Evolutionary Covariant Positions within Calmodulin EF-hand Sequences Promote Ligand Binding

Presented by Uma Vaidyanathan

In partial fulfillment of the requirements for graduation with the Health Science Scholars Honors Degree in Biochemistry

May 2018

Research Thesis supervised by Richard Aldrich, Ph.D.
Karl Folkers Chair in Interdisciplinary Biomedical Research II
Professor of Neuroscience, College of Natural Sciences
Center for Learning and Memory
The University of Texas at Austin

_____ _____

Richard W. Aldrich, Ph.D. Date
Supervising Professor

_____ _____

Jeffrey E. Barrick, Ph.D. Date
Honor's Advisor in Biochemistry

Evolutionary Covariant Positions within Calmodulin EF-hand Sequences Promote Ligand Binding

Department of Biochemistry

_____ _____

Uma Vaidyanathan Date

_____ _____

Richard W. Aldrich, Ph.D. Date
Supervising Professor

Table of Contents:

Abstract:

Intracellular calcium signaling is an essential regulatory mechanism through calcium-mediated signal transduction pathways involved in many cell processes, such as exocytosis, motility, apoptosis, excitability, transcription, and muscle contraction. The calcium-binding, ubiquitous, and highly conserved protein calmodulin (CaM) is an important regulator of hundreds of target proteins involved in cellular calcium signaling. CaM comprises of two pairs of EF-hand calcium-binding domains and these structural regions of the protein are highly conserved. Studying the molecular mechanisms underlying the binding of calcium to the EF-hands of CaM is critical in understanding the calcium-mediated cellular processes and how improper binding of calcium can lead to various human pathologies. Previous site-specific binding measurements indicate that each of the four EF-hands of CaM have distinct affinities for calcium. In this study, we have utilized covariance patterns and site-specific mutagenesis to analyze calcium affinity in the two EF-hands of the N-lobe of CaM in order to determine the specific amino acids that are evolutionarily conserved to coordinate calcium. The specific amino acids in CaM that we studied are theorized to coevolve, which means that in their protein coding genes, when a mutation occurs, a compensatory mutation is likely to follow to conserve structure and function of CaM. Since CaM is a highly conserved protein with a known structure, covariance analyses will help in understanding which amino acid contacts are most important for the coordination of calcium in the EF-hands of CaM and to determine which amino acids are under evolutionary constraint. Covariance algorithms, multiple sequence analyses and accompanied protein structure analyses were used to identify the two high scoring amino acid pairs in the N-lobe EF-hands: positions 22 and 24 in EF-hand site 1 and positions 58 and 60 in EF-hand site 2. The amino acids in these locations were mutated and accompanied calcium binding was measured to better understand the effects of the mutations on calcium binding. We have found that both the D24N mutation in site 1 and the D58N mutation in site 2 disrupt binding likely due to the removal of a necessary aspartate in the binding site. However, the combined D58N and N60D mutations restore binding in site 2 by providing the necessary aspartate in the covariant location. The N60D mutation by itself has little impact on calcium binding in site 2. Therefore, it is evident that evolution conserves at least one aspartate in the covariant positions of the binding site and the presence of two aspartates in the covariant positions of the binding site has little affect on calcium binding. We are currently studying the covariant positions in site 1 and future work includes structurally analyzing the covariant positions in the C-lobe of CaM and studying covariance patterns of other calcium-binding proteins with EF-hand binding domains.

1. Background

1.1. Cellular Calcium Signaling

Throughout evolution, eukaryotes have developed complex mechanisms to communicate between cells, regulate diverse functions, transport nutrients, and even commit programmed cell death utilizing highly regulated signaling pathways. Calcium (Ca^{2+}) is a regulated intracellular messenger that acts as a diverse signaling agent in the cell, activating a wide range of effector proteins, such as the protein calmodulin, ion channels, kinases and other enzymes. The binding of Ca^{2+} causes a conformational change in the proteins affecting their locations in the cell, associations with other proteins and molecules, and activities (Carafoli et al., 2001). The cell is extremely sensitive to small changes in cytoplasmic Ca^{2+} concentrations. Ca^{2+} can act as a secondary messenger and activate many different effector proteins, which will then go on to activate other proteins, resulting in a signal transduction pathway or an exponential activation of proteins. An example includes the G-protein coupled receptor and phospholipase C mediated pathway in which the secondary messenger inositol triphosphate (IP_3) binds to the IP_3-gated calcium channel on the endoplasmic reticulum membrane, resulting in Ca^{2+} release. Ca^{2+} then binds to the effector protein calmodulin, changing its conformation and allowing calmodulin to bind and activate CaM-kinase II, which then autophosphorylates to become fully activated. The CaM-kinase II protein family plays a large role in transcription factor regulation and metabolism. Other calcium-mediated signal transduction pathways can result in contraction, neurotransmitter release, membrane excitability, gene transcription, and enzyme activity (Carafoli et al., 2001). Ca^{2+}/calmodulin-dependent protein kinases are also found to play an important role in the progression of the cell cycle, and thus understanding the interactions between Ca^{2+}, calmodulin and the associated kinases can reveal how cancer cells disrupt these calcium-mediated responses

involved in the cell cycle to prompt rapid cell proliferation (Kahl and Means, 2003). Evidently, calcium signaling is vital to the cell for survival and apoptosis alike. The calcium-mediated cellular processes are complex with many intermediaries and activators to ensure that these important cell functions are highly regulated and controlled.

1.2. Calmodulin: Structure and Function

The calcium-binding, ubiquitous, and highly conserved protein calmodulin (CaM) is an important regulator of hundreds of target proteins involved in cellular calcium signaling and is central to the field of cell biology. CaM acts as an intermediary protein in signal transduction pathways by detecting intracellular calcium levels and activating enzymes, ion channels, and other proteins that serve various functions in the cell, such as exocytosis, motility, apoptosis, excitability, transcription, and muscle contraction. When CaM is activated by binding to Ca^{2+} in the event of an increase in calcium ions in the cell, CaM will undergo a conformational change and bind to other downstream effector proteins (Hoeflich and Ikura, 2002). This results in a calcium-mediated signal transduction pathway exponentially activating various proteins, such as transcription factors, to induce a rapid cellular response.

CaM from the human brain is a 148 amino acid protein, which has a molecular weight of approximately 16.7 kDa (Sasagawa, 1982). CaM is expressed in all eukaryotic cells and comprises of two pairs of EF-hand domains and these structural regions of the protein are highly conserved. The EF-hand domains serve as the calcium-binding sites in CaM and the calcium affinity at each binding site is studied to better understand the interactions of CaM in calcium signaling. CaM can bind up to four Ca^{2+} ions, as it contains two pairs of EF-hands: EF-hands site 1 and site 2 are located in the N-terminus of CaM and EF-hands site 3 and site 4 are located in

the C-terminus of CaM (Figure 1.2.1 below). The two globular domains, the N-lobe and C-lobe that contain the EF-hand binding site pairs, are connected by a central linker helix. When calmodulin binds to Ca^{2+}, it experiences a conformational change, or an opening between alpha-helices, exposing hydrophobic amino acid surfaces, which allows it to interact with other proteins (Haeseleer et al., 2002). Specifically, Ca^{2+} binding to calmodulin causes a conformational change in the protein's structure and results in an exposure of a hydrophobic cleft in the N-lobe of CaM (Vigil et al., 2001). Thus, when experimentally studying and purifying CaM, it is important to ensure that calcium is not exposed to CaM to change its characteristics, especially for binding to the hydrophobic column in the C18 purification experiments.

