UNIVERSITY OF CALIFORNIA, SAN DIEGO

Calculating the Kinetics of Calcium Ions Associating with Calmodulin Using Molecular Dynamics and Brownian Dynamics

A thesis submitted in partial satisfaction of the
requirements for the degree
Masters of Science

in

Bioengineering

by

Derrick Alonzo Buntin

Committee in charge:

Professor Terrence Sejnowski, Chair
Professor Andrew McCulloch, Co-Chair
Professor Rommie Amaro
Professor Gert Cauwenberghs

2017

The thesis of Derrick Alonzo Buntin is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____
Co-Chair

_____
Chair

University of California, San Diego

2017

DEDICATION

To my fantastic mentors who have provided abundant guidance and encouragement throughout the course of conducting this thesis research.

EPIGRAPH

If I have seen further

it is by standing on the shoulders of giants.

-Sir Isaac Newton

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Thanks to the committee members who have kindly taken the time to provide feedback for further progression and improvement of the project.

Thanks to Sophie Hirakis for providing ongoing mentorship throughout the Masters thesis project and supplying many of the templates necessary for performing NAMD simulations.

Thanks to Thomas Bartol for the development of MCell, as well as the involvement of teaching myself and many other students about the wide range of applications for this software.

Thanks to Robert Koneceny for giving a great deal of guidance and encouragement in making goals for this project and assisting with work flow of BrownDye implementation.

Thanks to Gary Huber for explaining the potential uses, limitations, and future developments of BrownDye to help define the scope for which this software can be used.

Thanks to Ben Jagger for elaborating on the uses of SEEKR and how this program can aid in collecting useful kinetic data for protein-ligand interactions.

VITA

2013 - 2014	3D Printer and Sugar Scaffold Technician at UCSD Center for Neural Repair under PI Dr. Mark Tuszynski and postdoc Dr. Kobi Koffler
2014 - 2015	Research Assistant at Salk Institute Computational Neurobiology Lab under PI Dr. Terrence Sejnowski, Postdoc Dr. Krishnan Padmanabhan, and Postdoc Dr. Thomas Bartol
2014 - 2015	Co-Founder and Co-President, Internal for BiomedWired UCSD Student Organization
2015	B. S. in Bioengineering: Biotechnology, University of California, San Diego
2016 - 2017	Graduate Teaching Assistant, University of California, San Diego

ABSTRACT OF THE THESIS

Calculating the Kinetics of Calcium Ions Associating with Calmodulin Using Molecular Dynamics and Brownian Dynamics

by

Derrick Alonzo Buntin

Masters of Science in Bioengineering

University of California, San Diego, 2017

Professor Terrence Sejnowski, Chair
Professor Andrew McCulloch, Co-Chair

Calmodulin is a ubiquitous calcium-sensing protein which activates many downstream cellular processes. Calmodulin cooperatively binds up to four calcium ions which can occupy the four EF-hand motif sites. As more calcium ions bind to calmodulin, this protein becomes more likely to change conformation from the Open/Inactive state to the Closed/Active state. Studying the kinetics of calcium binding to calmodulin can lead to further insight into how the conformational state of calmodulin depends on the local calcium concentration and the effect of cooperative binding due to calcium-induced

conformational changes. This project investigates the kinetics of calcium binding to calmodulin binding sites by combining Brownian Dyanmics and Molecular Dynamics to analyze how the presence of calcium ions affects the subsequent binding of additional calcium. Obtaining the kinetic information for these interactions can inform the mechanism for cooperative calcium binding influencing the conformational state of calmodulin.

Chapter 1

Background

1.1 Introduction to Calmodulin

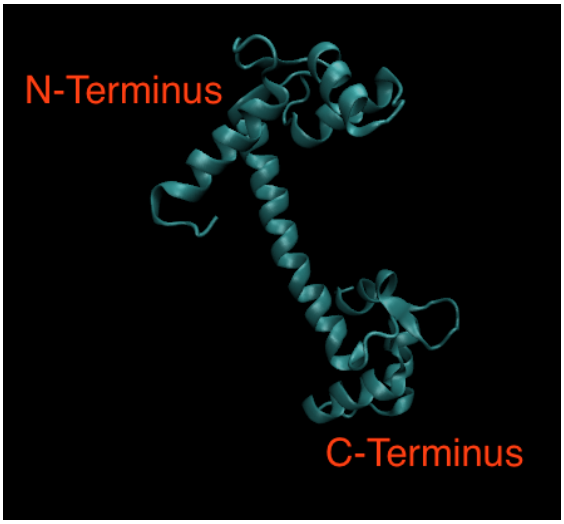

Figure 1.1: Calcium-free calmodulin (ribbon representation) with labeled N and C - terminal domains. The dumbbell-shaped molecule has two globular domains connected by a flexible alpha helix linker. Figure produced using VMD [1].

Calmodulin is a ubiquitous protein in Eukaryotic cells which functions to sense local calcium concentrations and interact with downstream proteins depending on these transient interactions. Up to four calcium ions cooperatively bind to calmodulin at

binding sites consisting of EF Hand motifs. The small calmodulin protein (16.8 kDa) has a dumbbell shape made up of two globular domains and a flexible alpha helix linker. Each of the globular domains houses two EF Hands which are helix-loop-helix motifs that are highly conserved in calcium binding proteins. Affinity for calcium is about 6-fold higher for the C-terminal binding sites than for the N-terminal sites due to high-off rates, but the binding kinetics tend to be more rapid at the N-terminal sites. The two low affinity sites on the N-terminal domain have a $K_D \sim 10 - 12\mu M$ and the two high affinity sites on the C-terminal domain have a $K_D \sim 1 - 2\mu M$ [3]. When calmodulin is in a tense (open) state, the N-lobe has a rapid association rate of $k_{ON} = 7.7$ x $10^8 M^{-1}s^{-1}$ and a dissociation rate of $k_{OFF} > 500s^{-1}$ [4]. The C-terminal domain has a greater affinity for calcium with slower kinetics of $k_{ON} \sim 0.05$ x $10^8 M^{-1}s^{-1}$ and $k_{OFF} \sim 10s^{-1}$ [3].

