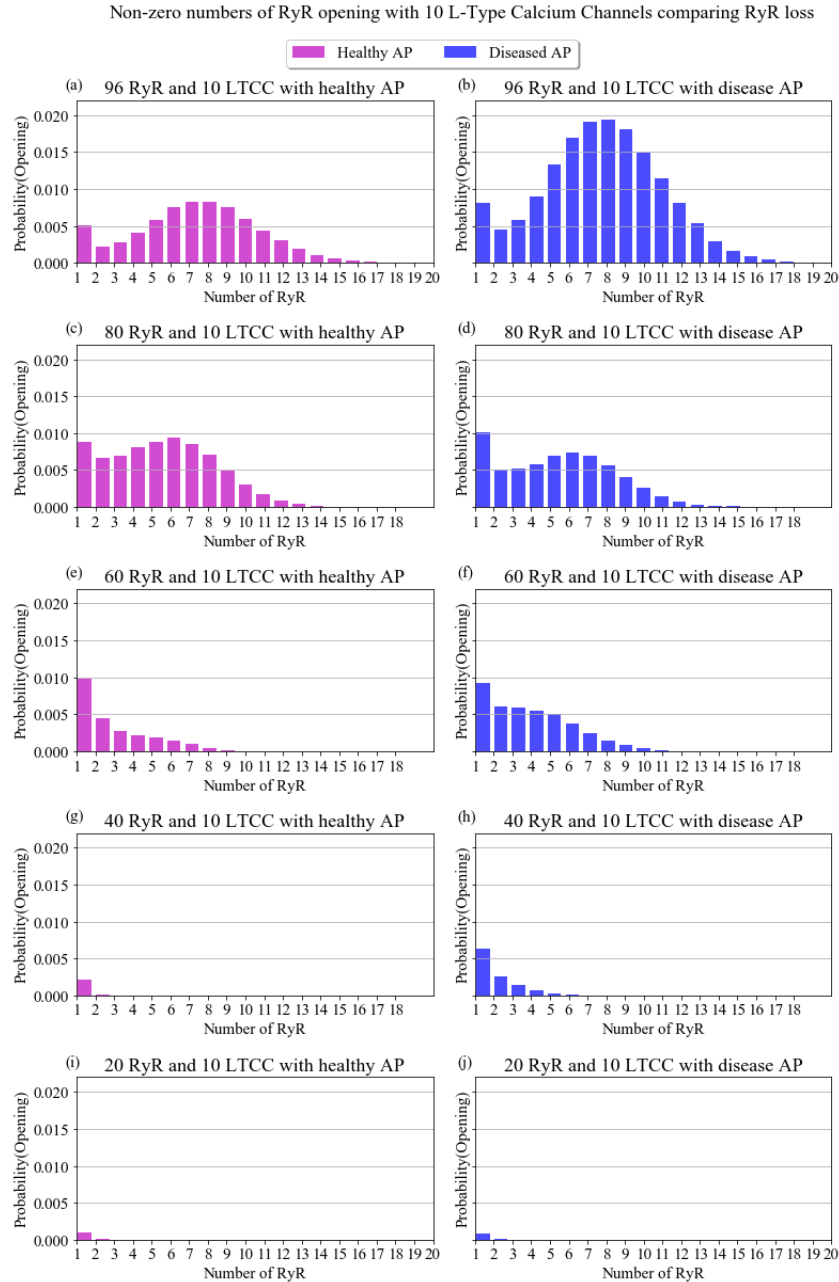


Figure 4.15. Non-zero numbers of RyR opening with 10 L-Type Calcium Channels comparing RyR loss

(a) Number of RyR open of 96 total with healthy AP; (b) Number of RyR open of 96 total with diseased AP; (c) Number of RyR open of 80 total with healthy AP; (d) Number of RyR open of 80 total with diseased AP; (e) Number of RyR open after 3 normal action potential stimuli in healthy t-tubules; (f) Number of RyR open after 3 diseased action potential stimuli in healthy t-tubules; (g) Number of RyR open after 4 normal action potential stimuli in healthy t-tubules; (h) Number of RyR open after 4 diseased action potential stimuli in healthy t-tubules; (i) Number of RyR open after 5 normal action potential stimuli in healthy t-tubules; (j) Number of RyR open after 5 diseased action potential stimuli in healthy t-tubules; Healthy action potential (pink), Diseased action potential (blue).

4.4.4 Simulating “heartbeats” with successive action potentials

The single action potential stimulation of Ryanodine receptors in the realistic geometry is a significant step towards realistic modeling of CICR. However, there are caveats to usage of single action potentials. Namely, these action potentials affect systems starting from equilibrium. In rapidly stimulated cells such as the mouse heart, which beats at a rate of 600 bpm, the system is hardly at equilibrium ever. The constant stimulation of these cells means that to accurately depict the complexity of the subcellular landscape, we must subject the cell to similar stimulation.

In order to model this successive stimulation that underlies the heartbeat of a mouse, we performed simulations that feature successive LTCC impulses for both healthy and diseased cases AP cases. We then examined the open probability of RyR opening at successive APs, the results of which are shown in figure 16. Similarly to our earlier experiments, the disease action potential is, overall, more pronounced than the healthy action potential (figure 16 a-i). Between the first and the second action potential, the differences are slightly increased, meaning that the CRUs are more sensitized after a second AP. By the third AP, the differences between the healthy and diseased APs become more significant, with the peak number of RyR firing shifting towards the range of number of RyR open shifting from a single RyR opening to a range of four to seven (figure 16 c,d). Additionally, the probability of firing increases significantly from under 0.01 after a single AP to about 0.02 as we move to the third AP (figure 16 e,f). By the third AP, the cell peaks at a range of seven to ten RyR open at a time of stimulation by a healthy AP. This peak number of open RyR is maintained as the action potentials increase (figure 16 g-i). Most interestingly, the diseased AP stimulated RyR are twice more likely to open in similar ranges of number of RyR activated (seven to ten). The successive action potentials (4AP and 5AP) only serve to sensitize the cell further. The ranges of RyR likely to fire after the 5th AP are at a maximal level in the diseased case. In future simulations, it will be important to gather our observations after the system has equilibrated, giving similar activation profiles after each AP. This will potentially require ten to fifteen successive APs because RyR are known to activate one-another in CICR, which at present, is computationally prohibitive (10 days/10 APs in a

Non-zero probability of RyR opening with 10 L-Type Calcium Channels during action potential train