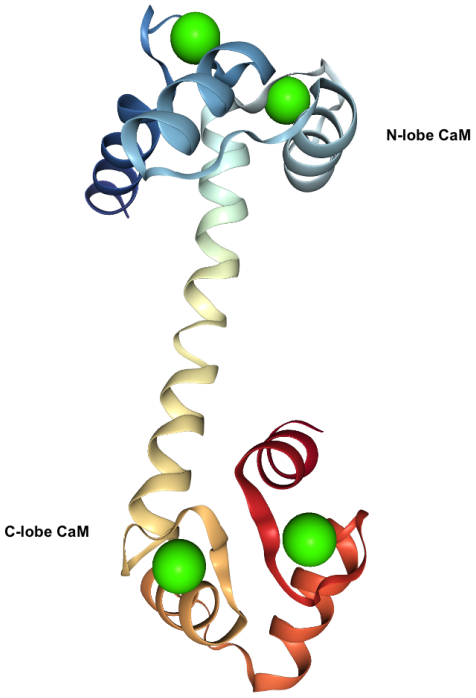

Figure 1.2.1: Ribbon diagram of the calmodulin protein representing the secondary structures and 3-dimentional configuration of the structure. EF-hands Site 1 and Site 2 are located in the N-lobe of CaM and EF-hands Site 3 and Site 4 are located in the C-lobe of CaM. Calcium ions are portrayed as green spheres. Image created using PyMOL with RCSB PBD entry 3CLN from Babu et al., 1988.

Each of the EF-hand binding sites have distinct sequences and affinities for Ca^{2+}, perhaps indicating that each binding site has a particular role in CaM functioning in calcium-mediated

pathways (Halling et al., 2016; Greeson-Bernier et al., 2013). Highly conserved amino acids are found at different positions when comparing the distinct sequences of the EF-hand binding sites, and the sequences of the EF-hand domains have been evolutionarily conserved perhaps to maintain the distinct structures and roles in calcium signaling of each of the EF-hand domains in CaM (Halling et al., 2016) (Figure 1.2.2 below).

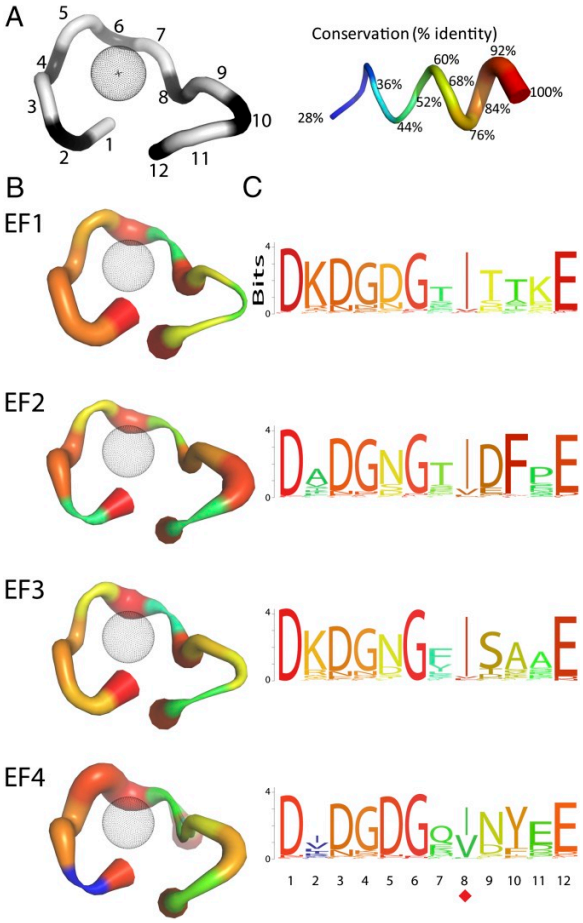

Figure 1.2.2: Comparison of the four EF-hand binding sites in terms of structure and conservation of amino acid sequences. All four of the EF-hand binding sites have distinct amino acid sequences that have likely been evolutionary selected for to conserve the structures of each binding site and function of CaM (Halling et al., 2016).

Additionally, the EF-hand binding sites of CaM are able to bind to luminescent lanthanide ions, such as terbium (Tb^{3+}) in the absence of Ca^{2+}, to measure and quantify binding of the metal ion to the EF-hand binding sites of CaM (Edington et al., 2018; Martin and

Richardson, 1979). However, recent studies have indicated that the binding of lanthanide ions to CaM makes the conformation of the protein more flexible and disrupts the geometric coordination of metal ions in the binding sites of CaM (Edington et al., 2018).

1.3. EF-hand Domains and Calcium-binding Proteins

There are 45 subfamilies of EF-hand proteins, and each of the protein families have distinct amino acid sequences and have a range of two to eight EF-hand domains; some of these protein families include calmodulin, troponin C, essential light chain of myosin, brain protein S-100, calpain, paralbumin, calcineurin, sorcin, and aequorin (Kawaski et al., 1998). By studying Ca^{2+}-binding proteins in the retina, researchers discovered neuronal Ca^{2+}-binding proteins (NCBPs) and Ca^{2+}-binding proteins (CaBPs) that have similar properties to CaM (Haeseleer et al., 2002). The EF-hand domain has been found in many organisms throughout the course of evolution and has a wide range of functions depending on the different calcium-binding proteins (Moncrief et al., 1990). Comparing the EF-hand binding domains of different protein families enables a better understanding of the important structural characteristics of EF-hands that contribute to its wide range of functions.

1.4. Medical Relevance: Learning and Memory

CaM is involved in calcium-mediated signal transduction pathways involved in learning and memory in the brain. The calmodulin-dependent kinase cascade is involved in memory consolidation in mammalian brains, as the kinases are able to function in a calcium-mediated signal transduction cascade to phosphorylate and activate many effector proteins (Giese and Mizuno, 2013). The calmodulin-dependent protein kinase cascade involves three kinases:

CaMKK, CaMKI and CaMKIV and these kinases respond to influxes of intracellular calcium levels in brain cells and T cells (Soderling, 1999). Specifically, CaMKIV plays a major role in regulating the transcription of CREB, which also regulates gene expression and is an important transcription factor in cell signaling, resulting in the expression of proteins involved in memory formation, consolidation, and extinction (Giese and Mizuno, 2013). These calmodulin-dependent protein kinases can also activate other kinases such as PKA and PKB, which act in memory processes and the signaling pathway for cell survival. In memory formation, kinases such as CaMKII help in the regulation of ion channel density and conductivity that aids in synaptic communication as well as the regulation of the translation of proteins contributing to synaptogenesis (Giese and Mizuno, 2013; Lisman et al., 2002).

Calmodulin has been found to play a critical role in the development of Alzheimer's disease evident in the analysis of CaM-binding proteins that are responsible for creating beta-amyloid plaques characteristic of Alzheimer's disease (O'Day, 2015). The CaM-binding proteins involved in the formation of these plaques include amyloid-β protein precursor, β-secretase, presenilin-1, and ADAM10. The understanding of calmodulin in relation to Alzheimer's disease may allow for the development of pharmaceuticals to potentially treat the symptoms of the disease.

Calmodulin has also been studied in relation to small conductance Ca^{2+}-activated potassium channels (SK channels), which are widely abundant in the central nervous system and are activated by changes in the intracellular Ca^{2+} concentrations in action potentials (Adelman, 2012). SK channels are characterized by their small single channel conductance and allow potassium to cross the cell membrane when activated, due to an increase in intracellular Ca^{2+}. The activation of SK channels causes a hyperpolarization of the cell, as potassium exits the cell

through the channels along their concentration gradients into the extracellular space. SK channels are characteristic of macromolecular complexes of pore-forming subunits bound to CaM. Calmodulin, when bound to the SK channels, acts as a Ca^{2+} gating subunit and regulates activity of the channel. SK channels are known to regulate somatic excitation of neurons and expression of synaptic plasticity, so they are directly involved with learning and memory in the brain. Small conductance Ca^{2+}-activated potassium channels are especially important in understanding calcium influx through voltage-gated calcium channels in cardiac cells. SK channels have been specifically characterized to have significant functions in human atrial fibrillation (Zhang, 2015). Thus, SK channels could be used as therapeutic targets in treating atrial arrhythmias and SK channels could be up-regulated in the case of heart failure. SK channels have been found to be both anti-arrhythmic and pro-arrhythmic. Discovering SK channel inhibitors could bring forth an understanding of possible treatments of atrial fibrillation. CaM is an intrinsic subunit of SK channels and function in the regulation of the channel when exposed to fluctuations in calcium levels (Li and Aldrich, 2009). Additionally, the interaction between the kinase CK2 and phosphatase PP2A serve to regulate the activity of calmodulin by phosphorylation, thus regulating the activity of the protein-channel complex (Adelman, 2012). SK channels are composed of four subunits that form a tetramer, and each of the subunits has six transmembrane domains that span the plasma membrane of the cell. The C-terminus of CaM (C-CaM) along with the linker region of CaM binds to SK channels. However, the EF-hands bound to calcium in the N-lobe of CaM is necessary for proper SK channel functioning and gating, but it is unknown whether the N-lobe physically binds to SK channels (Li et al., 2009; Halling et al., 2014).