1.2 The EF Hand Motif

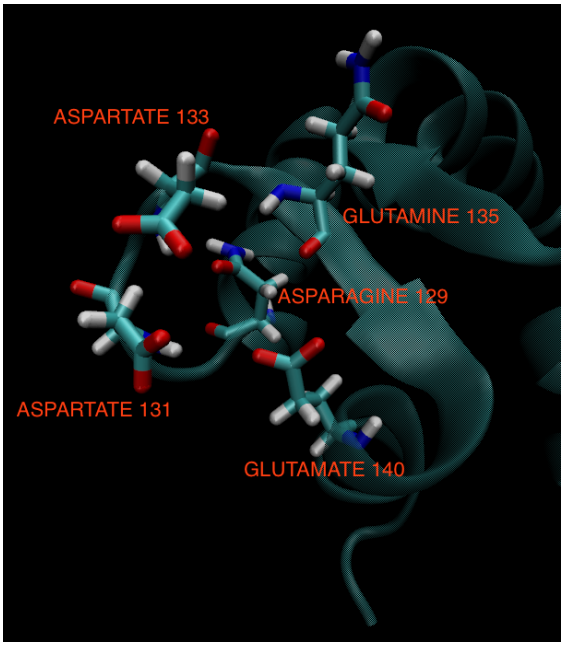

Figure 1.2: These five residues in the EF Hand coordinate a positively charged (2+) calcium ion. This domain corresponds with site D on the C-terminus of calmodulin. Red portions of the residue correspond to electronegative oxygen atoms.

This structure is highly conserved in calcium-sensing proteins because the size and shape of this pocket is well suited for non-covalently binding calcium. These EF Hands possess negatively charged amino acid residues (see Figure 1.2 and Table 2.1) which coordinate the positively charged calcium ion in the binding pocket. Each of the EF Hands is made up of about 30 amino acid residues, thereby making the majority of the protein function fit for calcium binding [5].

1.3 Calmodulin Interactions with Downstream Proteins

Calmodulin exhibits positive cooperativity between the two sites on each of its globular domains. The globular domain near the C-terminus has a higher affinity than the domain at the N-terminus [6]. At concentrations above $\sim 30\mu$M, calcium tends to bind to the C-terminus (as indicated by the thick red arrows in Figure 1.3) and at concentrations under $\sim 30\mu$M, calcium ions tend to bind to the N-terminus sites (indicated by yellow arrows in Figure 1.3) within one second of the initial calcium spike [2]. This binding sequence plays an increasingly important role in the activation of downstream proteins by calmodulin, primarily kinases and phosphatases. As a greater number of calcium ions bind to calmodulin, the calcium/calmodulin complex is much more likely to bind to a particular set of proteins. Figure 1.3 shows how calmodulin binds with calcium and then becomes more likely to complex with CaMKII, as indicated by the items including the 'K' term.

1.3.1 Calmodulin in Cardiac Cells

Intracellular calcium is an important secondary messenger which is responsible for regulating many vital biological processes [7]. Calmodulin is a small protein binds calcium ions and targets different effector proteins to mediate such processes. Calcium-

free calmodulin (apo-calmodulin) will interact with a different set of proteins than calcium-bound calmodulin. In cardiac cells, calmodulin interacts with molecules which are involved in calcium homeostasis and Excitation-Contraction Coupling (ECC), the process by which an membrane action potential is translated to muscle contraction. A trans-sarcoplasmic reticulum protein, type 2 Ryanodine Receptor (RyR) can bind at least four calmodulin molecules per tetramer. This process is critical for inducing calcium-induced calcium release in the sarcoplasmic reticulum. This action significantly increases the concentration of calcium in the cell by ECC to orchestrate a smooth propagation of the action potential throughout the cardiac muscle tissue [8].

1.3.2 Calmodulin in Neurons

In neuronal cells, proteins such as calcium/calmodulin-dependent protein kinase II (CaMKII) are associated with long-term potentiation (LTP) and synaptic plasticity. Once the calcium/calmodulin-CAMKII complex is formed, it can the be phosphorylated and join with other CAMKII subunits as an initiating step of the LTP mechanism. The phosphorylated CAMKII subunits are indicated by a 'pK' term in Figure 1.3. Calcium entering through NMDA receptors determines the local concentration of calcium in the post-synaptic density (PSD) where calmodulin proteins reside in the neuronal cells. The fluctuating calcium concentration in the PSD over time is the primary factor in determining whether the cells will experience LTP or long-term depression (LTD) at a particular synapse. The number of calcium ions bound to calmodulin is highly dependent on the temporal span of calcium influxes through the NMDA receptors, as well as the spatial proximity to local calcium ions. Gathering more detailed information on the association and dissociation rates of calcium binding to each of these calmodulin sites can therefore better inform simulations of the calcium/calmodulin complex interacting with proteins on longer time scales associated with LTP or LTD.

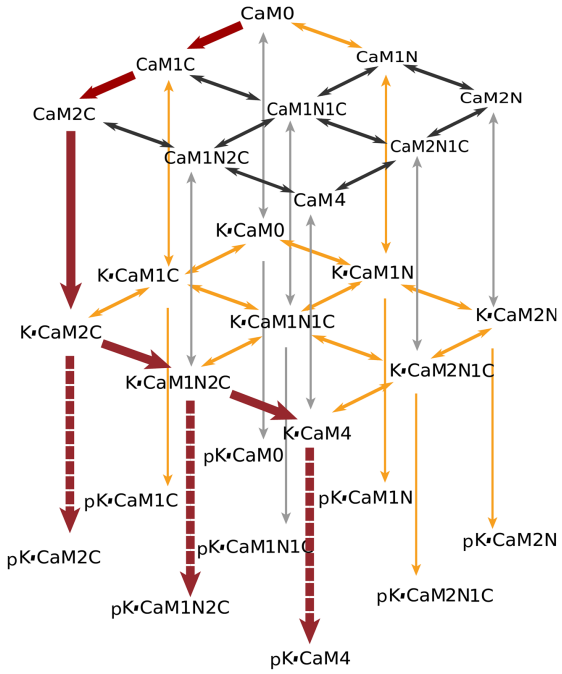

Figure 1.3: Pepke et. al. shows the general trend of calciums binding to calmodulin as a function of calcium concentration. [2]

This thesis will not examine the calcium/calmodulin complex binding to other proteins (such as the binding to CaMKII indicated by the 'K' prefix in Figure 1.3) or subsequent phosphorylation of said complexes. The major focus of this thesis is on the first layer of Figure 1.3 examining the kinetics of four calcium ions binding to calmodulin.