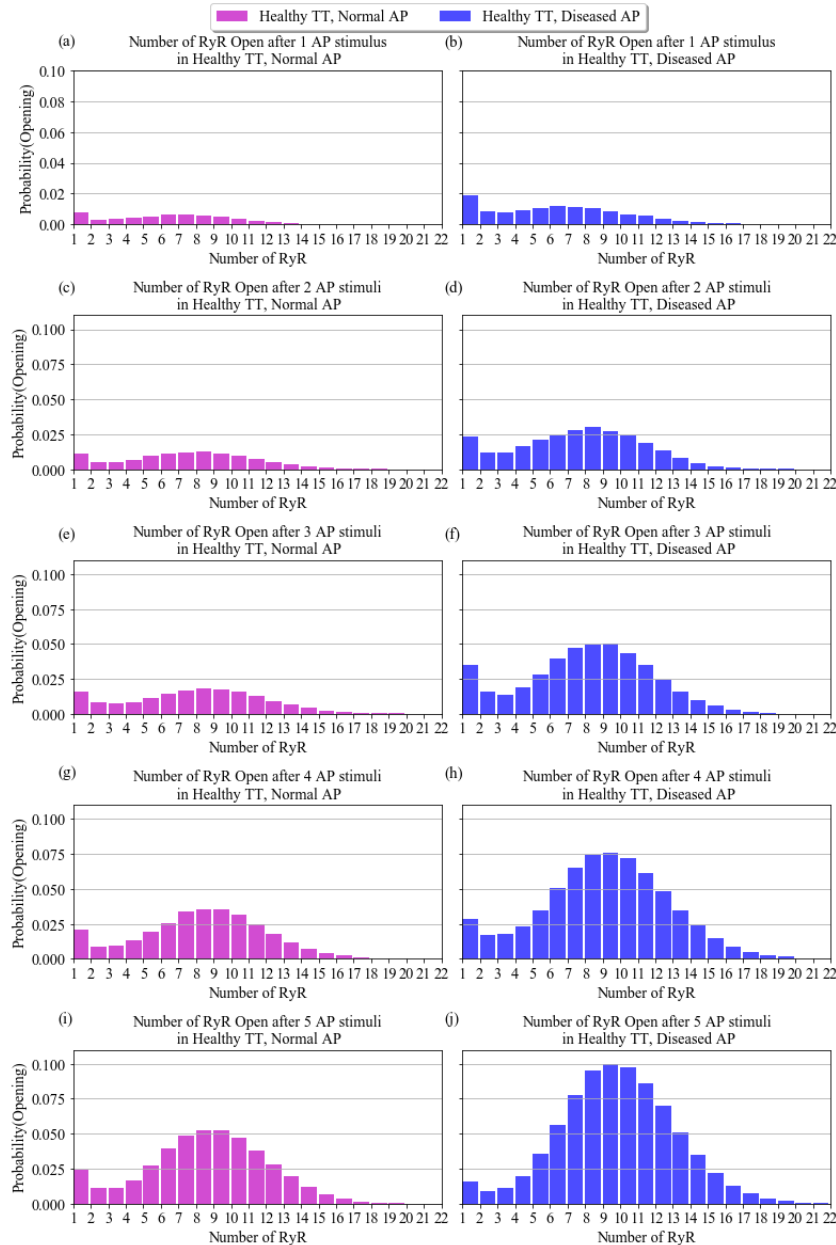


Figure 4.17. Non-zero probability of RyR opening with 10 L-Type Calcium Channels during an action potential train

(a) Number of RyR open after 1 normal action potential stimulus in healthy t-tubules; (b) Number of RyR open after 1 diseased action potential stimulus in healthy t-tubules; (c) Number of RyR open after 2 normal action potential stimuli in healthy t-tubules; (d) Number of RyR open after 2 diseased action potential stimuli in healthy t-tubules; (e) Number of RyR open after 3 normal action potential stimuli in healthy t-tubules; (f) Number of RyR open after 3 diseased action potential stimuli in healthy t-tubules; (g) Number of RyR open after 4 normal action potential stimuli in healthy t-tubules; (h) Number of RyR open after 4 diseased action potential stimuli in healthy t-tubules; (i) Number of RyR open after 5 normal action potential stimuli in healthy t-tubules; (j) Number of RyR open after 5 diseased action potential stimuli in healthy t-tubules; Healthy action potential (pink), Diseased action potential (blue)

single system).

4.4.5 Model caveats and future directions

Though our model yields significant insights into subcellular Ca^{2+} dynamics in myocardial CRUs, it is not without its shortcomings. First, the CRU geometry offers an $n=1$, meaning a low sample size with regards to spatial geometry variation, specifically the SR. An advantage of using MCell is the ability to import new segmented meshes within which the dynamics can be computed. With the development of new technologies to image [127] and reconstruct [31] cellular ultrastructure, this limitation can be remedied through the application of the system dynamics to new geometries.

Secondly, our model samples the dynamics of a small slice of the myocyte which is bound by a cytosolic region of size $1.43 \times 9.40 \times 4.06 \mu\text{M}$. This geometry is not large enough to fully encapsulate the effects that neighboring CRUs have on activation by CICR. To capture these dynamics, it is necessary to modulate the boundary condition of our cytosolic compartment. We plan to test different boundary conditions, whether they be reflective or absorptive, or even the addition of more cytosolic compartments as was done Hake et al. [33]. As of yet, our conversations with our continuum-modeling collaborators have not converged to what is “best-fit” for an explicit particle system like our own.

Because our system tracks the individual motions of every molecule in the system explicitly as well as all molecular reactions, large compute resources are required for sampling. Our present simulations are results from a total of 128 simulations per system (Normal T-Tubule, Healthy AP, etc). The systems that sampled only a single action potential equal to 3,840 independent simulations/30 systems. Adding in the simulations of RyR permutations sets us at over 5,000 individual simulations. Gathering accurate statistics for events that are rare to sample, such as single AP activated CICR, requires larger sampling of the system (on the order of 10,000 simulations or more). The results we report here are preliminary and likely to be exhibited in larger population sampling cases. Nevertheless, we admit that additional sampling is required to gain insight about confidence intervals and errors.

Molecular distribution is a key facet of spatial modeling methods. It is extremely likely that molecular distribution is non-linear. That is, molecular microdomains exist within the cytoplasmic region. At present, our model has not investigated the potential for protein localization alterations to influence our results. In future simulations, these effects can be directly interrogated using the special modeling techniques that we employ. The totality of the complex biochemical underpinnings of CICR are only partially explored by this model. Additional complexity is added when considering the regulation of the proteins on membranous surfaces like SERCA and RyR and of proteins in the cytosol like actin and myosin. To move towards a holistic and realistic understanding of CICR, it is necessary to include these species as they are known to play a vital role in cell function and disease states [159].

Our system required very little parameterization in order to produce the results that have been discussed. A study addressing the sensitivity of parameters in our system is required. To determine which variables are most likely to influence simulation results. For the purpose of peer review, it will be necessary to produce such reports.

Finally, to understand the effects of molecular crowding in systems with low molecular concentrations, it is not sufficient to represent molecular species as point particles. To address this, we have formed a collaboration with investigators at the Scripps Research Institute who have developed molecular packing algorithms that yield an accurate, atomic-level description of our system. The program we used for molecular packing is called cellPACK [162], and has been used in the past to visualize molecular dynamics simulations of an entire virus (unpublished). This cutting-edge visualization and packing technology can be combined with the reaction-diffusion powers of MCell to visualize the subcellular dynamics of our system at the finest level of molecular detail.

4.5 Conclusion

This chapter has laid out the details used to build the first-ever spatial, discrete model of subcellular calcium signaling mechanisms in realistic geometries of left ventricle mouse myocardium. The power of this model is not to be understated, as it can easily translate

and be applied to any segmented and meshed geometry obtained from electron microscopy methods. Arguably the most important facet of the model is the ability to count low numbers of calcium ions in parts of our geometry. Our simulations confirm the suspicions of those who intuit cardiomyocyte mechanics. Namely, that the average number of Ca^{2+} ions in the junctional dyadic cleft *are* zero at equilibrium [59].

Our model features a stochastic representation of Ryanodine Receptor dynamics that responds to local increases of cytosolic calcium. Coupling this advancement with the introduction of a voltage activated L-Type Calcium Channel model makes our model a suitable candidate for more-accurate description of the molecular-level details underlying cardiac calcium activation mechanisms. We incorporate discrete adaptations to existing SERCA model dynamics that satisfy microscopic reversibility [142]. Building upon the sophistication of the original CRU model by Hake et al. [33], we have moved towards towards a more realistic representation of the subcellular biochemistry.

Our investigations have determined that disease phenotypes are detectable on the level of single molecules in subcellular geometries. Our interrogation of disease action potentials resemble late calcium sparks that have been reported in the literature [157]. Using realistic geometries [32], we explored the effects of deformation on the junctional T-Tubule and demonstrated that drastic detubulation severely affects subcellular calcium handling, consistent with the observations of others [126,147]. Finally, we suggest that disease APs have a compensatory effect for deformations on the subcellular level.

4.6 Acknowledgments

Chapter 4 is currently being prepared for submission for publication of the material. **Hirakis, Sophia P.**; Bartol, Thomas M.; Sejnowski, Terrence, J; Amaro, Rommie E; The dissertation author was the primary investigator and author of this material.

This work would not have been possible without the scientific collaboration, expertise and aide of the following people, whom we wish to formally acknowledge. First, we thank Dr. Johan E. Hake for his help in translating his original model in our study. We thank Dr. Andrew G.

Edwards for his constant mentorship and advice in designing the simulations of the system. Dr. Andrew D. McCulloch also aided in the system design and contextualization of observations. We thank Dr. Eric E. Sobie for his frequent discussions on model scope and design. We acknowledge the lab of Dr. David D. Thomas, and especially Dr. J. Michael Autry for his frequent discussions on SERCA function. We acknowledge Dr. J. Andrew McCammon for his support of the project, and Dr. Rommie E. Amaro for pushing the model to completion. We acknowledge Dr. Terrence J. Sejnowski for his support and the entire lab for their feedback, especially Jorge Aldana for his support in computing resources. We thank Robert Kuczewski for his constant support and development of MCell. I would like to explicitly acknowledge Dr. Thomas M. Bartol for working tirelessly to develop this model with me for the greater half of the last decade; I would be nowhere without you.