1.5. Medical Relevance: Cardiac Channelopathies

Multiple channelopathies in human cardiomyocytes have been found to be associated with mutations in the genes that code for CaM. Defects in CaM significantly impact the protein's ability to bind to calcium, which negatively affects its roles in calcium signaling involved in ion channel functioning in cardiac cells, resulting in life-threatening cardiac arrhythmias. Specifically, the interaction between Ca^{2+}-bound CaM and L-type Ca^{2+} channels is disrupted. The inactivation of voltage-gated Ca^{2+} channels is disrupted when a mutation in CaM is present, which impacts the ability of calcium to bind to the C-lobe of CaM, leading to cardiac arrhythmias (Pipilas et al., 2016). Mutations in either the CALM1 or CALM2 gene, which code for CaM, cause ventricular arrhythmias and long QT syndrome early in life (Crotti et al., 2013). Long QT Syndrome affects the repolarization of cardiac cells and is characterized by a prolonged QT interval on the ECG, resulting in arrhythmias and the possibility of fatal heart attacks. Mutations in the CALM3 gene result in the manifestation of Long QT Syndrome as well (Reed et al., 2014). Overlapping clinical symptoms of both Long QT syndrome as well as catecholaminergic polymorphic ventricular tachycardia (CPVT) have been evident in patients with mutations in the CALM2 gene (Makita et al., 2013). Catecholaminergic polymorphic ventricular tachycardia (CPTV) is a rare heart condition in which an increase in heart rate can suddenly trigger ventricular tachycardia, and CPTV has been diagnosed for patients with mutations in the CALM1 gene (Nyegaard et al., 2012). Additionally, mutations in the CALM1 gene cause idiopathic ventricular fibrillation (IVF) in children and young adults (Marsman et al., 2014).

1.6. Covariance

A central concept in our experimental design and structural binding analyses is the idea of covariance. The basic principle of covariance or coevolving amino acids is that in the protein coding genes, when a mutation occurs, a compensatory mutation in another position is likely to follow to conserve structure and function of the protein of interest. These amino acids in the covariant positions can be seen in the amino acid sequences across eukaryotic species. The covariation signal is a statistical measure that is measured by comparing the sequences of a protein from a diverse range of species and across a large range of homologs (Ashenberg and Laub, 2013).

A multiple sequence alignment (MSA) is used to measure the coupling covariance pattern between amino acid pairs in the homologous protein sequences. By comparing the amino acid positions and changes in the MSA, the amino acids that are likely in contact in the folded spatial conformation of the protein can be determined. Two positions in the MSA that are correlated are likely in contact in the three-dimensional spatial structure of a protein (Ekeberg et al., 2013). Thus, by comparing amino acid positions with the MSA, we can determine which pairs might be in contact or coevolve (Figure 1.6.1 below). The covariation algorithm used in our study is conducted in two parts: the creation of the multiple sequence alignment (MSA) and the statistical comparison and coupling measurement of pairs of different positions in the MSA. We utilized a global probabilistic model or a maximum-entropy model that assessed the likelihood of covarying amino acids at each position or an inverse statistical analysis (Anishchenko et al., 2017; Ekeberg et al., 2013). This model algorithm is called the pseudolikelihood maximization procedure, or plmDCA (Ekeberg et al., 2013). This algorithm is a Potts model inference as it utilizes only the protein sequences and increases its accuracy with larger sets of multiple

sequence alignments. The algorithm accounts for the direct coupling scores by using parameters of a Markov random field (Ekeberg et al., 2013). When comparing four different covariance methods, the extent to which background conservation signals are accounted for is different between each of the methods (Fodor and Aldrich, 2004). Thus, it is highly possible that some algorithms may filter background signals more than others, which results in a variability in the amino acid positions that are defined as high scoring for coevolving (Fodor and Aldrich, 2004). Due to the limitations with covariance algorithms, it is important to experimentally structurally analyze the hypothesized covarying amino acids within the protein of interest, as we have performed in this study.

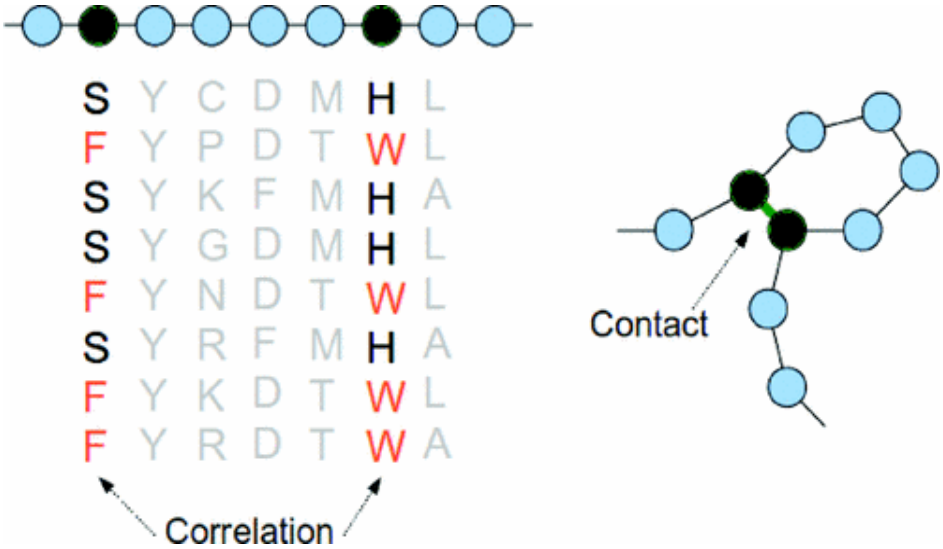

Figure 1.6.1: Multiple Sequence Alignment (MSA) with two correlating positions and a hypothetical spatial conformation (Ekeberg et al., 2013).

Covariance analyses can be very helpful in structural biology to roughly predict or infer which amino acids are in contact to help construct protein structures. Covariance studies on protein families have successfully predicted amino acids that contribute to protein folding, interactions between proteins and the amino acids necessary for proper enzymatic function

(Ashenberg and Laub, 2013). Covariance techniques are also very sensible because X-ray crystallography and NMR tomography have limitations, as it is sometimes very difficult and complicated to crystallize the fragile nature of some protein crystals and procedures may be costly and time-consuming (Ekeberg et al., 2013). Also, the number of known sequenced genomes is increasing exponentially with the advancements in sequencing technology. With proteins of known structures, covariance analyses can be used to determine which amino acids are important for the spatial configuration of the protein, inferring folding patterns, and determining which amino acids are under evolutionary constraint. However, correlation determined by covariance algorithms has limitations as well, especially considering distant coevolving amino acids that may not always be in contact in spatial protein structure. High scores for covariance between distant amino acids may not be because they are physically in contact with each other, but rather because of contacts through chaining effects, allosteric interactions, codon effects, or phylogenetic correlations (Anishchenko et al., 2017). When working with highly variant proteins, it is often difficult to sort through the false positive high scoring amino acid pairs (Anishchenko et al., 2017). However, evolution seems to operate more strongly on the physically interacting amino acids, than the ones involved in long-range allosteric networks (Anishchenko et al., 2017).

Additionally, covariance analyses comparing different protein sequences can reveal which amino acids in the different sequences might be interacting between proteins (Figure 1.6.2 below). Covariance analyses within protein families is helpful in understanding which amino acid pairs collaborate to support protein folding interactions, conserve the active site of enzymes, or uncover protein-protein interactions (Ashenberg and Laub, 2013).

