ABSTRACT

COMPUTATIONAL INVESTIGATION OF CALCIUM BINDING PROTEINS ANNEXIN A1 AND CARDIAC TROPONIN C

By

Kimberly A Lewis

July 2019

Chair: Dr. Andrew Morehead

Major Department: Chemistry

Calcium binding proteins are vital for many biological functions. Many undergo calcium induced conformation changes through sensing motifs, resulting in functional changes for the proteins. In this study computational investigations were performed for annexin A1 and cardiac troponin C (cTnC). Annexin A1 is known to induce membrane aggregation, and prior to interaction with membrane or proteins, must undergo specific calcium induced conformation change. This study focuses on investigating the conformation change pathway for annexin A1. Cardiac troponin C is involved in the calcium induced regulation of cardiac muscle contraction. This study investigates the impact of N-terminal mutations on the regulatory binding site for cTnC. The goal of this study is to use knowledge of these calcium binding proteins to determine the conformation pathway of annexin A1 and mutational effects of cTnC linked to cardiac diseases.

Multiple nudged elastic band (NEB) method simulations were performed for annexin A1. The trajectories for the conformation change were generated and examined for the full length annexin A1 protein. Our results suggest that the N-terminal domain of annexin A1 is

removed from repeat III of the core domain in a sliding motion. Previously, it has been unclear how the N-terminal removes itself from the core. The loop region of repeat III covering the N-terminal helix in the apo structure does not lift up allowing the N-terminal to swing out of the pocket. The process resembles a sliding motion, where the N-terminal pulls out from the bottom of the core domain. Our results also indicate that the folding of helix D in repeat III of the core domain folds in a two-step process, and during the conformation change calcium binding sites undergo secondary structure change. The results obtained using the NEB method provides an atomistic explanation for the complete conformation change pathway of annexin A1.

Molecular dynamics simulations totaling 1,425 ns were performed for wild type, D65A, S69C, A8V, L29Q, and C84Y mutations of cTnC. These mutations included five point mutations in the N-terminal domain, two of which are located at calcium binding site II. The simulation trajectories were analyzed using MMPBSA/MMGBSA, RMSD, and distance analysis. Our results showed that D65A, S69C, and L29Q have a decrease in calcium binding affinity. The A8V and C84Y mutations had an increase in calcium binding affinity. The loss of calcium binding affinity is detrimental when mutation of anchoring residues in site II occur, such as D65 and E76. Overall, trends in the effect of calcium binding affinity for these mutations and calcium binding to site II show consistencies with experimental studies. Mutations in the N-terminal domain of cTnC did not have as strong as an effect on calcium binding site III and IV of the C-terminal domain. These binding sites have higher calcium binding affinity than site II and are not affected much by the mutations tested in this work. The mutational studies of cTnC are important to investigate due to their link with cardiac diseases. cTnC's involvement in the muscle contracting process is vital, and many of these mutations lead to cardiac diseases that need to be studied in more detail to understand the mechanistic effect they have on muscle contraction.

COMPUTATIONAL INVESTIGATION OF CALCIUM BINDING PROTEINS ANNEXIN A1 AND CARDIAC TROPONIN C

A Thesis

Presented to

The Faculty of the Department of Chemistry

East Carolina University

In Partial Fulfillment

Of the Requirements for the Degree

Masters of Science in Chemistry

By

Kimberly A Lewis

July 2019

COMPUTATIONAL INVESTIGATION OF CALCIUM BINDING PROTEINS ANNEXIN A1 AND CARDIAC TROPONIN C

By

Kimberly A Lewis

APPROVED BY:

DIRECTOR OF THESIS: _____

Dr. Yumin Li

COMMITTEE MEMBER: _____

Dr. Libero J. Bartolotti

COMMITTEE MEMBER: _____

Dr. Anne Spuches

COMMITTEE MEMBER: _____

Dr. Xiaoping Pan

CHAIR OF THE DEPARTMENT OF CHEMISTRY:

_____

Dr. Andrew Morehead

DEAN OF THE GRADUATE SCHOOL:

_____

Dr. Paul J. Gemperline

ACKNOWLEDGEMENTS

I would like to acknowledge my research advisor, Dr. Yumin Li for her role in my research. When it got tough her words of encouragement helped me push forward and complete any task that was daunting. Her uplifting and positive approach made it a pleasure to work with. She is very caring and dedicated to the success of her students and encourages us to never give up and do our best. I will always remember her words of encouragement, and all the things she has taught me.

I would also like to acknowledge Dr. Libero Bartolotti for all the advice he has given me throughout my research. His expertise was vital for problem solving many issues that came up and was always willing to take the time to help me figure out where issues were coming from.

I would like to thank my committee members and the chemistry faculty that have heled along the way. I would like to thank my committee members, Dr. Anne Spuches and Dr. Xiaoping Pan for their support.

To the graduate students of the chemistry department I would like to thank you for making my time here at ECU enjoyable and I am thankful for the friendships I have made.

Finally, I would like to thank my family for all the love and support they have given me, especially my husband Zachary Sutphin who has encouraged me every step of the way and sat with me during late night trips to the office to finish work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Calcium Binding Proteins

Calcium binding proteins use the process of binding calcium, to regulate biological functions and interactions. Many proteins in biological systems require metal binding, such as calcium, to carry out function through signal transduction [1]. Binding loops located in calcium binding proteins contain negatively charged oxygen atoms which coordinate to the positively charged calcium ions [2]. Without proper coordination and binding of calcium, certain biological processes do not occur properly or might not occur at all. The calcium binding loops of proteins are vital, and changes to the binding loops may cause detrimental functional defects for the protein. Many proteins undergo calcium induced conformation changes, allowing for interactions with other proteins, ligands, or membranes [2]. Calcium binding proteins have different structures and functions, but all require calcium to perform certain functions for the proteins. In this work we studied two calcium binding proteins, annexin A1 and cardiac troponin C. These proteins do not have any structural or functional similarity, but they both require calcium induced conformation change to fulfill their function.

1.2 Annexins

The first member of the annexin family was discovered roughly 40 years ago. Synexin, now known as annexin A7, was discovered as a potential receptor for calcium in the process of exocytosis for adrenal medulla [3]. As the discovery of different annexin proteins progressed, naming of the proteins were chosen based on function. For example, annexins discovered early on were classified by names such as synexin, chromobindin, calcimedin, lipocortin, and calpactin [3-7]. At this time, it was unclear that these proteins were part of a much larger protein family. The realization that these proteins were part of a larger family came after biochemical studies, protein and cDNA sequencing, and gene cloning [8]. It was discovered that the proteins had similar biochemical properties, gene structures, and sequencing. Once it was discovered that these proteins belonged to the same multigene family, they were renamed with a common name of annexin.

The word annexin is derived from the Greek word annex, which means to bring or hold together. It was chosen because it accurately describes a property of annexins to bind to different biological structures such as membranes or other proteins. As of 2002, there were 160 annexin proteins identified over 65 different species [2]. Annexin proteins are most commonly found in eukaryotic organisms such as plants, fungi, and animals. To be considered an annexin protein, it must be able to bind to negatively charged phospholipids in a calcium dependent manner and have an annexin specific repeat. The annexin repeat is a 70 amino acid repeat sequence that they have in common. In humans there are 12 annexin proteins. These are named Annexin A1-A11 and A13, where annexin A12 is not assigned. The human annexin proteins are most commonly

found inside the cell near membranes, however, annexin A2 and A5 have also been found in the nucleus [9].

Annexins differ from each other in structure and function. Structurally human annexins differ mostly in the length and sequence of the N-terminal domains. Some human annexins have shorter N-terminals than others. Those annexins with shorter N-terminals are not able to perform certain functions, like annexins with longer N-terminals. It is known in the protein community that structure dictates function, and the annexin family is a good representation of this. Even though all human annexin are of the same family, they have slightly different structures lending to different functional characteristics such as phosphorylation, membrane aggregation and fusion, signal transduction, endocytosis, and exocytosis [10-15].

Structure

Annexins are composed of two domains, a unique N-terminal domain and a conserved core domain. The core domain houses calcium binding sites for the protein and the N-terminal domain is thought to regulate specific function for annexin proteins [1, 2]. Annexins have an alpha helical N-terminal domain. Each annexin has differences in their N-terminal, including differences in the length and sequencing. The annexin core domain are tightly packed helix-loop-helix motifs that are unique to annexin proteins [16]. An example structure for annexin with the N-terminal and core domain highlighted for distinction is expressed in **Figure 1.1** [17].

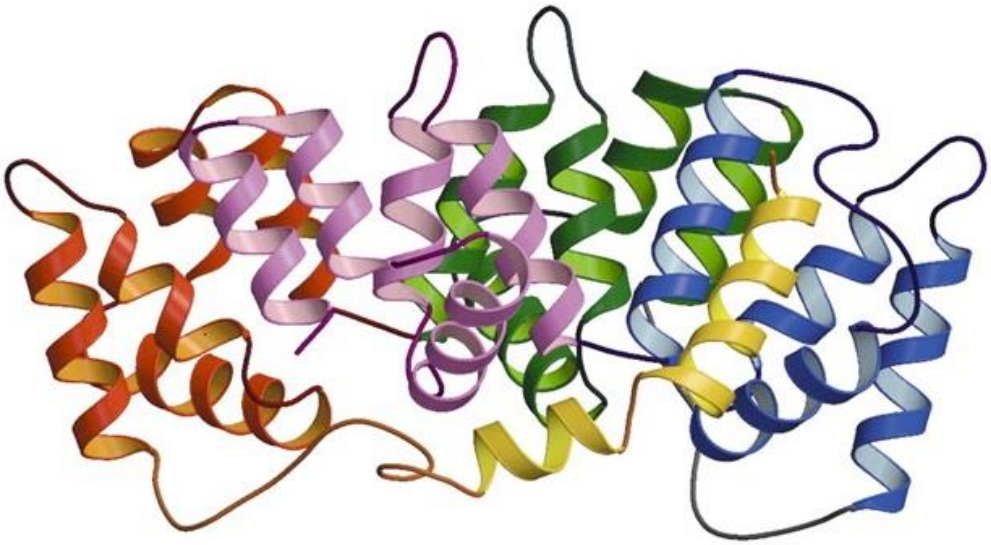


Figure 1.1. Full length structure of annexin A1 in a calcium free state. The N-terminal helix (yellow), and core domain repeat I (red), II (green), III (blue), and IV (purple) are highlighted for clarity.

N-terminal Domain

Previously we mentioned that the N-terminal domains of annexins vary in length. The members of the annexin family with large N-terminals range anywhere from 30 to 100 residues. Annexins with shorter N-terminals are not long enough to extend out from the protein to interact with proteins or membranes, like annexin with much longer N-terminals. Variation in the N-terminal domain is what makes each annexin unique and is believed to determine protein function [2]. In annexin proteins, the largest difference is the N-terminal, so this is the most reasonable cause for differences in protein function.

Annexins with long N-terminals are able to interact with membrane and protein and are involved in processes such as membrane aggregation [11]. When the N-terminal is extended from the core domain, there is free space around the N-terminal where other proteins can

interact, or membrane binding can occur. Annexins with shorter N-terminals are not able to extend from the core domain, so they have limitations if and how they interact with other species. **Figure 1.2** depicts the number of residues for a few of the human annexin proteins [18]. Annexin A1, A2, and A7 have the longest N-terminals as shown in **Figure 1.2**, with annexin A7 having the longest N-terminal by over one hundred residues.

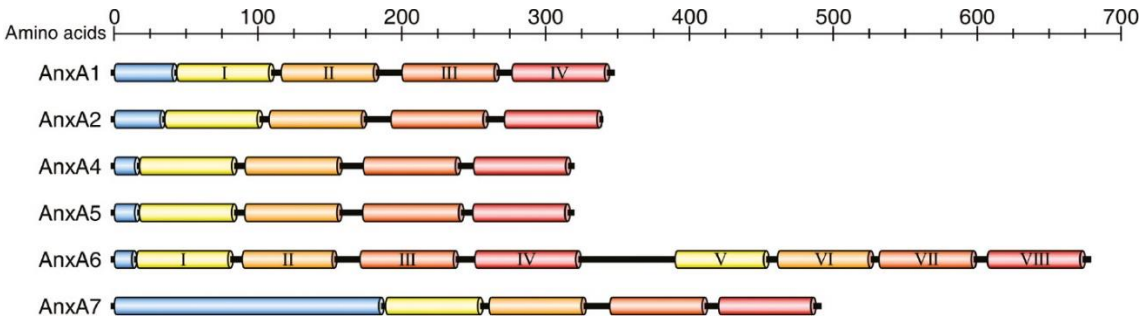


Figure 1.2. Amino acid counts for N-terminal domain and core repeats of annexin A1, A2, A4, A5, A6, and A7. The N-terminal residues are highlighted in blue, repeat I yellow, repeat II orange, repeat III pink, and repeat IV red. Annexin A6's repeat V-VIII are colored in the same order as the first four repeats.

In this work we focus on annexin A1, a member of the annexin family with a large N-terminal domain. Annexin A1 has a 28 residue N-terminal helix. The N-terminal of annexin A1 is composed of almost full helical structure. The first two residues of annexin A1 have loop structure and the remainder of the residues are alpha helical. The long N-terminal of annexin A1 can interact with membranes and other proteins. The N-terminal helix of annexin A1 is connected to the core domain by a 17 residue amino acid linker. This linker is flexible and allows for rotation and movement of the N-terminal helix in the cellular environment and is important for conformational change of annexin A1.

C-terminal Domain

Most annexins C-terminal, or core, domains are composed of four homologous repeats. However, annexin A6 is an exception and has eight core domain repeats as shown in **Figure 1.2**. Each repeat is composed of approximately 75 amino acid residues that fold into five alpha helices. The alpha helices are named A-E. The core domain house calcium binding sites on the convex side of the protein, as shown in **Figure 1.3** [17]**.** The convex side of the core domain curves up slightly, which allows for interaction of the protein and membrane. Once calcium binding occurs, the core domain of annexin is able to interact with negatively charged phospholipid membranes.

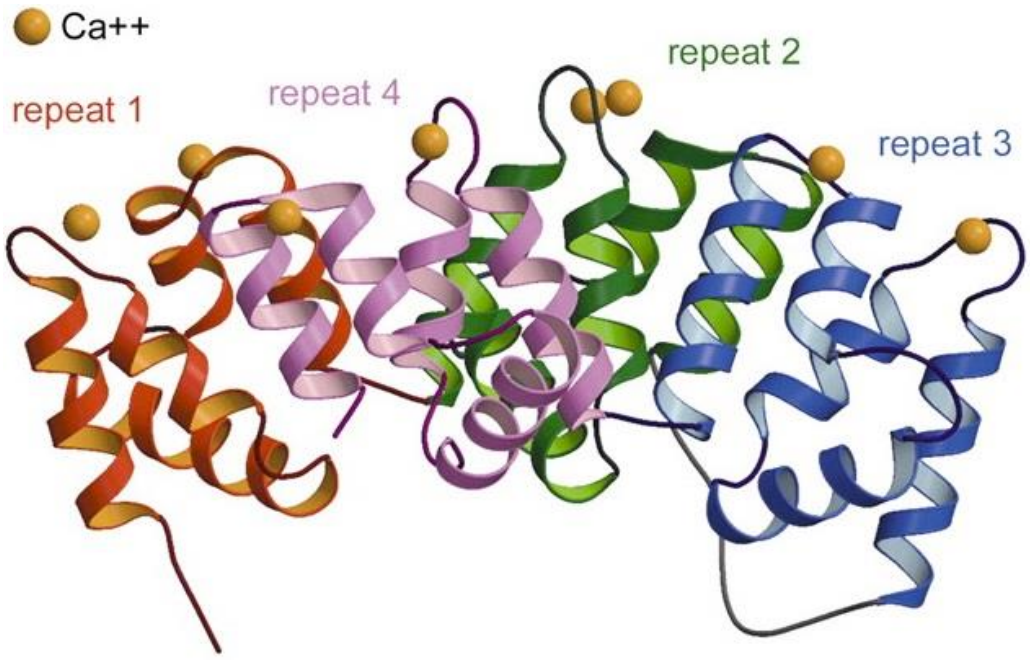


Figure 1.3. Core domain of annexin A1 with calcium bound to each binding site along the convex side of the core domain.

Annexin A1 has a core domain that houses eight calcium binding sites. These calcium's are coordinated to the protein through oxygen atoms located on the binding loops, as well as surrounding water molecules. Most calcium ions in the binding loops have seven ligand contacts, except for AB binding site I and II which have eight ligand contacts [19]. The concave side of the core domain is facing the N-terminal domain. In annexin A1, the N-terminal is tucked into repeat III of the core domain along the concave side of the protein. The annexin A1 core domain curvature resembles a curved disk, which allows for maximal binding and interaction of membrane.

Annexin Conformation Change

Calcium binding induces important conformational changes for many annexin proteins. In this work we focus on the conformation change of annexin A1. In the apo structure of annexin A1 the N-terminal domain is buried in repeat III of the core domain. The N-terminal helix holds the location of helix D of repeat III. The residues of repeat II that form helix D in the calcium bound structure, are a loop covering the N-terminal helix in the apo structure. A calcium induced conformation change occurs for annexin A1 ejecting the N-terminal helix from the core, and repeat III amino acids fold to form helix D. The N-terminal removal and folding of helix D in repeat III are the most notable changes in annexin A1 conformation due to calcium binding. **Figure 1.4** represents the calcium induced conformation change in X-ray structure for annexin A1 [17].

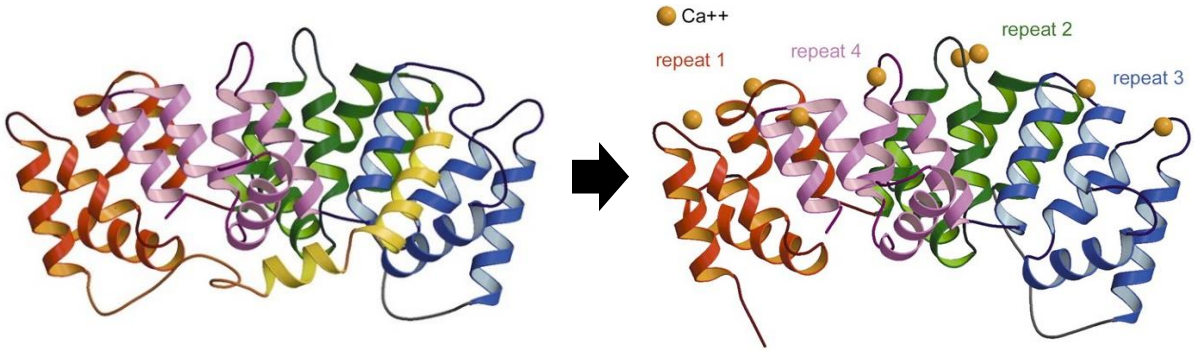


Figure 1.4. Structure change in annexin A1 for the calcium free and calcium bound structure.

Calcium induced conformation change is vital for the function of annexin A1. Calcium binding not only triggers protein interaction with membrane, but also initiates the conformation change for the N-terminal removal and helix formation of repeat III. Once the N-terminal is free from the core, it is able to interact with proteins such as S100 or membrane and induce membrane fusion and aggregation events.

Even though the conformation change for annexin A1 is vital for the function, the process in which this occurs is unclear. The N-terminal domain is connected by a flexible linker and in a calcium bound state is hard to crystalize without a binding partner [2]. This is evident since no structure for annexin A1 is present in the protein data bank with the N-terminal ejected from the core in the absence of other proteins. This is believed to be due to the mobility of the N-terminal. Without a protein bound to the N-terminal it is not stable enough to get a clear structural representation through X-ray crystallography. Experimentally it is hard to investigate the conformation change due to such challenges, however, with computational tools we try to determine structural change through the conformational process of annexin A1.

Membrane Aggregation and Fusion

Annexins have been shown to be involved in membrane aggregation processes, however, the process in which this occurs is debated. Annexins are involved regularly in membrane aggregation and fusion due to their ability to bind to negatively charged phospholipid membranes. The core domain is solely involved in interacting with membrane, while the N-terminal domain interacts with proteins to facilitate fusion of two membranes. Three different models for membrane aggregation have been proposed, as shown in **Figure 1.5** [19]. In each model the main difference consists with the location and orientation of the N-terminal helix.

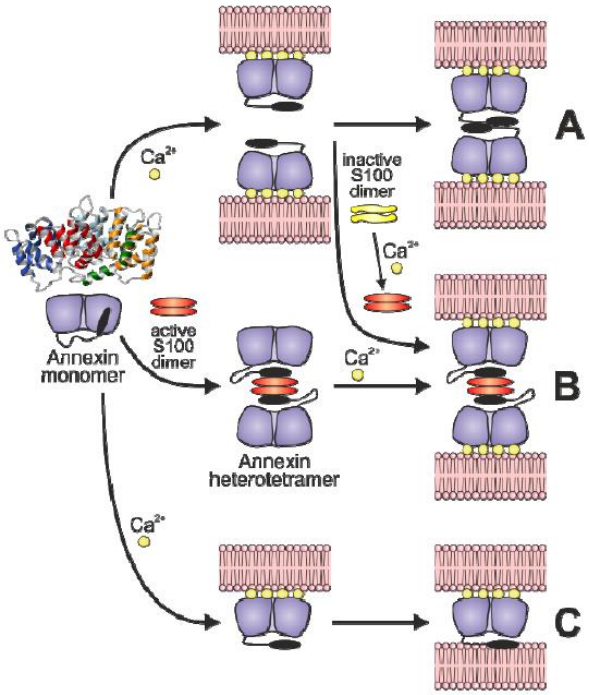


Figure 1.5. Three models for annexin interaction with phospholipid membranes. A) Two annexin monomers binding to two different membranes and N-terminal dimer formation. B) Annexin heterotetrametric formation and binding to membrane. C) Annexin monomer binding to membrane and N-terminal interaction with a second membrane.

In the first model it is purposed that after calcium binding to the convex side of two annexin monomers, the core binds to two different negatively charged membranes. After binding to the membranes, the N-terminal domains of each annexin come together and interact with each other bridging the membranes (**Figure 1.5 A**) [19]. In the second model, a similar process occurs for annexin binding to the membranes, except the N-terminal domains of annexin are bound and bridged by a S100 protein dimer (**Figure 1.5 B**) [19]. In the last model, calcium binds to the convex side of the core domain of an annexin monomer, which binds to one negatively charged membrane while the N-terminal helix binds to a second membrane (**Figure 1.5 C**) [19]. While all three models show different conformations of membrane interaction and binding, they all require both the convex side of the core and N-terminal domains to be involved in biding to bridge two negatively charged membranes. More studies on the exact process of membrane aggregation and fusion are required to really understand how this is done on an atomistic level.

Annexin-protein Interaction

As previously introduced, annexin proteins can interact with other proteins such as S100 proteins. Annexin bind to S100 proteins with their N-terminal domains, **Figure 1.6** [20]. Binding sites located on both proteins, facilitate the interaction and binding of annexin and S100. The N-terminal of annexin needs to be free to bind to the S100 protein, so for annexin A1 the calcium bound structure would be the only conformation able to bind to S100. S100 are a set of proteins that are involved in antitumor suppression [21, 22]. It is believed that their interaction with annexin may link the cell aggregation process to that of the signaling pathway for antitumor suppression [23].

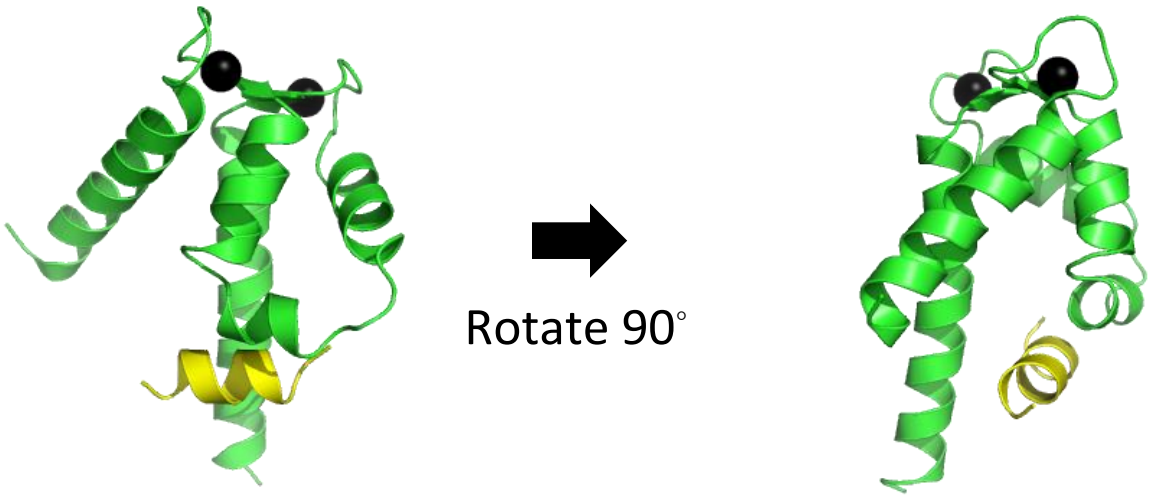


Figure 1.6. S100A11 dimer (green) bound to a portion of the N-terminal helix of annexin A1 (yellow). Calcium ions are shown in black.

Tetrameric Complex

Annexins are able to form tetrameric complexes to aid in membrane aggregation and fusion processes, as suggested in **Figure 1.5 B**. In these tetrameric complexes a dimer of proteins, such as S100 are located in-between two annexin monomers. The tetrameric complexes may be important for membrane fusion processes. Different annexin-S100 protein combinations can form such complexes, however only one has been determined through X-ray crystallography so far, **Figure 1.7** [24]. The N-terminal domain of annexin A2 (red) extends out away from the core domain and each one interacts with one monomer of S100A4. The flexibility of the linker region between the N-terminal domain and the core allow for movement of the N-terminal and ease of interaction with S100.

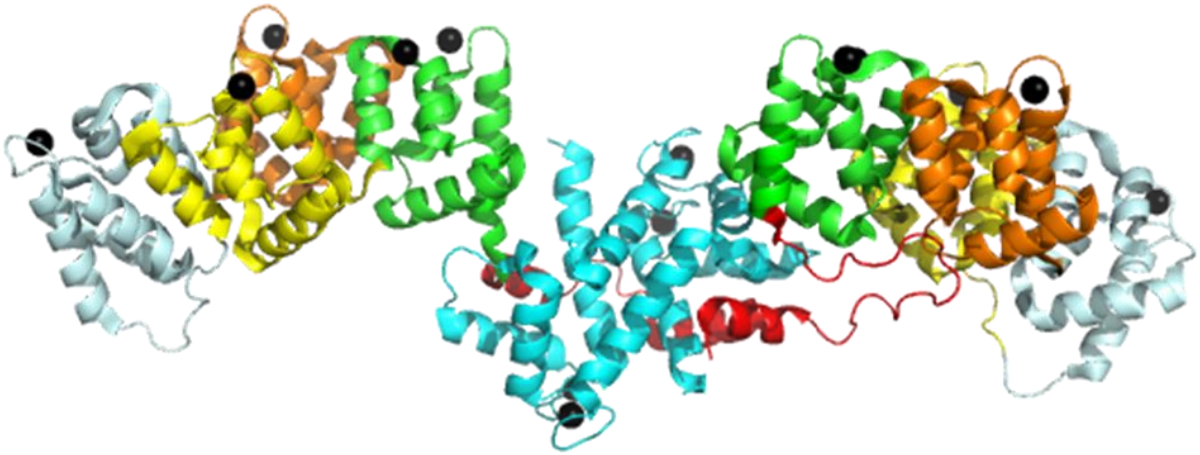


Figure 1.7. X-ray crystal structure of the tetrameric complex of annexin A2 and S100A4 dimer. The calcium ions are shown in black, S100A4 dimer in cyan, N-terminal domain of annexin A1 in red, core repeat I in green, core repeat II in yellow, core repeat III in grey, and core repeat IV in orange.