4.7 Supplementary Information

Table 4.1. Parameters of Calcium buffering species

| Parameter | Description | Value | Units | Reference |
|----------------------|--|--------------------------|----------------|---|
| D^{cyt}_{Ca} | Diffusion constant of Ca^{2+} (in cytosol and SR) | 2.2×10^6 | cm^2s^{-1} | Hake et al. (2012); Louch et al. (2010) |
| D^{cyt}_{ATP-Ca} | Diffusion constant of Ca^{2+} -bound ATP in cytosol | 1.4×10^{-6} | cm^2s^{-1} | Hake et al. (2012); Valent et al (2007) |
| $[Ca^{2+}]_{cyt}$ | Initial Concentration of Ca^{2+} in cytosol | 140×10^{-9} | M | Hake et al. (2012) |
| $[ATP]_{cyt}$ | Free ATP concentration in cytosol | 454.682×10^{-6} | M | Hake et al. (2012); Bers (2001) |
| $[ATP-Ca]_{cyt}$ | Ca^{2+} -bound ATP concentration in cytosol | 318×10^{-9} | M | Hake et al. (2012) |
| ATP_{kon} | ATP Ca^{2+} on rate constant | 2.25×10^8 | $M^{-1}s^{-1}$ | Hake et al. (2012); Picht et al. (2011); Bers (2001) |
| ATP_{koff} | ATP Ca^{2+} off rate constant | 450×10^2 | s^{-1} | Hake et al. (2012); Picht et al. (2011) |
| D^{cyt}_{CMDN} | Diffusion constant of Calmodulin in cytosol | 25×10^{-8} | cm^2s^{-1} | Hake et al. (2012); Michailova et al. (2002) |
| $D^{cyt}_{CMDN-Ca}$ | Diffusion constant of Ca^{2+} -bound Calmodulin in cytosol | $\times 10^{-6}$ | cm^2s^{-1} | Hake et al. (2012); |
| $[CMDN]_{cyt}$ | Free Calmodulin concentration in cytosol | 23.529×10^{-6} | M | Hake et al. (2012); Fabiato (1983) |
| $[CMDN-Ca]_{cyt}$ | Ca^{2+} -bound Calmodulin concentration in cytosol | 471×10^{-9} | M | Hake et al. (2012) |
| $CMDN_{kon}$ | Calmodulin Ca^{2+} on rate constant | 34×10^6 | $M^{-1}s^{-1}$ | Hake et al. (2012); Robertson et al (1981); Picht et al. (2011) |
| $CMDN_{koff}$ | Calmodulin Ca^{2+} off rate constant | 238 | s^{-1} | Hake et al. (2012); Robertson et al (1981) |
| D^{cyt}_{TRPN} | Diffusion constant of Troponin-C in cytosol | 0 | cm^2s^{-1} | Hake et al. (2012) |
| $D^{cyt}_{TRPN-Ca}$ | Diffusion constant of Ca^{2+} -bound Troponin-C in cytosol | 0 | cm^2s^{-1} | Hake et al. (2012) |
| $[TRPN]_{cyt}$ | Free Troponin-C concentration in cytosol | 56.8×10^{-6} | M | Hake et al. (2012); Bondarenko et al. (2004) |
| $[TRPN-Ca]_{cyt}$ | Ca^{2+} -bound Troponin-C concentration in cytosol | 13.2×10^{-6} | M | Hake et al. (2012) |
| $TRPN_{kon}$ | Troponin-C Ca^{2+} on rate constant | 32.7×10^6 | $M^{-1}s^{-1}$ | Hake et al. (2012); Bondarenko et al. (2004) |
| $TRPN_{koff}$ | Troponin-C Ca^{2+} off rate constant | 19.6 | s^{-1} | Hake et al. (2012); Bondarenko et al. (2004) |
| D^{cyt}_{Fluo4} | Diffusion constant of Fluo-4 in cytosol | 42×10^{-8} | cm^2s^{-1} | Hake et al. (2012); Picht et al. (2011) |
| $D^{cyt}_{Fluo4-Ca}$ | Diffusion constant of Ca^{2+} -bound Fluo-4 in cytosol | 42×10^{-8} | cm^2s^{-1} | Hake et al. (2012) |
| $[Fluo4]_{cyt}$ | Free Fluo-4 concentration in cytosol | 22.186×10^{-6} | M | Hake et al. (2012); Picht et al. (2011) |
| $[Fluo4-Ca]_{cyt}$ | Ca^{2+} -bound Fluo-4 concentration in cytosol | 2.82×10^{-6} | M | Hake et al. (2012) |
| $Fluo4_{kon}$ | Fluo-4 Ca^{2+} on rate constant | 110×10^6 | $M^{-1}s^{-1}$ | Hake et al. (2012); Picht et al. (2011) |
| $Fluo4_{koff}$ | Fluo-4 Ca^{2+} off rate constant | 110 | s^{-1} | Hake et al. (2012); Picht et al. (2011) |
| D^{cyt}_{Fluo5} | Diffusion constant of Fluo-5 in cytosol | 8×10^{-8} | cm^2s^{-1} | Hake et al. (2012); Picht et al. (2011) |
| $D^{cyt}_{Fluo5-Ca}$ | Diffusion constant of Ca^{2+} -bound Fluo-5 in cytosol | 8×10^{-8} | cm^2s^{-1} | Hake et al. (2012) |
| $[Fluo5]_{cyt}$ | Free Fluo-5 concentration in cytosol | 5.9×10^{-6} | M | Hake et al. (2012); Picht et al. (2011) |
| $[Fluo5-Ca]_{cyt}$ | Ca^{2+} -bound Fluo-5 concentration in cytosol | 19.1×10^{-6} | M | Hake et al. (2012) |
| $Fluo5_{kon}$ | Fluo-5 Ca^{2+} on rate constant | 110×10^6 | $M^{-1}s^{-1}$ | Hake et al. (2012); Picht et al. (2011) |
| $Fluo5_{koff}$ | Fluo-5 Ca^{2+} off rate constant | 110 | s^{-1} | Hake et al. (2012) |
| D^{sr}_{CSQN} | Diffusion constant of Calsequestrin in cytosol | 0 | cm^2s^{-1} | Hake et al. (2012) |
| $D^{sr}_{CSQN-Ca}$ | Diffusion constant of Ca^{2+} -bound Calsequestrin in SR | 0 | cm^2s^{-1} | Hake et al. (2012) |
| $[CSQN]_{cyt}$ | Free Calsequestrin concentration in SR | 56.8×10^{-6} | M | Hake et al. (2012); Bondarenko et al. (2004) |
| $[CSQN-Ca]_{cyt}$ | Ca^{2+} -bound Calsequestrin concentration in SR | 13.2×10^{-6} | M | Hake et al. (2012) |
| $CSQN_{kon}$ | Calsequestrin Ca^{2+} on rate constant | 32.7×10^6 | $M^{-1}s^{-1}$ | Hake et al. (2012); Bondarenko et al. (2004) |
| $CSQN_{koff}$ | Calsequestrin Ca^{2+} off rate constant | 19.6 | s^{-1} | Hake et al. (2012); Bondarenko et al. (2004) |

Table 4.2. Parameters for Sodium-Calcium Exchanger (NCX) and Plasma Membrane Calcium-ATPase (PMCA) pump models

| Parameter | Value | Units | Reference |
|------------------|-------------------|----------------|--|
| k_{P0P1} | 1.5×10^8 | $M^{-1}s^{-1}$ | Bartol et al. (2015); Brini and Carafoli (2009); Penheiter et al. (2003) |
| k_{P0P1} | 15 | s^{-1} | Bartol et al. (2015); Brini and Carafoli (2009); Penheiter et al. (2003) |
| k_{P0P1_pump} | 12 | s^{-1} | Bartol et al. (2015); Brini and Carafoli (2009); Penheiter et al. (2003) |
| k_{PMCA_leak} | 5.25 | s^{-1} | Yields 140nM Cytosolic Calcium |
| k_{N0N1} | 3.0×10^8 | $M^{-1}s^{-1}$ | Bartol et al. (2015); Hilgemann(1991) |
| k_{N1N0} | 300 | s^{-1} | Bartol et al. (2015); Hilgemann(1991) |
| k_{N1N0_pump} | 600 | s^{-1} | Bartol et al. (2015); Hilgemann(1991) |
| k_{NCX_leak} | 26.75 | s^{-1} | Yields 140nM Cytosolic Calcium |