Chapter 2

Methods

2.1 Motivation for Approach

The approach used in this thesis was inspired by the UCSD Bioengineering graduate course BENG 276 for Numerical Methods. In this course, a variety of multiscale computational software programs were demonstrated by instructors and were applied by graduate students. Among these programs were APBS, instructed by Dr. Andrew McCammon and Dr. Robert Koneceny; BrownDye, instructed by Dr. Gary Huber; and MCell, instructed by Dr. Thomas Bartol and Dr. Terrence Sejnowski. ABPS is the Adaptive Poisson-Boltzmann Solver used to describe electrostatic interactions in salty, aqueous media. BrownDye is a software package which uses Brownian Dynamics to perform simulations of ligands binding to larger, rigid biomolecules. MCell is a software which performs 3-D Monte Carlo simulations of ligand diffusion which can implement the kinetics of reactions and diffusion in cellular processes. The intent of studying the kinetics of calcium binding to calmodulin is to create a more detailed model for molecular interactions that can be implemented in MCell models.

2.2 Crystal Structure

The crystal structure used for this project is 3CLN from the Protein Data Bank [9]. This structure of calmodulin is refined at 2.2 Angstroms and originates from the species Rattus rattus. There were a total of 5 missing residues that were built in to the 3CLN.pdb file using the software Maestro [10]. The first four amino acid residues at the N terminus were missing (Alanine, Aspartate, Glutamine, Leucine) as well as the last amino acid at the C terminus (Lysine). Once the missing residues were built in, there was a total of 148 amino acid residues on the completed calmodulin chain. Maestro also did a brief energy minimization in order to place these residues in an energetically favorable conformation prior to solvation. This step allows for the stabilization of the apo-calmodulin structure before initialization by Molecular Dynamics simulations with NAMD.

2.3 Brownian Dynamics

In Brownian Dynamics simulations, two molecules are placed within a spherical volume (Q region or Q surface). Molecules are modeled as rigid bodies, with each unit making up the molecule being modeled as a hard sphere with electrostatic forces. The solvent is modeled using a continuum electric field to simulate the motion of water and presence of ions. When the molecule diffuses outside of the Q distance, the trajectory is counted as an ”escape.” If the molecules come into contact using specified binding criteria, an ”association” event is counted. The ratio of ”association” to ”escape” trajectories is used to calculate the probability and rate of association. BrownDye (BD) is a type of software used for performing Brownian Dynamics simulations of ligands binding to biomolecules. For this experiment, the associating element is a calcium ion. The .pqr file for calcium was manually created. The chief element being investigated with this

software was the differential binding probability among the four EF Hand binding sites on calmodulin.

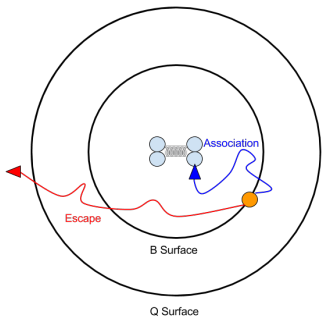

Figure 2.1: In BrownDye, the ion begins on the B surface and continues on a random walk trajectory. The trajectory will either reach the Q surface and escape or reach the reaction site and count as an association.

The four main experiment setups for BD were:

1) A calcium ion binding to one of the four sites of apo-calmodulin

2) A calcium ion binding to calmodulin with 1/4 calcium ions bound to each site

3) A calcium ion binding to calmodulin with 2/4 calcium ions bound to two sites

4) A calcium ion binding to calmodulin with 3/4 calcium ions bound to three sites

This allows for the comparison not just between four different binding sites on a single structure, but also for comparison of association rates when a different number of calciums ions are bound to the calmodulin structure. Upon preparing the crystal structure .pdb file for BD simulations, PROPKA was used on the PDB2PQR server in order to determine the proper protonation levels of the amino acids of calmodulin to ensure the charges are consistent with that of physiological pH levels near 7 [11]. There is a particular importance to check that any Histidine residues are at the correct protonation state since its side chain pKa of 6.10 is closest to this neutral pH.

2.4 Molecular Dynamics

A distinct advantage to performing molecular computations is the ability to obtain both dynamic information about a molecule as well as structural detail at the atomic level as a function of time [12]. The Molecular Dynamics (MD) simulations in this thesis were performed using AMBER and NAMD tools. AMBER (Assisted Model Building with Energy Refinement) refers to a set of molecular mechanics force fields for biomolecules as well as a package of molecular simulation programs including source code and tutorials [13]. The force fields in AMBER describe the potential energy function and parameters of the interactions of biomolecules being simulated. The force field used in this set of simulations was FF14SB, which includes updated torsion terms for the protein Phi-Psi angles and adjustments for the torsion terms for side chains from the previous FF12SB [14]. The AMBERTools16 package was used for preparing many of the relevant MD files [15]. The primary tools from AMBERTools16 which was utilized was tleap, for which scripts were written and executed to prepare the files required for performing simulations using the MD engine NAMD. NAMD is a scalable parallel molecular mechanics engine which works with AMBER and other forcefields to perform high-performance simulations of biomolecular systems [16].

The *.prmtop file, which describes the parameter and topology of the molecules in the system, as well as the *.inpcrd file, which describes the initial molecular coordinates of the system were prepared using tleap. Atomic coordinates of the 148AA calmodulin structure were prepared using tleap. A 10 Angstrom buffer was used to solvate the protein from the 9 Angstrom inter-atomic distance cutoff. The unsolvated calmodulin protein uses a negative charge to coordinate the positively changed calcium ions. After solvating the protein in TIP4P water molecules, the negative charge was counterbalanced by adding 23 sodium ions to neutralize the protein. An additional 43 sodium ions and

43 chloride ions were added to reach a concentration of 0.15M for each of the ions in the $\sim$ 480K Angstroms^3 box. All-atom simulations were carried out to 20 nanoseconds for calcium-bound systems and 25 nanoseconds for the apo-calmodulin system. The solvated system had just over 48,000 atoms and ran at 3.3 nanoseconds/day on 8 CPUs.