Continued research on the importance of membrane interaction of these tetrameric complexes needs to be performed. Due to computational complexity it is not easy to simulate systems of this size, and work in this area is slow.

Disease

To date there has been no evidence suggesting that annexins direct genome is involved in different biological diseases. However, annexins are indirectly linked to serious human diseases [11]. Changes in annexin expression levels can occur in certain conditions and these can be indicative of different disease pathophysiology. Diseases connected to certain annexin range from inflammatory disease to cancer. The variety of diseases range depending on the annexin protein, and certain diseases can be either up or down regulated with annexin protein expression.

Annexin A1 has an anti-inflammatory role due to its connection with COX-2 and PLA_2. The expression and secretion of annexin A1 is regulated by glucocorticoids and is suggested that annexin A1 can act as a messenger protein linking the glucocorticoid release as regulation of COX-2 and PLA_2. This process is supportive of the prostaglandin pathway [25]. For many of the inflammatory disease studied in connection to annexin, rheumatoid arthritis and myocardial infarct have been studied in more detail. In experimental studies fragments of annexin A1 have been shown to protect against myocardial reperfusion which suggests that during inflammatory responses annexin A1 may have an anti-inflammatory role [26].

Annexin A5 is expressed in high levels in the regulation of blood coagulation, and mutation of this protein are linked to myocardial infraction [27, 28]. This shows that annexin can have an impact in several diseases or conditions. In leukemia cells annexin A2 expression is up regulated which increases complications as a hemorrhagic disorder [29]. In different development stages of cancer cells there are different annexin expression profiles. For example, annexin A1, A2, A4, A6, and A11 are down regulated in prostate cancer and is suggested that the down regulation is dependent on DNA-methylation [28]. These are just a small list of the diseases that are associated with some of the annexin protein family members.

Membrane Repair for Cancer Cells

Recently, annexins involvement in membrane repair of cancer cells has become a popular area of study. The plasma membrane of cancer cells is less stable and are not as stiff as normal cellular membranes [30]. Due to the reduced stiffness of these cells' injury is prone to occur as the cells migrate through the cellular matrix [31]. The purposed mechanism for plasma

membrane repair in cancer cells involving annexin-S100 protein tetramers as well as actin, is shown in **Figure 1.8**.

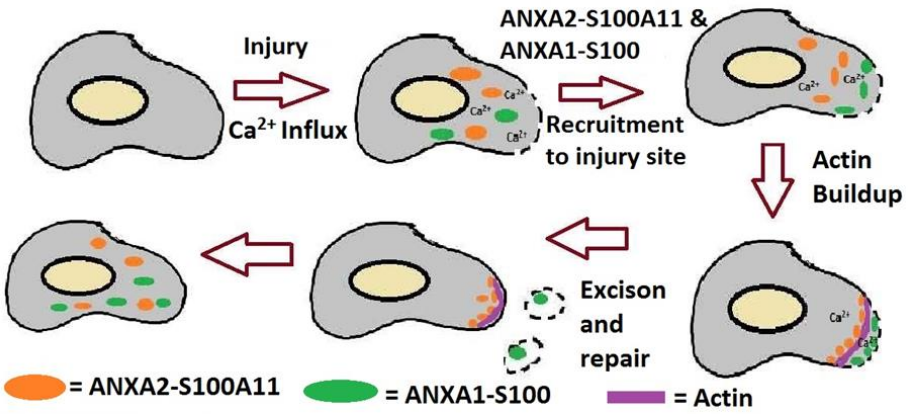


Figure 1.8. A purposed plasma membrane repair mechanism of cancer cells. Annexin-S100 tetramers gather at the injury site, removing and repairing the plasma membrane of cancer cells.

Upon membrane damage, a calcium influx triggers recruitment of ANXA2 and S100A11 to the repair site. After membrane injury ANXA2 and S100A11 are accumulated at the site of repair after 15-45 seconds [32]. ANXA1 is also recruited after injury, however, instead of being recruited to the site of repair ANXA1 is recruited to the site of injury. Rapid buildup of F-actin occurs at the site of membrane repair once ANXA2 and S100A11 are recruited to the site [33]. S100A11 and ANXA2 preserve the existing F-actin and allow new F-actin to form at the site of injury. ANXA1 is purposed to bind directly to the injury site and promote membrane fusion. Detailed information on the exact mechanism for membrane repair of cancer cells still needs to be performed to understand annexins exact role in the process.

1.3 Muscle Contraction

In the body muscle is the contractile system. Muscle is mainly composed of thin and thick filament. The thin filament is composed of proteins such as actin, tropomyosin and troponin, while the thick filament is composed of myosin. Both thin and thick filaments make the sarcomere, which is known as the basic unit of muscle tissue. The idea of a highly integrated protein composition for the muscle contracting mechanism was frequently debate, until the sliding filament model was introduced [34, 35]. In this model, contraction occurs as the thin and thick filament slide past each other. This process is shown in **Figure 1.9** [36].

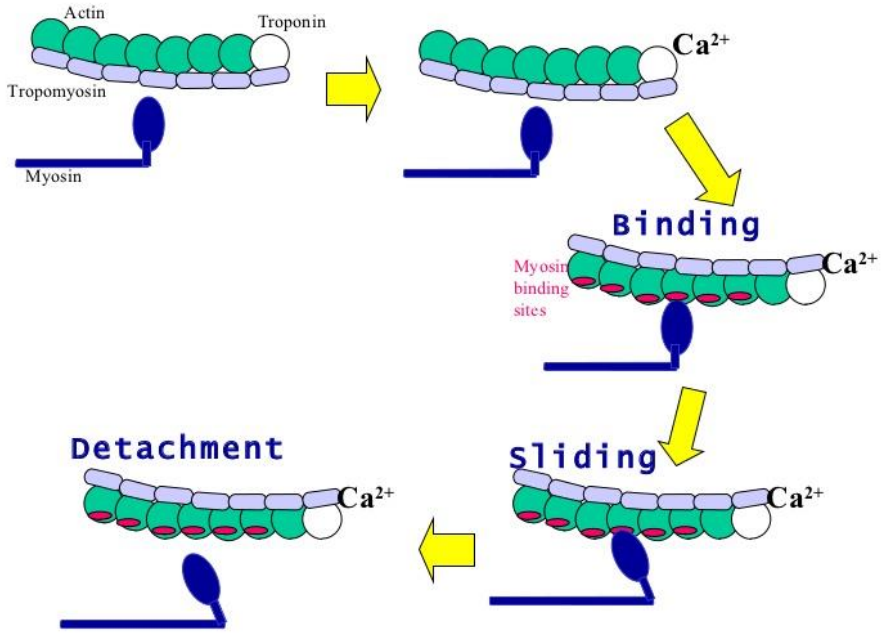


Figure 1.9. Cartoon depicting the involvement of the thin and thick filament in muscle contraction.

Calcium binding to troponin causes a series of conformational changes to occur. Troponin interacts with tropomyosin, lifting it away from actin exposing myosin binding sites. Myosin is then able to work its way along each myosin binding site facilitating contraction, when myosin detaches from actin and calcium is pumped out of the cell the tropomyosin lowers back over the myosin binding sites.

It was understood that calcium caused some sort of reaction that lead to muscle contraction; however, the exact role was debated. This debate continued until 1963 when a protein called component C was shown to restore calcium sensitivity to the muscle protein [37]. This protein was later renamed troponin and is part of the focus of this research.

Mechanism for Muscle Contraction

An influx of calcium to the sarcomere triggers the muscle contraction mechanism. Calcium binds to site II of cTnC acting as the initial step triggering contraction [38]. TnC undergoes conformational changes after calcium binding, which includes opening of a hydrophobic pocket in the N-terminal lobe of TnC as shown in **Figure 1.10**. When the hydrophobic pocket is exposed it can interact with the switch segment of TnI.

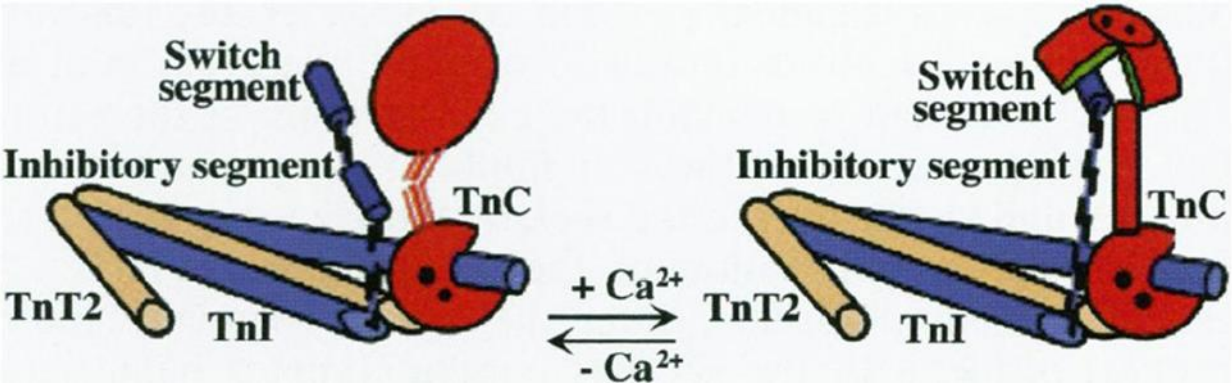


Figure 1.10. Cartoon depicting TnI switch segment interaction with the N-terminal domain of TnC after calcium binding.

The interaction of the switch segment of TnI and the N-terminal of TnC causes other segments of TnI to move. This series of movement weakens the inhibitory activity of TnI on the actin-tropomyosin filaments. With less strain the tropomyosin can move toward the helical groves in actin exposing myosin binding sites [39]. Exposure of the myosin binding sites allows myosin heads easy access to move along these sites facilitating muscle contraction, as shown in **Figure 1.9**. To relieve the contraction process, calcium ions are pumped out by sarcoplasmic reticulum calcium ATPase pumps. This process in known commonly as muscle relaxation. In the muscle contracting mechanism calcium plays a huge role, and further understanding of the mechanism will give more insight into each unit's specific role. The dynamic muscle contracting mechanism has many moving parts, and defects in one area of the mechanism can halt proper contraction from occurring.

1.4 Troponin EF Hand Protein

Troponin is a member of the "EF hand" protein family. The concept of the EF hand protein was developed by Kretsinger and his group when investigating a protein structure of parvalbumin [40]. The helix loop helix motif of this protein was unique and bound to calcium in a way similar to that of a pointer finger configuration, as shown in **Figure 1.11** [41]. The index finger and thumb act as the helices adjacent to the binding loop, and the fingers curved into the palm act as the binding site loop.

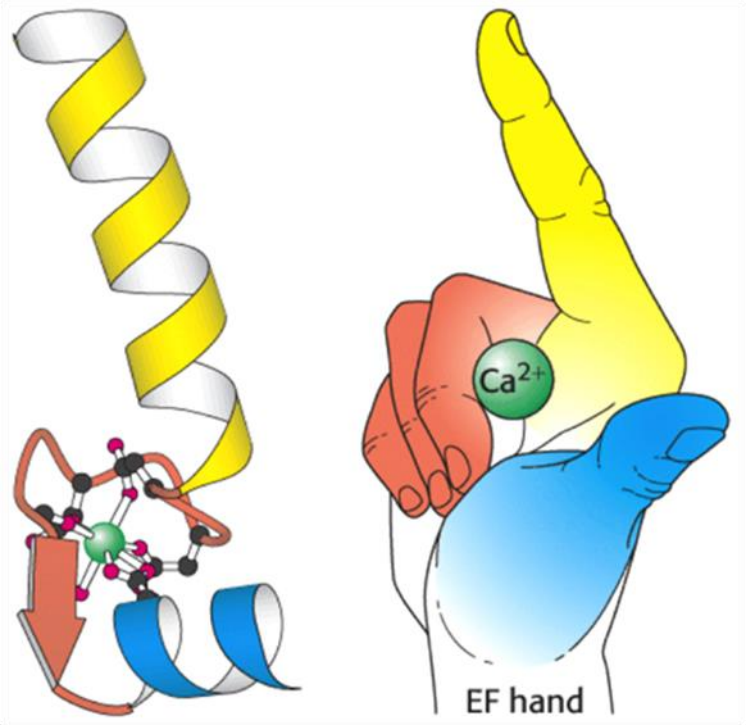


Figure 1.11. EF hand structural representation using an actual hand. The corresponding helices (yellow and blue) and loop (red) are shown colored on the hand.

EF hand proteins are classified as either calcium sensors or calcium buffers. Calcium sensors are tasked to carry out the signal of increasing calcium concentrations, which result in different biological responses such as conformation change. Calcium buffers modulate the signal to bring or remove calcium ions from the cells. Examples of EF hand calcium sensors are calmodulin and troponin, while calcium buffers are calbindin and parvalbumin.

Proteins can have various number of EF hand motifs, but troponin only has four. The motif has 12 highly conserved amino acid residues, where those at positions 1, 3, 5, 7, 9, and 12 coordinates with calcium [42]. All amino acids bind with monodentate ligands, except for the 12^{th} position that binds with a bidentate ligand. The seven oxygen atoms of these six residues give calcium its pentagonal bipyramidal geometry [43]. In this work cTnC has EF hand calcium binding motifs that we investigate through mutation.

1.5 Troponin

The two forms of striated muscle are skeletal and cardiac muscle. Troponin plays a major role in muscle contraction, and with different kinds of muscle there are different kinds of troponin. The forms of troponin are skeletal (sTn) and cardiac (cTn). Troponin is composed of three subunits, TnC, TnI, and TnT (**Figure 1.12)**. Troponin binds to tropomyosin in a 1:1 ratio, to help facilitate the muscle contraction process. In this work we focus on the TnC subunit of cardiac troponin (cTnC).

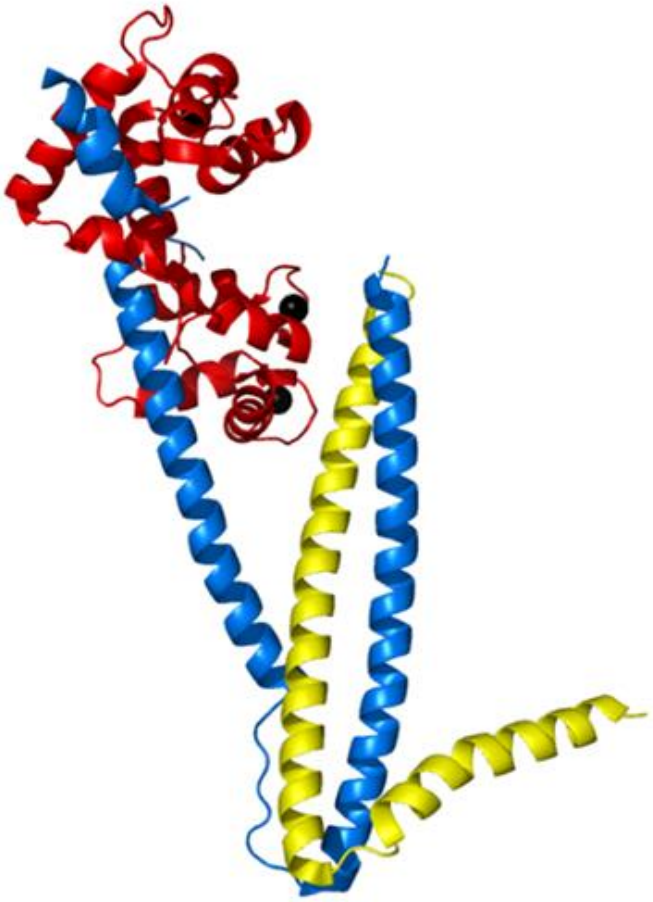


Figure 1.12. Cardiac troponin complex composed of TnC (red), TnI (blue) and TnT (yellow).

Troponin C: TnC

Troponin C (TnC) is the subunit of troponin that binds calcium. Calcium binding causes a chain reaction in conformation change of troponin leading to muscle contraction. This is the only subunit of troponin that will bind to calcium. TnC is composed of two globular domains connected by a flexible linker. TnC is also a member of the EF hand family, with its unique helix-loop-helix motif. There is a total of four EF hand motifs in TnC. Two motifs are located in the N-terminal domain and the other two the C-terminal domain. The EF hand motifs house the metal binding sites for the protein. TnC has four metal binding sites, however calcium only binds to three of the sites. Calcium binds to TnC metal binding site II, III, and IV. Site II is located in

the N-terminal domain, while sites III and IV are in the C-terminal domain and are shown in **Figure 1.13**.

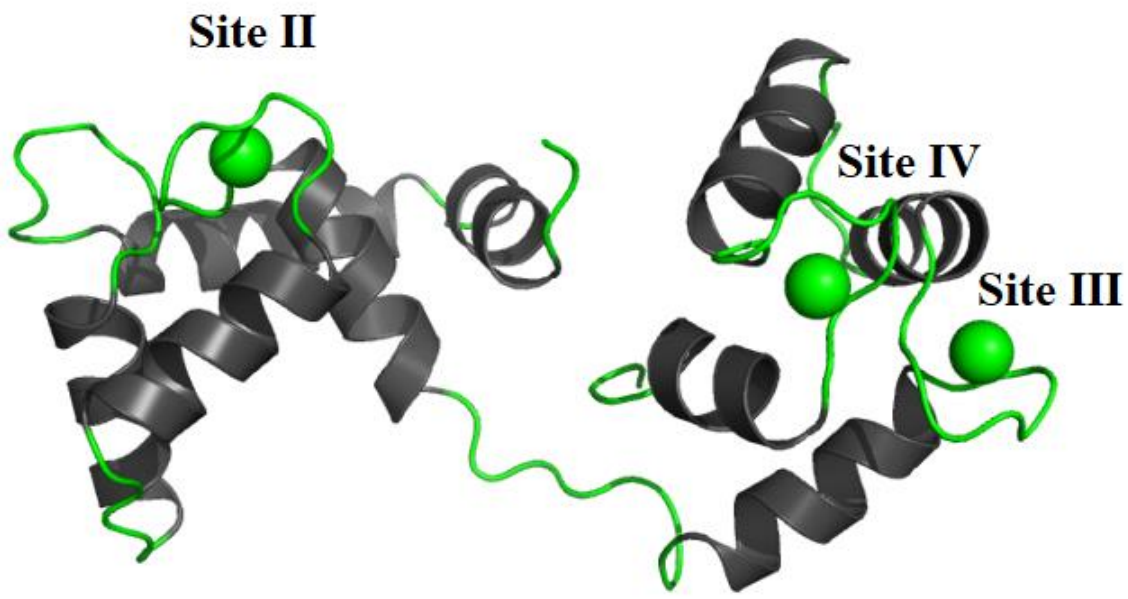


Figure 1.13. Wildtype cTnC protein with calcium ions and loop residues shown in green and helical residues shown in dark grey.

Site II has a low affinity binding site to calcium and site III and IV in the C-terminal domain are high affinity binding sites to calcium. The binding constants for N-terminal site II and C-terminal sites III and IV have been reported to be on the order of 10^5 M^{-1} and 10^7 M^{-1} respectively [44,45]. When TnC is in complex with TnI the binding constants increase by one order [44].

The difference between skeletal troponin C (sTnC) and cardiac troponin C (cTnC) are the number of calcium ions that bind to the N-terminal domain. In cTnC the first binding site is

non-functional due to amino acid mutations [46]. In sTnC calcium will bind to both site I and II. Calcium binding to the N-terminal domains of TnC are thought to be responsible for the regulation of muscle contraction. Studies that have investigated the calcium binding kinetics of the N-terminal domain have yielded data to support these claims [47-49], while the role for C-terminal sites are not as clear and seem to serve only a structural role [50].

TnI

Troponin I (TnI) is the subunit of troponin that has an inhibitory effect on the contraction process. There are three different kinds of TnI. In skeletal troponin there are slow skeletal troponin I (ssTnI) and fast skeletal troponin I (fsTnI). Cardiac troponin only has one form of troponin I (cTnI). The cTnI subunit is aligned in an anti-parallel fashion with cTnC, as shown in **Figure 1.12**. TnI has several segments including the N-terminal cardiac specific extension, C-terminal TnC binding region, TnT binding region, inhibitory region, switch region, and C-terminal region that binds actin-tropomyosin [51]. The region of TnI important for muscle contraction are the inhibitory and switch region. The inhibitory region has the ability to bind to actin-tropomyosin to inhibit contraction or TnC, and the switch region acts as a trigger for the calcium induced conformation changes for this part of the complex.

TnT

Troponin T (TnT) is the tropomyosin interacting subunit of troponin (**Figure 1.12**). TnT are divided into two sections, the T1 region that interacts with tropomyosin and T2 region that interacts with the other subunits of troponin. The roles of TnT's involvement in the muscle contraction process is not fully understood because the T1 region of the subunit is not available

in a high-resolution structure [38, 52]. At this point the role of TnT and muscle contraction is understood at a structural level where it serves a purpose to bind the complex to the thin filament [53].

Mutations

Mutations in any subunit of troponin can have detrimental effects on the contractile mechanism of muscle fibers. TnC, TnI, and TnT all have certain roles in the mechanism and the cooperative nature of protein interaction can be disrupted due to mutations in the subunits of troponin. Mutations in parts of troponin can lead to conformational changes effecting the calcium binding affinity, and lead to different diseases. Mutations have been discovered in both cardiac and skeletal troponin. Approximately 60 mutations have been found for cTn, while approximately 5 have been discovered for sTn leading to diseases [54, 55].

Disease

Skeletal troponin mutations are associated with disorders such as distal arthrogryposes [56]. These disorders are characterized with several congenital contractures of distal limb joints. Those mutations reported in sTn are linked to disorders like arthrogryposes. Mutations in sTn alter the calcium sensitivity in a process that is similar to that of mutations that lead to cardiomyopathies in cTn, which we will discuss in more detail [57].

Mutations in cTn will often times lead to diseases such as dilated cardiomyopathy (DCM) and hypertrophic cardiomyopathy (HCM). One of the most frequently inherited diseases are HCM which effects approximately 1 in 500 people and is a major contributor to sudden death in

heart disease patients [58]. A majority of the mutations associated with HCM occur in the TnT subunit of cardiac troponin, while fewer are associated in TnI and TnC. Many mutations in TnT cause HCM because mutations in the T1 region disrupt binding to tropomyosin, and mutations in the T2 region can affect TnT interaction with other parts of the troponin complex. There is also speculation that mutation at T2 of cTnT can affect the signaling capabilities for muscle contraction [51].

DCM can cause heart failure, arrythmias and blood clotting. Such events are speculated to occur in approximately 36.5 people out of 100,000 [59]. Fewer mutations are linked to DCM than HCM. So, a majority of people who contract a form of heart disease will likely get HCM. Because DCM is a less common disease, the lack of patients has made this more difficult to study than HCM. The muscle surrounding the ventricles of the heart are affected differently if you have HCM or DCM, as shown in **Figure 1.14** [60].

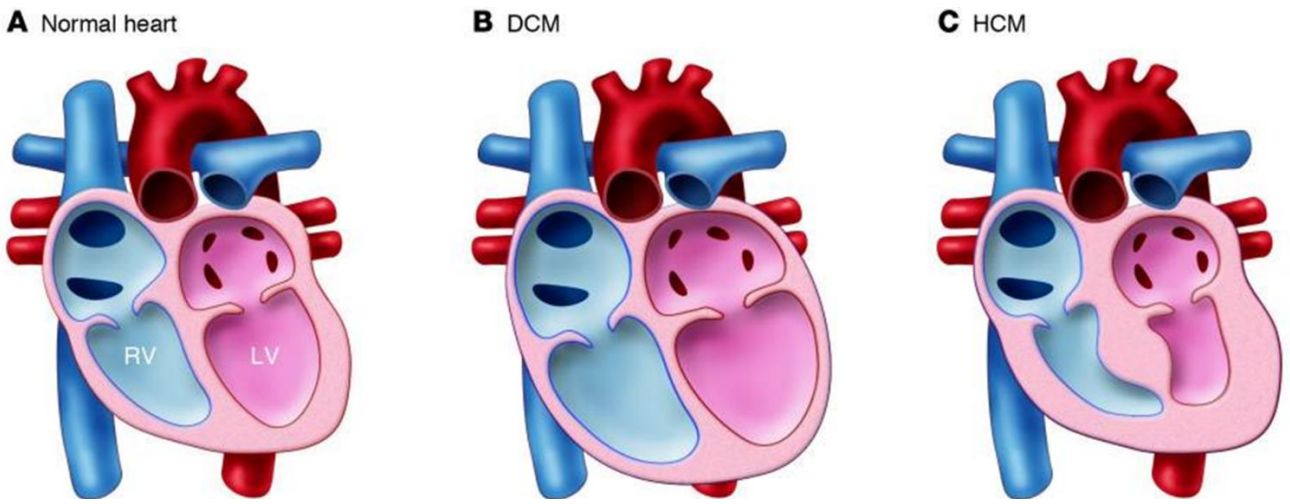


Figure 1.14. Comparison of the muscle and ventricle size for a human heart, and those with DCM and HCM type diseases.

In DCM the left and right ventricles of the heart become enlarged and are not able to pump blood as effectively to parts of the body. The enlarged ventricles make it hard to pump blood with the right amount of force as a normal heart. The heart muscle will also have difficulty contracting normally which leads to irregular heartbeats. In HCM however, the heart muscle is very thick. Increase in the thickness of the muscle shrinks the size of the ventricle. This makes it very hard to pump blood through the body.

A few of the mutations found in cTnC that are linked to diseases are displayed in **Table 1.1**. Mutations in cTnC can lead to downstream effects for the troponin complex effecting the ability of the sarcomere to regulate and be involved in muscle contraction. In this work we chose to investigate mutants D65A, S69C, A8V, L29Q, and C84Y in cTnC.

Table 1.1. Table expressing some of the point mutations of cTnC leading to HCM and DCM diseases.

Troponin Subunit	***Mutations***	***Disease***
TnC	Y5H	DCM
	A8V	HCM
	L29Q	HCM
	A32S	HCM
	E59D	DCM

	D65A	*
	S69C	*
	D75Y	DCM
	C84Y	HCM
	M103I	DCM
	E134D	HCM
	D145E	HCM
	I148V	DCM
	G159D	DCM

CHAPTER 2: COMPUTATIONAL METHODS

2.1 Molecular Dynamics Simulations

In the computational community molecular dynamics simulations are used frequently to investigate and study proteins. In this work molecular dynamics techniques have been used extensively to analyze the dynamics of troponin with point mutations. The calculation for molecular dynamics generates information at an atomic level and includes data for the position and velocity of atoms. In the classical molecular dynamic model atoms are thought of as balls connected by springs. This is where the "ball and spring" method connotation comes from. This is derived from Newton's second law of motion, F=ma. Where the force (F) applied to a particle is equal to the mass (m) and acceleration (a). This force can be related to the gradient of the potential energy of the system, shown in Equation 1, where m_i is the mass of particle i, r_i is the position of particle i, U is the potential energy of the system and N is the number of atoms.

$$m_i \frac{d^2\vec{r_i}}{dt^2} = -\nabla_i[U(\vec{r_1}, \quad \vec{r_2}, \quad \ldots, \quad \vec{r_N})] \quad i = 1, \ldots, N$$

Equation 1

The potential energy for the system is calculated using Equation 2, where the first four terms calculate the internal energy and the last three calculate the non-bonded energy contributions to the potential.

$$U = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} K_\chi[1 + \cos(n\chi - \sigma)] + \sum_{improper} K_\varpi(\varpi + \varpi_0)^2 + \sum_{electrostatic} \frac{q_i q_j}{\varepsilon_d r_{ij}} + \sum_{vdW} \left\{ \varepsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] \right\} + \sum_{H-bonds} \left[\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right]$$

Equation 2

The first term represents the covalent bond energy of two atoms using harmonic force as the approximation. K_b is the force constant, and $b\text{-}b_0$ is the change in bond length with reference to b_0. The second term represents the deviation in angles of θ to the reference angle of θ_0, where K_θ is the angles force constant. The third and fourth terms represent the contribution of the dihedrals and improper torsions to the potential energy. K_χ and K_ϖ are the force constants, σ is the reference angle at the maximum potential energy, and n is the multiplicity of the number of potential minima in a single rotation. The fifth term is the electrostatic interactions and are calculated using Coulomb's law, which is a summation of the interactions between pairs of point charges q_i and q_i within distance r_{ij} and dielectric constant ε_d. The sixth term represents the van der Waals interactions using Lennard-Jones potential function with the collision diameter $R_{min,\,ij}$, distance between atoms r_{ij} and depth of the energy well ε_{ij}. The last term represents the hydrogen bond contribution to the potential and is a part of the AMBER equation [61, 62].