Table 4.3. Parameters for L-Type Calcium Channel model

| Parameter | Value | Units | Reference |
|-----------|--------------------------|----------------|-------------------------------|
| a | 2 | - | Greenstein and Winslow (2002) |
| b | 1.9356 | - | Greenstein and Winslow (2002) |
| f | 850 | s^{-1} | Greenstein and Winslow (2002) |
| g | 2000 | s^{-1} | Greenstein and Winslow (2002) |
| f' | 5 | s^{-1} | Greenstein and Winslow (2002) |
| g' | 7000 | s^{-1} | Greenstein and Winslow (2002) |
| γ | 0.44×10^6 | $M^{-1}s^{-1}$ | Greenstein and Winslow (2002) |
| ω | 0.02158×10^3 | s^{-1} | Greenstein and Winslow (2002) |
| T | 310 | K | - |
| F | 96.485×10^3 | C/mol | - |
| N_A | 6.02214×10^{23} | #/mol | - |
| P_{Ca} | 9.13×10^{-16} | L/s | - |

Table 4.4. Parameters for Ryanodine Receptor Markov model

| Parameter | Value | Units | Reference |
|-----------------|--------------------|----------------|------------------------|
| k_{C1C2_H1} | 3.26×10^6 | $M^{-1}s^{-1}$ | Saftenku et al. (2001) |
| k_{C2C1_H1} | 116 | s^{-1} | Saftenku et al. (2001) |
| k_{C2C3_H1} | 0.66×10^6 | $M^{-1}s^{-1}$ | Saftenku et al. (2001) |
| k_{C3C2_H1} | 163 | s^{-1} | Saftenku et al. (2001) |
| k_{C1C2_L} | 1.24×10^6 | $M^{-1}s^{-1}$ | Saftenku et al. (2001) |
| k_{C2C1_L} | 13.6 | s^{-1} | Saftenku et al. (2001) |
| k_{C2C3_L} | 29.8×10^6 | $M^{-1}s^{-1}$ | Saftenku et al. (2001) |
| k_{C3C2_L} | 3867 | s^{-1} | Saftenku et al. (2001) |
| k_{C2C5_L} | 1.81 | s^{-1} | Saftenku et al. (2001) |
| k_{C5C2_L} | 3.63 | s^{-1} | Saftenku et al. (2001) |
| k_{C1C2_LH1} | 6.67×10^2 | $M^{-1}s^{-1}$ | Saftenku et al. (2001) |
| k_{C2C1_H1L} | 0.0833 | s^{-1} | Saftenku et al. (2001) |
| k_{C2C3_LH1} | 6.67×10^2 | $M^{-1}s^{-1}$ | Saftenku et al. (2001) |
| k_{C3C2_H1L} | 0.0833 | s^{-1} | Saftenku et al. (2001) |
| k_{C3C4_LH1} | 6.67×10^2 | $M^{-1}s^{-1}$ | Saftenku et al. (2001) |
| k_{C4C3_H1L} | 0.0833 | s^{-1} | Saftenku et al. (2001) |
| k_{C2O1_H1} | 7.86×10^6 | $M^{-1}s^{-1}$ | Saftenku et al. (2001) |
| k_{O1C2_H1} | 1480 | s^{-1} | Saftenku et al. (2001) |
| k_{C3O2_H1} | 7.77×10^6 | $M^{-1}s^{-1}$ | Saftenku et al. (2001) |
| k_{O2C3_H1} | 330 | s^{-1} | Saftenku et al. (2001) |
| k_{C3O1_L} | 731.2 | s^{-1} | Saftenku et al. (2001) |
| k_{O1C3_L} | 4185 | s^{-1} | Saftenku et al. (2001) |
| k_{C3O2_L} | 24.5 | s^{-1} | Saftenku et al. (2001) |
| k_{O2C3_L} | 156.5 | s^{-1} | Saftenku et al. (2001) |
| k_{C3O3_L} | 8.5 | s^{-1} | Saftenku et al. (2001) |
| k_{O3C3_L} | 111.7 | s^{-1} | Saftenku et al. (2001) |
| k_{C4O2_L} | 415.3 | s^{-1} | Saftenku et al. (2001) |
| k_{O2C4_L} | 1995 | s^{-1} | Saftenku et al. (2001) |
| k_{C4O3_L} | 43.3 | s^{-1} | Saftenku et al. (2001) |
| k_{O3C4_L} | 253.3 | s^{-1} | Saftenku et al. (2001) |
| k_{C4O2_H1} | 2390 | s^{-1} | Saftenku et al. (2001) |
| k_{O2C4_H1} | 298 | s^{-1} | Saftenku et al. (2001) |
| k_{RyR_flux} | 1.09×10^9 | $M^{-1}s^{-1}$ | Guo et al. (2012) |

Table 4.5. Parameters for Sarco/Endoplasmic Reticulum Calcium-ATPase (SERCA) pump

| Parameter | Value | Units | Reference |
|------------------|-----------------|----------------|---|
| k_{X0X1} | 2×10^8 | $M^{-1}s^{-1}$ | Bartol et al. (2015); Higgins et al. (2006) |
| k_{X1X0} | 146.775 | s^{-1} | Bartol et al. (2015); Higgins et al (2006) |
| k_{X1X2} | 1×10^8 | $M^{-1}s^{-1}$ | Bartol et al. (2015); Higgins et al. (2006) |
| k_{X2X1} | 293.551 | s^{-1} | Bartol et al. (2015); Higgins et al. (2006) |
| k_{X2Y2} | 0.6 | s^{-1} | Bartol et al. (2015); Higgins et al. (2006) |
| k_{Y2X2} | 0.097 | s^{-1} | Bartol et al. (2015); Higgins et al. (2006) |
| k_{Y2Y1} | 60.03 | s^{-1} | Bartol et al. (2015); Higgins et al. (2006) |
| k_{Y1Y2} | 1×10^5 | $M^{-1}s^{-1}$ | Bartol et al. (2015); Higgins et al. (2006) |
| k_{Y1Y0} | 30.015 | s^{-1} | Bartol et al. (2015); Higgins et al. (2006) |
| k_{Y0Y1} | 2×10^5 | $M^{-1}s^{-1}$ | Bartol et al. (2015); Higgins et al. (2006) |
| k_{Y0X0} | 0.4 | s^{-1} | Bartol et al. (2015); Higgins et al. (2006) |
| k_{X0Y0} | 0.0012 | s^{-1} | Bartol et al. (2015); Higgins et al. (2006) |

Bibliography

- [1] G. E. Moore. Cramming More Components Onto Integrated Circuits, *Electronics*, April 19, 1965. *Electronics*, 38(8):82–85, 1965.
- [2] A. Caretta and C. Mucignat-Caretta. Protein Kinase A in cancer, 2011.
- [3] A. Horvath, J. Bertherat, L. Groussin, M. Guillaud-Bataille, K. Tsang, L. Cazabat, R. Libe, E. Remmers, F. Rene-Corail, F. R. Faucz, E. Clauser, A. Calender, X. Bertagna, J. A. Carney, and C. A. Stratakis. Mutations and Polymorphisms in the Gene Encoding Regulatory Subunit Type 1-Alpha of Protein Kinase A (PRKAR1A): An Update. *Human Mutation*, 31(4):369–379, 2010.
- [4] C. Kim, N. H. Xuong, and S. S. Taylor. Crystal structure of a complex between the catalytic and regulatory (RI α) subunits of PKA. *Science*, 2005.
- [5] C. Kim, C. Y. Cheng, S. A. Saldanha, and S. S. Taylor. PKA-I holoenzyme structure reveals a mechanism for cAMP-dependent activation. *Cell*, 130(6):1032–1043, 2007.
- [6] C. Y. Cheng, J. Yang, S. S. Taylor, and D. K. Blumenthal. Sensing domain dynamics in protein kinase A-I{alpha} complexes by solution X-ray scattering. *J Biol Chem*, 284(51):35916–35925, 2009.
- [7] L. S. Kirschner, J. A. Carney, S. D. Pack, S.E. Taymans, C. Giatzakis, Y. S. Cho, Y.S. Cho-Chung, and C. A. Stratakis. Mutations of the gene encoding the protein kinase A type I- α regulatory subunit in patients with the Carney complex. *Nature Genetics*, 2000.
- [8] D. Calebiro, A. Hannawacker, S. Lyga, K. Bathon, U. Zabel, C. Ronchi, F. Beuschlein, M. Reincke, K. Lorenz, B. Allolio, C. Kisker, M. Fassnacht, and M. J. Lohse. PKA catalytic subunit mutations in adrenocortical Cushing’s adenoma impair association with the regulatory subunit. *Nature Communications*, 2014.
- [9] W. L. Bragg. The diffraction of X-rays by crystals, 2014.
- [10] H. C. Urey, F. G. Brickwedde, and G. M. Murphy. A hydrogen isotope of mass 2 and its concentration. *Physical Review*, 1932.
- [11] John R. Engen and David L. Smith. Peer Reviewed: Investigating Protein Structure and Dynamics by Hydrogen Exchange MS. *Analytical Chemistry*, 2001.
- [12] L. Konermann, J. Pan, and Y. H. Liu. Hydrogen exchange mass spectrometry for studying protein structure and dynamics, 2011.