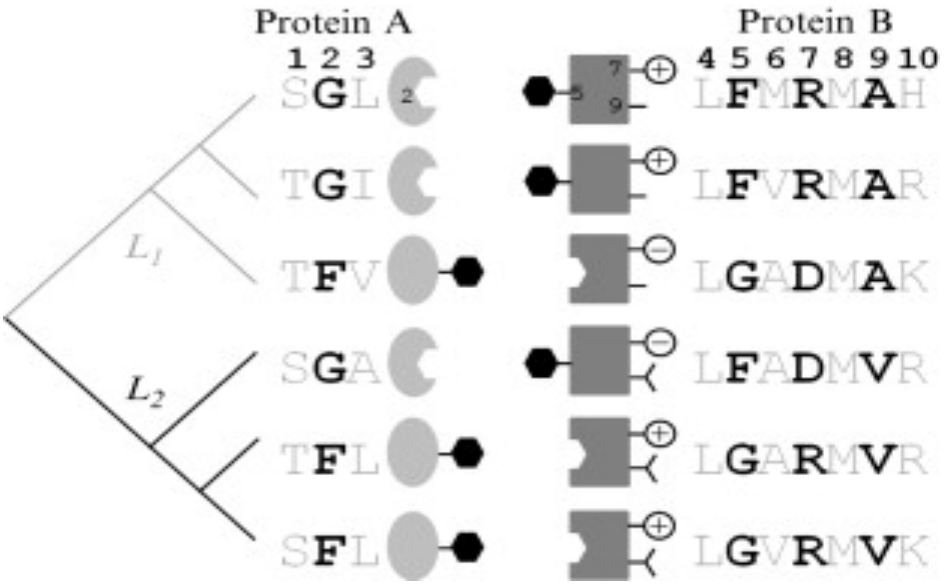

Figure 1.6.2: Covariation patterns between two interacting and homologous proteins, protein A and protein B. The multiple sequence alignment (MSA) indicates that the proteins come from similar species and that positions 2 and 5 covary to conserve hydrophobic interactions (Ashenberg and Laub, 2013).

2. Purpose: Specific Aims of the Project

Understanding the molecular mechanisms underlying how calmodulin interacts with other molecules and proteins will allow for a better understanding of how calmodulin mutations cause various pathologies and to better understand ion channel regulation and other calcium-mediated cellular processes. The Aldrich lab conducts a significant amount of research regarding the biophysical mechanism by which calcium ions bind to the EF-hand binding domains of CaM and to analyze the site-specific manner and cooperativity of binding at these sites. Previous work has indicated that the four EF-hand binding sites in CaM have distinct affinities for Ca^{2+} and it is likely that there are evolutionary selective pressures to conserve the distinct sequences of each EF-hand domain in CaM (Halling et al., 2016; Greeson-Bernier et al., 2013).

A specific aim of our project in the lab is to understand how mutations in covariant positions in the EF-hand binding domains of CaM impact the structure and function of the protein. Covariance patterns and site-specific mutagenesis are used to measure changes in calcium affinity of the mutated EF-hand domains of the N-lobe of CaM. The different binding measurements allow us to compare different mutations at covariant positions within an EF-hand domain and to determine the specific amino acids in that particular EF-hand that are evolutionarily conserved to coordinate Ca^{2+}. The amino acids that were studied received high scores in the covariance analysis and thus are theorized to coevolve to preserve the distinct EF-hand structures. The goal of our research is to test whether the basic principles of covariance apply to CaM EF-hand binding domains, in which when a mutation occurs in the CaM encoding gene, a compensatory mutation follows to conserve calcium binding. Our research is unique because it applies the concept of covariance, which is usually used to predict protein structure, on a protein with a known crystallized structure. Since CaM is a highly conserved protein with a

known structure, covariance analyses indicate which amino acid contacts are under selective pressure and are most important for Ca^{2+} coordination in the EF-hand binding domains. Our specific questions include: do amino acids that are structurally close coevolve in order to coordinate Ca^{2+} in the EF-hand binding sites of CaM? Does disrupting the covariance pattern disrupt Ca^{2+} binding? Can we restore binding by restoring the covariance pattern?

By utilizing covariance algorithms and multiple sequence analyses, the four amino acids pairs that theoretically coevolve in the EF-hand binding domains of CaM were identified as positions 22 and 24 in EF-hand site 1, positions 58 and 60 in EF-hand site 2, positions 138 and 140 in EF-hand site 3, and positions 159 and 160 in EF-hand site 4. The covariant positions in EF-hand site 1 and EF-hand site 2 of the N-lobe of CaM were studied because the lab has conducted a significant amount of work on the N-lobe of CaM, and thus expression and purification techniques for N-lobe mutant CaM proteins are known to be relatively successful. Consequently, the amino acids in positions 22, 24, 58 and 60 were mutated to understand the structural effects of these mutations in relation to our hypotheses from analyzing the covarying positions (Figure 2). For site 1, we hypothesize that the D24N mutation and D22N mutation would disrupt binding due to the removal of a necessary aspartate amino acid; however, the combined D24N & T28D mutations as well as the combined D22N & T28D mutations would restore binding due to the restoration of the aspartate amino acid. Similarly, for site 2, we hypothesize that the D58N mutation would disrupt binding due to the removal of an aspartate amino acid in the 58 position of the EF-hand protein sequence; however, binding would be restored with inducing both the D58N & N60D which restores the necessary aspartate amino acid at the covariant position 60. To test our hypotheses, site-specific mutagenesis at positions

22, 24, 58 and 60 and accompanied binding measurements are conducted to better understand the effects of the mutations on Ca^{2+} binding.