The potential energy equation requires that force constants and equilibrium bond parameters are derived from a force field. A forcefield contains certain parameters required to calculate energies for a given geometry. In molecular dynamic simulations common forcefields are AMBER, CHARMM, GROMOS, MMFF, CFF, and OPLS. The AMBER program and all

atom forcefield were used in our simulations [63, 64]. Before the simulation can begin, starting coordinates and the forcefield must be defined, and initial velocities are calculated. Once this is done the simulation can be started. The simulation integrates Newton's equation of motion using a finite difference method. This method separates each integration step using a fixed time-step, or dt. The force on each particle during the time-step is calculated as the vector sum of its interaction with other particles. The acceleration from the force is combined with the starting position and velocity to calculate the new position and velocity at time t + dt. This process is repeated over the time of the simulation, and the atom positions and velocities are recorded in a trajectory file which can be used for analysis. Time-steps are commonly chosen to have an order of 1 femtosecond (fs) for molecular dynamic simulations. Time-steps can be increased to 1.5 fs when incorporating the SHAKE algorithm. SHAKE will fix fast atoms, such as hydrogen, in place [65].

In this work, we used AMBER 16 simulation package to complete our simulations. AMBER is a collection of programs used to prepare, simulate, and analyze a system and its trajectories. AMBER has been used for many projects to carry out simulations in implicit and explicit solvents. The forcefield for proteins, carbohydrates, and nucleic acids makes AMBER desirable to many kinds of computational studies. Other major program modules used were Leap, SANDER, and PTRAJ. LEap is used for system preparation before the simulation is started, SANDER is used during minimization and simulation, and PTRAJ is used in the analysis of trajectories.

Steps Required to Run Molecular Dynamic Simulations

1) System Preparation

The first step in simulation protocol is to prepare the coordinate and topology files using LEap. These files are necessary to run the simulation. The topology file contains information for the atoms such as, atom types, bond parameters, connections of atoms etc. The topology is derived from the specified forcefield. The coordinate file contains the 3-d position of each atom. This represents each atoms location in Cartesian space. The steps to perform the generation of these files in the LEap module is describe below, and a visual representation of outcome for these steps are show in **Figure 2.1**.

- AMBER forcefield parameters (parm99) were obtained with the modification file leaprc.ff03 to accommodate parameters for proteins. The charge parameters for the ligands were created using the xLEap module of AMBER [66].
- The system structures from X-ray or NMR pdb format are used in LEap. The hydrogens are removed and then the pdb structure is loaded into LEap. The hydrogens are removed to preserve the naming system used by AMBER.
- Any missing hydrogen atoms are added by LEap. This is done based on the geometry specified in the residue database.
- If there is an overall charge present in the system, it is neutralized. This is completed by adding counter ions in a shell around the system using a Coulombic potential. If the total charge for the system was negative Na^+ atoms were added, and if the charge was positive Cl^- atoms were added. Adding the counter ions stabilizes the charge of the system.

- The next step is to solvate the system. In our molecular dynamic simulations, the systems were solvated using the TIP3P water model [67]. The TIP3P water model are rigid bodies of water used to solvate the protein to mimic a biological environment. The protein is solvated in a water box with a distance of 14 Å between the edge of the protein and the water box.
- Once solvated the topology and coordinate files are saved and used for the simulation.

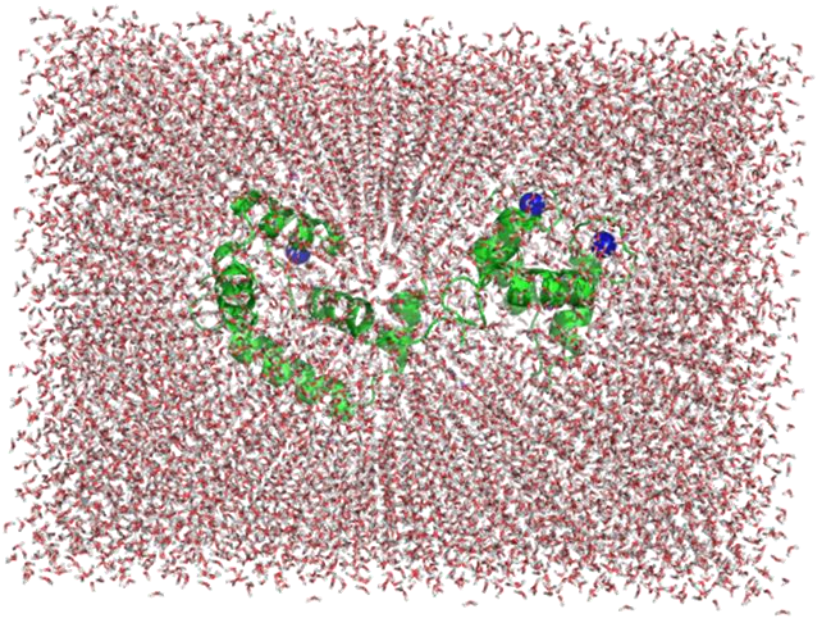


Figure 2.1. Example system preparation using cardiac troponin C. The blue spheres represent the calcium ions and the green structure is cTnC. Using LEap counter ions and TIP3P waters (red) are added surrounding the protein in a 14 Å box.

Periodic boundary conditions are used to avoid water molecules pushing out of the system due to the kinetic energy of certain temperatures. Images of the solvated unit are duplicated around the water box, and if one water molecule leaves its image will take its place. This ensures that the protein remains in a fully solvated system. All molecular dynamic

simulations were run using periodic boundary conditions with a 10 Å non bonded cut off. The particle mesh Ewald method was also used to treat any long-distance electrostatic contribution of the force field [68, 69]. All steps involved in the molecular dynamics simulation was performed on a 128 processor SGI Altix 47000 cluster at East Carolina University.

2) Energy Minimization

The initial structures were minimized to remove any energetic strain and bad contacts. Minimization is completed prior to the simulation and brings down the energy gradient. In all simulations two minimizations were completed. First a constrained minimization was completed keeping the protein fixed, the second minimization included the whole system, including the protein. The minimization for the fixed protein was performed for 3000 cycles and the second minimization for the whole system was performed for 5000 cycles. 500 cycles of the final minimization used the steepest decent algorithm. This algorithm removes bad steric overlaps before moving to the conjugate gradient minimization. Energy data was recorded every 25 cycles, and the final coordinates were wrapped in a box and written to restart and trajectory files. The restart file is used for the next step of the simulation.

3) Heating and Equilibration: Warming

Warming of the system includes a heating and equilibration process. This is done to heat the minimized structures to physiological temperatures. Slow warming was performed for 16,000 steps, increasing the temperature from 0 K to 300 K. This was done by heating in 50 K intervals of 2,000-5,000 steps. A time-step of 0.0015 ps was used with SHAKE. The energy data was

stored every 50 steps and the last coordinate data set was written to the restart file which is used in the next step.

4) Constant Pressure Dynamics (NPT)

NPT dynamics were run using an isothermal-isobaric ensemble, where the moles (N), pressure (P), and temperature (T) of the system are held constant. The NPT calculation is used to equilibrate the density of the system. This is required because when LEap adds water molecules to the protein the density in one area may not be the same as another. In this run the box size is reduced to equilibrate the density to approximately 1 atm. The flag for constant pressure was incorporated with an anisotropic pressure scaling to the 1 atm pressure target. The NPT run is stopped when the density stabilizes at a constant temperature of 300 K. Most of the NPT runs in this work were run for 100,000 steps. Every 100 steps the energy data was stored in the output file. Every 1,000 steps the coordinates were recorded to the trajectory and restart files.

5) Constant Volume Dynamics (NVT)

When the density of the system is stabilized the volume and temperature can be fixed, thus allowing for variations in the pressure. NVT dynamics are run keeping the number of moles (N), volume (V), and temperature (T) constant. This portion of the molecular dynamic simulation is the most time-consuming and will vary depending on the size of the system. For every 1,000 steps the coordinate set was recorded to the trajectory and restart file. Each simulation was run for at least 200 ns for the NVT step. The trajectories obtained from the NVT runs were used in analysis of the systems.

2.2 Trajectory Analysis

Analysis of the trajectory were carried out using the PTRAJ module of AMBER. The water molecules and counter ions were stripped from the trajectory to save analysis time. The new trajectory without water and counter ions can then be used to compute RMSD, RMSF, distance measurements etc. Preparation of the trajectory file for the molecular dynamic simulations are vital because the trajectory file is used in the analysis process.

Root Mean Square Coordinate Deviation (RMSD)

Before analyzing the trajectory and deriving information from the simulation, it is important to check to see if the system has equilibrated. To do this, the RMSD is checked for the simulation. RMSD measures the magnitude of displacement for atoms in reference to their previous structures. The RMSD is calculated as a function of time using Equation 3.

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(v_{ix} - w_{ix})^2 + \left(v_{iy} - w_{iy}\right)^2 + (v_{iz} - v_{iz})^2}$$

Equation 3

The terms v and w represent the position of two compared points in the x, y, and z axes. The total number of atoms in the protein is represented using, n. The RMSD backbone atoms were calculated for all simulated systems with the initial coordinate set at the reference to

evaluate stability of the system over the simulation. Proteins of larger size may have fluctuation due to changes in domains.

MM-PBSA/GBSA Estimation of Binding Energies

Data obtained from the simulation can be used to estimate the free energies of binding using the Molecular Mechanics-Poisson Boltzmann Surface Area (MM-PBSA) method and interaction energies using the Molecular Mechanics Generalized Borne Surface Area (MMGBSA) method. MMPBSA and MMGBSA methods combine MM with continuum models to analyze a series of snapshots taken from the simulation trajectory [70]. The thermodynamic equations used to calculate the free energy of a system are expressed in Equations 4 and 5.

$$\Delta G = \Delta H - T\Delta S$$

Equation 4

$$\Delta Gtot = \Delta E_{MM} + \Delta G_{sol} - T\Delta S$$

Equation 5

ΔG is the free energy, ΔH is the enthalpy, and ΔS is the entropy at a certain temperature T. The enthalpy of binding is composed of the gas phase energies and solvation free energy, ΔE_{MM} and ΔG_{solv} respectively. This expansion of the enthalpic term in show in Equation 5 and solved in this method. The binding energy of a ligand in a complex is calculated by summing the

free energy of the receptor and ligand, then subtracting it by the free energy of the complex as shown in Equation 6.

$$\Delta Gbind = \Delta Gcom - (\Delta Grec + \Delta Glig)$$

Equation 6

It is ideal to calculate the free energy using the solvated models of the three subunits (complex, receptor, and ligand) as shown in **Figure 2.2**, then use equation 6 to determine the free energy of binding [71]. This calculation takes a large amount of time to converge so MMPBSA approximates the calculation by breaking it down.

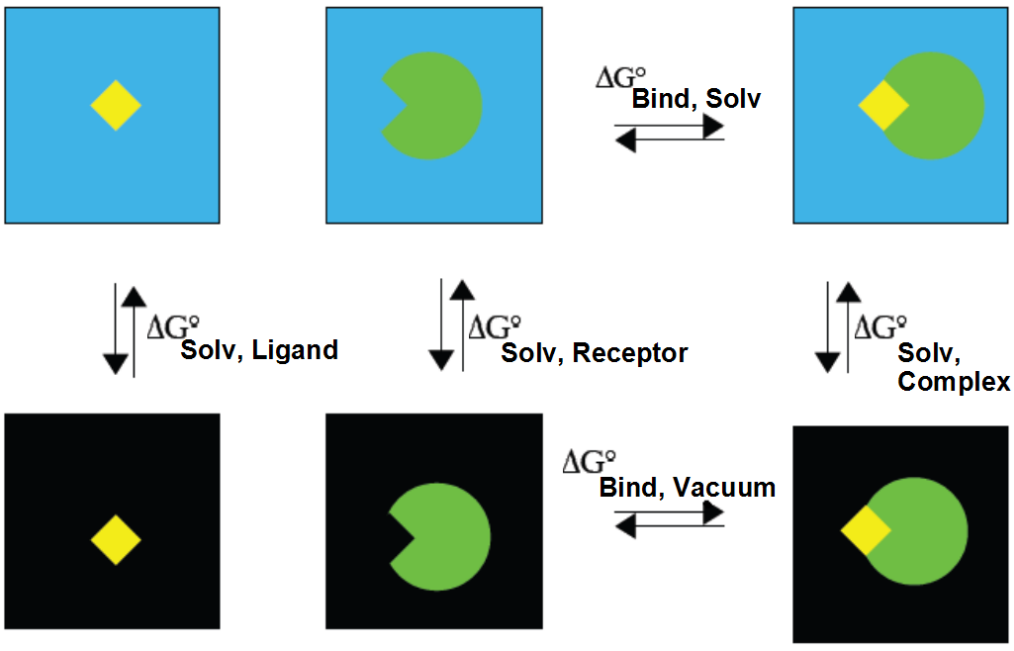


Figure 2.2. The approximation of MMPBSA by calculating gas phase energies (black) and solvation energies (blue).

MMPBSA calculates the gas phase energy of the complex, receptor and ligand using MM based on the AMBER forcefield consisting of van der Waals and electrostatic energy. The solvation energy, ΔGsol, includes the electrostatic and nonpolar interactions. The electrostatic portion is calculated using the Poisson-Boltzmann (PB) or Generalized Borne (GB) method. The nonpolar contribution is estimated by the solvent-accessibility surface method. The entropy contribution to the free energy, TΔS, can be estimated using normal mode analysis. In this study the entropic contribution was not calculated because it is computationally expensive and can have large error bars [72, 73]. The approximation can be used for relative comparison of binding energies of similar ligands binding to the same area of the protein.

For the binding free energy calculations in this work approximately 250 snapshots were used from the trajectories to get a series of structures. For all ligand-calcium binding estimations MMPBSA method was used. For residue-residue interactions the MMGBSA method was used, which includes an approximation to the PB equation and includes implicit water for solvation [74]. MMPBSA and MMGBSA have been tested to give good approximations for binding energies and are correlated with actual binding energies [73].

2.3 Nudged Elastic Band Method

In this work we use the nudged elastic band method (NEB) to obtain coordinates to map out the conformation change of annexin A1. The NEB method is implemented in the AMBER package [75]. It is used to calculate low energy paths for conformational changes. This method is novel in the fact that it does not require a hypothesis for the potential change in conformation. NEB could be a good method to use for large proteins such as annexin A1. This method uses a timescale-independent method to examine conformation change. MD simulations are limited to nanosecond timescales. This is unfavorable because many vital protein conformations tend to occur at a microsecond or millisecond timescale. Hence the NEB method can be favorable to detect such conformational changes in a timely manner.

The NEB method was derived from the plain elastic band method introduced by Elber and Karplus [76]. This method included a spring potential to the energy surface and minimized the total energy for the system. In the plain elastic band method low-energy paths are determined but corners tend to get cut in the energy landscape. Implementation of the NEB method avoids cutting corners by applying spring forces perpendicular to the tangent of the path [77].

The conformation change in NEB is determined using a series of protein structures distributed along the path, which represents the trajectory that the protein undergoes during conformational change [78]. The structures located at the endpoints of the pathway are representative of the reactant and product structure. These structures are fixed in space and do not explore the energy landscape. Each structure along the path is connected by springs, which serve to keep the structures from sliding down the energy landscape. If structures were to slide on the landscape, they can slide onto adjacent structures in the pathway. To keep structures

evenly spaced, forces from the molecular potential are condensed along the path. This assists in keeping a smooth conformation change pathway.

The path for conformation change is approximated using a series of structures in the NEB calculations. Forces applied to each structure are governed by virtual springs. The springs serve to fix the position of each structure along the path in reference to its adjacent structures. This allows the structures to interact, so they follow a path and remain spaced along this path. The total force exerted on image, i, in the pathway are composed of two components, a perpendicular and parallel force, as show in equation 7.

$$F_i = F_i^{\perp} + F_i^{\parallel}$$

Equation 7

The perpendicular and parallel forces are represented by their respective equations, shown in equation 8 and 9.

$$F_i^{\perp} = -\Delta V(P_i) + \Big(\big(\Delta V(P_i)\big)\tau\Big)\,\tau$$

Equation 8

$$F_i^{\parallel} = \big[\big(k_i + 1(P_{i+1} - P_i) - k_i(P_i - P_{i-1})\big)\tau\big]\tau$$

Equation 9

F_i represents the force exerted on each image i, P_i is the 3N-dimensional position vector of image i, k_i is the spring constant between image i-1 and i. The potential described by the forcefield is noted using V, and the 3N dimensional tangent unit vector that described the path is noted by τ. The force for the endpoint structures are set to zero, so they remain in the same conformation [77].

The spring constants can be the same between each image or can be scaled by using climbing image to move the images closer together when reaching transition states. For this work we use the same spring constants, and do not turn on climbing image. Many NEB calculations are performed using the generalized Born implicit solvation model which has been shown to approximate explicit solvation calculation of nucleic acids and proteins [79-81]. Simplifying the calculation by using the implicit solvation model reduced computation time and complexity of the path space. For our NEB simulation we used the implicit solvation model.

Steps Required to Run NEB

1) Minimization

Before a NEB simulation can begin, minimization of the protein systems must be performed. Before running the NEB calculation, the annexin proteins must be in an optimal conformational representation. Topology and coordinate files for the corresponding structures were used during the minimization to calculate the optimal structure. For minimization an optimal structure is one that has been relieved of as many negative or strained contacts as possible. This ensures the structure is in a relaxed and stable conformational state. For this work we were determining whether the NEB method is reliable, and if we could obtain consistent

results over multiple runs. Thus, the number of cycles for minimization for each run was different and are specified in the simulation setup of chapter three.

2) Heating

The first step for the NEB calculation is heating the system from 0 to 300K. For the NEB calculations performed in this work, the simulation setup allowed for a total of 32 images. This includes the starting (apo) and ending (calcium bound) structure, and 30 images between expressing the changes that occur going from an apo to a calcium bound state. To get a total of 32 images, the first 16 string of images or coordinates for the heating step used the calcium free structure obtained from the minimization step. The last 16 images or coordinates used the calcium bound structure from the minimization step. As the NEB heating calculation is performed the images are changing to adapt to the potential energy pathway for the conformation change. Specific simulation parameters for the heating stage of NEB are described in more detail in chapter three, and each simulation used differing parameters.

3) Equilibration

The NEB equilibration is performed after heating. The last coordinates obtained in the heating step is used for the initial coordinates of the equilibration. At this step the system is held at an isothermal temperature of 300K. The equilibration is very important because the coordinates obtained from the heating step were the best resolved images during the simulation, however allowing equilibration of the coordinates ensures the transition from the heating to the

annealing step is smooth. Each simulation used different parameters for the equilibration step, and the parameters are described in chapter three.

4) Annealing

The final coordinates obtained in the equilibration step are the initial coordinates used for the annealing NEB calculation. In this step of the NEB simulation the system is heated from 300K to 500K and back down to 300K. This step allows the structures to explore the energy landscape and find those structures that will later be isolated and examined in the potential conformational pathway.

In each simulation the annealing calculation was the longest step. This is important because there should be optimal time allowed for each image of the conformation pathway to explore the energy landscape. The more steps completed for the annealing process allows more calculations to be completed and can ensure that the coordinates obtained from the annealing are those that are accurately representing conformational change along the energy landscape. Specific parameters for the annealing step for the three simulations are described in chapter three.

5) Cooling

The final step of the NEB calculation is the cooling step, which cools the system from 300 K to 0 K. In this calculation the last coordinates obtained from the annealing step are used as the initial coordinates for cooling. This step of the NEB method cools the system down, so the

images freeze along the potential energy pathway. Cooling is performed slowly so the images frozen along the pathway should have the lowest potential energy. At this stage the energy associated with each image of the conformation pathway and corresponding structures can be analyzed. The parameters used for the cooling step for each simulation are described in chapter three.

CHAPTER 3. RESEARCH PROJECT 1: MOLECULAR DYNAMICS STUDY ON CALCIUM INDUCED CONFORMATION PATHWAY FOR ANNEXIN A1

3.1 Purpose

The purpose of this study was to investigate the conformation pathway for annexin A1 going from an apo to a calcium bound state. Using the NEB method, we provide a potential conformation pathway giving insight into the process of the N-terminal domain of annexin A1 removing itself from the core, and other cooperative conformation changes. This process of N-terminal removal and conformation change has not yet been described at an atomistic level. We will also determine if the NEB method is reasonable in determining conformation change for large protein systems. NEB has not yet been used for proteins of this size (343 residues). By studying the structures produced after the NEB simulation we hope to provide insight into the conformation change.

3.2 System Preparation

The initial coordinates for annexin A1 were obtained from the X-ray crystal structure of full length annexin without the presence of calcium (1HM6.pdb) [11]. These coordinates were used as the reactant for the NEB simulation. There are no crystal structures available for full length annexin A1 with the N-terminal exposed outside of the core domain. Both structures were needed for this work to study the conformation change from the apo to calcium bound annexin structure. To obtain the calcium bound structure, the calcium bound core domain were obtained

from the PDB X-ray crystal structure (1MCX.pdb) and the missing 41 residues of the N-terminal domain were fused to the core domain using Insight II software [82]. This construction was part of previous molecular dynamics work and was available at the start of this study [83]. Both conformations of annexin have 343 residues and are shown overlapped in **Figure 3.1**.

The crystal structures for annexin in the calcium free and calcium bound states were not solvated with an explicit water model. NEB runs the simulation in an implicitly solvated system so there is no explicit water model added around the protein structures. Three NEB simulations were performed using varying parameters. The steps for performing NEB as well as the parameters used in each simulation are detailed below.

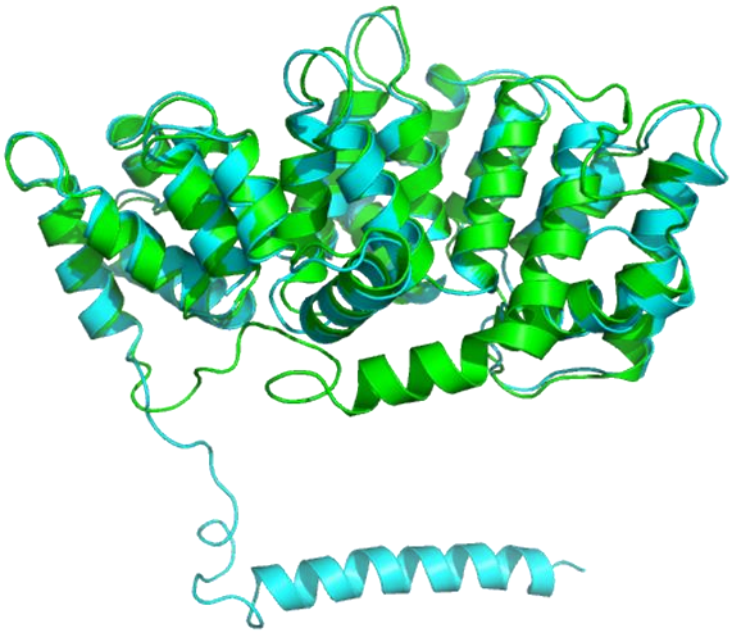


Figure 3.1: X-ray structures of annexin AI calcium free and calcium bound state used in the NEB simulation. The calcium free (green) and calcium bound (cyan) structures represent the reactant and product for the pathway.

Minimization

In this work minimization was completed using a steepest decent and conjugate gradient method. Two minimizations were performed for each simulation, the first minimization is run holding all atoms fixed except for hydrogen (min0), and the second minimization is run without holding any atoms fixed (min). The number of steps were varied to ensure that slightly different starting structures were used for each simulation. **Table 3.1** expresses the different minimization cycles that were allowed for each of the three NEB simulations. The first simulation totaled 8,500 cycles of minimization, while in the second and third simulation minimization cycles totaled 12,000 and 14,000 respectively. In theory the longer a minimization step is ran, the more energetically stable it can be. Once the minimization is complete, the structures obtained are used in the first step of the NEB calculation, which is the heating step.

Heating

Figure 3.2 expresses an example of the initial potential energy for each image and correlates with the first 16 structures of the reactants coordinates and last 16 structures as the products coordinates. In this simulation the reactant is the calcium free structure and the product is the calcium bound structure of annexin A1.

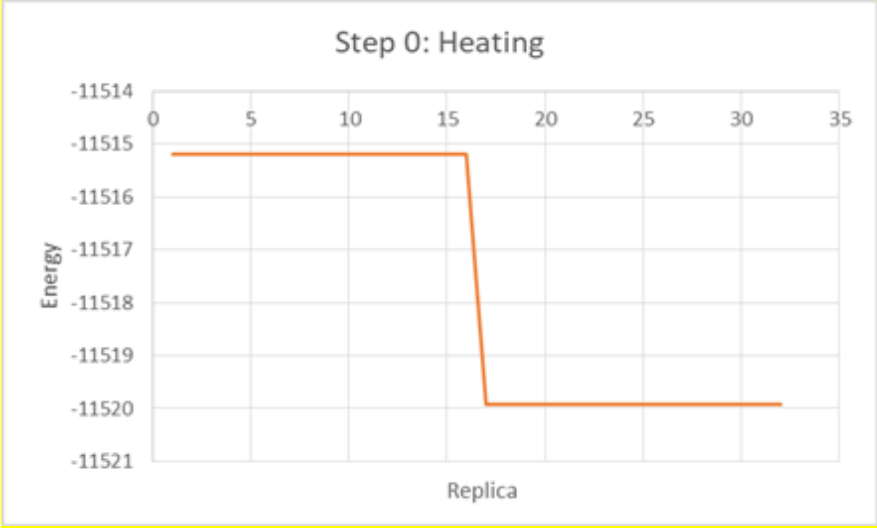


Figure 3.2. Example of the energy diagram at step 0 for the first implicit simulation of annexin A1.

Similar to the minimization step, each simulation performed used different simulation parameters and is expressed in **Table 3.1**. For the first and second implicit simulations the parameters for the number of steps performed, the time step, and the spring constant were 40,000, 0.5 fs, and 10 kcal/mol$Å^2$. The third simulation had alterations to the time step and was increased to 1.5 fs. The heating step for the NEB simulations performed in this work had similarities in many of the simulation parameters, however in the proceeding steps for equilibrium, annealing, and cooling more changes were made in the parameter set up.

Equilibrium

For all three NEB simulations the different simulation parameters used are expressed in **Table 3.1**. Variations in step size for the simulations range from 100,000 to 200,000 steps, and the time step ranged from 1.0 to 1.5 fs, while all three simulations used different spring constants

of 10, 20, or 30 kcal/molÅ^2. Allowing more changes to these parameters allows for comparison of the methods used once all three simulations are completed.

Annealing

Similar to the previous NEB simulation parameters the step size, time step, and spring constants were varied across the three simulations. However, for the annealing calculation there was drastic changes in the set up. Specific changes for each simulation are shown in **Table 3.1**. The number of steps allowed for the simulations ranged from 420,000 to 900,000, while the time step range from 0.5 to 1.5 fs, and the spring constants used ranged from 10 to 50 kcal/molÅ^2. In comparison to the steps for NEB that were previously completed, the annealing step has the largest change in parameters. This variation will allow us to determine if the conformation change of annexin A1 is consistent when using large changes in simulation parameters.