Analysis for MD trajectories was carried out by calculating the relative displacement of atoms in the calmodulin structure over time. The root mean square deviation (RMSD) is the most common calculation for the comparison of two sets of superimposed atomic coordinates. The RMSD for a molecule with a MD trajectory is calculated with the equation below.

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - x_{i,ref})^2}$$

2.5 Calcium Parameters

An important factor for determining the association rate of calcium to calmodulin is ensuring that the parameters for calcium are sufficient for performing both BrownDye and Molecular Dynamics simulations. Ions will have different radii and charges depending on the number of electrons in the outer valence shell of the atom, also indicated by their position on the Periodic Table of Elements. The size of the ion is rather important for simulating the interaction with surrounding water molecules. Calcium parameters were investigated from Gail Bradbrook's studies of Concanavalin A, which possesses EF Hand motifs similar to that of calmodulin [17]. Concanavalin A is legume lectin homotetramer, with each of its four subunits containing a binding site for one calcium and one magnesium ion to bind. The calcium ions used in these simulations are derived from Bradbrook's single atom simple sphere model. The simple sphere model is defined with the charge of the ion as well as the two parameters describing the Lennard-Jones potential. The Lennard-Jones parameters are chosen such that the correct hydration free

energy is reproduced for the calcium ion. Additionally, Akansha Saxena and David Sept have examined calcium to more accurately model the electrostatic interactions, free energy, and interaction with surrounding water molecules [18].

2.6 Defining Calcium-Calmodulin Associations

For BrownDye simulations, reaction criteria are defined in order to determine which trials concluded in an association or successful binding event. Up to six coordinating atoms were used to define a reaction for each site. Table 2.1 includes a list of atoms which coordinate each of the binding sites. A minimum of 3 of the coordinating atoms must be within the specified reaction distance in order for BrownDye to count the trajectory event as an association. The reaction lengths used for different trials were 2.5, 3.5, and 4.5 Å. For a given trajectory, if a calcium ion does not come within the reaction length of at least 3 of these atoms simultaneously and diffuses outside of Q-surface, BD will consider the trajectory as 'escaped'.

The distance from the coordinating oxygen atom to calcium will fluctuate as the conformation of calmodulin changes over time. For Table 2.1, these distances were measured using VMD using the static holo-calmodulin as a reference conformation structure [1].

Table 2.1: Binding sites for calcium were defined by the coordinating atoms in close proximity. Although distances to calcium may vary, these distances were obtained from the 3CLN crystal structure.

Binding Site	Coordinating Residue & Atom	Distance to Calcium (Å)
A	ASP22:OD1	2.42
	ASP24:OD2	2.61
	THR26:O	2.45
	GLU31:OE1	2.27
	GLU31:OE2	2.38
B	ASP56:OD2	2.21
	ASP58:OD2	2.48
	ASN60:OD1	2.46
	THR62:O	2.17
	GLU67:OE1	2.49
	GLU67:OE2	2.31
C	ASP93:OD2	2.14
	ASP95:OD2	2.22
	TYR99:O	2.05
	GLU104:OE1	2.76
	GLU104:OE2	2.32
D	ASN129:OD1	2.17
	ASP 131:OD2	2.56
	ASP133:OD2	2.07
	GLN135:O	2.38
	GLU140:OE1	2.57
	GLU140:OE2	2.32

Chapter 3

Results

3.1 Structure Preparation

Starting from the 3CLN.pdb structure provided from the Protein Data Bank, a total of 5 missing amino acids were built in order to obtain the full amino acid chain for calmodulin. Figure 3.1 was produced with PyMOL in order to visualize the molecule and ensure that all 148 amino acid residues were correctly built onto the molecule using Maestro [19]. This is particularly important for the ends of the protein chain where missing residues were built in manually. The naming scheme used for binding sites uniquely refers to sites by A, B, C, or D. Site A is nearest to the N terminus with the lowest numbers for coordinating amino acid residues. Sites B, C, and D have increasingly higher numbers for the amino acid residues which surround the binding sites.

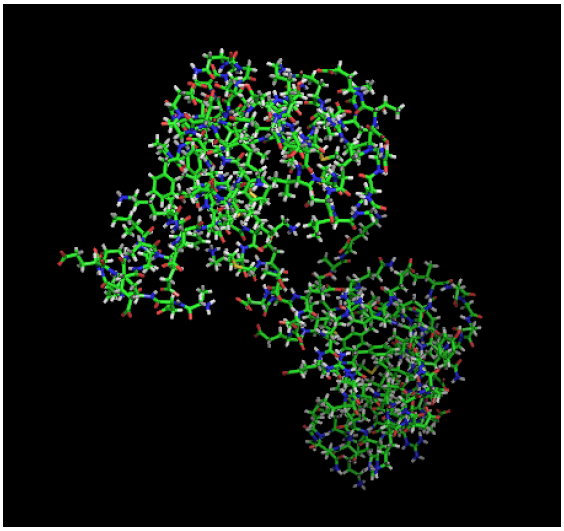

Figure 3.1: Apo-calmodulin is visualized with PyMOL after adding 5 missing residues using Maestro software. This ensures that the residues are properly connected as a single polypeptide chain which is ready for molecular dynamics and BrownDye simulations.