- [13] R.A. Harrison and J. R. Engen. Conformational insight into multi-protein signaling assemblies by hydrogen-deuterium exchange mass spectrometry, 2016.
- [14] L. Boldon, F. Laliberte, and L. Liu. Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application. *Nano Reviews*, 2015.
- [15] J. W. P. Carvalho, P. S. Santiago, T. Batista, C. E. G. Salmon, L. R. S. Barbosa, R. Itri, and M. Tabak. On the temperature stability of extracellular hemoglobin of *Glossoscolex paulistus*, at different oxidation states: SAXS and DLS studies. *Biophysical Chemistry*, 2012.
- [16] J. Dubochet, J. Lepault, R. Freeman, J. A. Berriman, and J. ÅR̈C Homo. Electron microscopy of frozen water and aqueous solutions. *Journal of Microscopy*, 1982.
- [17] J. Lepault, F. P. Booy, and J. Dubochet. Electron microscopy of frozen biological suspensions. *Journal of Microscopy*, 1983.
- [18] J. Lepault and J. Dubochet. Electron Microscopy of Frozen Hydrated Specimens: Preparation and Characteristics. *Methods in Enzymology*, 1986.
- [19] N. Sartori, K. Richter, and J. Dubochet. Vitrification depth can be increased more than 10-fold by high-pressure freezing. *Journal of Microscopy*, 1993.
- [20] R. Ishima and D. A. Torchia. Protein dynamics from NMR, 2000.
- [21] J. Frank, J. Zhu, P. Penczek, Y. Li, S. Srivastava, A. Verschoor, M. Radermacher, R. Grassucci, R. K. Lata, and R. K. Agrawal. A model of protein synthesis based on cryo-electron microscopy of the E. coli ribosome. *Nature*, 1995.
- [22] H. Lee, S. A. Brendle, S.M. Bywaters, J. Guan, R. E. Ashley, J. D. Yoder, A. M. Makhov, J. F. Conway, N.D. Christensen, and S. Hafenstein. A Cryo-Electron Microscopy Study Identifies the Complete H16.V5 Epitope and Reveals Global Conformational Changes Initiated by Binding of the Neutralizing Antibody Fragment. *Journal of Virology*, 2015.
- [23] M. Sevana, F. Long, A. S. Miller, T. Klose, G. Buda, L. Sun, R. J. Kuhn, and M. G. Rossmann. Refinement and Analysis of the Mature Zika Virus Cryo-EM Structure at 3.1 Å Resolution. *Structure*, 2018.
- [24] W. Peng, H. Shen, J. Wu, W. Guo, X. Pan, R. Wang, S. R. W. Chen, and N. Yan. Structural basis for the gating mechanism of the type 2 ryanodine receptor RyR2. *Science*, 2016.
- [25] J. Frank. *Untersuchungen von elektronenmikroskopischen Aufnahmen hoher Auflösung mit Bilddifferenz- und Rekonstruktionsverfahren*[*Studies of Electron Micrographs Using Image Difference and Reconstruction Methods*]. PhD thesis, Technische Universität München, 1970.
- [26] W. O. Saxton and J. Frank. Motif detection in quantum noise-limited electron micrographs by cross-correlation. *Ultramicroscopy*, 1976.

- [27] J. Frank. Single-particle reconstruction of biological macromolecules in electron microscopy-30 years. *Quarterly Reviews of Biophysics*, 2009.
- [28] K. Fridman, A. Mader, M. Zwerger, N. Elia, and O. Medalia. Advances in tomography: Probing the molecular architecture of cells, 2012.
- [29] R. I. Koning, A. J. Koster, and T. H. Sharp. Advances in cryo-electron tomography for biology and medicine, 2018.
- [30] Z. Y. Yu, M. J. Holst, and J. A. McCammon. High-fidelity geometric modeling for biomedical applications. *Finite Elements in Analysis and Design*, 44(11):715–723, 2008.
- [31] C. T. Lee, J. B. Moody, R. E. Amaro, J. A. McCammon, and M. Holst. The Implementation of the Colored Abstract Simplicial Complex and its Application to Mesh Generation. 0(0), 2018.
- [32] T. Hayashi, M. E. Martone, Z. Y. Yu, A. Thor, M. Doi, M. J. Holst, M. H. Ellisman, and M. Hoshijima. Three-dimensional electron microscopy reveals new details of membrane systems for Ca²⁺ signaling in the heart. *Journal of Cell Science*, 122(7):1005–1013, 2009.
- [33] J. E. Hake, A. G. Edwards, Z. Y. Yu, P. M. Kekenus-Huskey, A. P. Michailova, J. A. McCammon, M. J. Holst, M. Hoshijima, and A. D. McCulloch. Modelling cardiac calcium sparks in a three-dimensional reconstruction of a calcium release unit. *Journal of Physiology-London*, 590(18):4403–4422, 2012.
- [34] M. Levitt and A. Washel. Computational simulation of protein folding. *Nature*, 1975.
- [35] J. A. McCammon and M. Karplus. Dynamics of activated processes in globular proteins. *Proc Natl Acad Sci U S A*, 76(8):3585–3589, 1979.
- [36] D. E. Shaw, K.J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. L. Klepeis, J. S. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, M. A. Moraes, R. O. Dror, S. Piana, Y. Shan, B. Towles, J. K. Salmon, J. P. Grossman, K.M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, and B. Batson. Millisecond-scale molecular dynamics simulations on Anton. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis - SC '09*, 2009.
- [37] L. C. T. Pierce, R. Salomon-Ferrer, C. A. F. de Oliveira, J. A. McCammon, and R. C. Walker. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *Journal of Chemical Theory and Computation*, 8(9):2997–3002, 2012.
- [38] D.A. Case, R.M. Betz, D.S. Cerutti, T.E. III Cheatham, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, and J.T. Berryman. AMBER 2015, 2015.
- [39] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry*, 30(10):1545–1614, 2009.

- [40] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J.C. Berendsen. GROMACS: Fast, flexible, and free, 2005.
- [41] W. Damm, A. Frontera, J. Tirado-Rives, and W. L. Jorgensen. OPLS all-atom force field for carbohydrates. *Journal of Computational Chemistry*, 1997.
- [42] M. Christen, P. H. Hunenberger, D. Bakowies, R. Baron, R. Burgi, D. P. Geerke, T. N. Heinz, M. A. Kastenholtz, V. Krautler, C. Oostenbrink, C. Peter, D. Trzesniak, and W. F. Van Gunsteren. The GROMOS software for biomolecular simulation: GROMOS05. *Journal of Computational Chemistry*, 26(16):1719–1751, 2005.
- [43] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, and P. Kollman. Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Computer Physics Communications*, 91(1-3):1–41, 1995.
- [44] M. T. Nelson, W. Humphrey, A. Gursoy, A. Dalke, L. V. Kale, R. D. Skeel, and K. Schulten. NAMD: A parallel, object oriented molecular dynamics program. *International Journal of Supercomputer Applications and High Performance Computing*, 10(4):251–268, 1996.
- [45] K. J. Bowers, F. D. Sacerdoti, J. K. Salmon, Y. Shan, D. E. Shaw, E. Chow, H. Xu, R. O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, Istvan Kolossvary, and Mark A. Moraes. Molecular dynamics—Scalable algorithms for molecular dynamics simulations on commodity clusters. In *Proceedings of the 2006 ACM/IEEE conference on Supercomputing - SC '06*, 2006.
- [46] R. Anandakrishnan, A. Drozdetski, R. C. Walker, and A. V. Onufriev. Speed of conformational change: Comparing explicit and implicit solvent molecular dynamics simulations. *Biophysical Journal*, 2015.
- [47] P. Florová, P. Sklenovský, P. Banáš, and M. Otyepka. Explicit water models affect the specific solvation and dynamics of unfolded peptides while the conformational behavior and flexibility of folded peptides remain intact. *Journal of Chemical Theory and Computation*, 2010.
- [48] A. Hospital, J. R. Goñi, M. Orozco, and J. L. Gelpí. Molecular dynamics simulations: Advances and applications, 2015.
- [49] G. A. Huber and J. A. McCammon. Browndye: A software package for Brownian dynamics. *Computer Physics Communications*, 181(11):1896–1905, 2010.
- [50] S. H. Northrup, S. A. Allison, and J. A. McCammon. Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *The Journal of Chemical Physics*, 1984.
- [51] R. R. Gabdouliline and R. C. Wade. Simulation of the diffusional association of barnase and barstar. *Biophysical Journal*, 1997.
- [52] R. R. Gabdouliline and R. C. Wade. Brownian dynamics simulation of protein-protein diffusional encounter. *Methods*, 14(3):329–341, 1998.