Cooling

In the cooling step for each NEB simulation there were variations in the parameters that were used and are specifically show in **Table 3.1**. For the cooling step all simulations used 120,000 steps as the step size. Alterations in the time step ranged from 1.0 to 1.5 fs, and the spring constants used were either 50 or 10 kcal/molÅ^2. The first two simulations used the same parameters, and the third simulation used different time step and spring constant values. The coordinates produced from this step of the simulation were used in analysis for the conformation change of annexin A1.

Table 3.1. NEB simulation parameters for all three implicit water simulations. Differing parameters were used for the minimization cycles, number of steps performed, time step (DT), and spring constant (k).

	IMP 1	IMP 2	IMP 3
MIN0: CYCLES	3500	6000	7000
MIN: CYCLES	5000	6000	7000
HEAT			
NUM STEPS	40,000	40,000	40,000
DT (fs)	0.5	0.5	1.5
k (kcal/molÅ²)	10	10	10
EQUILIBRATION			
NUM STEPS	100,000	100,000	200,000
DT (fs)	1.0	1.0	1.5
k (kcal/molÅ²)	20	30	10
ANNEALING			
NUM STEPS	420,000	900,000	800,000
DT (fs)	0.5	1.5	1.5
k (kcal/molÅ²)	50	20	10
COOLING			
NUM STEPS	120,000	120,000	120,000
DT (fs)	1.0	1.0	1.5
k (kcal/molÅ²)	50	50	10

As previously discussed, different parameters were used for each step of the NEB simulation for each run. This was important because we are trying to determine the conformation change in the pathway from the calcium free structure to the calcium bound structure. Using

different number of steps, spring constant, and time steps will give variation in the simulations. This will ensure slight variation in the structures being used in each proceeding step.

Increasing the number of steps in any part of the NEB calculation allows more time for the simulation to run and gives more opportunity for the minimum potential energy pathway to be determined. In combination with the number of steps the calculation can run, altering the time step can also have an effect on obtaining a smooth pathway. A larger time step will give more time for each step to be performed. For example, if a 0.5 fs time step is used each step will be performed much faster than if a 1.5 fs time step is used. Using a larger time step will also increase the length of time that the simulation was ran. The last parameter that was altered was the spring constant, which correlates with the spring force for the NEB calculation. In different steps of the NEB simulation there were different spring constants used. The smallest spring constant used was 10 kcal/mol$Å^2$, while the largest spring constant used was 50 kcal/mol$Å^2$. For all NEB steps the spring constants were applied to the entire protein. Different spring constants were applied to examine if smaller spring constants would be able to produce the conformational changes expected, or if larger spring constants would be needed to force the change to occur.

After all simulations were completed, comparison of the potential energy pathways and structural conformations were performed. If they are similar, then we can suggest that this is a reasonable prediction of the conformation change. Performing multiple calculations also gives insight into the reproducibility of the NEB method.

3.3 Results and Discussion

Potential Energy Plots

Three simulations totaling 121 ns were performed to determine the conformation change of annexin A1 from an apo to a calcium bound state. This conformation change includes N-terminal domain removal from the third repeat of the core domain, which is a vital step required for the completion of many biological functions. The minimized coordinates for each annexin structure were incorporated in the NEB and all 32 images along the conformation pathway were used to map out the reaction pathway and track conformational changes. Tabulated data for the energy associated with each image along the conformation pathway are provided in **Tables 3.2-3.4.** For each simulation the energy associated with each image is displayed on a potential energy plot. These plots are used to connect the potential energy of each image on the conformation pathway. Connecting the images gives insight into the potential energy of each structure along the pathway. The theoretical potential energy barrier needed to go from a calcium free to a calcium bound state of annexin A1 and can be determined by looking at the energy plot and determining the structures that lie at the highest energy.

A. First Implicit Simulation

The first simulation for annexin A1 totaled 14.4 ns. This simulation time includes the total simulation of all 32 images over all NEB steps. In **Table 3.2** each structure along the NEB pathway and their corresponding energy values are presented. The initial calcium free structure (image 1) has a potential energy of -11515.1901 kcal/mol and the final calcium bound structure (image 32) has a potential energy of –11519.9307 kcal/mol. These structures have the lowest

potential energy along the conformational pathway. The structure that has the highest potential energy long the pathway is image 14. This structure has a potential energy of -11198.9909 kcal/mol. For the first implicit NEB simulation the range in potential energy for the structures along the conformation pathway were 320.9398 kcal/mol. The potential energy plot for this simulation is expressed in **Figure 3.3**. The energy plot gives more insight into the visual fluctuations in the energy for each structure.

Table 3.2. Corresponding potential energy for each structure along the NEB conformation pathway for the first implicit simulation totaling 14.4 ns.

Structure/ Image Number	Potential Energy (kcal/mol)
1	-11515.1901
2	-11431.8254
3	-11399.9598
4	-11364.4609
5	-11379.5926
6	-11351.7213
7	-11308.0427
8	-11322.7243
9	-11305.7723
10	-11321.1917
11	-11280.7855
12	-11212.3516
13	-11204.4432
14	-11198.9909
15	-11199.7581
16	-11215.0124
17	-11201.7992
18	-11201.7992
19	-11215.3753
20	-11202.4479
21	-11236.6885
22	-11282.9965
23	-11337.5354

24	-11350.8852
25	-11409.7519
26	-11432.6191
27	-11446.6358
28	-11476.8354
29	-11480.2105
30	-11502.6681
31	-11472.2379
32	-11519.9307

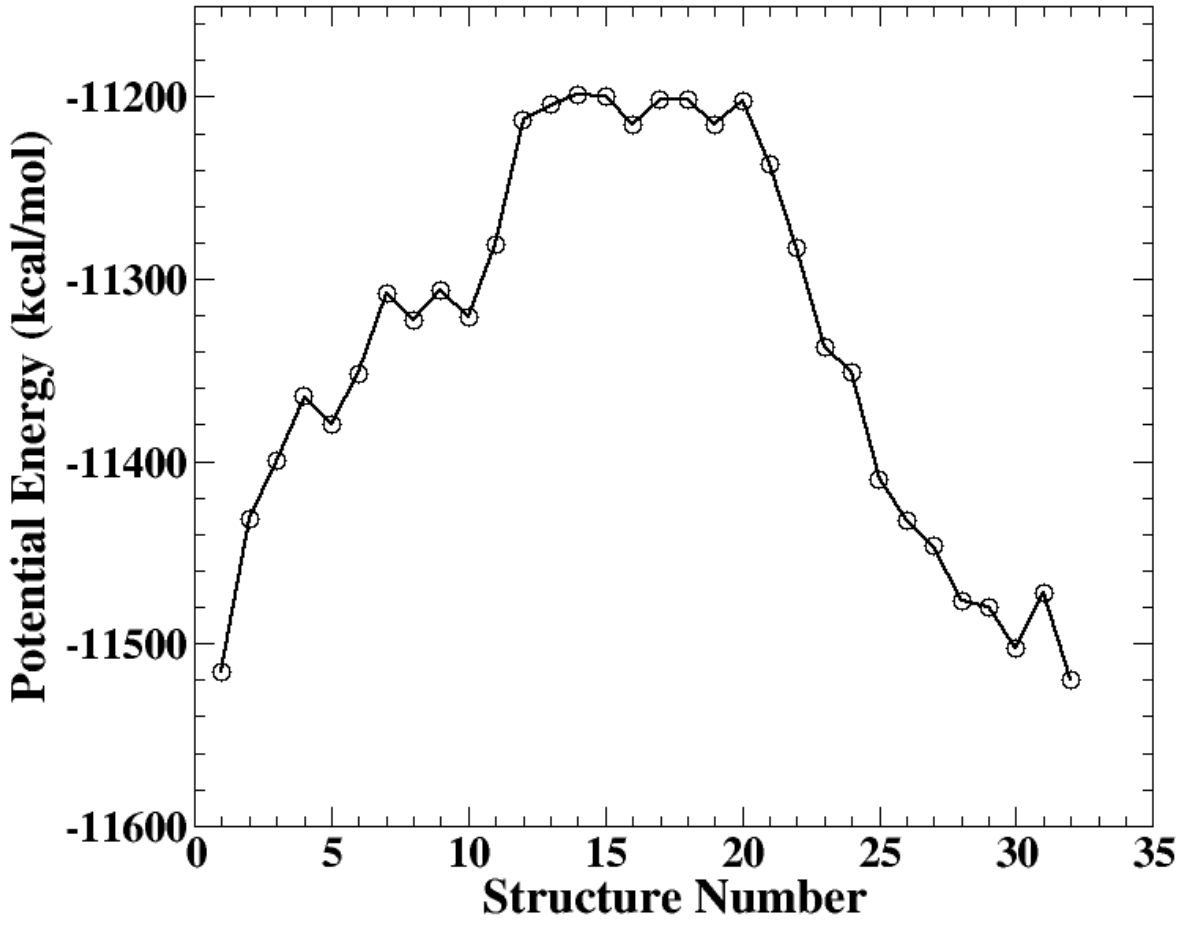


Figure 3.3. The potential energy plot for the 14.4 ns implicit NEB simulation of annexin A1 conformation change over all 32 images.

In the potential energy diagram, we see visual representation of the structures that have the lowest and highest potential energy. Image 1 and 32 were minimized prior to beginning the NEB simulation and were held fixed during the NEB simulation. As the NEB simulation progressed the 30 images in-between the initial and final images explored the energy landscape. **Figure 3.3** gives a clear representation of the changes in potential energy for the conformation change of annexin A1.

The shape of the potential energy curve is representative of what is expected. Going from two stable conformations (calcium free and calcium bound structure) we expect to see these structures having more stability, with lower potential energies. Both the initial and final structure along the pathway have the lowest energy values. As the simulation is run the structures mapping out the conformation change are spread uniformly over the energy plot. There are no major spikes in energy for structures in close proximity. **Figure 3.3** also gives some insight into the distribution of the structures along the energy plot. When looking at **Table 3.2** it is hard to see where these structures are by simply looking at the values. However, with the energy plot we are able to see that 9 structures in the middle of the conformation change pathway have very similar potential energies. Having so many structures lying at the top of a potential energy curve like this are most likely due to structural similarities. These structures are all located on the highest point of the energy plot and are less stable in energy than other structural conformations.

For the first implicit NEB simulation of annexin A1 we are see a potential energy curve that is representative of conformation change. Based on the shape of the curve and the distribution of the images along the energy landscape, these results were promising and lead to interest in structural studies.

B. Second Implicit Simulation

The second simulation of annexin A1 totaled 50.9 ns. This simulation was run using more steps and larger step sizes as shown in **Table 3.1**, thus contributing to a longer simulation run time we are able to compare results. Similar to the first simulation, the second simulation was run using a total of 32 images. The corresponding potential energy values for each structure are shown in **Table 3.3**. Similar to the first simulation, the second simulation also had the initial and final structures of the pathway as the lowest in potential energy at -11520.3038 kcal/mol and -11522.2444 kcal/mol respectively. In the second simulation the structure with the highest potential energy was structure 14. This is also comparable to the first simulation. Structure 14 had a potential energy of -11285.1325 kcal/mol for the second simulation.

Table 3.3. Corresponding potential energy for each structure along the NEB conformation pathway for the second implicit simulation totaling 50.9 ns.

Structure/ Image Number	Potential Energy (kcal/mol)
1	-11520.3038
2	-11492.2710
3	-11473.1044
4	-11433.4963
5	-11440.1659
6	-11436.6439
7	-11441.1529
8	-11409.4850
9	-11383.8401
10	-11375.8209
11	-11313.0959
12	-11338.4972
13	-11333.6301
14	-11285.1325
15	-11285.7452

16	-11349.2049
17	-11303.8384
18	-11328.8831
19	-11300.555
20	-11302.8930
21	-11334.7913
22	-11361.9541
23	-11412.9971
24	-11443.7144
25	-11478.9653
26	-11461.6903
27	-11460.3858
28	-11467.2537
29	-11472.7885
30	-11445.2369
31	-11447.0134
32	-11522.2444

Comparing the potential energy values for both simulations gives insight into the reproducibility of the NEB method. For the initial structure (image 1) there is approximately a 5 kcal/mol difference in energy, while for the last structure (image 32) there is a 2 kcal/mol difference. This is very similar and suggests that the structures used for both simulations are very similar. In both simulations the structure with the highest potential energy was structure 14. However, the difference in energy for these structures are approximately 86 kcal/mol. This difference is much larger than those found for the initial and final structures. This could be due to the fact that these are different structures along the NEB pathway, and due to the variation in NEB parameters different structures are produced for the same image number.

The potential energy plot for the second implicit simulation is shown in **Figure 3.4**. This figure gives more insight into the location of the structures along the potential energy curve. In the second implicit simulation the initial and final coordinates have the lowest potential

energy. This is comparable to the first simulation where the range in potential energy for the structures are similar. The potential energy curve for the first and second simulation are somewhat similar in shape, however there are a few differences. In the second simulation, there is not as much of an even distribution in energy as the previous simulation. The second simulation, **Figure 3.4,** shows a more rigid increase in energy. At the top of the potential energy curve there is no leveling of energy, however the energy values for these structures act as a pulse where the structures are alternating in lower and higher energy fluctuations.

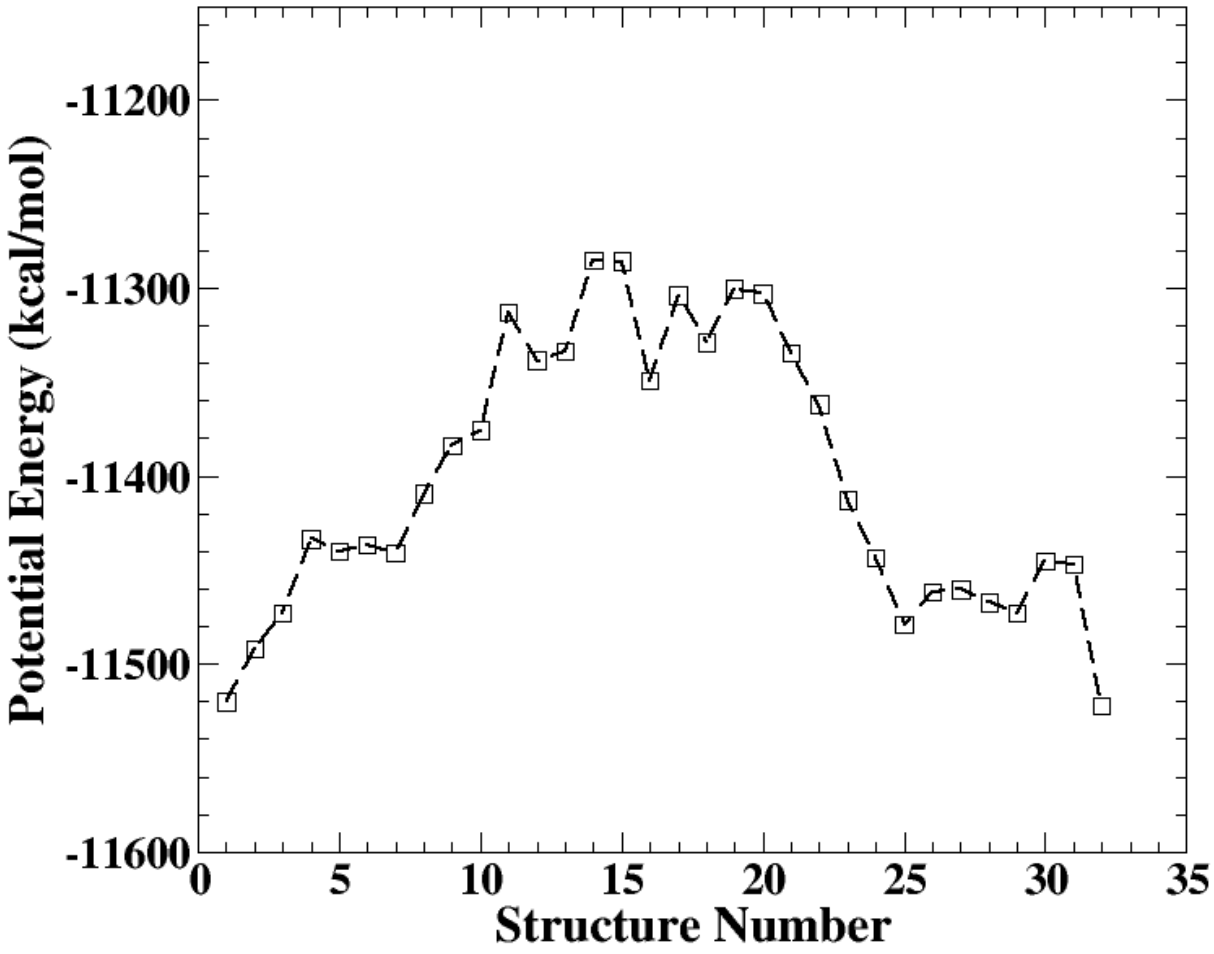


Figure 3.4. Potential energy plot for the second implicit NEB simulation of annexin A1 conformation change totaling 50.9 ns.

The largest differences in the parameters for the first and second simulation are the equilibration and annealing steps. Changing the simulation parameters for the NEB calculations can affect the overall energy plot along the conformational pathway. While there are slight differences in the potential energy curve there are still enough similarities between the curves to suggest similar pathways are being produced.

C. Third Implicit Simulation

The last simulation of annexin A1 totaled 55.88 ns, which was the longest of the three simulations. Similar to the second simulation this one was run with a large number of steps and step size. The major difference between the second and third simulation, is the spring constant. In the previous simulations, variations to the spring constant were performed at each step. However, for this simulation a spring constant of 10 kcal/mol$Å^2$ was used for all NEB steps. These variations in the parameters are show in **Table 3.1**. The potential energy values for each structure along the third NEB conformation pathway is shown in **Table 3.4**.

Table 3.4. Corresponding potential energy of each structure along the NEB conformation pathway for the third implicit simulation totaling 55.88 ns.

Structure/ Image Number	Potential Energy (kcal/mol)
1	-11522.1979
2	-11484.6743
3	-11436.7344
4	-11432.3093
5	-11421.3033
6	-11398.2515
7	-11418.9161
8	-11384.7658

9	-11286.7162
10	-11263.5422
11	-11260.8694
12	-11263.5159
13	-11247.6662
14	-11248.7462
15	-11301.7596
16	-11258.6360
17	-11283.5969
18	-11254.0563
19	-11295.0921
20	-11288.0482
21	-11315.9894
22	-11318.5779
23	-11421.7549
24	-11537.9865
25	-11488.8256
26	-11475.5538
27	-11492.0789
28	-11490.7281
29	-11490.7712
30	-11447.7251
31	-11446.2226
32	-11525.8447

For the third implicit simulation the initial (image 1) and final (image 2) structures of the conformation pathway have some of the lowest potential energies of -11522.1978 and -11525.8447 kcal/mol. However, unlike the previous simulations for annexin A1 there is a structure that had a lower potential energy than the initial and final structures. This was image 24, which had a potential energy of -11537.9865 kcal/mol. The potential energy for image 24 is approximately 12 kcal/mol lower than that of image 32. It could be possible that image 24 for this simulation would be a better structure to use as the product due to the lowered potential energy. In the previous simulations the initial and final coordinates for annexin A1 had the

lowest potential energy values. Comparing the potential energy of the first image to the previous simulations there was a range of approximately 7 and 2 kcal/mol for the first and second simulation. The final image in the previous simulations had a range in potential energy of approximately 5 and 3 kcal/mol for the first and second simulation. In all simulations there is less variation in the range of potential energy of the final image. However, all simulations had very similar potential energy values for these structures and are considered to be consistent over all three simulations.

The image with the highest potential energy in the third simulation was 13. This structure had a potential energy of -11247.6662 kcal/mol. The range in potential energy values for the third simulation was 290.3203 kcal/mol. This range lied between the ranges for the previous simulations. By comparing the numerical values, all three simulations were similar in potential energy. The potential energy values were the closest for the second and third simulation which had more similarities in the simulation parameters.

The potential energy plot for the third implicit simulation is shown in **Figure 3.5**. The initial and final images on the NEB pathway have some of the lowest potential energy values, other than image 24. The potential energy curve resembles the curve for the second simulation.

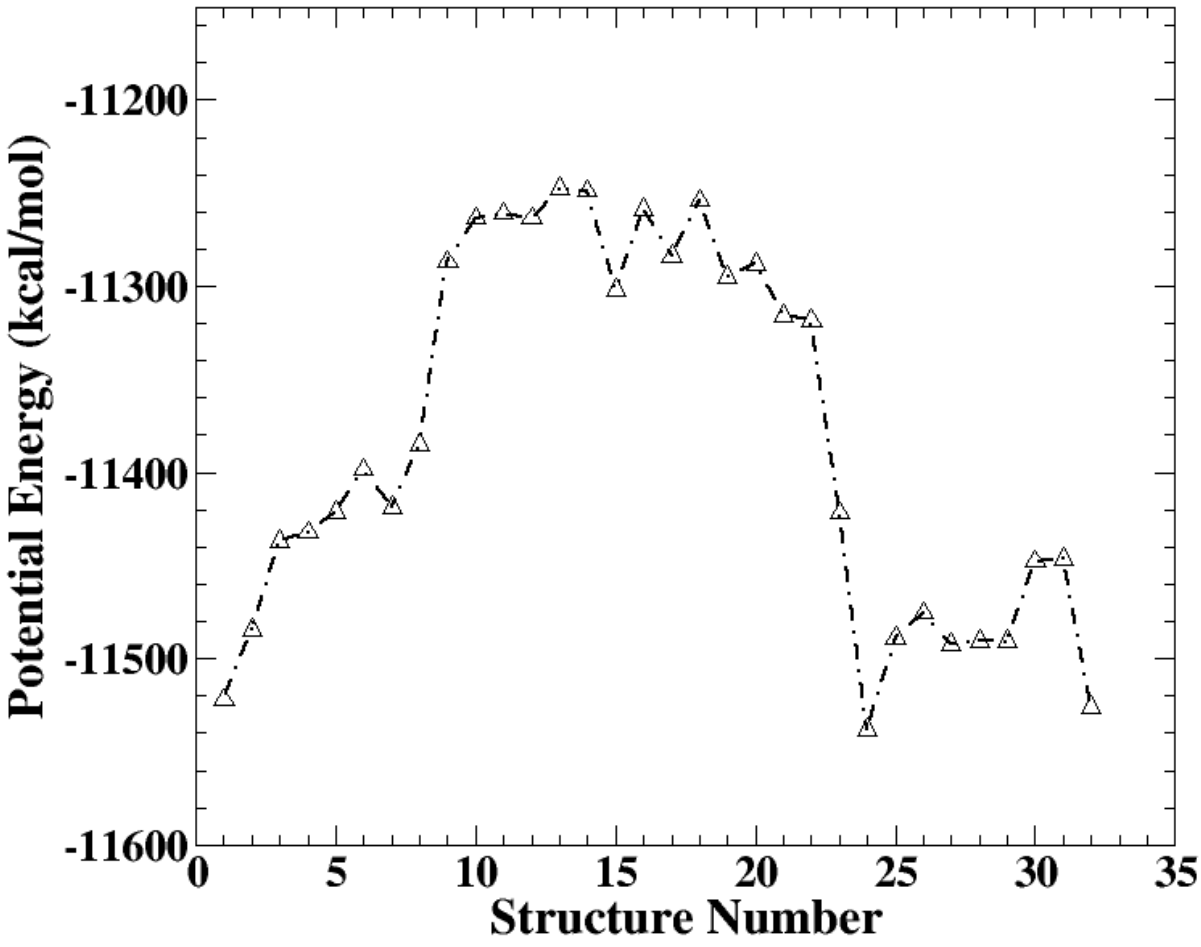


Figure 3.5. Potential energy curve for the third implicit NEB simulation for the conformation change of annexin A2 totaling 55.88 ns.

The third simulation has a similar fluctuation in potential energy at the top of the energy curve as the second simulation. There are 11 structures that are along this highest energy area. Structures 10-12 are along the peak of the energy curve, but do not fluctuate in energy as drastically as structures 14-19. This suggests that structures 10-12 are closer in energy and possibly have more conformational similarities than structures 14-19. Another difference between the second and third simulation are the structures at the end of the energy curve. Both

curves have the same phenomena of two small peaks in energy of structures 25-30. However, in the third simulation the peaks are more profound. Overlapping the three potential energy curves gives more insight into the differences for the three simulations.

D. Potential Energy Curve Comparison

Overlapping the three annexin A1 potential energy curves gives insight into the simulation parameters that lower the energy barrier going from a calcium free to a calcium bound state of annexin A1. **Figure 3.6** show the overlapping of the energy curves previously presented. The simulations that were run for a longer period of time showed a reduced barrier for the potential energy of the conformation change. The barrier was reduced by approximately 80 kcal/mol. The shape of the energy curves for all three simulations are similar. They have similar increases and decreases in energy for corresponding images, and similar image numbers and values sit at the top of the energy curve. This is promising information that can suggest similarity in all three simulations. The NEB method used in this work produced structures that seem to be similar based on their energy curves. While some simulations were able to lower the energy, with different parameters the overall conformation change that is occurring in all three simulations may be similar. To investigate the conformational similarities in more detail cluster analysis was performed.

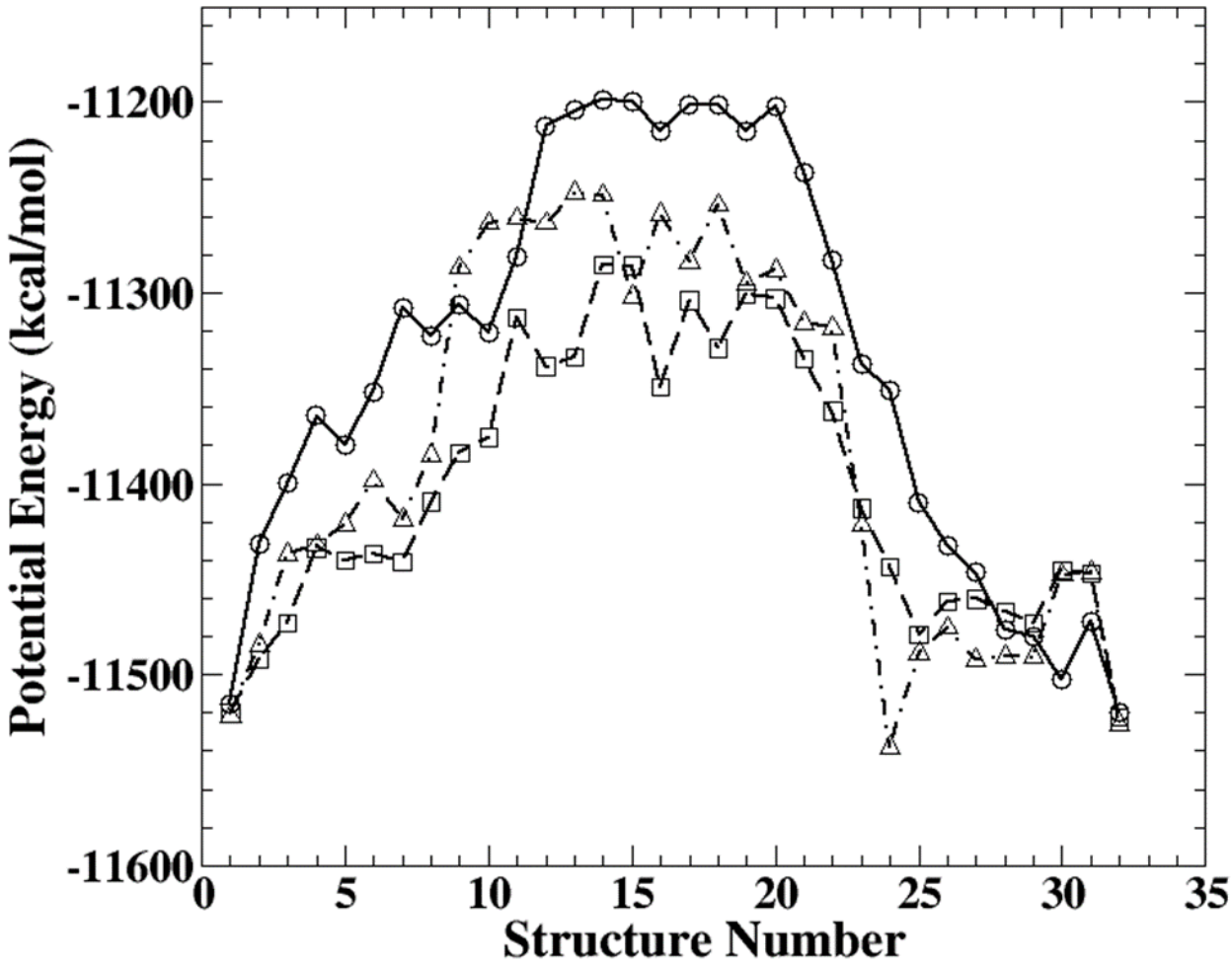


Figure 3.6. Potential energy diagram for each image of the implicit NEB simulation. All three implicit water simulations are present (simulation 1: circle, simulation 2: square, simulation 3: rectangle).