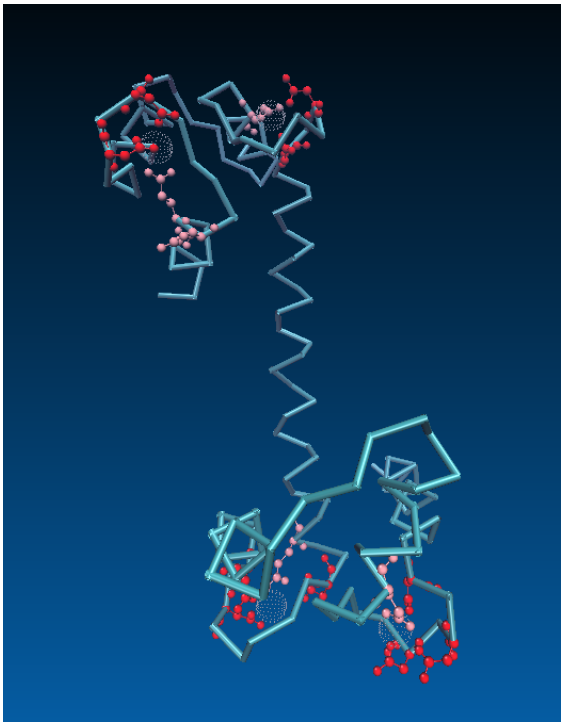

Figure 3.2: 3CLN with 148 residues including spheres for coordinated calcium ions. This holo-calmodulin structure includes negatively charged residues (CPK style) responsible for coordinating calcium.

Figures 3.3, 3.4, 3.5, and 3.6 were produced in order to show the stabilization of

total pressure in the system over time. In this figure, there is an initial rise in pressure in the first steps of the simulation, but this rise quickly finds a stable energy level just below 0 kcal/mol.

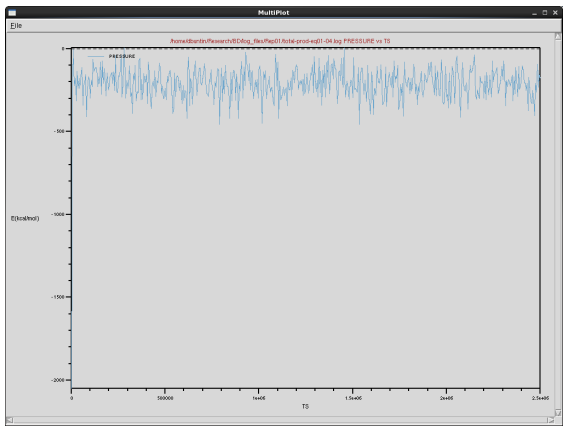

Figure 3.3: Four consecutive MD production runs energetic outputs were concatenated and analyzed to verify that the pressure of the system stabilizes and maintains a relatively constant value as the simulation progresses.

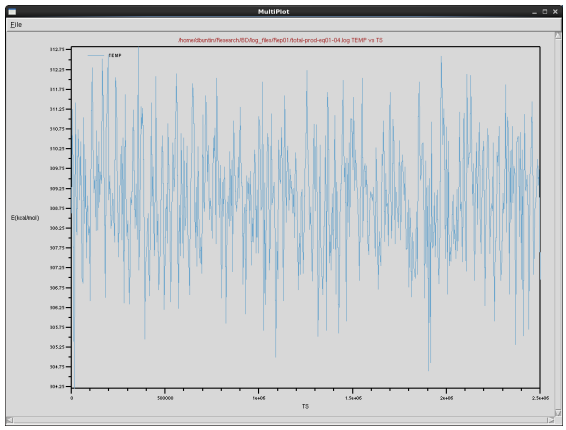

Figure 3.4: Four consecutive MD production runs were concatenated and analyzed to verify that the temperature of the system stabilizes and maintains a relatively constant value as the simulation progresses.

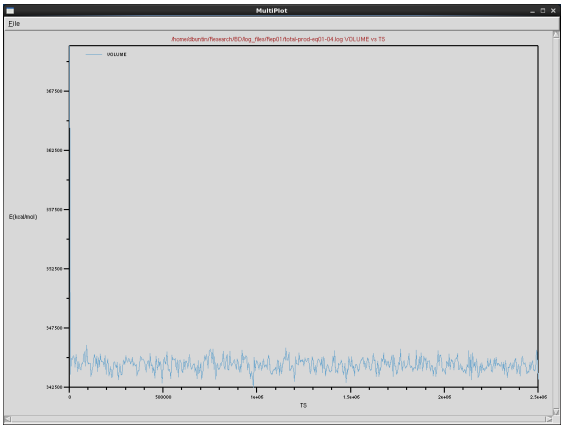

Figure 3.5: Four consecutive MD production runs were concatenated and analyzed to verify that the volume of the system stabilizes and maintains a relatively constant value as the simulation progresses.

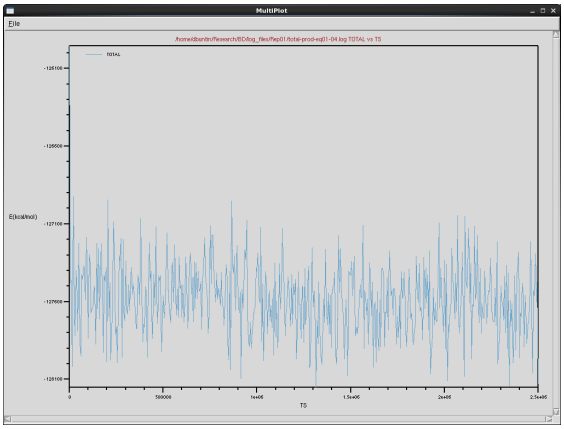

Figure 3.6: Four consecutive MD production runs were concatenated and analyzed to verify that the total energy of the system stabilizes and maintains a relatively constant value as the simulation progresses.

3.2 Brownian Dynamics Results

The first set of BD simulations performed were for calcium binding to apo-calmodulin with a reaction distance of 2.5 $\AA$ to the designated coordinating atoms in each EF Hand. There were no associations for this reaction length, implying that this distance is too small for the calcium ion to be within 2.5 $\AA$ of three coordinating atoms

Table 3.1: Table for a single calcium ion associating to apo-calmodulin at each binding site with a distance of 3.5 Angstroms.