- [53] J. Schöneberg and F. Noé. ReaDDy—a software for particle-based reaction-diffusion dynamics in crowded cellular environments. *PLoS one*, 2013.
- [54] T. Geyer. Many-particle Brownian and Langevin Dynamics Simulations with the Brownmove package. *BMC Biophysics*, 2011.
- [55] M. Długosz, P. Zielinski, and J. Trylska. Brownian dynamics simulations on CPU and GPU with BD-BOX. *Journal of Computational Chemistry*, 2011.
- [56] B. Honig and A. Nicholls. Classical electrostatics in biology and chemistry. *Science*, 1995.
- [57] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, 98(18):10037–10041, 2001.
- [58] J. R. Faeder, M. L. Blinov, and W. S. Hlavacek. Rule-based modeling of biochemical systems with BioNetGen. *Methods in Molecular Biology*, 2009.
- [59] D. M. Bers. Cardiac excitation contraction coupling. *Nature*, 415:198–205, 2002.
- [60] J. R. Stiles and T. M. Bartol. Monte Carlo Methods for Simulating Realistic Synaptic Microphysiology using MCELL. In *Computational Neuroscience: Realistic Modeling for Experimentalists*. 2001.
- [61] R. A. Kerr, T. M. Bartol, B. Kaminsky, M. Dittrich, J. C. J. Chang, S. B. Baden, T. J. Sejnowski, and J. R. Stiles. Fast Monte Carlo Simulation Methods for Biological Reaction-Diffusion Systems in Solution and on Surfaces. *Siam Journal on Scientific Computing*, 30(6):3126–3149, 2008.
- [62] S. Gupta, J. Czech, R. Kuczewski, T. M. Bartol, T. J. Sejnowski, R. E. C. Lee, and J. R. Faeder. No Title. In *Quantitative Biology: Theory, Computational Methods, and Models*, chapter 24, pages 485–512. 2018.
- [63] J. R. Carapetis, A. C. Steer, E. K. Mulholland, and M. Weber. The global burden of group A streptococcal diseases. *Lancet Infect Dis*, 5(11):685–694, 2005.
- [64] J. N. Cole, T. Barnett, V. Nizet, and M. J. Walker. Molecular insight into invasive group A streptococcal disease, 2011.
- [65] J. B. Dale, T. A. Penfound, B. Tamboura, S. O. Sow, J. P. Nataro, M. Tapia, and K. L. Kotloff. Potential coverage of a multivalent M protein-based group A streptococcal vaccine. *Vaccine*, 31(12):1576–1581, 2013.
- [66] M. F. Good, M. Pandey, M. R. Batzloff, and G. J. Tyrrell. Strategic development of the conserved region of the M protein and other candidates as vaccines to prevent infection with group A streptococci, 2015.
- [67] C. McNamara, A. S. Zinkernagel, P. Macheboeuf, M.W. Cunningham, V. Nizet, and P. Ghosh. Coiled-coil irregularities and instabilities in group A Streptococcus M1 are required for virulence. *Science*, 2008.

- [68] P. Ghosh. The nonideal coiled coil of M protein and its multifarious functions in pathogenesis. *Adv Exp Med Biol*, 715:197–211, 2011.
- [69] C. Sandin, F. Carlsson, and G. Lindahl. Binding of human plasma proteins to Streptococcus pyogenes M protein determines the location of opsonic and non-opsonic epitopes. *Molecular Microbiology*, 2006.
- [70] T. A. Penfound, E. Y. Chiang, E. A. Ahmed, and J. B. Dale. Protective efficacy of group A streptococcal vaccines containing type-specific and conserved M protein epitopes. *Vaccine*, 28(31):5017–5022, 2010.
- [71] J. Lannergrd, M. C.U. Gustafsson, J. Waldemarsson, A. Norrby-Teglund, M. Stålhammar-Carlemalm, and G. Lindahl. The hypervariable region of streptococcus pyogenes M protein escapes antibody attack by antigenic variation and weak immunogenicity. *Cell Host and Microbe*, 2011.
- [72] J. B. Dale, T. A. Penfound, E. Y. Chiang, and W. J. Walton. New 30-valent M protein-based vaccine evokes cross-opsonic antibodies against non-vaccine serotypes of group A streptococci. *Vaccine*, 29(46):8175–8178, 2011.
- [73] D. J. McMillan, P. A. Dreze, T. Vu, D. E. Bessen, J. Guglielmini, A. C. Steer, J. R. Carapetis, L. Van Melderren, K. S. Sriprakash, and P. R. Smeesters. Updated model of group A Streptococcus M proteins based on a comprehensive worldwide study. *Clin Microbiol Infect*, 19(5):222–9, 2013.
- [74] J. Persson, B. Beall, S. Linse, and G. Lindahl. Extreme sequence divergence but conserved ligand-binding specificity in Streptococcus pyogenes M protein. *Plos Pathogens*, 2(5):e47, 2006.
- [75] J. Persson, B. Beall, S. Linse, and G. Lindahl. Extreme sequence divergence but conserved ligand-binding specificity in Streptococcus pyogenes M protein. *PLoS Pathogens*, 2006.
- [76] D. Ermert and A. M. Blom. C4b-binding protein: The good, the bad and the deadly. Novel functions of an old friend, 2016.
- [77] J. D. Lambris, D. Ricklin, and B. V. Geisbrecht. Complement evasion by human pathogens, 2008.
- [78] A. M. Blom, T. Hallström, and K. Riesbeck. Complement evasion strategies of pathogens-Acquisition of inhibitors and beyond, 2009.
- [79] F. Carlsson, K. Berggård, M. Stålhammar-Carlemalm, and G. Lindahl. Evasion of Phagocytosis through Cooperation between Two Ligand-binding Regions in Streptococcus pyogenes M Protein. *The Journal of Experimental Medicine*, 2003.
- [80] D. Ermert, J. Shaughnessy, T. Joeris, J. Kaplan, C. J. Pang, E. A. Kurt-Jones, P. A. Rice, S. Ram, and A. M. Blom. Virulence of Group A Streptococci Is Enhanced by Human Complement Inhibitors. *PLoS Pathogens*, 2015.
- [81] D. H. Fremont, M. Matsumura, E. A. Stura, P. A. Peterson, and I. A. Wilson. Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. *Science*, 1992.