Cluster Analysis

For each simulation the 32 images that created the reaction pathway were clustered into eight representative images. Because some images are very similar to each other and there is very little conformation change, it is easier to examine conformational change over eight images. The clustering process groups images with similar conformations together. Those within a small RMSD difference are clustered into one group. All 32 images are examined, and eight clusters

were made. A single image for each cluster represents the conformation for the images of that cluster. The image that is chosen as the representative structure is the one that most accurately represents the characteristics of the protein for that stage of the conformation change.

A. First Implicit Simulation

For the first simulation we have separated the clusters with their images in reference to the potential energy curve, represented in **Figure 3.7**. Here we are able to visualize which structures are similar on the potential energy curve. Each cluster is separated by a vertical line on the potential energy curve. Cluster 4 and 5 have the most images associated with their clusters. This means at this portion of the conformation pathway there is less structural change going on. It is also interesting to note that these structures are located at the highest point on the potential energy curve. The location of the images for cluster 4 and 5 suggest that these structures are less energetically stable. Fifteen out of the thirty-two structures for the reaction pathway are located in these two clusters. This is just less than half the structures.

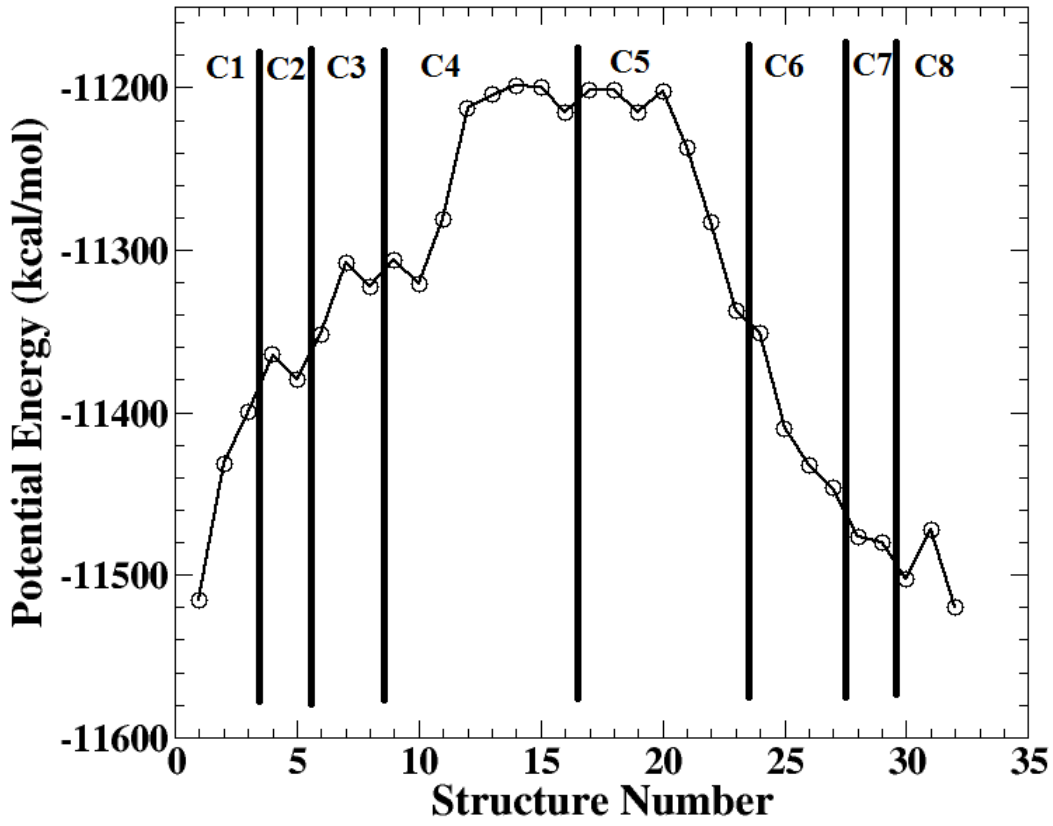


Figure 3.7. Clustered images for implicit NEB simulation 1, overlapping the potential energy curve.

B. Second Implicit Simulation

Figure 3.8 expresses the results obtained after cluster analysis of the second simulation for the images along the conformation pathway. Comparing the location of the lines associated with the cluster it is clear that they are not exactly the same as the clusters of the first simulation. In **Figure 3.8** there are less images represented in the first four clusters, while in the last four there are more images represented. This clustering suggests that the second simulations conformational change might occur quicker for the first half of the pathway than the second. Similar to the first simulation the images located at the top of the potential energy curve are

grouped into larger clusters. In the second simulation cluster 5 and 6 are located along the top of the potential energy curve. These clusters incorporate 14 out of the 32 images.

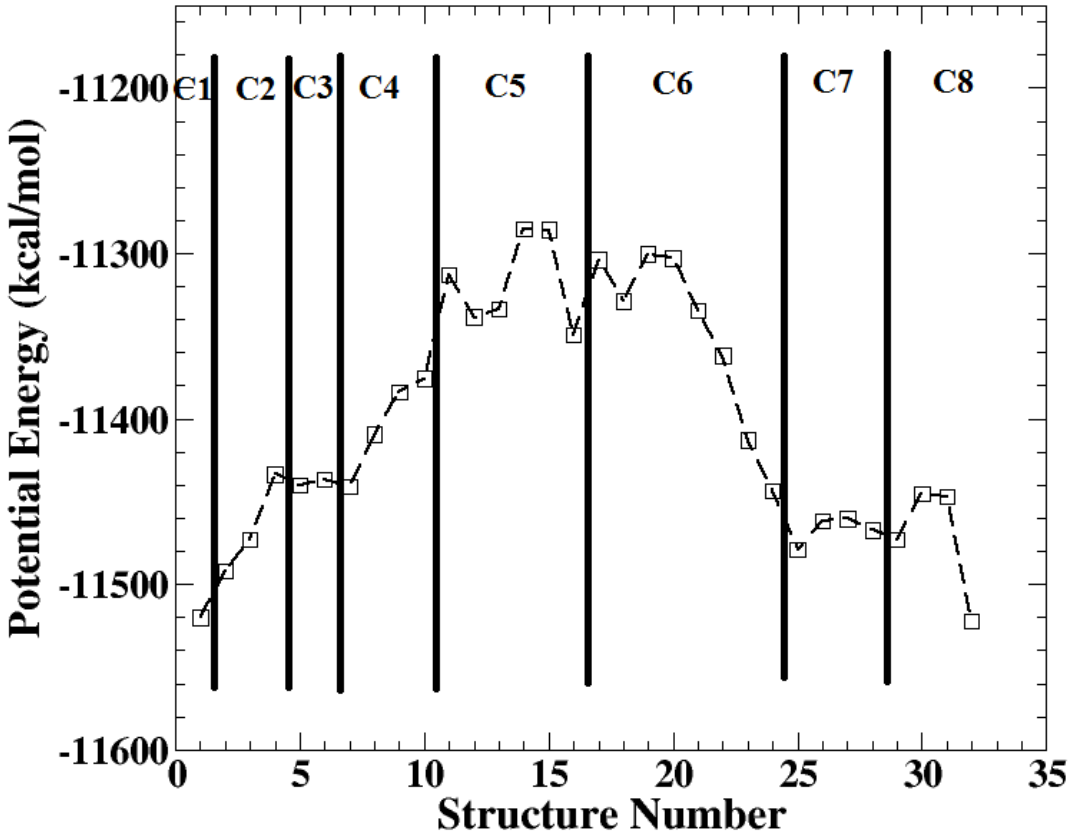


Figure 3.8. Clustered images for implicit NEB simulation 2, overlapping the respective potential energy curve.

C. Third Implicit Simulation

In the final simulation the cluster analysis looks like the opposite of the second simulation. **Figure 3.9** shows the structures represented in each cluster. Here the first four clusters have more structures represented than the last four. This suggested that for this simulation there is more similarity to the structures of the first half of the simulation than the last. In this simulation cluster 3 and 4 are located at the top of the potential energy curve. These

clusters incorporate 14 out of the 32 images in the pathway. Similar to the previous reactions those images with the highest energy have more similarities in structure.

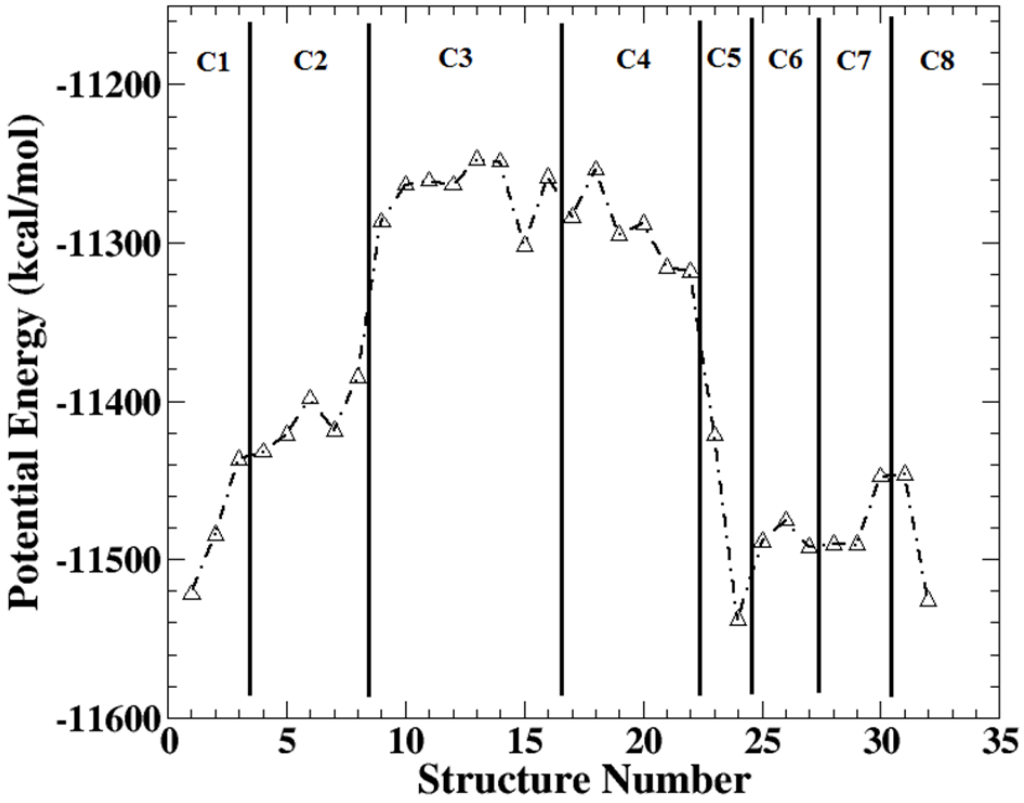


Figure 3.9. Clustered images for implicit NEB simulation 3, overlapping the respective potential energy curve.

There are similarities in cluster analysis over all three simulations. In all three simulations structures with the highest potential energy are divided into two clusters. These clusters incorporate more structures than other clusters along the pathway. For all simulations approximately fourteen structures are included in these clusters, which is almost half of the structures along the conformation pathway. The NEB simulations are run using differing parameters and starting structures, so we do not expect to see clusters that are exactly the same.

Cluster analysis of the images in the simulations simplify the number of structures that need to be used during analysis of the conformation pathway. Without performing cluster analysis, it would not be clear which structures are similar to each other. Cluster analysis selects a representative structure from each cluster. Instead of having to compare 32 structures for each simulation, we are now able to compare 8. Due to the method used for statistical clustering of the structures there is confidence that the representative structures chosen are sufficient for analysis. Comparing the clustered images for each simulation will give insight into the consistency of the conformation pathway using different NEB simulations.

Structure Overlap of the Representative Clusters

Cluster analysis has provided eight simplified structures to use for analysis of the conformation change. Before performing detailed analysis of the conformation change, it is important to determine if the conformation change over all three simulations are consistent. To do this each representative cluster along all simulations are compared.

Figure 3.10 overlaps the representative cluster 1 structures for the three simulations. The core domain in all three overlap nicely and do not have much variation. There are slight differences in the location of the loop region connecting the third and fourth repeat of the core domain. In the structures for cluster 1 the major differences are in the linker between the N-terminal domain and the core domain. This linker is flexible due to the loop secondary structure and has freedom to move in all simulations. Overall the overlapped structures are very similar to each other.

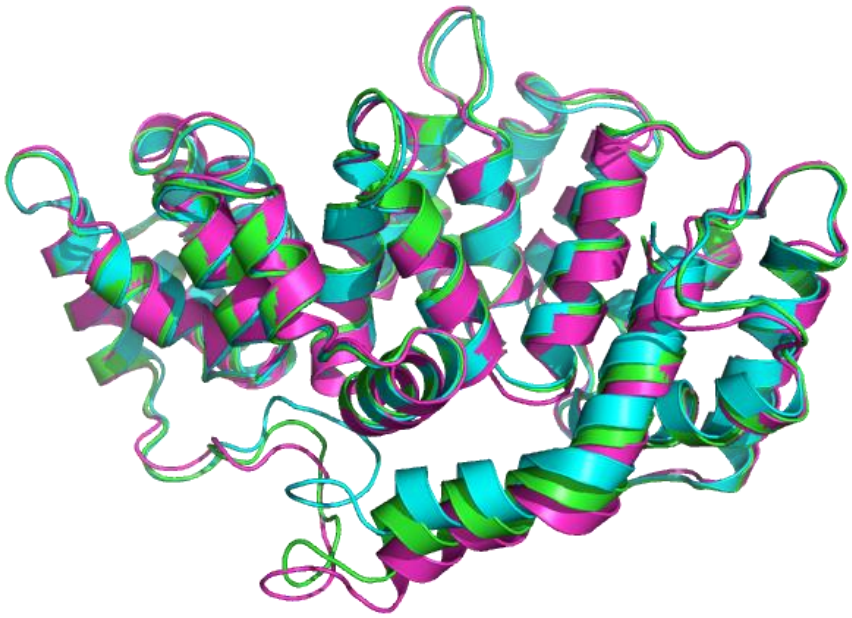


Figure 3.10. Overlap of representative cluster 1 structures for all three implicit NEB simulations (1: green, 2: cyan, and 3: magenta).

The second representative clustered structures also have a lot of similarities, as shown in **Figure 3.11**. In these structures the core domain overlaps without much difference. The loop region connecting the third and fourth repeat of the core domain align better than the first cluster. The clustered structure for the third simulation has an N-terminal that is slightly lower than the other two and has structure loss in the center of the N-terminal helix. Overall the structures for the second cluster are similar.

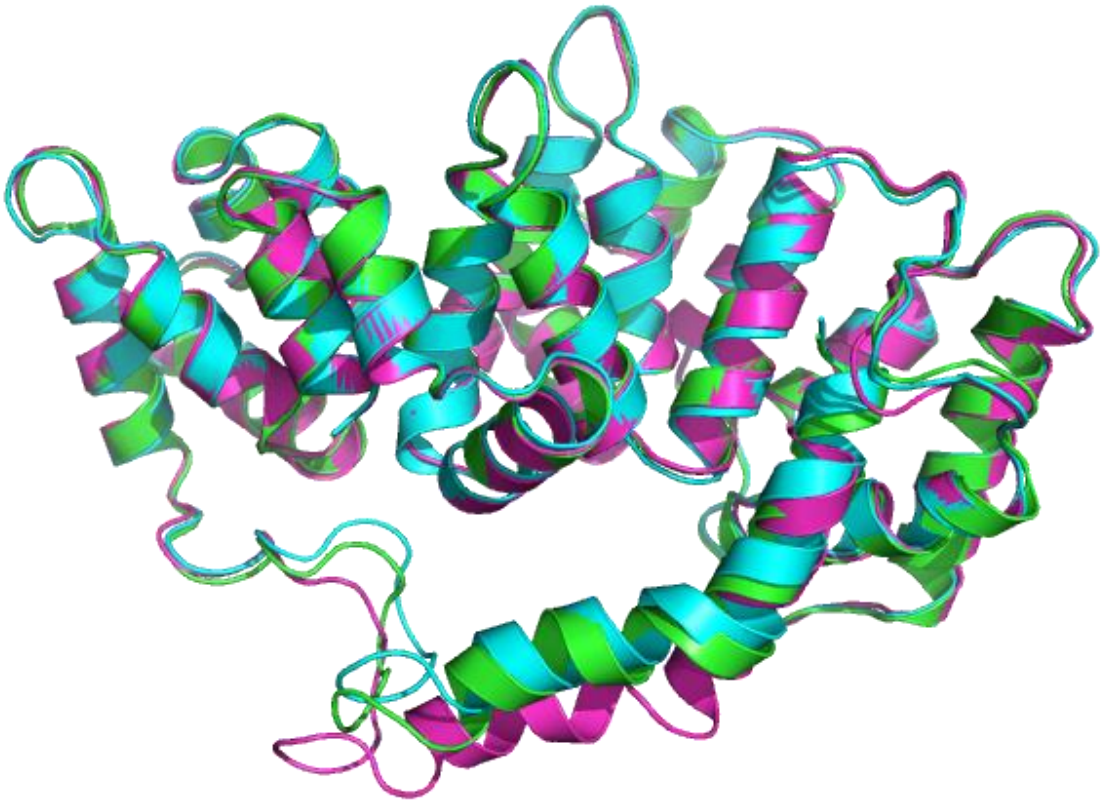


Figure 3.11. Overlap of representative cluster 2 structures for all three implicit NEB simulations (1: green, 2: cyan, and 3: magenta).

In the third clustered images there are slightly more variation, as shown in **Figure 3.12**. The core domains do not align as nicely as the first few clustered structures. Even though the alignment is not as accurate, the overall structural representation of the core is the same. The N-terminal domain for the three simulations are in a similar location. The secondary structure loss that was previously observed for the third simulation is not observed for the second simulation. Similar conformational changes are occurring with the simulations, they are just occurring at different stages.

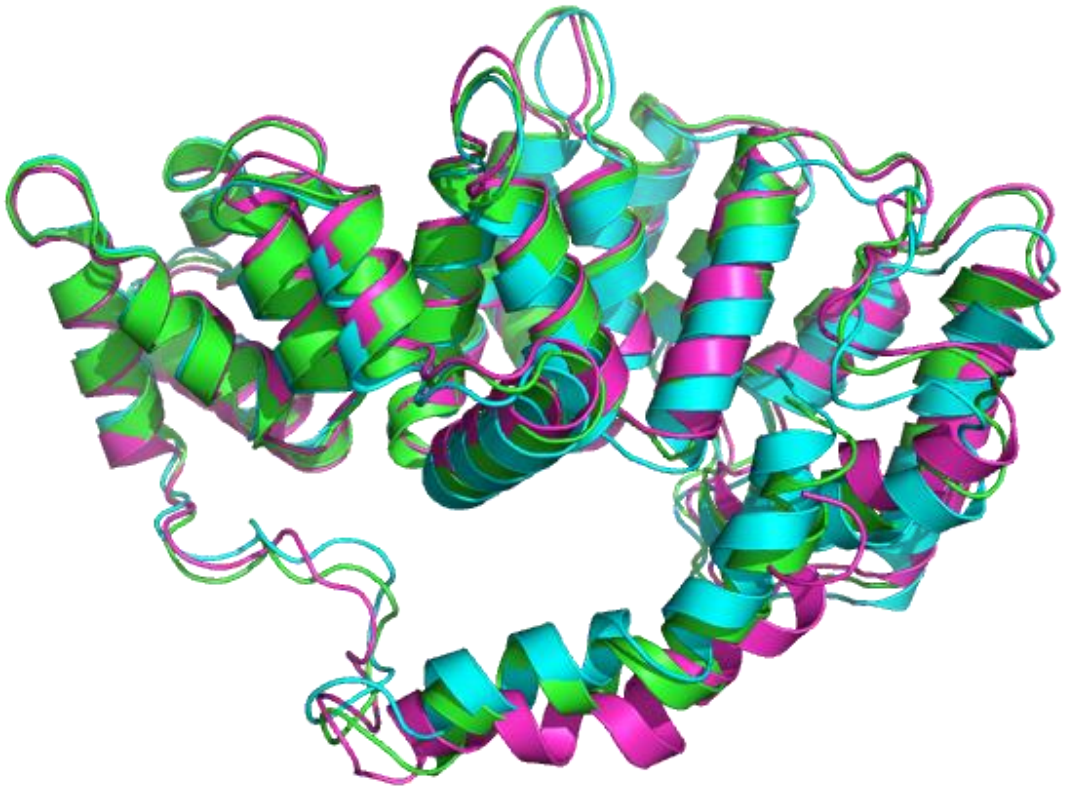


Figure 3.12. Overlap of representative cluster 3 images for all three implicit NEB simulations (1: green, 2: cyan, and 3: magenta).

Figure 3.13 represents the overlapping of structures obtained in the fourth cluster for each simulation. The core domain for each structure are aligned with more precision than the third cluster. Certain structures along the conformation pathway align better than others, but each alignment for the core domains are similar. The major differences noted for this comparison is the helical formation in repeat III of the core domain.

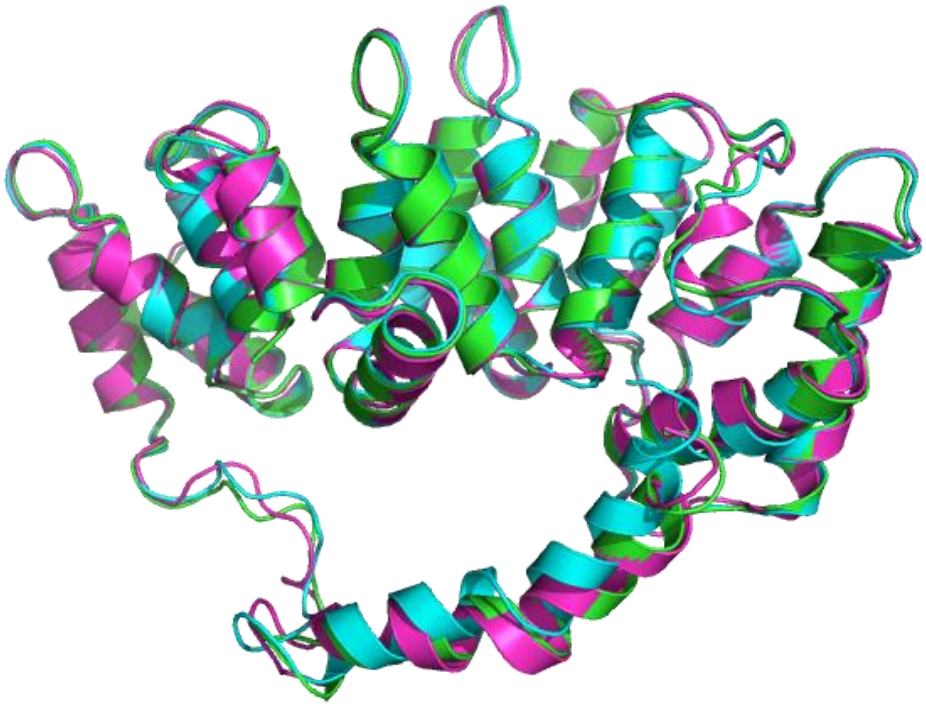


Figure 3.13. Overlap of representative cluster 4 structures for all three implicit NEB simulations (1: green, 2: cyan, and 3: magenta).

Over the remainder of the clusters (5-7), shown in **Figures 3.14-3.16**, there is similar conformational change occurring. While the change is not occurring at the exact same time or location on the potential energy curve, the change is occurring in the same sequential order. For the remainder of the clustered structures the comparison between them are similar, as we have previously distinguished.

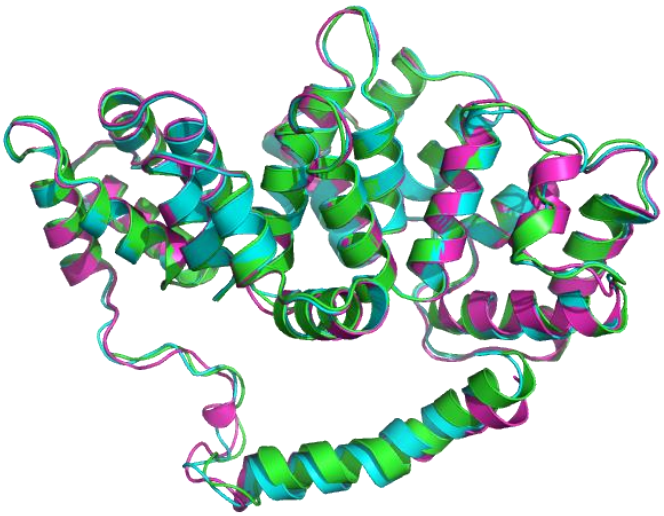


Figure 3.14. Overlap of representative cluster 5 structures for all three implicit NEB simulations (1: green, 2: cyan, and 3: magenta).

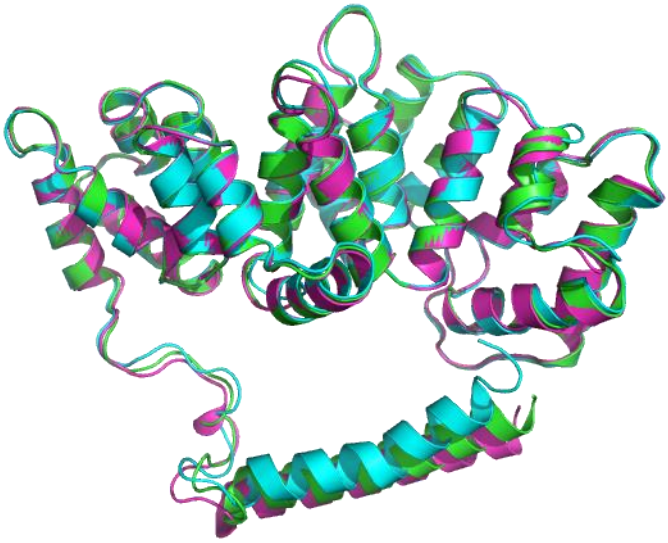


Figure 3.15. Overlap of representative cluster 6 structures for all three implicit NEB simulations (1: green, 2: cyan, and 3: magenta).

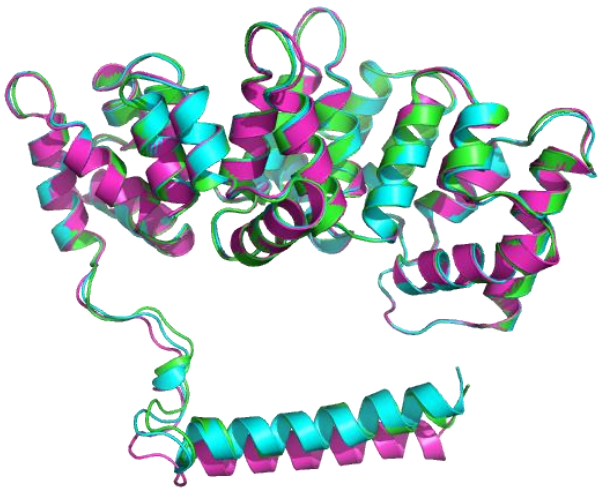


Figure 3.16 Overlap of representative cluster 7 structures for all three implicit NEB simulations (1: green, 2: cyan, and 3: magenta).