Binding Site	Mean Reaction Probability	Mean Rate Constant
A	0.00784102	1.14461e+09
B	0.00514098	7.50465e+08
C	0.00404821	5.90947e+08
D	0.00722834	1.05518e+09

simultaneously. Associations likely failed because of steric clashes in the reaction criteria or in the path to association. The reaction distance was then increased to 3.5 $\AA$ and 4.5 $\AA$ in order to show the increase in associations leading to successful binding reactions. Table 3.1 shows the BD simulation statistics for a single calcium ion binding to one specified EF Hand binding site on apo-calmodulin. All rate constants are within the same order of magnitude. Furthermore, sites A and D have very similar rates as well as sites B and C being very close in magnitude. Tables 3.2 and 3.3 list the kinetic information for calcium associating to the one-calcium-bound and two-calcium-bound calmodulin structures, respectively. Tables 3.1 and 3.5 contain information for calcium associating to the 3-calcium-bound structures of calmodulin. Table 3.5 differs from Table 3.1 in that the duration of trajectories is increased from 1,000,000 to 5,000,000 and the reaction distance was decreased from 4.5 to 3.5 Angstroms. Increasing the trajectory duration was sufficient to eliminate any stuck trajectories which resulted in simulations for Table 3.1.

Table 3.2: BD results for a single calcium ion associating to calmodulin structures with one calcium ion bound. Reaction distance is 3.5 Angstroms and there are no stuck trajectories. The three remaining sites are all possible reaction for each trajectory.

Occupied Site	Mean Reaction Probability	Mean Rate Constant
A	0.015234	2.22406e+09
B	0.0192292	2.80692e+09
C	0.01866	2.66848e+09
D	0.0162914	2.33811e+09

Table 3.3: BD results for a single calcium ion associating to calmodulin structures with two calcium ions already bound. Reaction distance is 3.5 Angstroms and there are no stuck trajectories. The two remaining sites are both possible reaction for each trajectory.

Reaction Sites	Mean Reaction Probability	Mean Rate Constant
A B	0.0109105	1.59275e+09
A C	0.0117278	1.67737e+09
A D	.00900841	1.29361e+09
B C	0.0151433	2.16639e+09
B D	0.0124294	1.78381e+09
C D	0.0123382	1.72817e+09

Table 3.4: This table shows the binding statistics of a fourth calcium binding when 3/4 calcium ions are already bound to the calmodulin protein. The required reaction distance is 4.5 Angstroms within at least 3 of the coordinating negatively charged calmodulin atoms.

Binding Site	Reaction Probability	Stuck Probability	Mean Rate Constant
A	0.034724	0.0002126	4.8635e+09
B	0.0251998	0.0002604	3.564e+09
C	0.0193917	0.00038476274	2.78445e+09
D	0.022982	0.0003217368	3.287e+09

Table 3.5: Table for a single calcium ion associating to calmodulin with 3 calcium already bound. Reaction distance is 3.5 Angstroms and stuck trajectories were eliminated by increasing run duration.

Binding Site	Mean Reaction Probability	Mean Rate Constant
A	0.0071745	1.02614e+09
B	0.00389892	5.59847e+08
C	0.00545823	7.71956e+08
D	0.00827527	1.15906e+09

3.3 Molecular Dynamics Results

Even lacking any calcium ions, the apo-calmodulin structure shown in Figure 3.8 has a relatively high degree of mobility. Although calmodulin in known to change into a closed conformation as additional calcium ions bind to the structure, there is still a partial unraveling of the central alpha helix which appears to be an intermediate between the open and closed conformations. The globular domains of the protein come into closer proximity than the starting structure, and this altered conformational state provides a new static structure to conduct further BD simulations. All RMSD calculations were performed after aligning the first 30 AA residues for the molecular structure. This allows the N-terminal globular domain to remain fixed in position while allowing the free motion of the flexible linker and neighboring C-terminal domain. The average RMSD for the apo-calmodulin structure in Figure 3.8 is 3.716 Angstroms, which is quite a large deviation for this small molecule given that it has not been exposed to any calcium ions to affect the conformation of the molecule. The RMSD of this structure is plotted over time in Figure 3.7.

Table 3.6: RMSD values for three replicate runs of apo-calmodulin are shown. Simulations were carried out for 25 ns each. RMSD values are provided in Angstroms.

Replicate	Average RMSD	RMSD Standard Deviation
1	2.722	0.585
2	3.716	0.904
3	2.839	0.713

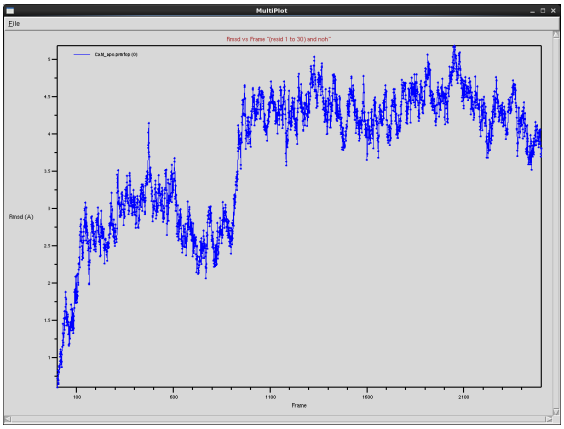

Figure 3.7: The apo-calmodulin (Replicate 2) structure had an a RMSD of 3.716 Angstroms. Over the span of 25 ns, the RMSD tends to increase rapidly for about 10 ns and then stabilizes at a RMSD of nearly 5 Angstroms.

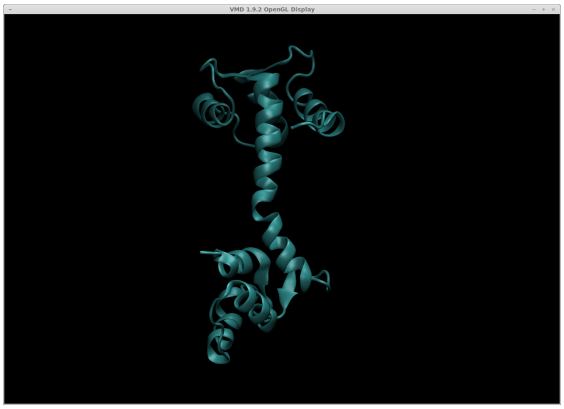

Figure 3.8: Within 25ns of simulation time, apo-calmodulin (Replicate 2) took on this configuration with a bent linker alpha helix domain. This time frame had a high average RMSD of 3.716 Angstroms.