- [82] D. R. Madden, J. C. Gorga, J. L. Strominger, and D. C. Wiley. The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight peptide binding to MHC. *Cell*, 1992.
- [83] H. T. Jenkins, L. Mark, G. Ball, J. Persson, G. Lindahl, D. Uhrin, A. M. Blom, and P. N. Barlow. Human C4b-binding protein, structural basis for interaction with streptococcal M protein, a major bacterial virulence factor. *J Biol Chem*, 281(6):3690–3697, 2006.
- [84] P. Accardo, P. Sanchez-Corral, O. Criado, E. Garcia, and S. Rodriguez de Cordoba. Binding of human complement component C4b-binding protein (C4BP) to *Streptococcus pyogenes* involves the C4b-binding site. *J Immunol*, 157(11):4935–4939, 1996.
- [85] A. M. Blom, K. Berggard, J. H. Webb, G. Lindahl, B. O. Villoutreix, and B. Dahlback. Human C4b-binding protein has overlapping, but not identical, binding sites for C4b and streptococcal M proteins. *J Immunol*, 164(10):5328–5336, 2000.
- [86] I. André, J. Persson, A. M. Blom, H. Nilsson, T. Drakenberg, G. Lindahl, and S. Linse. Streptococcal M protein: Structural studies of the hypervariable region, free and bound to human C4BP. *Biochemistry*, 2006.
- [87] M.C. Lawrence and P.M. Colman. Shape complementarity at protein/protein interfaces. *J Mol Biol*, 234:946–950, 1993.
- [88] M. Sanderson-Smith, D.M.P. De Oliveira, J. Guglielmini, D.J. McMillan, T. Vu, J. K. Holien, A. Henningham, A. C. Steer, D. E. Bessen, J. B. Dale, N. Curtis, B. W. Beall, M.J. Walker, M. W. Parker, J. R. Carapetis, L. Van Melderen, K. S. Sriprakash, P. R. Smeesters, M. Batzloff, R. Towers, H. Goossens, S. Malhotra-Kumar, L. Guilherme, R. Torres, D. Low, A. McGeer, P. Krizova, S. El Tayeb, J. Kado, M. Van Der Linden, G. Erdem, A. Moses, R. Nir-Paz, T. Ikebe, H. Watanabe, S. Sow, B. Tamboura, B. Kittang, J. Melo-Cristino, M. Ramirez, M. Straut, A. Suvorov, A. Totolian, M. Engel, B. Mayosi, A. Whitelaw, J. Darenberg, B. H. Normark, C. C. Ni, J. J. Wu, A. De Zoysa, A. Efstratiou, S. Shulman, and R. Tanz. A systematic and functional classification of *Streptococcus pyogenes* that serves as a new tool for molecular typing and vaccine development. *Journal of Infectious Diseases*, 2014.
- [89] C. McNamara, A. S. Zinkernagel, P. Macheboeuf, M. W. Cunningham, V. Nizet, and P. Ghosh. Coiled-coil irregularities and instabilities in group A *Streptococcus* M1 are required for virulence. *Science*, 319(5868):1405–1408, 2008.
- [90] W. Kabsch, A. T Brünger, P. A. Karplus, K. Diederichs, S. McSweeney, R. B. G. Ravelli, P. Evans, S. French, K. Wilson, W. Kabsch, P. Kursula, and M. S. Weiss. XDS. *Acta Crystallographica Section D Biological Crystallography*, 2010.
- [91] Z. Otwinowski and W. Minor. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, 1997.
- [92] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R.J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 2010.

- [93] P. Emsley and K. Cowtan. Coot: Model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, 2004.
- [94] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 2010.
- [95] J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004.
- [96] M. H. M. Olsson, C. R. Sondergaard, M. Rostkowski, and J. H. Jensen. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK(a) Predictions. *Journal of Chemical Theory and Computation*, 7(2):525–537, 2011.
- [97] C. R. Sondergaard, M. H. M. Olsson, M. Rostkowski, and J. H. Jensen. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK(a) Values. *Journal of Chemical Theory and Computation*, 7(7):2284–2295, 2011.
- [98] P. F. Li, B. P. Roberts, D. K. Chakravorty, and K. M. Merz. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *Journal of Chemical Theory and Computation*, 9(6):2733–2748, 2013.
- [99] T Darden, D York, and L Pedersen. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics*, 98(12):10089–10092, 1993.
- [100] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A Smooth Particle Mesh Ewald Method. *Journal of Chemical Physics*, 103(19):8577–8593, 1995.
- [101] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–1802, 2005.
- [102] J. A. Izaguirre, D. P. Catarello, J. M. Wozniak, and R. D. Skeel. Langevin stabilization of molecular dynamics. *Journal of Chemical Physics*, 114(5):2090–2098, 2001.
- [103] W. Jiang, D. J. Hardy, J. C. Phillips, A. D. MacKerell, K. Schulten, and B. Roux. High-Performance Scalable Molecular Dynamics Simulations of a Polarizable Force Field Based on Classical Drude Oscillators in NAMD. *Journal of Physical Chemistry Letters*, 2(2):87–92, 2011.
- [104] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, and J. R. Haak. Molecular-Dynamics with Coupling to an External Bath. *Journal of Chemical Physics*, 81(8):3684–3690, 1984.
- [105] L. S. D. Caves, J. D. Evanseck, and M. Karplus. Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Science*, 7(3):649–666, 1998.
- [106] D. R. Roe and T. E. Cheatham. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*, 9(7):3084–3095, 2013.

- [107] D.A. Case, III Cheatham, T.E, Simmerling C.L., J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Gołtzy, I. Kolossvai, Ary, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, and Darden T A Wu. AMBER 12, 2012.
- [108] B. G. Levine, J. E. Stone, and A. Kohlmeyer. Fast analysis of molecular dynamics trajectories with graphics processing units-Radial distribution function histogramming. *Journal of Computational Physics*, 230(9):3556–3569, 2011.
- [109] D. R. Burton and J. R. Mascola. Antibody responses to envelope glycoproteins in HIV-1 infection, 2015.
- [110] L. Pappas, M. Foglierini, L. Piccoli, N. L. Kallewaard, F. Turrini, C. Silacci, B. Fernandez-Rodriguez, G. Agatic, I. Giacchetto-Sasselli, G. Pellicciotta, F. Sallusto, Q. Zhu, E. Vicenzi, D. Corti, and A. Lanzavecchia. Rapid development of broadly influenza neutralizing antibodies through redundant mutations. *Nature*, 2014.
- [111] A. C. Steer, I. Law, L. Matatolu, B.W. Beall, and J. R. Carapetis. Global emm type distribution of group A streptococci: systematic review and implications for vaccine development, 2009.
- [112] S. Ringer. A further Contribution regarding the influence of the different Constituents of the Blood on the Contraction of the Heart. *The Journal of Physiology*, 1883.
- [113] S. A. Bernstein and G. E. Morley. Gap junctions and propagation of the cardiac action potential, 2006.
- [114] H. Cheng, M. B. Cannell, and W. J. Lederer. Propagation of excitation-contraction coupling into ventricular myocytes. *Pflügers Archiv European Journal of Physiology*, 1994.
- [115] J. T. Lanner, D. K. Georgiou, A D. Joshi, and S. L. Hamilton. Ryanodine Receptors: Structure, Expression, Molecular Details, and Function in Calcium Release. *Cold Spring Harbor Perspectives in Biology*, 2(11), 2010.
- [116] M. Fill and J. A. Copello. Ryanodine receptor calcium release channels. *Physiological Reviews*, 82(4):893–922, 2002.
- [117] X. Lu, L. Xu, and G. Meissner. Activation of the skeletal muscle calcium release channel by a cytoplasmic loop of the dihydropyridine receptor. *Journal of Biological Chemistry*, 1994.
- [118] M. D. Stern, L. S. Song, H. Cheng, J. S. Sham, H. T. Yang, K. R. Boheler, and E. Ríos. Local control models of cardiac excitation-contraction coupling. A possible role for allosteric interactions between ryanodine receptors. *The Journal of general physiology*, 1999.
- [119] M. D. Stern. Theory of Excitation-Contraction Coupling in Cardiac-Muscle. *Biophysical Journal*, 63(2):497–517, 1992.
- [120] H. Cheng, W. J. Lederer, and M. B. Cannell. Calcium sparks: Elementary events underlying excitation-contraction coupling in heart muscle. *Science*, 1993.
- [121] M. B. Cannell, H. Cheng, and W. J. Lederer. Spatial non-uniformities in $[Ca^{2+}]_i$ during excitation-contraction coupling in cardiac myocytes. *Biophysical Journal*, 1994.