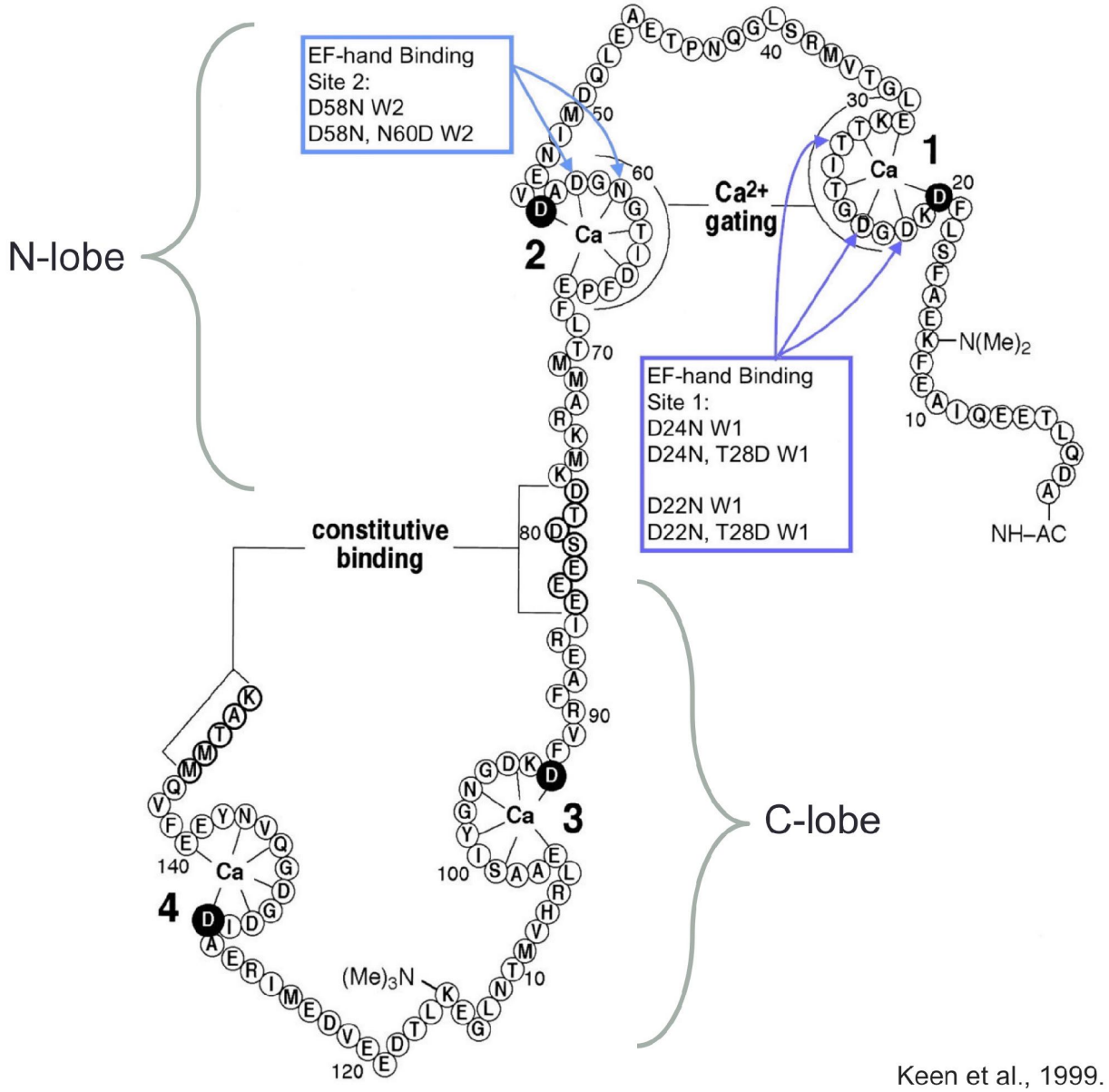

Figure 2: The point mutations chosen after analyzing the four highest covariance-scoring positions within the EF-hand binding sites of the N-lobe of CaM are positions. Calmodulin structural and sequence figure is adapted from Keen et al., 1999.

3. Materials, Methods and Data Collection

3.1. Multiple Sequence Alignment and Covariance Analysis

CaM is expressed in all eukaryotic cells and we aimed to study CaM amino acid sequences by using a multiple sequence alignment (MSA) representing species across all eukaryotes. The CaM amino acid sequences were obtained using the protein sequences from the evolutionary conserved amino acid study on CaM previously performed in the Aldrich lab (Halling et al., 2016). Halling et al. used the phylogenetic tree presented by Torruella et al. to collect CaM sequences from a diverse set of eukaryotes with a slight preference for yeast because yeast are evolving CaM the fastest and have more diversity in their CaM sequences (Torruella et al., 2011; Halling et al., 2016) (Figure 3.1.1 below).

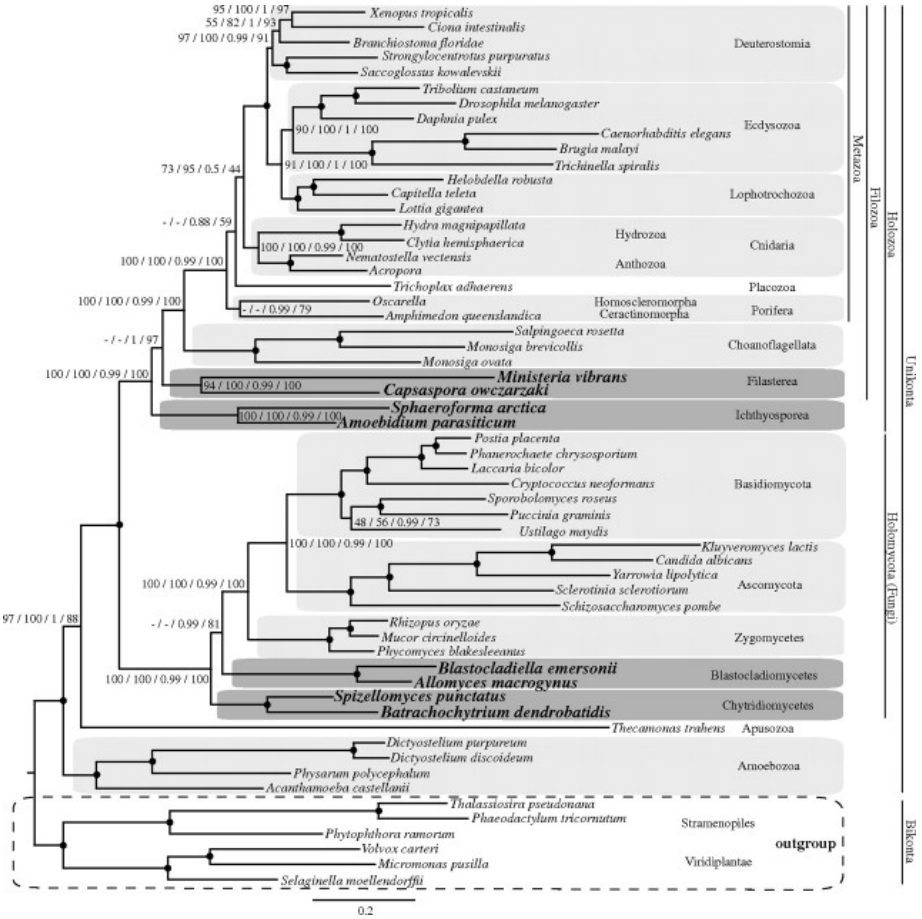

Figure 3.1.1: Species used for the creation of the multiple sequence alignment utilizing the phylogenetic tree constructed by Torruella et al., 2011.

The National Center for Biotechnology Information (blast.ncbi.nlm.nih.gov/Blast.cgi) gene databank was used to obtain the CaM protein sequences. A forward BLAST of the human CaM protein sequence against eukaryotic species and a following reverse BLAST of the potential CaM protein sequences from eukaryotic species against the human CaM sequence were performed to obtain the CaM protein sequences. Our criteria for the CaM sequences collected were different from the Halling et al., 2016 study in that we aimed to obtain CaM sequences that only contained four EF-hand domains and were functional CaM genes under the same selective pressures. For example, gene duplicates of CaM genes that were under different selective pressures were undesirable.

Then, we trimmed down the multiple sequence alignment of 442 sequences and we checked each sequence for only four EF-hand domains in the BLAST gene databank for conserved domains. We also ensured that the BLAST searches only returned CaM protein sequences rather than other similar proteins.

The statistical comparison and covariance of pairs of different positions in the sequences were computed using MATLAB and the plmDCA program, based on the pseudolikelihood maximization algorithm proposed by Ekeberg et al., 2013. The algorithm utilizes a maximum-entropy model that assessed the likelihood of covarying amino acids at each position and the direct coupling scores were measured by using parameters of a Markov random field (Anishchenko et al., 2017; Ekeberg et al., 2013). The algorithm calculates the frequencies of the amino acids at each position and the frequencies of every possible amino acid pair in different positions. The equation $f_i(k) = \left(\frac{1}{B}\right)\sum_{b=1}^{B}\delta(\sigma_i^{(b)}, k)$ describes the pairwise frequency calculation with B number of aligned sequences and σ representing the amino acid sequence of a protein domain (Ekeberg et al., 2013). The frequencies determined by this equation are used to create a

covariance matrix with the correlations using the equation $c_{ij}(k,l) = f_{ij}(k,l) - f_i(k)f_j(l)$, where $f_i(k)$ is the fraction of sequences that the possibility of position i is amino acid k (Ekeberg et al., 2013). The MSA of all of the processed CaM protein sequences was entered into the program. The algorithm compares the identities of the amino acids in each position column. The outputs of the covariance algorithm were scores based on each position pairing. The algorithm utilizes a pairwise column comparison and assesses the amino acid identities in each column and whether positions depend on one another. If the two positions vary independently, then the output covariance score is zero. The distribution of the scores was used and a histogram was created to identify the highest scores that would indicate which amino acids were most likely coevolving and which amino acid positions were covariant positions (Figure 3.1.2 below). Most amino acid pairs do not covary, which is why the data in the histogram in Figure 3.1.2 clusters around zero; however, the amino acid pairs with the highest covariance scores are seen as outliers in the histogram. The top four amino acids pairs in the entire CaM gene sequence that theoretically coevolve according to the covariance statistical tests performed were positions 22 and 24 in EF-hand site 1, positions 58 and 60 in EF-hand site 2, positions 138 and 140 in EF-hand site 3, and positions 159 and 160 in EF-hand site 4.