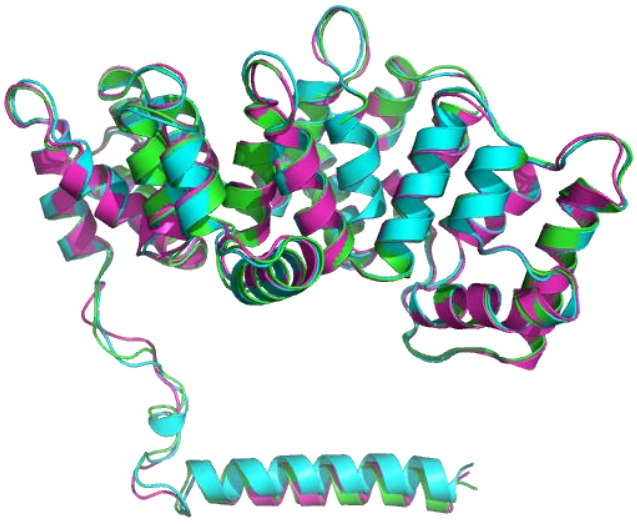


Figure 3.17 Overlap of representative cluster 8 structures for all three implicit NEB simulations (1: green, 2: cyan, and 3: magenta).

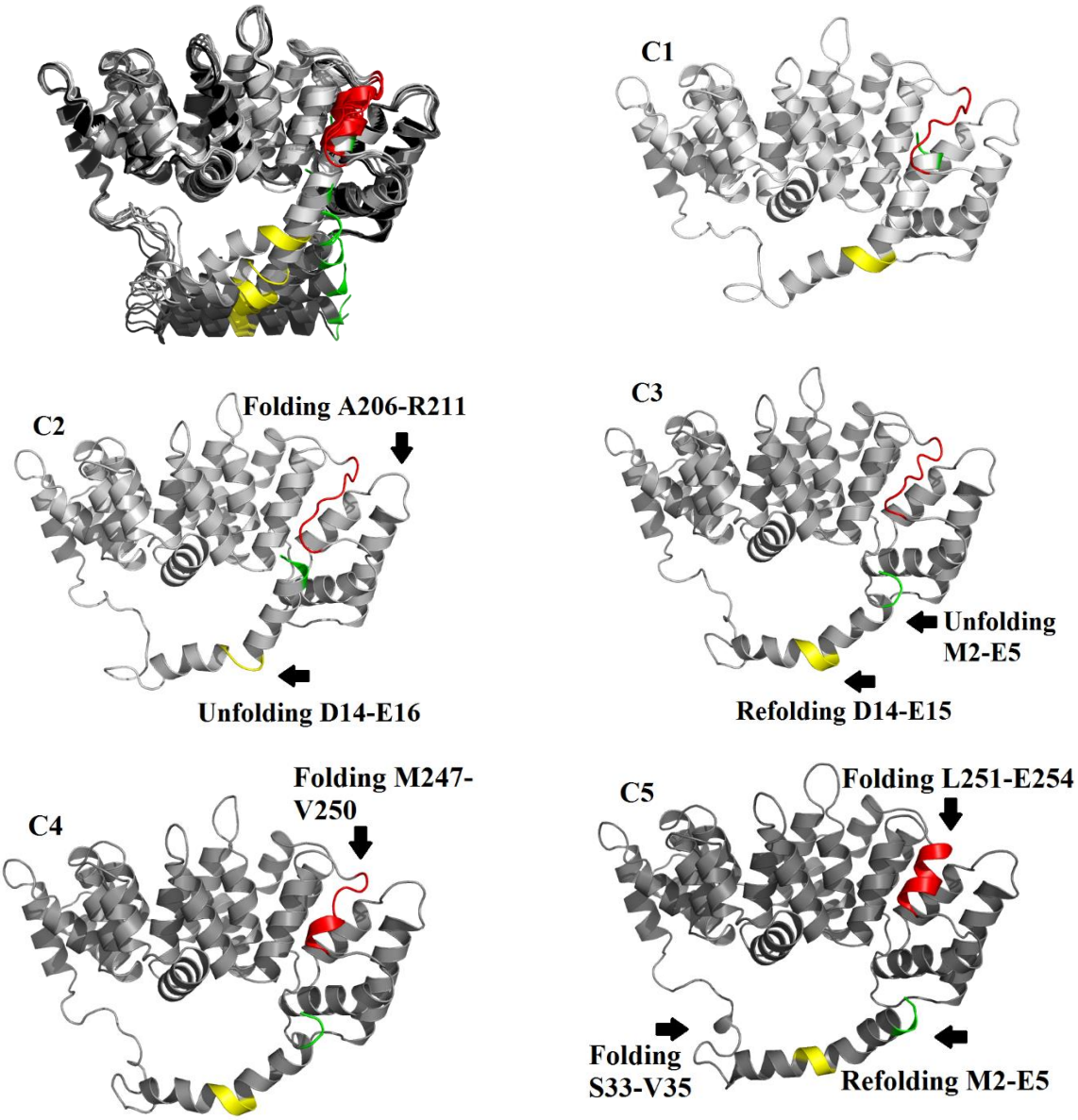
Overlapping the structures for each cluster it can be concluded that the structures are consistent over all three simulations. This suggests that the NEB method is consistent over all three simulations. Each simulation used different simulation parameters and produced similar structures along the conformation pathway. Due to the consistency of the data obtained up to this point, we are confident that the conformation change along all simulations are the same, and we will be going into detail for the conformation change for only one simulation. In the remainder of the work we will be investigating the conformation change for annexin A1 using results obtained from the third simulation.

N-terminal Conformation Change

N-terminal removal from the core domain of annexin A1 has been confirmed using many experimental methods. However, we examined conformational change at an atomistic level which cannot be done with many experimental methods. **Figure 3.18** expresses the conformation change that is occurring from the calcium free to calcium bound structure of annexin A1. **Figure 3.18** includes an overlapped image of all the clustered structures, allowing for clear representation of how the N-terminal domain is removed from the core.

In the calcium free structure of annexin A1 the N-terminal helix is buried inside of the third repeat of the core domain. In **Figure 3.18 (C1)** the tip of the N-terminal helix (green) is being covered by the loop region of repeat three (red). Using the NEB method, we are able to see that the loop is covering the N-terminal helix. The N-terminal helix slides down from its previous location in the second clustered structure **Figure 3.18 (C2)**. This shifting of the N-terminal helix could be facilitated by unfolding of residues D14-E16. Theses residues have lost their secondary structure and are now loops. Loss in their secondary structure will make these

residues more flexible than they were as a helix. It is possible that the flexibility of residues D14-E16 are needed to help pull the tip of the N-terminal helix out of the core domain.



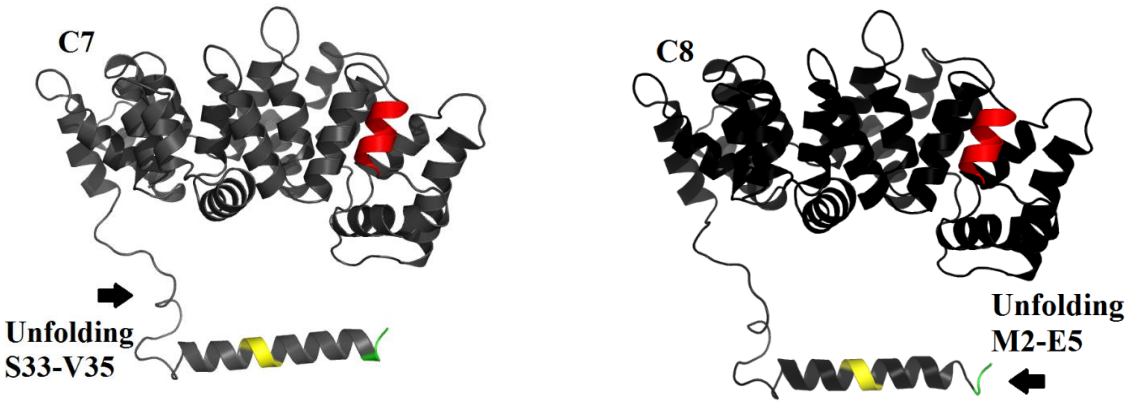


Figure 3.18 Representative clustered images showing how the calcium free structure of annexin A1 may undergo conformation change to a calcium bound structure using the NEB method. Major conformational changes are labeled with their respective cluster representations.

In the third clustered structure, **Figure 3.18 (C3)**, residues D14-E15 have reformed their original helical structure. The N-terminal helix is shifted even further down from the third repeat, which solidifies the potential involvement of the loss of secondary structure in residues D14-E15. In cluster 3 there is unfolding of residues M2-E5. At this point in the conformation change the N-terminal helix is almost fully removed from the N-terminal domain. Secondary structure change of the first few residues in the N-terminal helix might be required to fully remove the N-terminal helix from the core domain. Helical shaped structures tend to take up more space than structures with loop like conformations. When the N-terminal is a helix it is more ridged and is not able to slide out of the pocket as easily. Losing some conformation change right before its removal from the core domain may assist the tip of the N-terminal to come out of the pocket, by making it more flexible. After the N-terminal is removed from the third repeat of the core domain the N-terminal continues to shift away from the core.

In **Figure 3.18 (C4-C5)** the N-terminal continues to move away from the core domain and the N-terminal helix begins to straighten. In previous structures, the N-terminal was bent due

to its association inside of the core domain. Once free, this helix is able to straighten itself out. In the 5th clustered structure there is also refolding of the N-terminal helix at residues M2-E5. The N-terminal has completely removed itself from the core domain and reforms its helical structure. This process of refolding suggests that once fully removed from the core the end of the helix can now refold. This structure also has folding of residues S33-V35 in the N-terminal linker region. This linker is very flexible and is able to move around during the simulation. In this process it is possible that its movement puts certain areas of the linker in proper conformational orientation to fold. That could be an explanation why this area of the linker has slight folding. There is no experimental evidence suggesting why folding in the linker could occur. In the structure for 7th cluster, the folded portion of the linker region losses its structure and goes back to a loop conformation. The N-terminal helix straightens out even more in this structure.

In the final 8th clustered structure, the N-terminal helix is fully straightened and resembles the structure for the calcium bound annexin A1. Unfolding of M2-E5 occurs again in this structure. In previous studies there has been speculation that the N-terminal helix is unable to remain a helix without binding of a partner protein, such as S100. If this N-terminal helix is supposed to lose all secondary structure once removed from the core domain, we are unable to see this here because our reference coordinate keeps the N-terminal as a helix. Loss in secondary structure of M2-E5, could be the start of loss in secondary structure for the N-terminal, but is inconclusive for this study.

Overall the N-terminal removal from the third repeat of the core domain is vital for further protein and membrane interaction. Here we are able to see that the N-terminal gradually pulls itself out of the third repeat. There are also some changes in secondary structure for the N-terminal during this process that seem to aid in proper removal of the N-terminal. Without

flexibility of the linker region connecting the N-terminal and core domain this process would not occur properly, or at all.

Core Conformation Change

In the core of annexin A1, the most notable change that occurs is the formation of helix D in repeat III. The detailed process in which this conformation change occurs has not yet been introduced. Here we provide insight into a detailed process on how conformation change in the third repeat can occur. When examining the overlapped annexin A1 clustered images in **Figure 3.18** there is less visible change that is occurring than the N-terminal helix. However, there is noticeable conformation change occurring in the third repeat of the core domain. This conformational change is one that we expect to see and need to look at in detail to determine how conformational change from a calcium free to calcium bound annexin A1 structure occurs.

In the first three clustered structures of annexin A1, **Figure 3.18 (C1-C3)**, the portion of the third helix involved in folding is initially a loop. This area, shown in red, is undergoing movement along the conformation pathway. In **C1** this loop is overlapping the tip of the N-terminal helix which is in the location where helix D will form. As the N-terminal helix is removed from the third repeat, the loop begins to compact itself and create more waves in the loop. Comparing **C1** to **C3** there is noticeable difference in orientation of the loop. Orientation of the loop in this fashion can put it in proper orientation to form helix D, as shown in **Figure 3.18 (C4).** In **C2** there is folding of A206-R211. These residues are located on helix A of repeat III. Initially these residues are loops connecting helix A and B, but early in the conformation pathway these residues fold making helix A longer. These residues remain folded so it can be assumed that these residues are needed for proper conformational change to occur in the

remainder of the core domain. The first set of residues to start formation of helix D in repeat III are M247-V250. These residues start forming the bottom of the helix. As soon as the N-terminal helix is removed from the binding pocket, the loop of repeat III is not blocked by the N-terminal and can form the helix. The second step of folding for helix D, are residues L251-E154 **Figure 3.18 (C5)**.

To make sure all folding steps were identified, all 32 structures were examined. In all simulations performed folding of helix D in repeat III is a two-step process. This conformation change is consistent over all three simulations. In the remainder of the structures **Figure 3.18 (C7-C8)** the core domain does not undergo any major conformational change. Due to imperfections in visual interpretation of conformation change, RMSD and secondary structure analysis was performed to see if any other changes occurred over the conformation pathway.

Root Mean Square Coordinate Deviation

Performing RMSD analysis for each cluster, areas of conformation change can be detected. This analysis technique is important because visual inspection of the structures will often times overlook minute details. This method looks in closer detail at each residue and compares structural alignments between two structures. In **Figure 3.19** RMSD comparisons of each cluster are given, to shown changes in certain areas of annexin A1 core domain over the entire conformation pathway.

Spikes in RMSD at the binding site located in repeat II at residue 130 are present most prominently in comparisons of C1 and C2, and C4 and C5. These two RMSD plots have the largest peaks for this area. Residue 130 is a site II binding loop for calcium in repeat II. The loops are more flexible than helical portions of the protein and could be attributing to the spikes

in RMSD change. In close proximity to this binding site there is also movement for residues 190-200. These residues are part of the linker connecting repeat II and III. This region is flexible and its proximity to repeat III helix D formation may be a result in its increased movement. The more flexible an area is, the higher the RMSD change will be due to either movement of the atom locations or structure change.

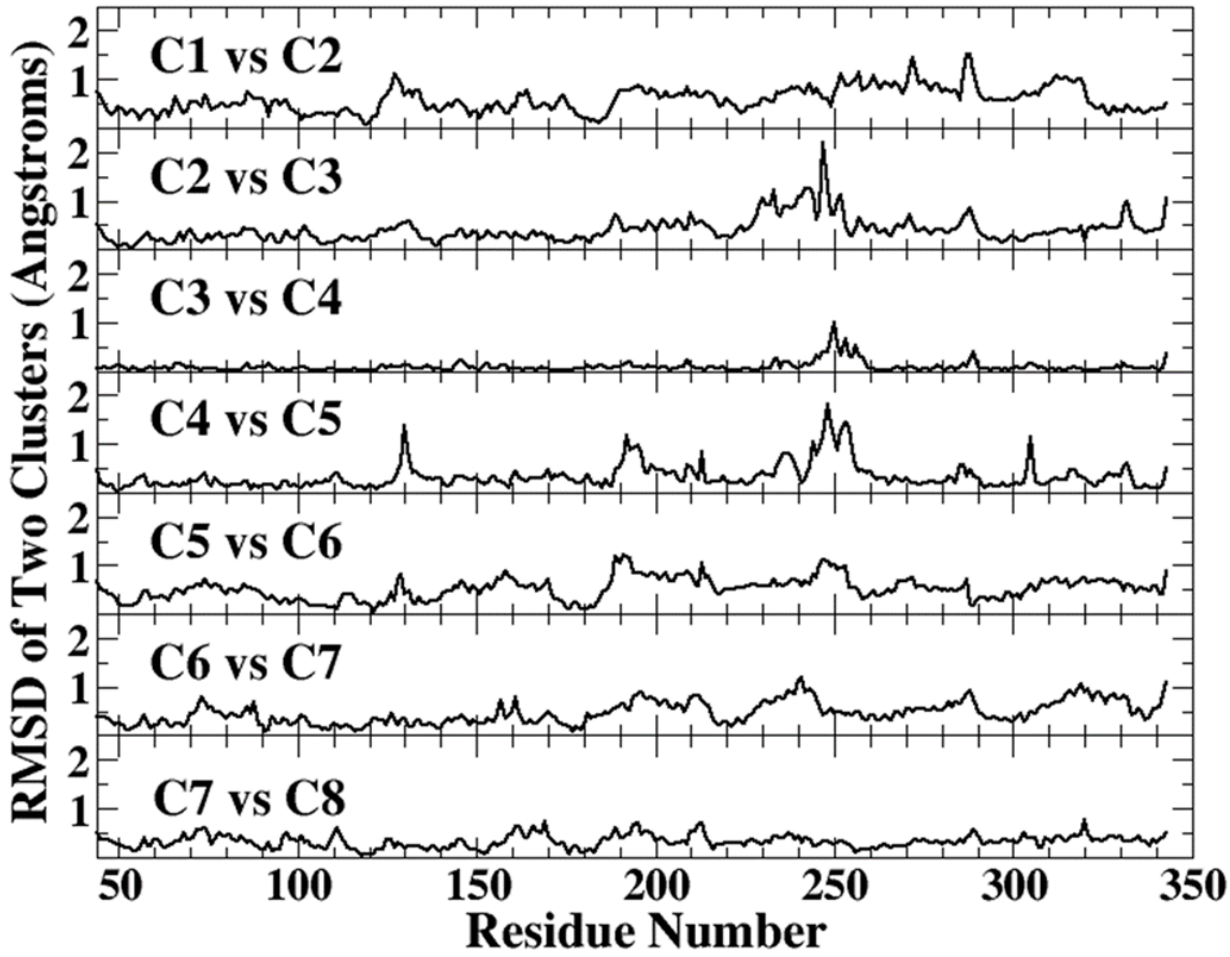


Figure 3.19 RMSD comparisons of the clusters going from a calcium free to a calcium bound structure, including residues of the core domain.

The area of the protein with the most amount of change is located in repeat III. This is expected due to the conformation change and formation of helix D. Residues surrounding helix D also have changes in RMSD through the conformation change. These fluctuations can be attributed to changes in orientation of repeat III to facilitate proper folding of the loop. It is not reasonable to assume that only folding of helix D will be occurring. Residues in repeat III will have some cooperative changes to ensure the protein will fold properly. In repeat IV there are changes in RMSD for calcium binding site II in the area around residue 290. **C1 vs C2** has the largest spike for this area of the core domain. However, there are smaller spikes and fluctuations occurring in **C2 vs C3**. RMSD differences for this binding site could be caused by movement of the residues or conformational change.

By performing RMSD analysis we are able to determine that more change is occurring rather than just the formation of helix D in repeat III. Some of these fluctuations were not noticed by visual investigation. For most of the RMSD changes they occur anywhere between 1-2 Å, which can be hard to detect visually. Further analysis is needed to determine what kind of change is occurring. If there is any kind of secondary structure change occurring in the core domain, we can detect that by performing secondary structure analysis. This will help investigate which regions of the protein are moving slightly or which areas are having conformational change.

Secondary Structure Analysis

The third repeat of annexin A1 has drastic structure change for residues M247-E254. These residues participate in folding to form helix D. To determine if other areas of the core domain are undergoing cooperative secondary structure change, we have investigated the helical percent of each residue of the core for all clustered structures, as shown in **Figure**

3.20. Overlapping the helical percent data gives insight into areas of the protein that are undergoing structure change.

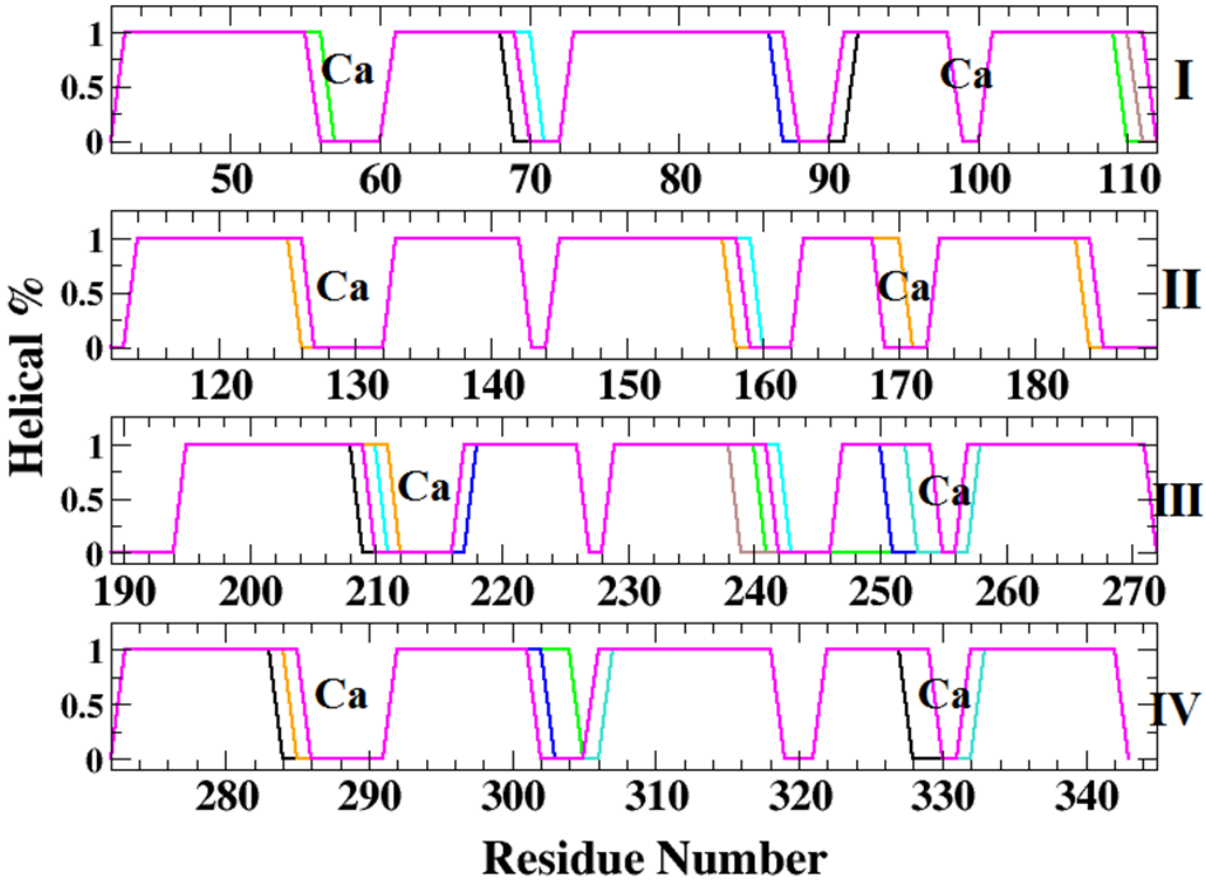


Figure 3.20 Secondary structure helical percent for each repeat of annexin A1 core domain. Each NEB cluster is represented and overlapped on top of each other (C1: black, C2: cyan, C3: green, C4: blue, C5: turquoise, C6: brown, C7: orange, C8: magenta). Calcium binding sites determined from x-ray structure 1MCX are expressed with Ca labels.

The residues with the most structure change are M247-V254. In **Figure 3.20** this is shown by variation in helical percent for these residues. Overlapping the data for all eight clusters shows that this area is cluttered with variations in helical percent. The secondary structure plot expresses that C4 starts to form the helix of repeat III first and continues to form

until all residues have formed the helix. In this plot, areas of the protein that have no secondary structure change will not show fluctuations as all helical percent and will be the same.

Areas of the core domain that seem to have structure change other than helix D of repeat III, are areas where calcium binding would occur. In **Figure 3.20** these binding sites fluctuate in the amount of secondary structure change that is occurring. For example, in the first calcium binding site of repeat I there is fluctuation in secondary structure between two residues, while in the first calcium binding site of repeat III there is change over four residues. The first calcium binding site of repeat III, has increase in helical structure from residues 208-212 over the first seven clusters, then in cluster eight there is loss of most of the secondary structure. For some binding sites there are fluctuations in secondary structure.

Overall, the majority of change in secondary structure for the protein is occurring in the third repeat. However, over the course of the conformation change secondary structure change is occurring at calcium binding sites. It is important to keep in mind that for the NEB simulation, calcium ions were not used.

CHAPTER 4. RESEARCH PROJECT 2: MOLECULAR DYNAMIC STUDY OF CTNC POINT MUTATIONS

4.1 Purpose

The purpose of this study was to investigate the result of point mutations with cardiac troponin C using molecular dynamics simulations. The importance in cTnC calcium binding in the muscle contracting mechanism is one that is known, but at an atomistic level calcium binding interactions due to mutations are elusive. By studying the dynamics of the point mutations, we hope to provide insight into the effect these mutations have on calcium binding for the protein.

4.2 System Preparation

The initial coordinates for wild type cTnC in the presence of calcium were obtained from the X-ray crystal structure 1J1D [38]. The cTnI and cTnT subunits were removed from the structure, and cTnC were used in the simulations. Modifications to the crystal structure were made to accommodate for various point mutations. In this work five point mutations were investigated including D65A, S69C, A8V, L29Q, and C84Y. Five separate cTnC structures were saved, in each structure the corresponding mutation was modified using pymol. **Figure 4.1** shows a collective combination of all point mutations, to give clarity of their location on cTnC.

Once the mutations were made, the individual structures were neutralized and explicitly solvated using TIP3P waters in a 14 Å box using LEap. The solvent box including the protein,

TIP3P water and counter ions for the WT, D65A, S69C, A8V, L29Q, and C84Y mutations consisted of 42555, 42558, 42552, 42558, 42550, and 42556 atoms respectively.

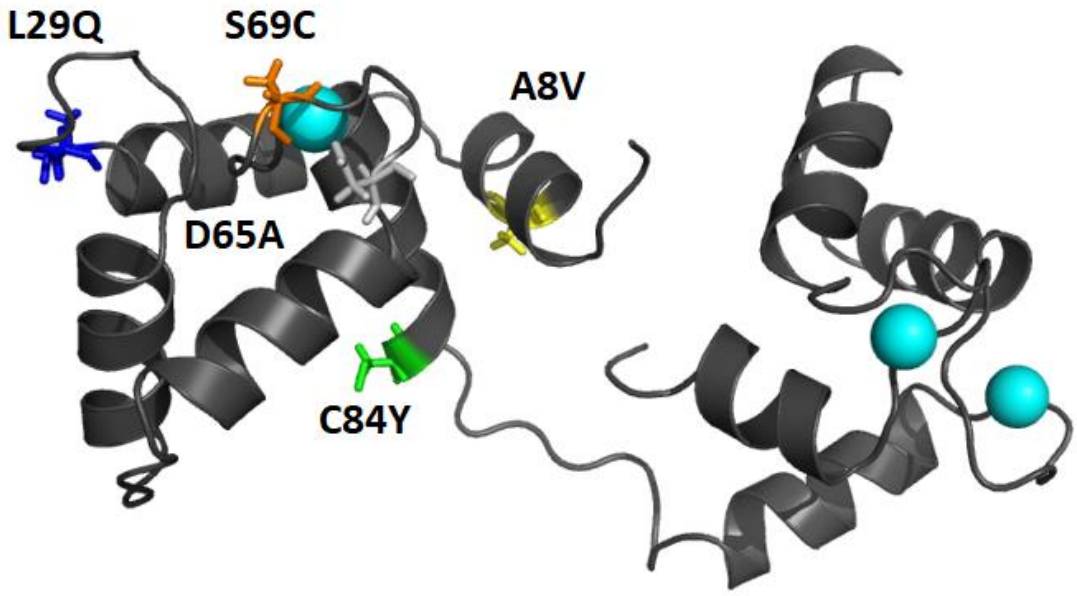


Figure 4.1. Wild type cTnC protein with point mutations shown as sticks.