- [122] M. B. Cannell, H. Cheng, and W. J. Lederer. The control of calcium release in heart muscle. *Science*, 1995.
- [123] M. B. Cannell and C. Soeller. Numerical analysis of ryanodine receptor activation by L-type channel activity in the cardiac muscle diad. *Biophys J*, 1997.
- [124] M. M. Maleckar, A. G. Edwards, W. E. Louch, and G. T. Lines. Studying dyadic structure-function relationships: A review of current modeling approaches and new insights into Ca²⁺(mis)handling. *Clinical Medicine Insights: Cardiology*, 11, 2017.
- [125] X. Koh, B. Srinivasan, S. C. Hwee, and A. Levchenko. A 3D Monte Carlo analysis of the role of dyadic space geometry in spark generation. *Biophysical Journal*, 2006.
- [126] W. E. Louch, J. E. Hake, G. F. Jølle, H. K. Mørk, I. Sjaastad, G. T. Lines, and O. M. Sejersted. Control of Ca²⁺release by action potential configuration in normal and failing murine cardiomyocytes. *Biophysical Journal*, 2010.
- [127] T. R. Kolstad, J. van den Brink, N. MacQuaide, P. K. Lunde, M. Frisk, J. M. Aronsen, E. S. Norden, A. Cataliotti, I. Sjaastad, O. M. Sejersted, A. G. Edwards, G.T. Lines, and W. E. Louch. Ryanodine receptor dispersion disrupts Ca²⁺ release in failing cardiac myocytes. *eLife*, 2018.
- [128] S. Morotti, A. G. Edwards, A. D. McCulloch, D. M. Bers, and E. Grandi. A novel computational model of mouse myocyte electrophysiology to assess the synergy between Na⁺loading and CaMKII. *Journal of Physiology*, 592(6):1181–1197, 2014.
- [129] D. N. Mastrorade. Correction for non-perpendicularity of beam and tilt axis in tomographic reconstructions with the IMOD package. In *Journal of Microscopy*, 2008.
- [130] J. Czech, M. Dittrich, and J. R. Stiles. Rapid creation, Monte Carlo simulation, and visualization of realistic 3D cell models. *Molecular Modeling of Proteins: 2nd Edition*, 500:237–287, 2009.
- [131] D. M. Bers. *Excitation-contraction coupling and cardiac contractile force*. Number 237. Kluwer Academic Publishers, Dordrecht ; Boston, 2nd edition, 2001.
- [132] E. Picht, A. V. Zima, T. R. Shannon, A. M. Duncan, L. A. Blatter, and D. M. Bers. Dynamic Calcium Movement Inside Cardiac Sarcoplasmic Reticulum During Release. *Circulation Research*, 108(7):847–U188, 2011.
- [133] S. P. Robertson, J. D. Johnson, and J. D. Potter. The time-course of Ca²⁺ exchange with calmodulin, troponin, parvalbumin, and myosin in response to transient increases in Ca²⁺. *Biophysical Journal*, 1981.
- [134] A. Fabiato. Calcium-induced release of calcium from the cardiac sarcoplasmic reticulum. *American Journal of Physiology-Cell Physiology*, 1983.
- [135] A. Michailova, F. DelPrincipe, M. Egger, and E. Niggli. Spatiotemporal features of Ca²⁺ buffering and diffusion in atrial cardiac myocytes with inhibited sarcoplasmic reticulum. *Biophysical Journal*, 2002.

- [136] V. E. Bondarenko. Computer model of action potential of mouse ventricular myocytes. *AJP: Heart and Circulatory Physiology*, 2004.
- [137] T. R. Shannon and D. M. Bers. Assessment of intra-SR free [Ca] and buffering in rat heart. *Biophysical Journal*, 1997.
- [138] A. R. Penheiter, Z. Bajzer, A. G. Filoteo, R. Thorogate, K. Török, and A. J. Caride. A Model for the Activation of Plasma Membrane Calcium Pump Isoform 4b by Calmodulin. *Biochemistry*, 2003.
- [139] M. Brini and E. Carafoli. Calcium Pumps in Health and Disease. *Physiological Reviews*, 2009.
- [140] T. M. Bartol, D. X. Keller, J. P. Kinney, and C. L. Bajaj. Computational reconstitution of spine calcium transients from individual proteins. 7(October):1–24, 2015.
- [141] D. W. Hilgemann, D. A. Nicoll, and K. D. Philipson. Charge movement during Na⁺translocation by native and cloned cardiac Na⁺/Ca²⁺exchanger. *Nature*, 1991.
- [142] E. R. Higgins, M. B. Cannell, and J. Sneyd. A buffering SERCA pump in models of calcium dynamics. *Biophysical Journal*, 2006.
- [143] E. Saftenu, A. J. Williams, and R. Sitsapesan. Markovian models of low and high activity levels of cardiac ryanodine receptors. *Biophysical Journal*, 80(6):2727–2741, 2001.
- [144] J. L. Greenstein and R. L. Winslow. An integrative model of the cardiac ventricular myocyte incorporating local control of Ca²⁺ release. *Biophysical Journal*, 2002.
- [145] B. Hoch, R. Meyer, R. Hetzer, E. G. Krause, and P. Karczewski. Identification and expression of δ -isoforms of the multifunctional Ca²⁺/calmodulin-dependent protein kinase in failing and nonfailing human myocardium. *Circulation Research*, 1999.
- [146] M. Ibrahim, J. Gorelik, M. H. Yacoub, and C. M Terracciano. The structure and function of cardiac t-tubules in health and disease. *Proceedings. Biological sciences / The Royal Society*, 278(1719):2714–23, 2011.
- [147] D. J. Crossman, A. A. Young, P. N. Ruygrok, G. P. Nason, D. Baddeley, C. Soeller, and M. B. Cannell. T-tubule disease: Relationship between t-tubule organization and regional contractile performance in human dilated cardiomyopathy. *Journal of Molecular and Cellular Cardiology*, 2015.
- [148] E. Poláková and E. A. Sobie. Alterations in T-tubule and dyad structure in heart disease: Challenges and opportunities for computational analyses, 2013.
- [149] S. Sine and P. Taylor. Functional consequences of agonist-mediated state transitions in the cholinergic receptor. Studies in cultured muscle cells. *Journal of Biological Chemistry*, 1979.
- [150] S. M. Sine and P. Taylor. The relationship between agonist occupation and the permeability response of the cholinergic receptor revealed by bound cobra α -toxin. *Journal of Biological Chemistry*, 1980.

- [151] S. M. Sine and P. Taylor. Relationship between reversible antagonist occupancy and the functional capacity of the acetylcholine receptor. *Journal of Biological Chemistry*, 1981.
- [152] D. W. Song, J. G. Lee, H. S. Youn, S. H. Eom, and H. Kim. Ryanodine receptor assembly: a novel systems biology approach to 3D mapping. *Prog Biophys Mol Biol*, 105(3):145–161, 2011.
- [153] T. Guo, D. Gillespie, and M. Fill. Ryanodine receptor current amplitude controls Ca²⁺ sparks in cardiac muscle. *Circulation research*, 2012.
- [154] J. R. Lopez-Lopez, P. S. Shacklock, C. W. Balke, and W. G. Wier. Local, stochastic release of Ca²⁺ in voltage-clamped rat heart cells: visualization with confocal microscopy. *The Journal of Physiology*, 1994.
- [155] E. A. Sobie, K.W. Duly, J. Dos Santos Cruz, W. J. Lederer, and M. S. Jafri. Termination of cardiac Ca²⁺ sparks: An investigative mathematical model of calcium-induced calcium release. *Biophysical Journal*, 2002.
- [156] A. G. Edwards, E. Grandi, J. E. Hake, S. Patel, P. Li, S. Miyamoto, J. H. Omens, J. H. Brown, D. M. Bers, and A. D. McCulloch. Nonequilibrium reactivation of Na⁺ current drives early afterdepolarizations in mouse ventricle. *Circulation: Arrhythmia and Electrophysiology*, 7(6):1205–1213, 2014.
- [157] E. D. Fowler, C. H. Kong, J. Hancox, and M. B. Cannell. Late Ca²⁺ Sparks and Ripples During the Systolic Ca²⁺ Transient in Heart Muscle Cells. *Circulation Research*, 2017.
- [158] T. Guo, T. Zhang, K. S. Ginsburg, S. Mishra, J. H. Brown, and D. M. Bers. CaMKII δ slows [Ca]_i decline in cardiac myocytes by promoting Ca sparks. *Biophysical Journal*, 2012.
- [159] S. Ather, J. L. Respress, N. Li, and X. H. T. Wehrens. Alterations in ryanodine receptors and related proteins in heart failure. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1832(12):2425–2431, 2013.
- [160] J. A. Vest, X. H. T. Wehrens, A. Wronska, A. R. Marks, S. E. Lehnart, and S. Reiken. Ryanodine receptor/calcium release channel PKA phosphorylation: A critical mediator of heart failure progression. *Proceedings of the National Academy of Sciences*, 2006.
- [161] M. J. Bround, R. Wambolt, H. Cen, P. Asghari, R. F. Albu, J. Han, D. McAfee, M. Pourrier, N. E. Scott, L. Bohunek, J. E. Kulpa, S. R. W. Chen, D. Fedida, R.W. Brownsey, C. H. Borchers, L. J. Foster, T. Mayor, E. D.W. Moore, M. F. Allard, and J. D. Johnson. Cardiac ryanodine receptor (Ryr2)-mediated calcium signals specifically promote glucose oxidation via pyruvate dehydrogenase. *Journal of Biological Chemistry*, 291(45):23490–23505, 2016.
- [162] G. T. Johnson, D.S. Goodsell, L. Autin, S. Forli, M. F. Sanner, and A.J. Olson. 3D molecular models of whole HIV-1 virions generated with cellPACK. *Faraday Discuss.*, 169:23–44, 2014.