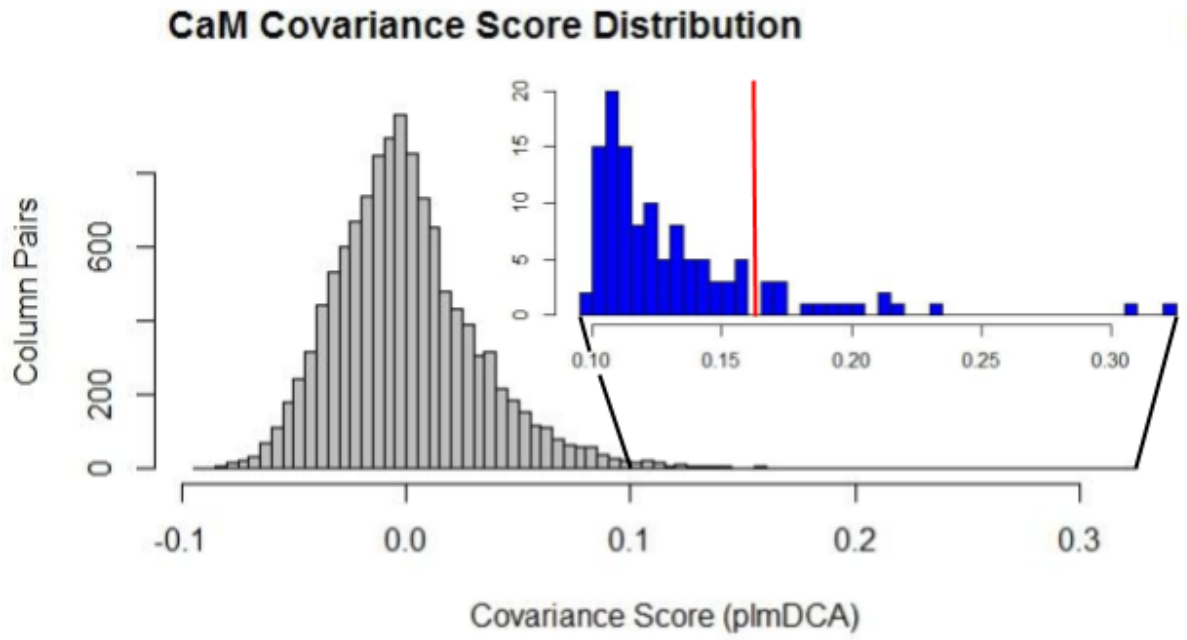

Figure 3.1.2: The histogram of the covariance scores of the multiple sequence alignment of CaM protein sequences. The inlaid portion is a display of the highest scores, the red bar indicating the top 0.1% of covariance scores. Four out of the highest scoring five pairs are positions 3 and 5 within each of the four EF hands; this corresponds to positions 22 and 24 in EF-hand site 1, positions 58 and 60 in EF-hand site 2, positions 138 and 140 in EF-hand site 3, and positions 159 and 160 in EF-hand site 4. Covariance scores obtained by Suzanna Bennett.

3.2. Site-specific Mutagenesis and Structural Pattern of Covarying Amino Acids

In order to analyze the structural pattern of covarying amino acids in the EF-hand site 1 and site 2 of CaM, site-specific mutagenesis utilizing polymerase chain reaction was performed. Specific primers were designed to ensure that the point mutations desired were induced and that the reporter tryptophan amino acid was also incorporated into the binding site. A combination of phusion buffer, dNTPs, primers, base DNA, OmniPure water, and DNA polymerase phusion enzyme were combined to perform the PCR reaction. The PCR product was transformed into E. Coli Top10 cells, the DNA plasmids were purified, and the DNA sequence was verified using Sanger sequencing. Sequencing was particularly important because it identified the constructs

that had the desired point mutation and tryptophan amino acid in the EF-hand binding site of CaM that would be used in protein expression.

We then used a bacterial expression system to generate the mutated CaM. The verified sequenced DNA was transformed into E. Coli BL21 cells. We utilized a specific protocol for the expression to ensure there was little contamination and the cells achieved optimum growth after a few trial and errors. We prepared two 20 mL starter cultures in LB media with carbenicillin by selecting a single BL21 colony transformed with the desired CaM construct with a toothpick and placing the toothpick into the starter cultures. The starter cultures were placed in a 37 °C incubator and shaker. After 7 hours, the starter cultures were placed into the fridge at 4 °C overnight in order to halt growth. The next day, the starter cultures were used to inoculate two 750 mL of LB and carbenicillin media and the cultures were placed in the 37 °C incubator and shaker. A spectrophotometer, measuring absorbance at a wavelength 600nm, was used to monitor the growth of the bacterial cultures every half hour until the optical density of the culture was between 0.6 and 0.7. When growth had reached an optical density of 0.6-0.7, 0.5 mM IPTG was added to each 750 mL bacterial culture. Isopropyl β-D-1-thiogalactopyranoside (IPTG) is a reagent that binds to the lac repressor protein on the DNA, causing it to release from the DNA binding site, which allows the T7 RNA polymerase gene to be accessible for transcription or the production of mRNA. The T7 RNA polymerase is then able to start the transcription process to create the mRNA by binding to the promoter and beginning the production of CaM by translation of the mRNA. The cultures grew for two to three hours exposed to IPTG in the induction step. The cells were then harvested using a series of centrifugation steps and the cell pellets were kept frozen at -80 °C.

Next, we purified the mutant CaM proteins by utilizing a series of high performance liquid chromatography (HPLC) experiments. We performed CaM purifications using an initial desalting step, anion exchange column, C18 column, and a final desalting step. First, we prepared a bacterial protein extraction reagent (B-Per) buffer solution for solubilizing the CaM cell pellets. The B-Per solution is composed of protease inhibitors, 1 mM EDTA, 0.1 mM EGTA, 2 mM $MgCl_2$, and 1 mM DTT, and Benzonase (or a universal nuclease). The frozen cell pellets were resuspended in the B-Per solution and were rotated for half an hour to fully lyse the cells and break up the nucleic acids present in the cell lysate. The lysate was then centrifuged at 20 kg for 20 minutes and the supernatant was extracted. Ammonium sulfate precipitation was then performed to separate the CaM protein based on its different solubility properties in 60% and 80% ammonium sulfate; CaM is soluble in 60% ammonium sulfate and insoluble in 80% ammonium sulfate. Dry ammonium sulfate was added to the supernatant to achieve 60% saturation and the solution was stirred for 1 hour at 4 °C and then centrifuged at 30 kg for 30 minutes. Then, dry ammonium sulfate was again added to increase the saturation to 80% and the total supernatant was collected and pH adjusted to 4.2 using glacial acetic acid. The solution was stirred for 1 hour at 4 °C and then centrifuged at 30 kg for 30 minutes. The extracted protein pellets were then collected. The pellet was suspended in 100 mM Tris at a pH of 7.4 with protease inhibitors and the solution was centrifuged at 50 kg for 30 minutes. The next step was to perform the initial desalt step on the HPLC with 20 mM Tris at a pH of 7.4 to initially separate the protein from contaminating salts from the ammonium sulfate precipitation steps (Figure 3.2.1 below). The desalting step is necessary for the protein to bind to the Q-HP column in the following anion exchange column. The desalt column is made of porous beads made from the sephadex material and operates as a molecular sieve. The beads have small pores, which allow

the small molecules to pass slowly through the small pores due to friction. Whereas, the large protein molecules move quickly past the spaces between the beads, rather than being stuck in the small pores. The proteins fall through the column fastest and have the least distance to travel and the salts move through the matrix very slowly. This is evident in the initial desalt data, as the conductivity trace, representing salts, falls after the absorbance at 280nm trace, representing the protein. The initial desalt allows for the protein to be eluted and the salts to be discarded.