4.3 Results and Discussion

Root Mean Square Deviation

The NVT simulations for wild type cTnC and each mutation was carried out for at least 200 ns each. The total simulation time including all simulations were approximately 1,425 ns. To investigate protein stability the backbone root-mean square coordinate deviations (RMSD) for all the entire protein were calculated and shown in **Figure 4.2**. The RMSD for the full protein was based on the simulated structural snapshots and aligned to determine if structural variation occurred.

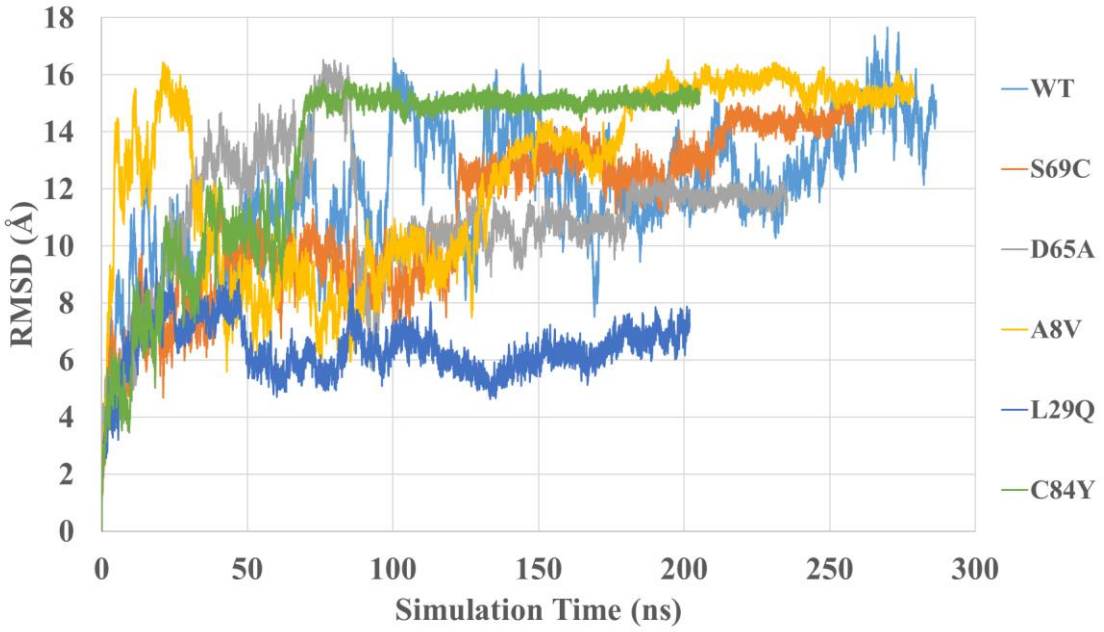


Figure 4.2. Backbone RMSD alignment for each cTnC simulation (full protein) run for at least 200 ns each.

As shown in **Figure 4.2** the RMSD for the full length cTnC proteins seem to be level for a lot of the simulations. The simulation with the most amount of fluctuation was the wild type (light blue). All simulations with mutations have RMSD plots which have stabilized after about 150 ns. The largest spikes in RMSD occurred until 50 ns of simulation time. This is important for computational purposes because it shows that failure to run simulations for a sufficient period of time may result in misinterpreted data. It is interesting that for all simulations, the tail end of the RMSD curves are converged in approximately the same area between 12-16 Å. The large change in RMSD for the wild type simulation suggest, that the structure was not fully stabilized with respect to its relative position. This is most likely due to flexibility in the linker region connecting the N and C terminal domains. To investigate this in more detail, individual RMSD plots for the individual domains were constructed.

RMSD analysis of the N-terminal domain of the cTnC structures gives insight into the flexibility of this domain over the simulation time. The wild type structure had the most flexibility in the N-terminal, as shown in **Figure 4.3**. For the first 150 ns, the wild type simulation was stable around 8 Å. However, after 150 ns the RMSD spiked and did not stabilize over the rest of the simulation. The N-terminal domain of cTnC does have a hydrophobic pocket that opens up as part of its conformation change when bound by calcium. It is possible that such flexibility in the RMSD of the N-terminal is affected by this property.

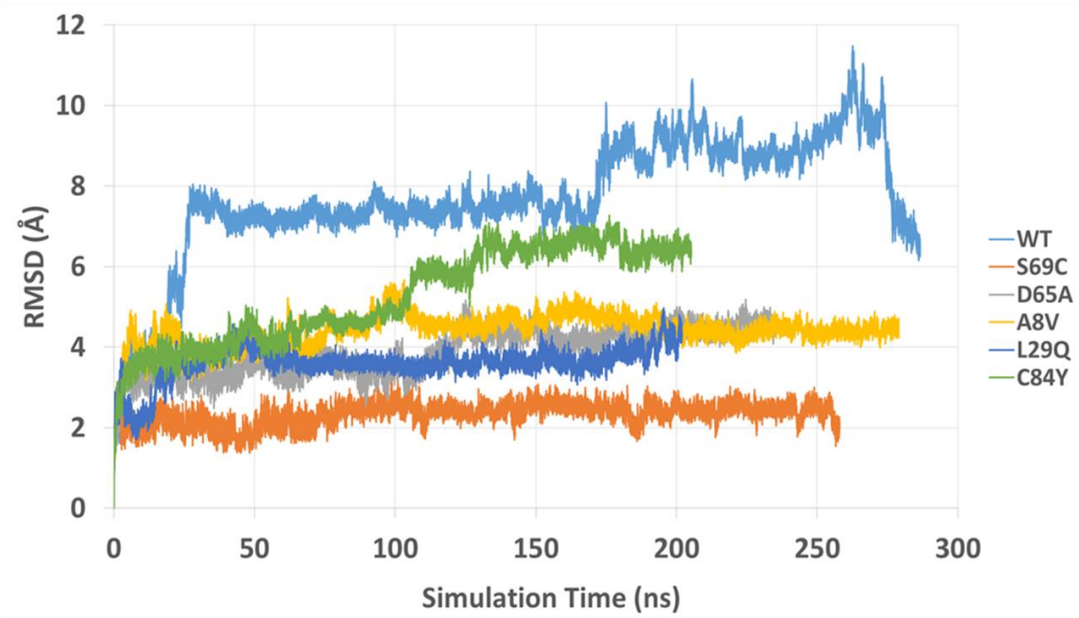


Figure 4.3. Backbone RMSD alignment for each cTnC simulation (N-terminal) run for at least 200 ns each.

The wild type simulation has the largest RMSD difference, compared to the mutations. Most mutations range in RMSD differences from 2-4 Å. The only mutation that does not have a stabilized RMSD over the entire simulation was the L29Q mutation. After 100 ns there is a

gradual increase in the RMSD from 5-6 Å, for the remainder 100 ns. This mutation does not have and large spikes in RMSD like the wild type simulation. For the most part the N-terminal domains are stable for the simulations.

The C-terminal RMSD differences are more stable than the N-terminal, **Figure 4.4**. In this figure all simulations have RMSD differences below 5 Å. For the C-terminal domain there are no large fluctuations in the RMSD. The C-terminal domain for the wild type simulation is more stable because it is not fluctuating over the simulation as greatly as it does in **Figure 4.3**. The N-terminal domain is expected to have more flexibility due to its involvement in calcium induced conformation change. Studying the RMSD analysis over the simulation we can see that the N-terminal domain has more flexibility than the C-terminal domain.

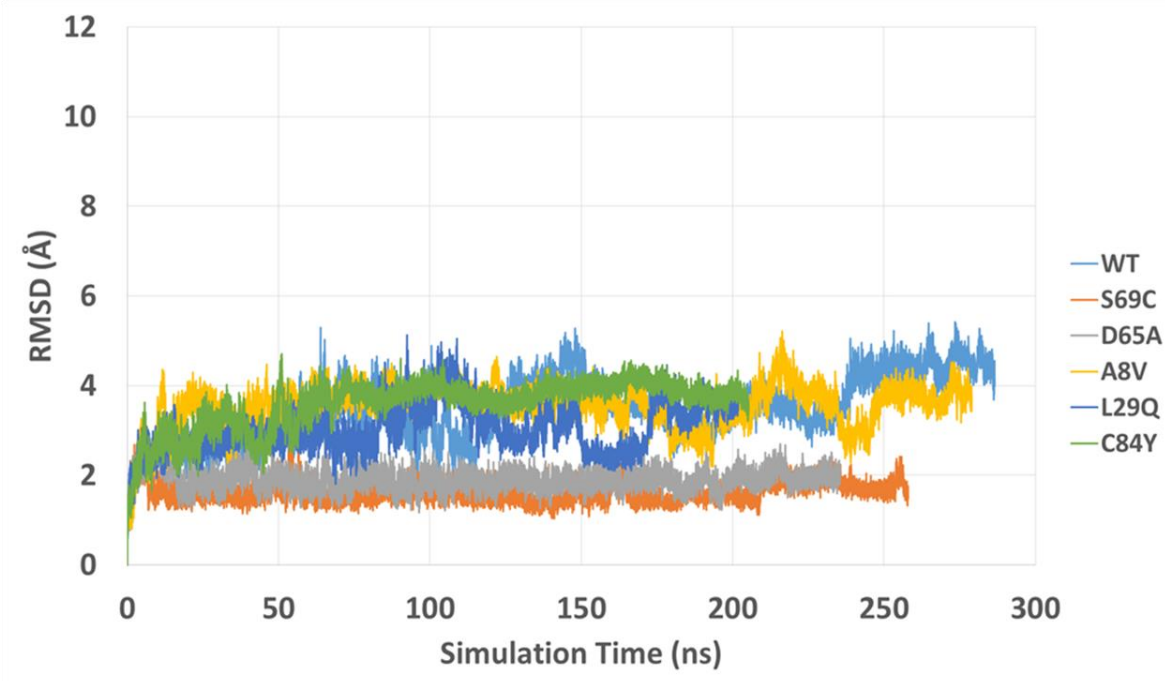


Figure 4.4. Backbone RMSD alignment for each cTnC simulation (c-terminal) run for at least 200 ns each.

Closer examination of the RMSD plots for the N and C terminal domains, it was concluded that the simulations were equilibrated. If the structures were not equilibrated there would be large variations in the RMSD for the simulations. All mutations had stable RMSD values and did not have any major spikes. The wild type N-terminal domain did have a little bit of flexibility after 150 ns, but for the majority of the simulation it was stable. The RMSD for the C-terminal domain was stable so it was concluded that the simulation was run for a sufficient period to perform the rest of the analysis.

MMPBSA/ MMGBSA

The calcium binding loop on the N-terminal domain of cTnC is composed of 12 amino acid residues from D65-E76. Previously the importance of the residues in this binding have been investigated, however, at an atomistic level they have not been determined. Previous studies suggest that residues D65, D67 and E76 are important in coordination of the calcium ion upon binding. We have investigated the effect of 5 different mutations on the binding of calcium to site II.

MMBSA analysis was performed for each simulation to investigate the change in calcium binding when point mutations occur. Here we have shown the interaction energy in kcal/mol of the calcium at binding site II for all simulations in **Table 4.1**.

Table 4.1. Interaction energy for site II and calcium, for wild type and five different point mutations.

Mutation	*Interaction (kcal/mol)*	*Standard Deviation*	*Comparison to WT*
Wild Type (WT)	-90.63	33.74	—
D65A	-16.84	14.20	Decrease
S69C	-62.09	5.37	Decrease
A8V	-113.18	23.10	Increase
L29Q	-47.21	17.76	Decrease
C84Y	-123.00	25.39	Increase

Based on the molecular dynamics simulation we are able to see which mutations have increased or decreased calcium binding in comparison to the wild type simulation. The wild type mutation had an interaction energy of -90.63 kcal/mol with a standard deviation of 33.74 and is used as the baseline to compare to the mutations. Mutations D65A, S69C, and L29Q had a decrease in calcium binding. These results are comparable to the trends found in experimental studies of cTnC mutations. D65A and S69C are mutations located in the calcium binding loop of site II, while L29Q is located in the center of a loop in the N-terminal. The D65A mutation has the largest decrease in binding compared to the other two mutations, however the deviation for this theoretical method must be taken into account. The D65A and S69C mutants had a standard deviation of 14.20 and 5.37 respectively, for the MMPBSA calculation.

Mutations A8V and C84Y have an increase in calcium binding. The A8V mutation is located at the very beginning of the N-terminal domain, while the C84Y mutation is located at the end of the N-terminal domain, right before the linker region. The interaction energy of the A8V and C84Y mutations were -113.18 and -123.00 kcal/mol, with deviations of 23.10 and 25.39. Mutations in different regions of the protein have different effects on the protein. Proteins are dynamic systems and making an alteration in one area, like a mutation, has an effect in

another area of the protein. We were able to see this dynamics process occur in the fact that altering one single amino acid will change the calcium binding ability.

Site II Calcium Interaction

Performing MMPBSA we were able to see the effect point mutations have on calcium binding of site II. To investigate the effect of the mutations at an atomistic level, MMGBSA was performed. The interacting energy for site II residues and the calcium at this site is shown in **Figure 4.5**. Mutations S69C, D65A, and L29Q had loss in calcium interaction. This is consistent with the decrease in binding found for the MMPBSA calculation. The S69C mutant lost calcium interaction at three residues within site II. These residues include slight loss in D67, loss in C69, and loss of D73, as shown in row two of **Figure 4.5**. In the wild type interaction energy plot, there is not a lot of interaction between the calcium and S69. When mutated to C69, there is loss, but it is not as drastic as some of the other mutations.

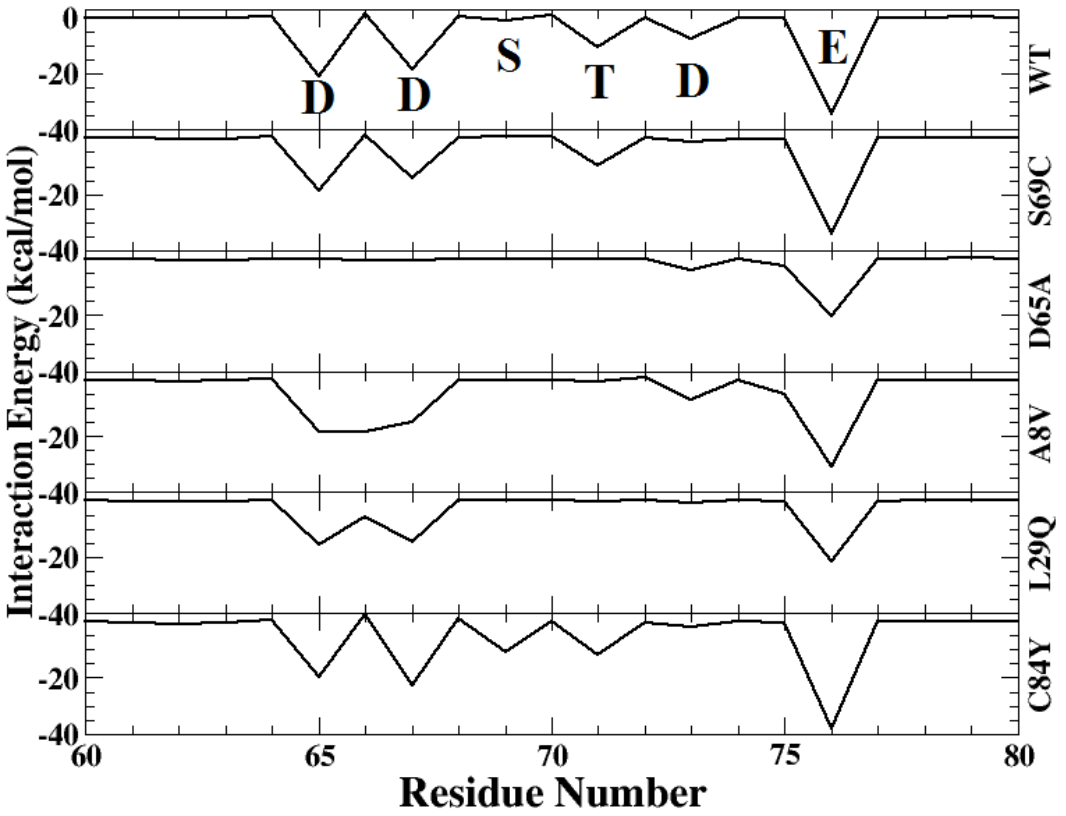


Figure 4.5 Interaction energy of calcium and site II residues for wild type and all five mutations in the N-terminal domain.

The D65A mutation has the greatest loss for site II calcium binding. Loss in binding occurred for four residues of site II. This mutant had significantly more loss in binding than the S69C mutant. Loss of binding was present at A65, D67, S69, T71, and E76, as shown in row three of **Figure 4.5**. Almost all residues that lost calcium interaction, had complete loss of interaction. The only residue that lost slight interaction was E76. This coordinating residue is the strongest within site II, as it has two coordinating oxygen atoms instead of one, as shown in **Figure 4.6**. Both D65A and S69C mutants are located inside site II binding site, however the effect of the mutations are very different for overall binding. The D65A mutation is located at the first coordinating residue in the calcium binding loop. The wild type protein of cTnC has the

greatest interaction between D65 and E76, which are the anchoring residues with the binding site, as shown in row one of **Figure 4.5**. The distance between the oxygen atoms and calcium of these two residues are 2.3 Å and 2.4 Å for D65 and E76, as shown in **Figure 4.6**. Mutating an anchoring residue of the binding site has a detrimental effect on the calcium interaction. S69C mutant is located in the center of the binding loop, and alteration in this mutation only has loss in a few residues.

The last mutant inducing a loss of calcium coordination was L29Q. This mutant is located at site I in cTnC. Complete loss in calcium interaction occurred for T71 and D73, while decreases in binding occurred at D65, D67, and E76, as shown in row five of **Figure 4.5**. The calcium interaction of this mutant shows how cooperative proteins are, and how sensitive they can be to changes. For all mutants resulting in a decrease in calcium interaction energy, the anchoring residues of site II are vital for calcium coordination. In S69C and L29Q mutant, the coordination of D65 and E76 are still present and with decent strength. When the anchoring residue D65 is mutated the interaction energy for most of the sites are lost.

Two mutations that had an increase in calcium interaction energy were A8V and L2Q, as shown in row four and six of **Figure 4.5**. Both mutants were located a sizeable distance away from site II, where A8V was at the beginning of cTnC N-terminal and C84Y was at the end of cTnC N-terminal. A8V mutant had increased interaction for E66 and D75, while other residue interaction energies in the site were comparable to the wild type. This mutant had slight loss in T71, however in respect to the overall interaction there is increased interaction with calcium when compared to wild type. C84Y mutant had increased interaction at S69 and E76, with slight decrease in interaction at D73. The other residues calcium interaction in site II are comparable to wild type. This mutant has increased interaction in E76 which increased the interaction of the

anchor residue of site II. This could contribute to the large increase in calcium interaction for this mutant as compared to wild type, as shown in **Table 4.1**.

Performing MMGBSA analysis for the calcium interaction of site II residues we are able to see in more detail the residues that are contributing to either a decrease or increase in calcium interaction. Mutations in differing parts of cTnC N-terminal affect calcium binding differently and show the proteins dynamic characteristic.

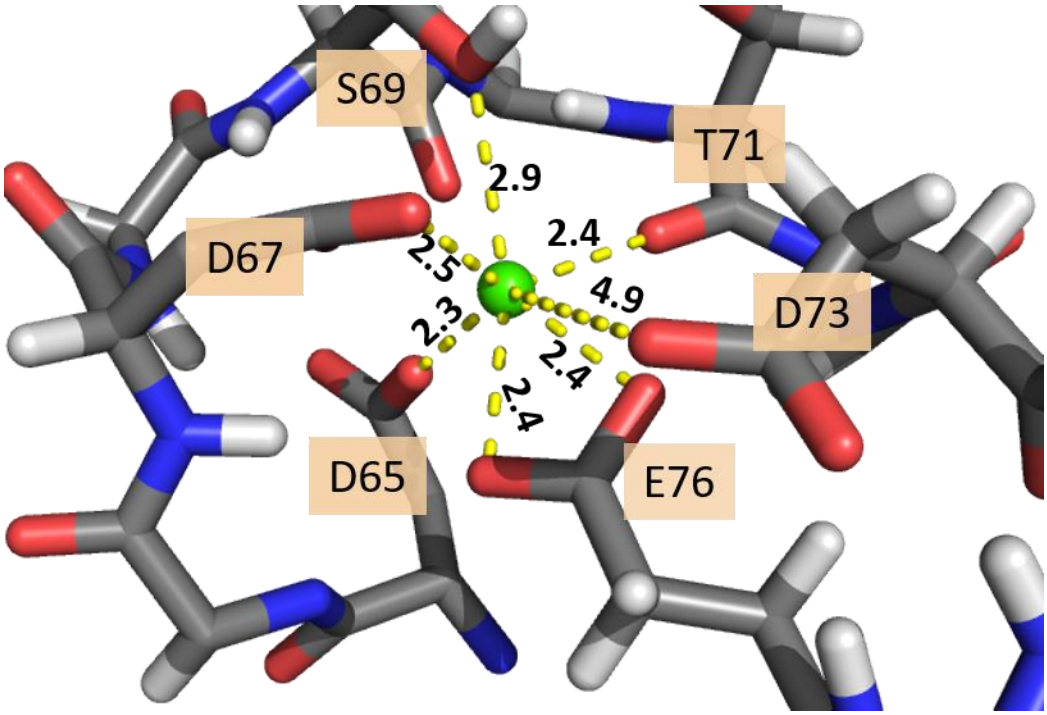


Figure 4.6. Site II binding loop with calcium coordinating residues and distance (Å) values.

C-terminal Calcium Interaction

Due to the cooperative and dynamic nature of proteins, we investigated whether N-terminal mutations in cTnC would affect calcium binding sites III and IV. Similar to the previous analysis technique, we used MMGBSA to calculate the interaction energies for the binding loop residues and calcium.

For a majority of the mutants there was not much alteration in calcium interaction for site II, as shown in **Figure 4.7**. There were slight variations in interaction for two of the mutations. D65A mutant showed loss in calcium interaction of N107 and C84Y mutant had loss in Y111, as shown in row three of **Figure 4.**7 and row six of **Figure 4.7** respectively. For site III calcium interaction there was not much difference for the mutants.

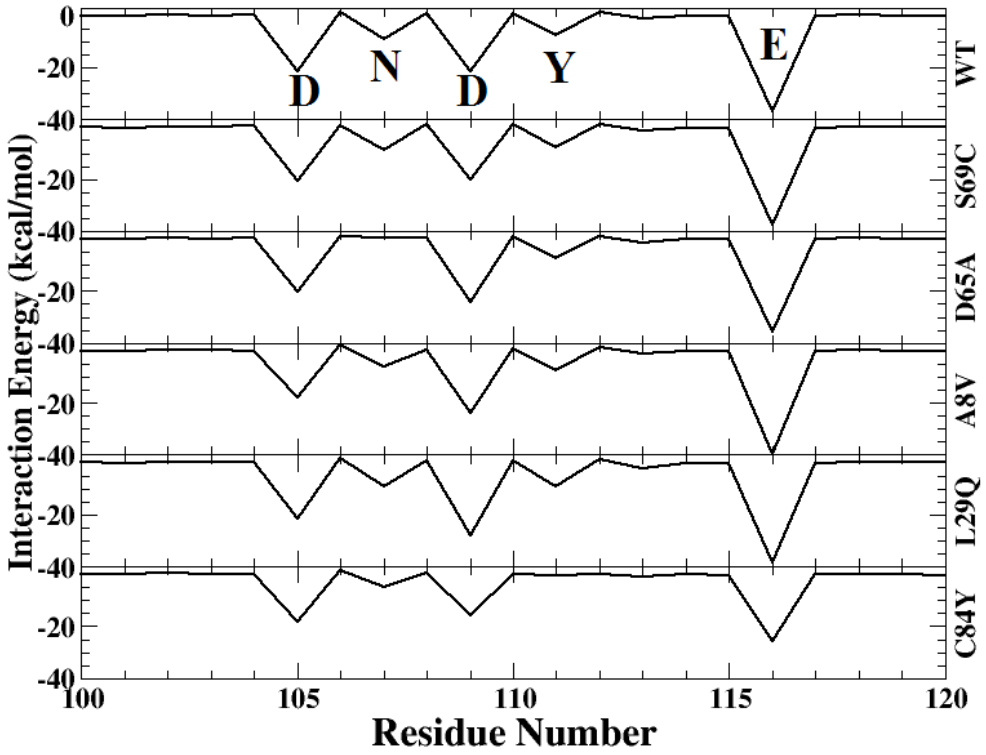


Figure 4.7. Site III calcium interaction energy for wild type and S69C, D65A, A8V, L29Q, and C84Y mutants.

Similar to site III, site IV calcium interaction does not have many changes for the mutants, as shown in **Figure 4.8**. In all mutants there is loss at D149, however the interaction for wild type initially is moderate. The interaction for this residue is not very intense so loss at this residue would not affect the overall binding much. For all other residues of site IV, the calcium binding energies are similar for the mutants and wild type.

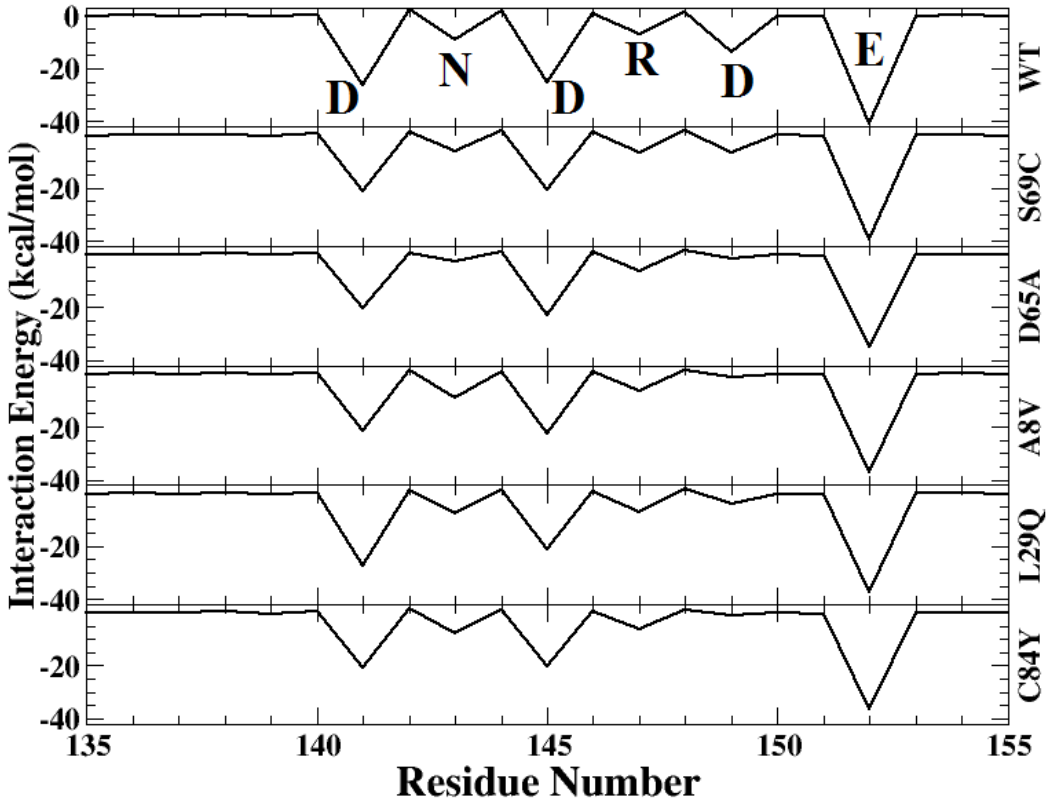


Figure 4.8. Site IV calcium interaction energies for wild type and S69C, D65A, A8V, L29Q, and C84Y mutants.