RMSD calculations are shown in Table 3.6 for all apo-calmodulin MD runs at 25 ns each. The four calcium-bound configurations can be seen in Table 3.7. These three calcium bound structures were run for 20 ns each. Only MD runs with 20 ns or more of MD data are reported here since shorter runs are unlikely to have a stable RMSD value by the end of the run.

Table 3.7: RMSD values for three calcium bound structures of calmodulin are shown for NAMD runs of 20 ns duration. Values are given in Angstroms and show relatively little fluctuation on average.

Bound Sites	Average RMSD	RMSD Standard Deviation
A B C	2.801	0.540
A B D	2.570	0.435
A C D	2.583	0.427
B C D	3.419	0.612

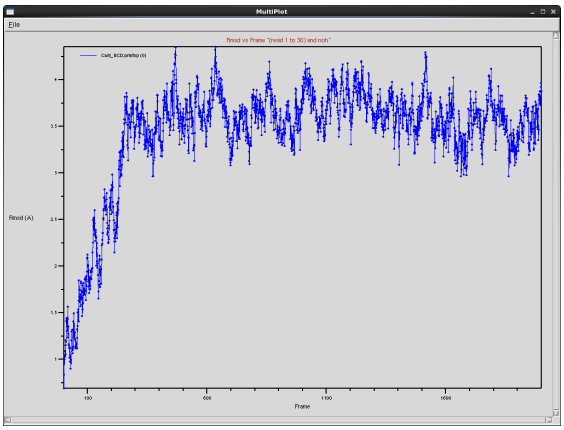

Figure 3.9: This is a plot of RMSD over a 20 ns time span for one of calcium bound structure. Here, calcium is bound at sites B, C, and D. This structure has the highest average RMSD for the three calcium bound configurations.

On average, the apo-calmodulin structure has a slightly higher RMSD value than the calcium-bound structures. In addition, the apo-calmodulin has the production run with the highest RMSD out of all MD trajectories.

Chapter 4

Discussion

4.1 BrownDye

The BrownDye simulations were performed without any waters included in the calmodulin protein structure. The implicit water molecules, sodium ions, and chlorine ions were modeled with dielectric constants for the surrounding solution. This solution was formulated to have a physiological concentration similar to that of the microenvironment of calmodulin inside the cell. Magnesium is another important ion which competitively binds to the calmodulin binding pocket, but this ion was not included in the solution nor as a ligand in order to isolate the kinetics of calcium alone. It would be interesting, however, to include magnesium ions in future simulations in order to analyze the different conformational changes that can be induced by magnesium when compared to the changes induced by calcium ions.

This BD software operates rapidly based on a few assumptions and parameters which can greatly alter the outcome results for simulation runs. Typically, BD is operated to calculate the kinetics of a ligand interacting with a large, rigid protein structure. Here, we are using a relatively small protein and implement a single calcium ion in place of a

ligand. Calmodulin has the fastest kinetics of the major calcium binding proteins and is reliant upon calcium to induce conformational changes as a key part of its protein functionality. When calmodulin is simulated in BD, the main limitation is that only a single protein structure must be used rather than an ensemble of conformational states. Furthermore, the kinetics of calcium binding to calmodulin is dependent upon the surrounding ions in physiological solution and whether or no calmodulin is bound to another protein. The kinetic rates of binding additional calcium ions can be influenced by the non-covalent interactions between calmodulin and surrounding proteins or ions. For example, magnesium is antagonistic to calcium at each site and decreases binding affinity for calcium by a factor of 4 [20].

Although a single static structure may not provide a physiologically accurate estimation for the association rate, conducting BD runs with multiple conformations will give more revealing relative kinetic information. There is a greater potential to yield meaningful results with by comparing the kinetics for calcium binding to calmodulin with zero, one, two, or three calcium ions already bound to the structure. Furthermore, there are multiple configurations in which calmodulin can hold these ions. There are four configurations in which calmodulin can hold a single ion, six configurations for holding two ions, and four more configurations for holding three ions.

Simulations were carried out with 5,000,000 trajectories each in order to obtain statistically accurate information for the number of successful associations vs. escaped trajectories. Table 3.4 has a portion of the trajectories getting stuck and failing. In order to reduce these stuck trajectories, two methods were used. First, the reaction distance was increased from 3.5 Angstroms to 4.5 Angstroms. Although this method slightly reduced the number of stuck trajectories, there were still a significant fraction getting stuck. The next technique which solved the issue of obtaining stuck trajectories was to increase the max number of steps for each trajectory from 1,000,000 to 5,000,000 steps. This allows

each trajectory to run five times longer and allow the calcium ion the chance to either associate or escape before the trial ends with a stuck result.

4.2 Molecular Dynamics

Molecular Dynamics trajectories were generated for both calcium-bound and calcium-free forms of calmodulin. The calcium-bound conformations included all four possible combinations of three-calcium-bound calmodulin. Each of the calcium-bound configurations were run for 20 ns of production MD while three replicates of apo-calmodulin were run for 25 ns. Performing MD runs on the scale of microseconds to even milliseconds can greatly enhance the range of conformations sampled in the given time. Longer MD trajectories have an increasingly larger likelihood of experiencing a significant bend in the flexible alpha helix linker of the molecule, similar to that of Figure 3.8. Further insight can be gained into these conformational changes over time by also performing MD on a crystal structure for holo-calmodulin to verify that similar results are found. RMSD values were calculated by aligning the first 30 residues of the protein structure. These RMSD values can have a different range of fluctuation depending on which amino acid residues are aligned for the RMSD calculations. The range of 1 to 30 was chosen in order to fix the N-terminal domain and allow the mobility of the flexible helix to be taken into account for the RMSD fluctuations.

4.3 Moving Forward

A logical step for proceeding with the investigation of the interaction between calcium and calmodulin would be using additional software which is suited for changing the conformation of the target protein. One useful tool that was investigated in the course

of this research was a package for Simulation Enabled Estimation of Kinetic Rates, or SEEKR. SEEKR is a collection of open-source scripts which determines binding kinetics using a combination of molecular dynamics, Brownian dynamics, and milestoning theory. A preliminary test run can be completed to benchmark the system and give an estimation for the amount of ns/day that can be performed in the simulation run. Obtaining kinetic information for a single reaction site of similarly sized systems has taken upwards of one million CPU hours [21]. These results would incorporate the ability to have a protein with a conformation changing over time. This method is a computationally expensive method to obtain kinetic information and led the project to pursue simulations using more rapid computation using the relatively quick software for BrownDye. The primary reason for using apo-calmodulin and a three calcium bound model is for preparation to run SEEKR, which can calculate the kinetics of a single binding site at a time. This is beneficial for obtaining the association rate, dissociation rate, and binding energy for each reaction.