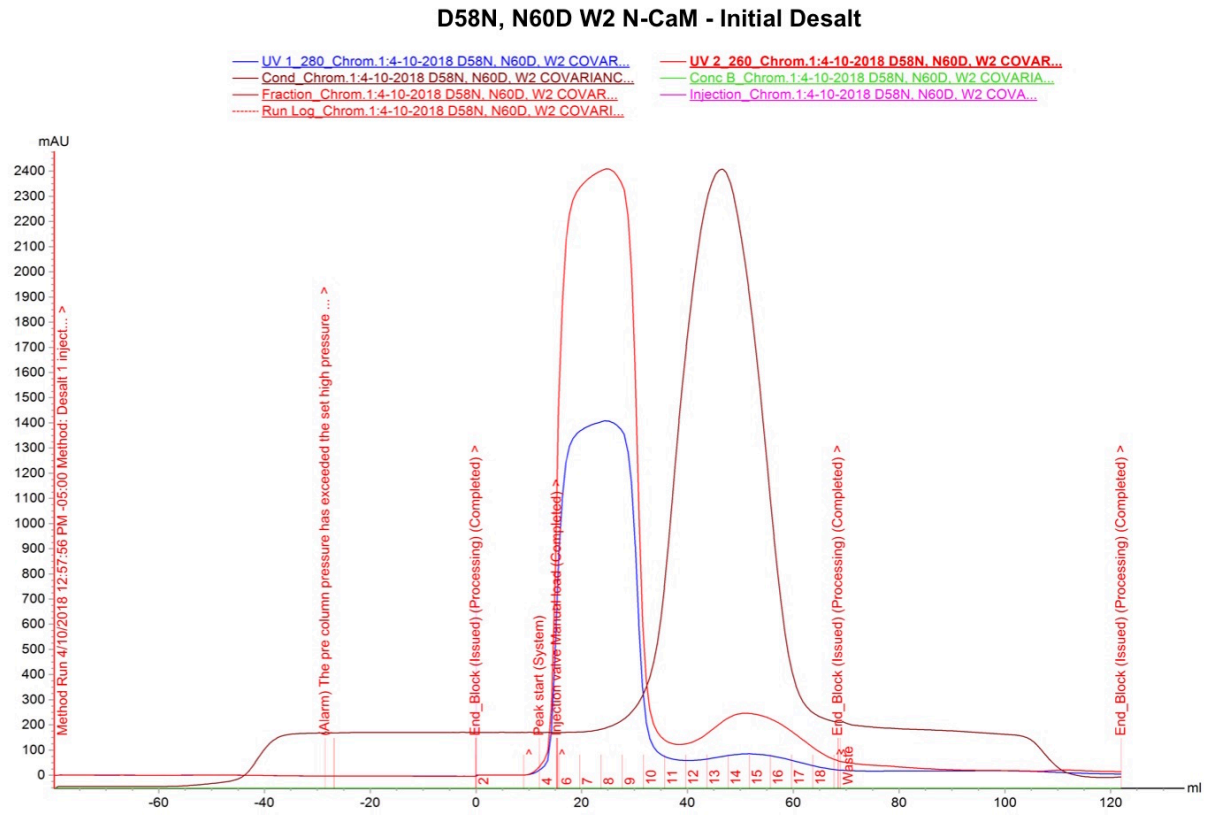

Figure 3.2.1: Initial Desalt for CaM construct D58N, N60D W2 N-CaM. The red trace corresponds to the absorbance at a wavelength of 260nm, representing contaminating nucleic acids. The blue trace corresponds to the absorbance at a wavelength of 280nm, representing the CaM construct protein. The brown curve is the conductivity to show the presence of salts. Evidently, there are contaminating nucleic acids present, however after the anion exchange and C18 purification steps, the red trace decreases substantially, indicating a more purified protein sample.

Next, the desalted protein was injected into the Q-HP anion exchange column, in which the protein was eluted with an increasing salt concentration using 20 mM Tris buffer and Tris+

1.5 M NaCl (Figures 3.2.2 and 3.2.3 below). CaM is slightly negatively charged, so it is able to bind to the positively charged quaternary amino group that is coupled to the anion exchange resin. An increasing concentration of NaCl salt is used to elute the protein because the negative chloride ions compete with calmodulin for binding to the anion exchange column. Calmodulin is eluted with the NaCl salt concentration of 16% of 1.5 M NaCl and CaM falls off early in the salt gradient, indicating that CaM binds to the column, but has a weak affinity for the positively charged column.

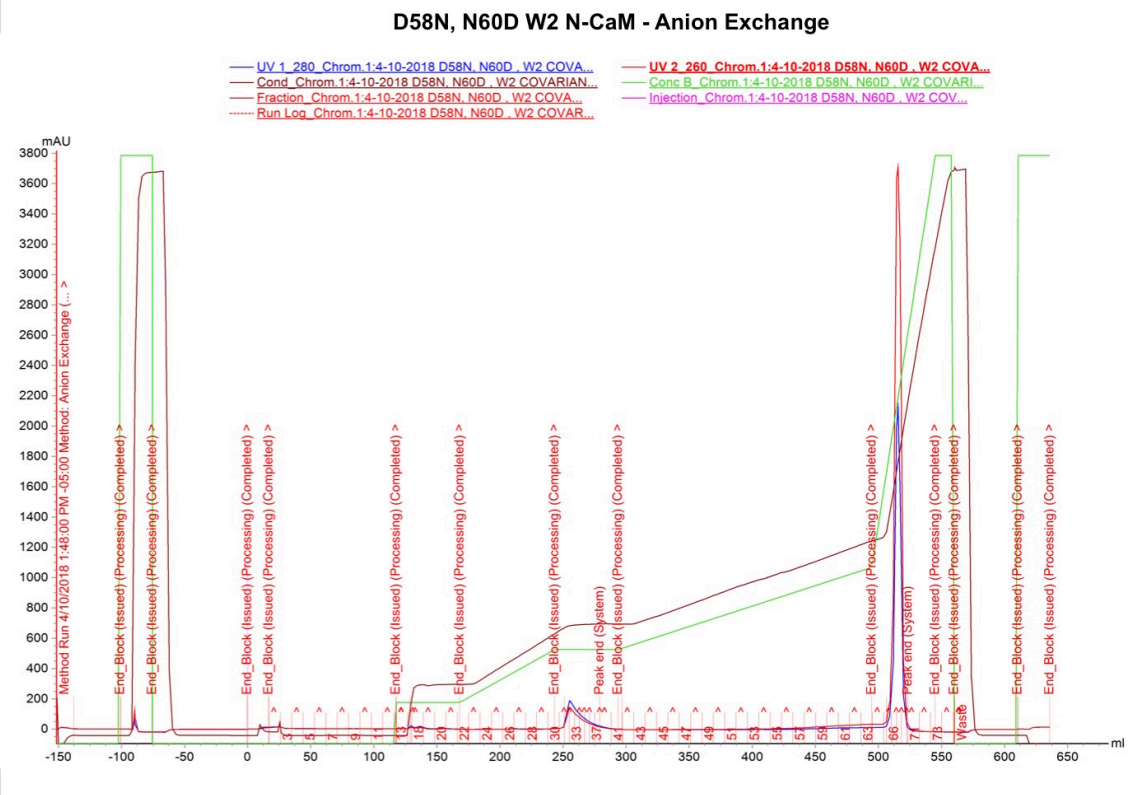

Figure 3.2.2: Large-scale view of the anion exchange step for the purification of CaM construct D58N, N60D W2 N-CaM, in which fractions 31-37 were collected. The protein, represented by the blue trace at 280nm, was eluted in fractions 31-37. An increasing salt concentration using 20mM Tris and 1.5 M NaCl was used. The plateau of the green curve is the salt concentration where CaM is expected to fall off the column. The place where CaM is eluted off the anion exchange column is subject to change slightly based on the construct. The blue trace has reached a higher absorbance than the red trace, indicating that the protein is more concentrated in the sample.