Mutations in the N-terminal residues presented in this work do not have much effect on the calcium interaction of the C-terminal domain. Site III and IV have stronger calcium coordination than site II. Site II needs to be able to bind and unbind calcium to induce

contraction and relaxation of muscle, so the effect of mutations near that site will be impacted more than site III and IV. Studies in mutations near site III and IV may give more insight into how c-terminal mutation effect the binding of these sites.

Distance Analysis

The coordination of calcium at site II for wild type and all five mutations were analyzed using the molecular dynamic simulation trajectories. The distances between the calcium and each oxygen atom in the binding loop for residues 65-76 are shown in **Figure 4.9**.

In the wild type simulation, the oxygen atoms of D65, D67, T71, D73, and E76 are coordinated to calcium with an approximate distance of 2.5 Å. Here you can see that E76 has two oxygen atoms coordinated with the calcium at this distance, as shown in column one, row seven of **Figure 4.9**. This bidentate coordination represents the strong anchoring of calcium at the binding loop. For the entire wild type simulation, the distance of the calcium and oxygen atoms are consistent.

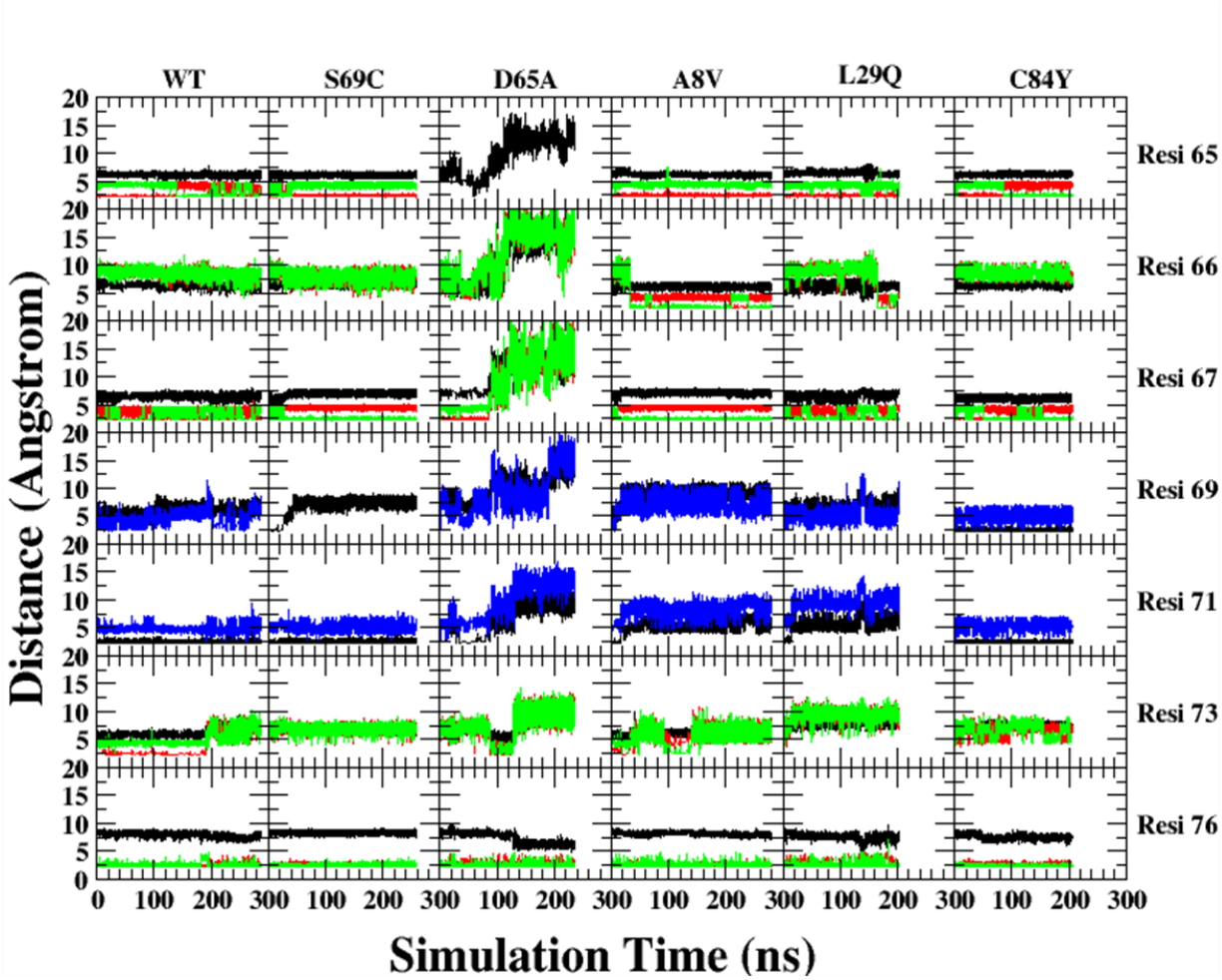


Figure 4.9. Distance of oxygen atoms at site II residues and calcium over the full-length simulations. Each simulation is composed of a column, the corresponding residue is represented by the rows.

The largest change in coordination of calcium and oxygen for site II is for the D65A mutant. All residues except for E76, show loss in coordination after approximately 100 ns, as shown in column three of **Figure 4.9**. The longer the simulation is run the less interaction residues at site II have with calcium. In the interaction energy plot of site II with the D65A mutation we were able to see which residues lost binding. The distance analysis plot allows us to see when this loss occurs. The other mutation at site II has very little change in coordination distance. The only loss is residue 69, as shown in column two, row 4 of **Figure 4.9**. In the wild

type protein, there are two oxygen atoms in the serine residue (black and blue), however mutation to cysteine only has one (black) as shown in **Figure 4.9.** All other residue coordination in site II for the S69C mutation are comparable to the wild type simulation.

L29Q has noticeable loss in calcium coordination of residues S69, T71, and D73, as shown in column five, rows four-six of **Figure 4.9**. The coordination of calcium to these residues is affected soon after the simulation begins. There is a larger distance between the oxygen atoms and calcium for these residues than the wild type simulation. The distance for these three residues approximately doubles in response to the L29Q mutation. The last two mutations had an increase in calcium interaction for the MMGBSA calculation. A8V showed increase in E66, interaction and this is correlated with the distances, as shown in column four, row two of **Figure 4.9**. The wild type simulation has coordination of E66 oxygen atoms around 5 Å. The A8V mutant causes this coordination in oxygen atoms to increase to approximately 2.5 Å. This mutation also causes loss in coordination of T71 and D73, where the distance in these atoms increases as compared to wild type, as shown in column four, row five and six of **Figure 4.9**. Increase in coordination is also found in the C84Y mutant. Residues S69 and E76 showed an increase in calcium interaction after the MMGBSA calculation. Distance analysis shows that at S69 there is a decrease in the distance between the coordinating atoms, as shown in column six, row four of **Figure 4.9**. The coordinating oxygen (black) has decreased its coordination distance by approximately half. It goes from 5 Å to 2.5 Å. For residue E76 there is strong consistent coordination of the two oxygen atoms, similar to that of the wild type. In the C84Y mutation the oxygen (black) with the furthest distance to calcium gets a litter closer over the course of the simulation, as shown in column six, row seven of **Figure 4.9**. However, based on the comparison with the wild type distance plot, they look very similar.

The analysis techniques used on the molecular dynamic simulation trajectories for the wild type and mutations, gave insight into the effect they have on calcium binding. Different forms of analysis were preformed including MMPBSA, MMGBSA, and distance analysis, and all yielded results comparable to each other. D65A. S69C, and L29Q mutations result in loss of calcium binding of site II, while A8V and C84Y mutations increase calcium binding.

CHAPTER 5. CONCLUSIONS

In this work calcium binding proteins were investigated using computational techniques. Simulations were performed for annexin A1 and cTnC. We were able to confirm the reliability of the NEB method using a larger protein then used before. All three NEB simulations yielded results consistent with each other. Therefore, this method can be used to obtain a conformation pathway for large proteins. We also performed MD simulations on cTnC to investigate the effect of point mutations on calcium binding to site II. We determined which residues in site II were affected by the different mutations and how the coordination to the oxygen atoms of these residues are affected over the length of the simulation.

5.1 CONCLUSIONS FROM PROJECT 1: Molecular dynamics study on calcium induced conformation pathway for annexin A1

In this work the NEB method was used to investigate conformation change for annexin A1. Three simulations were performed to track the conformation change of annexin A1 from the calcium free to the calcium bound state. The conformation pathway for all three simulations were compared, and due to structural consistency, the conformation change for all three simulations were concluded to be consistent. Deeper analysis into the conformation change gave insight into N-terminal and core domain conformation changes.

In the calcium free structure of annex A1 the N-terminal helix is buried inside of the core domains repeat III. The process in which the N-terminal helix removes itself from the core domain has not been fully investigated. Based on our results the N-terminal comes out of the

core in a “sliding” motion. There is no evidence that the loop region of the third repeat is swinging away from the helix allowing space for it to come out. This is interesting because residue D14-E16 of the N-terminal helix lose secondary structure, possibly allowing for bending of the N-terminal and proper removal from the core. After removal from the core the N-terminal helix reforms its helical structure at D14-E16.

In the core domain residues M247-V250 start helical formation of helix D in repeat III. The last residues to complete helix D formation are L251-E254. Over all three simulations we were able to see this two-step process for helical formation and consider the results consistent and a reliable possibility for the conformation change. The last form of conformation change was noticed at the calcium binding loops of the convex side of the core domain. Here secondary structure analysis showed that fluctuations in the secondary structure were occurring throughout the conformation change of annexin. In our study there was no calcium present and could have been a cause for these fluctuations.

Overall we were able to determine a possible mechanism for annexin A1’s conformation change from a calcium free to calcium bound structure. In detail we see the N-terminal remove itself, and repeat III of the core domain fold to form helix D. Without such conformational tools it is not possible to investigate such conformation changes. Based on our results we feel that the NEB method is reasonable to use for proteins of this size (343 residues) and confirm the ability to obtain consistent experimental results.

5.2 CONCLUSIONS FROM PROJECT 2: Molecular dynamic study of cTnC point mutations

In this work multiple MD simulations totaling 1,425 ns were performed on cTnC. These simulations included the wild type and five point mutations, D65A, S69C, A8V, L29Q, and C84Y. N-terminal domain mutations have adverse effects on the coordination of calcium at site II. The anchoring residues for the loop, D65 and E76, are vital for proper calcium coordination and have the strongest interaction with calcium. Mutation of D65 resulted in almost complete loss of binding for site II. Mutating the center of the binding loop (S69C) did not have as large of an impact on the binding. Loss in coordination for this mutation occurred at C69 and D73. Mutations in other portions of the N-terminal had effects on calcium binding of site II. This expresses the cooperative nature of proteins. A8V and C84Y mutants lead to an increase in calcium binding, while L29Q a decrease. A8V mutant increased binding at E66 and D75, while losing binding at T71. C84Y increased binding at S69 and E76, while decreasing binding at D73. For both of these mutations, the overall increase in calcium interaction outweighed the small loss in calcium binding of site II. All N-terminal mutations were consistent with experimental results for their effect on calcium binding.

The mutations performed in this work did not have an impact on the calcium binding of site III and IV. Mutating residues in the N-terminal domain, had only minor effects on the interaction was not enough of an effect to decrease the binding. Site III and IV does not have the same binding characteristic as site II, because these sites are not calcium regulated for muscle contraction. These binding loops are a lot stronger than site II. Alterations in calcium binding at site III and IV may be affected when mutations occur closer to these binding loops.

REFERENCES

1. Raynal, P. and Pollard, H. B. Annexins: the problem of assessing the biological role for a gene family of multifunctional calcium- and phospholipid-binding proteins. Biochim Biophys Acia. 1994, 1197: 63-93.
2. Gerke V. and Moss, S.E. Annexins: from structure to function. Physiol. Rev. 2002, 82: 331–371.
3. Creutz CE, Pazoles CJ, Pollard HB: Identification and purification of an adrenal medullary protein (synexin) that causes calcium-dependent aggregation of isolated chromaffin granules. J Biol Chem 1978, 253:2858-2866.
4. Creutz CE, Zaks WJ, Hamman HC, Crane S, Martin WH, Gould KL, Oddie KM, and Parsons SJ. Identification of chromaffin granule binding proteins. J Biol Chem 1987, 262: 1860–1868.
5. Moore PB, Kraus Friedmann N, and Dedman JR. Unique calcium dependent hydrophobic binding proteins: possible independent mediators of intracellular calcium distinct from calmodulin. J Cell Sci 1984, 72: 121–133.
6. Flower RJ. Background and discovery of lipocortins. Agents Actions 1986, 17: 255–262.
7. Glenney JR, Tack B, and Powell MA. Calpactins: two distinct Ca2-regulated phospholipid and actin-binding proteins isolated from bovine lung and placenta. J Biol Chem 1987, 104: 503–511.
8. Crumpton MJ and Dedman JR. Protein terminology tangle. Nature 1990, 345: 212.
9. Gerke V, Creutz CE, Moss SE. Annexins: linking Ca2+ signaling to membrane dynamics. Nat. Rev. Mol. Cell Biol., 2005. 6 (6): 449–61.

10. De Coupade C, Gillet R, Bennoun M, Briand P, Russo-Marie F, and Solito E. Annexin I expression and phosphorylation are upregulated during liver regeneration and transformation in antithrombin III SV40 large T antigen transgenic mice. Hepatology 2000, 31: 371–380.
11. Rosengarth A, Gerke V, and Luecke H. X-ray structure of full-length annexin I and implications for membrane aggregation. J Mol Biol, 2001. 306: 489–498.
12. Pollard HB, Guy HR, Arispe N, De La Fuente M, Lee G, Rojas EM, Pollard JR, Srivastava M, Zhang-Keck Z-Y, Merezhinskaya N, Caohuy H, Burns AL, and Rojas E. Calcium channel and membrane fusion activity of synexin and other members of the annexin gene family. Biophys J 1992, 62: 15–18.
13. Biener Y, Feinstein R, Mayak M, Kaburagi Y, Kadowaki T, and Zick Y. Annexin II is a novel player in insulin signal transduction. J Biol Chem 1996, 271: 29489–29496.
14. Larsson M, Majeed M, Ernst JD, Magnusson KE, Stendahl O, and Forsum U. Role of annexins in endocytosis of antigens in immature human dendritic cells. Immunology 1997, 92: 501–511.
15. Sarafian T, Pradel L-A, Henry J-P, Aunis D, and Bader M-F. The participation of annexin II (Calpactin I) in calcium-evoked exocytosis requires protein kinase C. J Cell Biol 1991, 114: 1135–1147.
16. Isas JM, Cartailler JP, Sokolov Y, Patel DR, Langen R, Luecke H, Hall JE, and Haigler HT. Annexins V and XII insert into bilayers at mildly acidic pH and form ion channels. Biochemistry 2000, 39: 3015– 3022.
17. Rosengarth, A. and Luecke, H. A calcium-driven conformational switch of the N-terminal and core domains of annexin A1. J. Mol. Biol. 2003, 326: 1317-1325.

18. Grewal T, Wason S, Enrich C, and Rentero C. Annexins- insights from knockout mice. Biol. Chem. 2016, 397(10): 1031-1053.
19. Lizarbe MA, Barrasa JI, Olmo N, Gavilanes F, Turnay J. Annexin-Phospholipid Interactions. Functional Implications. International Journal of Molecular Sciences. 2013, 14(2):2652-2683.
20. Rety S, Osterloh D, Arie JP, Tabaries S, Seeman J, Russo-Marie F, Gerke V, and Lewit-Bentley A. Structural basis of the Ca (2+)-dependent association between S100C (S100A11) and its target, the N-terminal part of annexin 1. 2000, 8(2): 175-184.
21. Garrett SC, Varney KM, Weber DJ, Bresnick AR. J Biol Chem. 2006, 281(2):677-80.
22. Salama I, Malone PS, Mihaimeed F, Jones JL. Eur J Surg Oncol. 2008, 34(4):357-64.
23. Grigorian M, Andresen S, Tulchinsky E, Kriajevska M, Carlberg C, Kruse C, Cohn M, Ambartsumian N, Christensen A, Selivanova G, Lukanidin E. J Biol Chem. 2001, 276(25):22699-708.
24. Ecsedi P, Kiss B, Gogl G, Radnai L, Buday L, Koprivanacz K, Liliom K, Leveles I, Vertessy B, Jeszenoi N, Hetenyi C, Schlosser G, Katona G, and Nyitray L. Regulation of the equilibrium between closed and open conformations of annexin A2 by N-terminal phosphorylation and S100A4 binding. Eub 2017, 25 (8): 1195-1207.
25. Vago JP, Nogueira C, Tavares L, Soriani F, Lopes F, Russo R, Pinho V, Teixeira M, and Sousa L. Annexin A1 modulates natural and glucocorticoid-induced resolution of inflammation by enhancing neutrophil apoptosis. Jour. Leukocyte Biol. 2012, 92: 249-258.
26. Stuqui B, Paula-Silva M, Carols C, Ullah A, Arni RK, Gill CD, and Oliani SM. Ac2-26 mimetic peptide of annexin A1 inhibits local and systemic inflammatory processes

induced by bothrops moojeni venom and the Lys-49 phospholipase A2 in a rat model. PLoS ONE 2015, 10(7): 1-18.

27. Rothhut B, Coméra C, Cortial S, Haumont PY, Diep Le KH, Cavadore JC, Conard J, Russo-Marie F, Lederer F. A 32 kDa lipocortin from human mononuclear cells appears to be identical with the placental inhibitor of blood coagulation. The Biochemical Journal. 1989, 263 (3): 929–35.
28. Hayes, M., Moss, S. Annexins and disease. Biochemical and Biophysical Research Communications. 2004, 322(4):1166-1170.
29. Madoiwa S, Someya T, Hironaka M, Kobayashi H, Ohmori T, Mimuro J, Sugiyama Y, Morita T, Nishimura Y, Tarumoto T, Ozawa K, Saito K, and Sakata Y. Annexin 2 and hemorrhagic disorder in vascular intimal carcinomatosis. Thrombosis Research 2007, 119(2): 229-240.
30. Swaminathan V, Mythreye K, O'Brien ET, Berchuck A, Blobe GC, Superfine R. Mechanical stiffness grades metastatic potential in patient tumor cells and in cancer cell lines. Cancer Res 2011, 71:5075–80.
31. Jaiswal JK, Lauritzen SP, Scheffer L, Sakaguchi M, Bunkenborg J, Simon SM, Kallunki T, Jaattela M, Nylandsted J. S100A11 is required for efficient plasma membrane repair and survival of invasive cancer cells. Nat. Commun. 2014, 5:3795.
32. Lauritzen, S., Boye, T., Nylandsted, J. Annexins are instrumental for efficient plasma membrane repair in cancer cells. Seminars in cell & developmental biology 2015, 45: 32-38.
33. Jaiswal JK, Nylandsted J. S100 and annexin proteins identify cell membrane damage as the Achilles heel of metastatic cancer cells. Cell Cycle 2015, 14:502–509.

34. Huxley AF and Niedergerke R. Structural changes in muscle during contraction-interference microscopy of living muscle fibers. Nature 1954, 173: 973-976.
35. Huxley H and Hanson J. Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation. Nature 1953, 172: 530-532.
36. Retrieved from https://www.slideshare.net/vajira54/muscle-contraction-7505328. Accessed 2019.
37. Ebashi S. Third component participating in super precipitation of actomyosin. Nature 1963, 200: 1010.
38. Takeda S, Yamashita A, Maeda Y. Structure of the core domain of human cardiac troponin in the Ca2+ saturated form. Nature 2003, 424: 35-41.
39. Wakabayashi T. Mechanism of the calcium regulation of muscle contraction- in pursuit of its structural basis. Proc Jpn Acad Ser B Phys Biol Sci 2015, 97(7): 321-350.
40. Kretsinger RH and Nockolds CE. Carp muscle calcium-binding protein. 2. Structure determination and general description. J. Biol. Chem 1973, 248: 3313-3326.
41. Retrieved from https://www.quora.com/What-is-the-difference-between-EF-Hand-and-helix-turn-helix-protein-motifs. Accessed 2019.
42. Gifford JL, Walsh MP, Vogel HJ. Structures and metal ion binding properties of the Ca2+ binding helix-loop-helix EF hand motifs. Biochem J 2007, 405: 199-221.
43. Schaub MC and Heizmann CW. Calcium, troponin, calmodulin, S100 proteins: from myocardial basics to new therapeutic strategies. Biochem Biophys Res Comm 2008, 369: 247-264.

44. Potter JD and Gergely J. The calcium and magnesium sites on troponin and their role in the regulation of myofibrilliar adenosine triphosphate. J. Biol. Chem. 1975, 250(12): 4628-4633.

45. Leavis PC and Kraft EL. Calcium binding to cardiac troponin C. Arch. Biochem. Biophys. 1978, 186:411-415.

46. Van Eerd JP and Takahashi K. The amino acid sequence of bovine of cardiac troponin-C. Comparison with rabbit skeletal troponin-C. Biochem. Biophys. Res. Comm. 1975, 64:122-127.

47. Robertsson SP, Johnson JD and Potter JD. The time-course of calcium exchange with calmodulin, troponin, paravalbumin, and myosin in response to transient increases in calcium. Biophys. J. 1981, 34:559-569.

48. Leavis PC, Rosenfield SS, Gergely J, Grabarek Z, and Drabikowski W. Proteolytic fragments of troponin C, Localization of high and low affinity calcium binding sites and interactions with troponin I and troponin T. J. Biol. Chem. 1978, 253:39-43.

49. Andersson JD, Drakenberg T, Forsen S, and Thulin E. A 43Ca NMR and 25Mg NMR study of rabbit skeletal muscle troponin C: exchange rates and binding constants. FEBS Lett. 125:39-43.

50. Zot HG and Potter JD. A structural role for the Ca^{2+}-Mg^{2+} sites on troponin-C in the regulation of muscle contraction- preparation and properties of troponin-C depleted myofibrils. J. Biol. Chem. 271(14): 8381-8386.

51. Li MX, Wang X, and Sykes BD. Structural based insights into the role of troponin in cardiac muscle pathophysiology. Journal of Muscle Res and Cell Mot. 2004, 25: 559-579.

52. Vinodgradova MV, Stone DB, Malanina GG, Karatzaferi C, Cooke R, Mendelson RA, and Fletterick RJ. Ca^{2+} regulated structural changes in troponin. PNAS 2005, 102:5038-5043.

53. Harada L and Morimoto S. Inherited cardiomyopathies as a troponin disease. Japanese journal of physiology 2004, 54:307-318.

54. Watkins H. Hypertrophic cardiomyopathy: from molecular and genetic mechanisms to clinical management. European Heart Journal Supplements 2001, 3: L43-L50.

55. Robinson P, Lipscomb S, Preston LC, Altin E, Watkins H, Ashley CC, and Redwood CS. Mutations in fast skeletal troponin I, troponin T, and β-tropomyosin that cause distal arthrogryposis all increase contractile function. FASEB 2007, 21:896-905.

56. Leblanc L, Bennett A, and Borgford T. Calcium affinity of regulatory sites in skeletal troponin-C is attenuated by N-cap mutations of helix C. Archives of Biochemistry and Physics 2000, 384: 296-304.

57. Kimber E, Tajsharghi H, Kroksmark AK, Oldfors A, and Tulinius M. A mutation in the fast-skeletal muscle troponin I gene causes myopathy and distal arthrogryposis. Neurology 2006, 67: 597-601.

58. Maron BJ, Gardin JM, Flack JM, Gidding SS, Kuosaki TT, and Bild DE. Prevalence of hypertrophic cardiomyopathy in a general population of young adults: echocardiographic analysis of 4111 subjects in the CARDIA study: coronary artery risk development in (young) adults. Circulation 1995, 92: 785-789.

59. Cohn JN, Bristow MR, Chien KR, Colucci WS, Frazier OH, Leinwand LA, Lorell BH, Moss AJ et al. Report of the national heart, lung and blood institute special emphasis panel on heart failure research. Circulation 1997, 95: 766-770.

60. McNally EM, Golbus JR, and Puckelwartz MJ. Genetic mutations and mechanisms in dilated cardiomyopathy. J Clin Invest 2013, 123(1): 19-26.

61. Weiner SJ, Kollmann PA. An all atom force field for simulations of proteins and nucleic acids. J. Comp. Chem. 1986, 7: 230-252.

62. Kollman PA et al. A new force field for molecular mechanical simulation of nucleic acids and proteins. J. Amer. Chem. Soc. 1984, 106: 765-784.

63. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The amber biomolecular simulation programs. J. Comp. Chem. 26: 1668-1688.

64. Norberg J, Nilsson L (2003) Advances in biomolecular simulations: methodology and recent applications. Quart. Revs. Biophys. 36: 257-306.

65. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical-integration of cartesian equations of motion of a system with constraints – molecular-dynamics of N-alkanes. J.Comp. Phys. 1977, 23: 327-341.

66. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J. Comp. Chem. 2004, 25: 1157-1174.

67. Jorgensen, WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 1983, 79: 926-935.

68. Crowley MF, Darden TA, Cheatham III TE, Deerfield II DW (1997) Adventures in improving the scaling and accuracy of a parallel molecular dynamics program. Journal of supercomputing 11:255-278.

69. Toukmaji A, Sagui C, Board J, Darden T (2000) Efficient particle-mesh Ewald based approach to fixed and induced dipolar interactions, Journal Chemical Physics 113:10913-10927.

70. Stoica I, Sadiq SK, Coveney PV. Rapid and accurate prediction of binding free energies for saquinavir-bound HIV-1 proteases. J. Amer. Chem. Soc. 2008, 130:2639-2648.

71. Retrieved from http://www.hotelsrate.org/binding-free-energy-equation/. Accessed 2019.

72. Marques AT, Fernandes PA, Ramos MJ. Molecular dynamics simulations of the amyloid-beta binding alcohol dehydrogenase (ABBAD) enzyme. Biorg. Med. Chem. 2008, 16:9511-9518.

73. Ferrari AM, Degliesposti G, Sgobba M, Rastelli G. Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors. Biorg Med Chem. 2007, 15: 7865-7877.

74. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990, 112 (16): 6127-6129.

75. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, DeBolt S et al. AMBER, a package of computer programs for applying molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Compu.t Phys. Commun. 1995, 91: 1-41.

76. Elber R, Karplus M. A method for determining reaction paths in large molecules: application to myoglobin. Chem. Phys. Letters 1987, 139:375-380

77. Matthews DH and Case DA. Nudged elastic band calculation of minimal energy paths for the conformational change of a CG non-canonical pair. J. Mol. Biol 2006, 357: 1683-1693.

78. H. Jónsson, G. Mills, K.W. JacobsenNudged elastic band method for finding minimum energy paths of transitions. B.J. Berne, G. Ciccoti, D.F. Coker (Eds.), Classical and Quantum Dynamics in Condensed Phase Simulations. World Scientific, Singapore 1998, 385-404.

79. Feig M, Onufriev A, Lee MS, Im W, Case DA and Brooks CL. Performance comparison of generalized born and poisson methods in the calculation of electrostatic solvation energies for protein structures. J. Comput. Chem. 2004, 25:265-284.

80. Tsui V and Case DA. Theory and applications of the generalized born solvation model in macromolecular simulations. Biopolymers 2001, 56:275-291.

81. Feig M and Brooks CL. Recent advances in the development and application of implicit solvent models in biomolecule simulations. Curr. Opin. Struct. Biol. 2004, 14:217-224.

82. Luecke H and Rosengarth A. A calcium-driven conformational switch of the N-terminal domains of annexin A1. J. Mol. Biol. 2003, 326:1317-1325.

83. Donohue M, Bartolotti L, and Li Y. The N-terminal of annexin A1 as a secondary membrane binding site: a molecular dynamics study. Proteins 2014, 82(11): 2936-2942.