Results can further be compared by performing similar simulations using more recent calcium parameters such as those from Saxena and Sept. Implementing multiple ”dummy atoms” allows for the distribution of electrostatic and steric interaction points around an uncharged central atom. This helps alleviate some issues with the simple sphere calcium model used in these simulations, such as the inner shell waters being held too tightly and disallowing exchange with bulk solvent [18].

Chapter 5

Conclusion

This project has explored how many different factors can affect simulation results for calculating kinetic rate constants for association. BrownDye, with its architecture for hosting a rigid molecule and performing association trials with a ligand, rapidly computes a association rates with a 95% confidence interval. A few important factors which contribute to this calculation are defining the reaction pairs, reaction lengths, and number of steps for each trajectory run. Molecular dynamics is a good pairing with BrownDye because the all-atom simulation gives the molecule flexibility over time and can introduce another available conformation to analyze using BrownDye. This project has laid a foundation for investigating the molecular mechanism for conformational changes in calmodulin and how the addition of calcium changes subsequent binding kinetics. These methods can be further extrapolated to use on larger systems such as how the calcium/calmodulin complex binds to CAMKII, RyR, or other downstream effector proteins.

Bibliography

[1] W. Humphrey, A. Dalke, and K. Schulten, "Vmd - visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14, pp. 33–38, 1996.

[2] S. Pepke, T. Kinzer-Ursem, S. Mihalas, and M. B. Kennedy, "A dynamic model of interactions of ca 2+, calmodulin, and catalytic subunits of ca 2+/calmodulin-dependent protein kinase ii," *PLoS Computational Biology*, vol. 6, no. 2, 2010.

[3] L. Hoffman, A. Chandrasekar, X. Wang, J. A. Putkey, and M. N. Waxham, "Neurogranin alters the structure and calcium binding properties of calmodulin," *The Journal of Biological Chemistry*, vol. 289, no. 21, pp. 14644–14655, 2014.

[4] G. C. Fass, S. Raghavachari, J. E. Lisman, and I. Mody, "Calmodulin as a direct detector of ca 2+ signals," *Nature Neuroscience*, vol. 14, no. 3, pp. 3389–3397, 2011.

[5] S. W. Vetter and E. Leclerc, "Novel aspects of calmodulin target recognition and activation.," *European Journal of Biochemistry*, vol. 270, no. 3, pp. 404–414, 2003.

[6] S. E. Brown, S. R. Martin, and P. M. Bayley, "Kinetic control of the dissociation pathway of calmodulin-peptide complexes," *The Journal of Biological Chemistry*, vol. 272, no. 6, pp. 3389–3397, 1997.

[7] D. L. Cioffi, J. Barry, Christina, and T. Stevens, "Role of calcium as a second messenger in signaling: A focus on endothelium," in *Textbook of Pulmonary Vascular Disease* (J. X. J. Yuan, J. G. N. Garcia, J. B. West, C. A. Hales, S. Rich, and S. L. Archer, eds.), ch. 15, pp. 261–272, Springer US, 2010.

[8] D. Yang, L.-S. Song, W.-Z. Zhu, W. Wang, K. Chakir, C. Wu, Y. Wang, R.-P. Xiao, S. R. Wayne Chen, and H. Cheng, "Calmodulin regulation of excitation-contraction coupling in cardiac myocytes," *Circulation Research*, vol. 92, no. 6, pp. 659–667, 2003.

[9] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.

[10] "Maestro." Version 10.2. Schrodinger, LLC, NewYork, NY, 2015.

[11] T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielson, J. H. Jensen, G. Klebe, and N. A. Baker, "Pdb2pqr: expanding and upgrading automated preparation of biomolecular structures for molecular simulations," *Nucleic Acids Research*, vol. 35, no. 2, pp. W522–W525, 2007.

[12] D. Shukla, C. X. Hernandez, J. K. Weber, and V. S. Pande, "Markov state models provide insights into dynamic modulation of protein function," *Accounts of Chemical Research*, vol. 48, no. 2, pp. 414–422, 2015.

[13] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham III, 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*, vol. 91, no. 1-3, pp. 1–41, 1995.

[14] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, "ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb," *Journal of Chemical Theory and Computation*, vol. 11, no. 8, pp. 3696–3713, 2015.

[15] D. A. Case, D. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, D. M. York, and P. A. Kollman, "Amber 2017," 2017.

[16] 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*, vol. 26, no. 16, pp. 1781–1802, 2008.

[17] G. M. Bradbrook, T. Gleichmann, S. J. Harrop, J. Habash, J. Raftery, J. Kalb, J. Yariv, I. H. Hillier, and J. R. Helliwell, "X-ray and molecular dynamics studies of concanavalin-a glucoside and mannoside complexes relating structure to thermodynamics of binding.," *Journal of the Chemical Society, Faraday Transactions*, vol. 94, no. 11, pp. 1603–1611, 1998.

[18] A. Saxena and D. Sept, "Multisite ion models that improve coordination and free energy calculations in molecular dynamics simulations," *Journal of Chemical Theory and Computation*, vol. 9, no. 8, pp. 3538–3542, 2013.

[19] "The pymol molecular graphics system." Version 1.8. Schrodinger, LLC, NewYork, NY, 2015.

[20] J. A. Cox, "Interactive properties of calmodulin," *Biochemistry Journal*, vol. 249, no. 3, pp. 621–629, 1988.

[21] L. W. Votapka, B. R. Jagger, A. L. Heynemen, and R. E. Amaro, "Seekr: Simulation enabled estimation of kinetic rates, a computational tool to estimate molecular kinetics and its application to trypsin-benzamidine binding," *The Journal of Physical Chemistry*, vol. 121, no. 15, pp. 3597–3606, 2017.