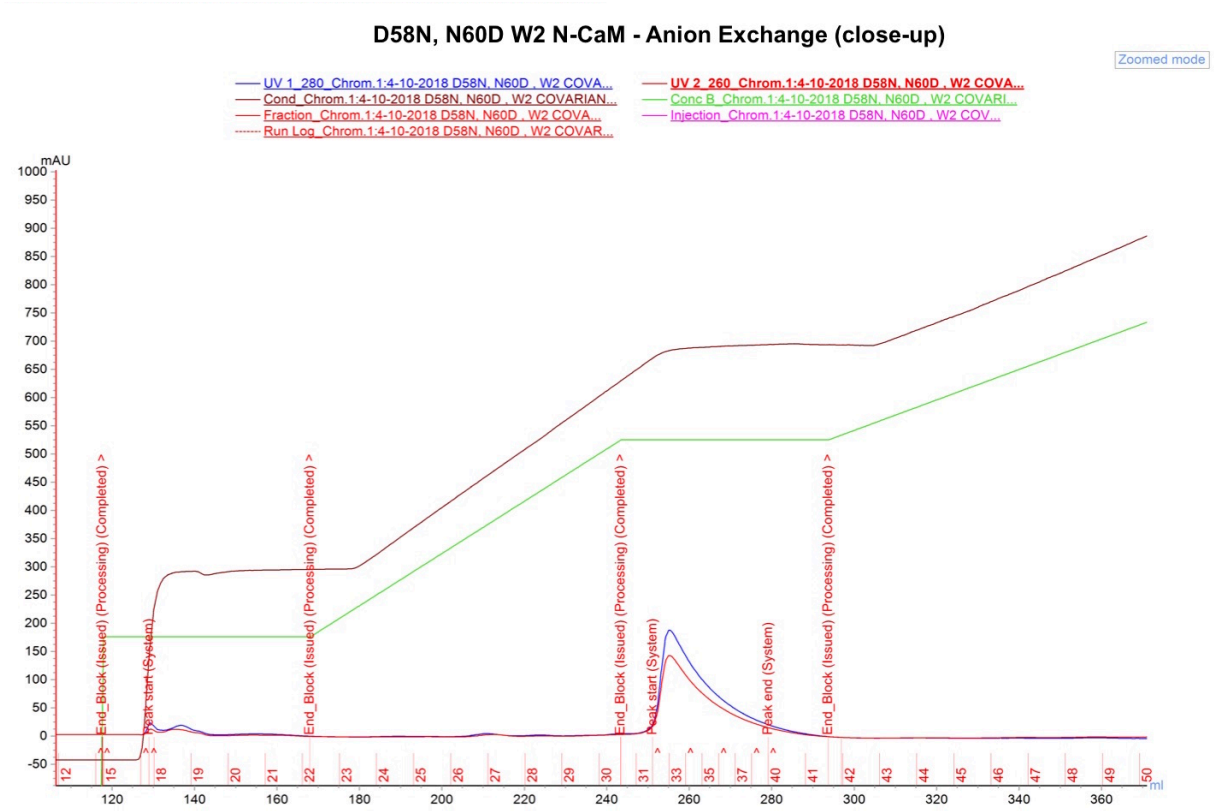

Figure 3.2.3: Close-up view of the anion exchange step for the purification of CaM construct D58N, N60D W2 N-CaM. Fractions 31-37 were collected.

Next, the fractions with eluted protein from the anion exchange were injected into the C18 Reverse Phase HPLC column for further purification (Figures 3.2.4 and 3.2.5). The C18 column is composed of an octadecyl-silica stationary phase, in which each molecule contains an eighteen-carbon chain that creates a very hydrophobic column. CaM has few hydrophobic regions that bind to the column through hydrophobic interactions. CaM is eluted with 30% acetonitrile, which is an organic solvent. The acetonitrile competes with CaM in binding to the very hydrophobic C18 column.

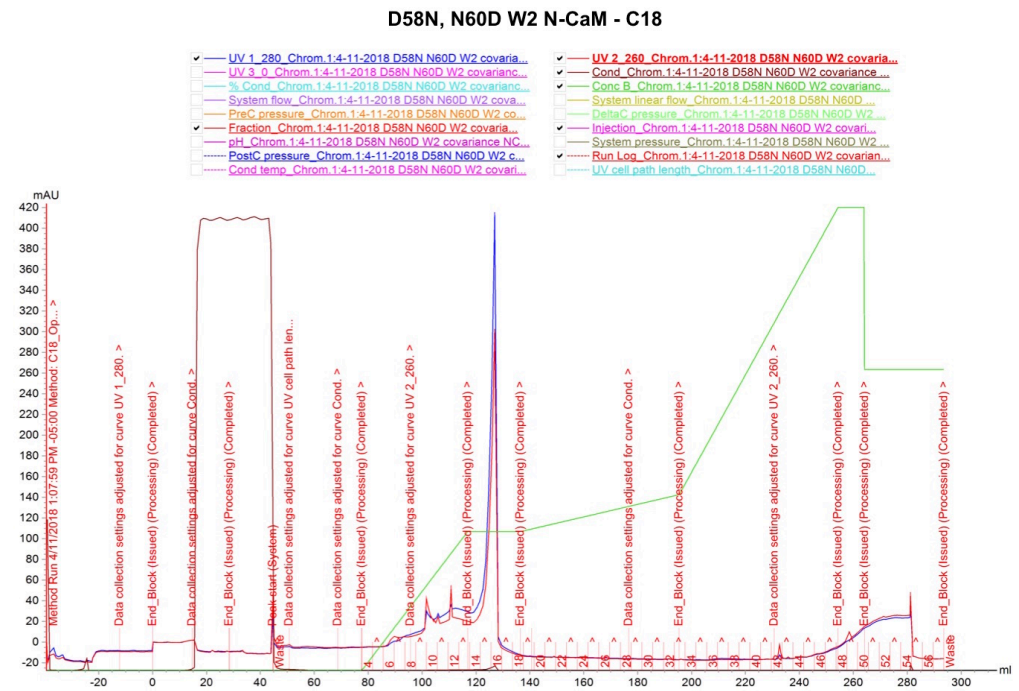

Figure 3.2.4: Large-scale view of the C18 column purification of the construct D58N, N60D W2 N-CaM, in which fractions 15 and 16 were collected. N-CaM has few exposed hydrophobic regions that attach to the eighteen carbon chained molecule attached to the column and the protein is eluted with 30% acetonitrile. The blue trace corresponds to the absorbance at a wavelength of 280nm, representing the purified protein.

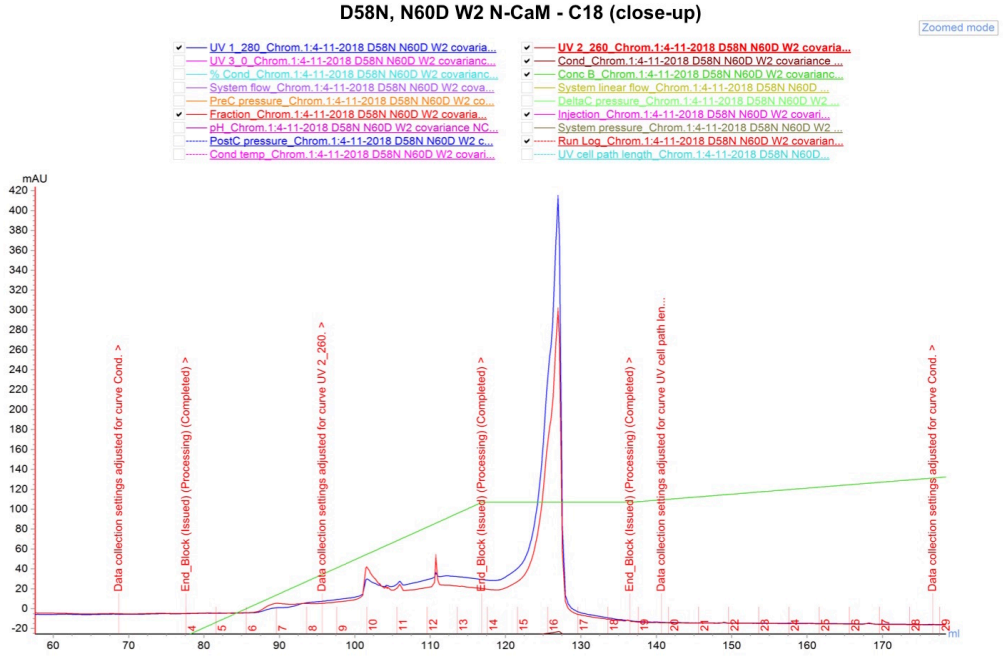

Figure 3.2.5: Close-up view of the C18 column purification of the construct D58N, N60D W2 N-CaM. Fractions 15 and 16 were collected.

Next, a final desalt step was performed to prepare the purified CaM construct in 5 mM HEPES buffer and to discard any contaminating salts (Figure 3.2.6). After the final desalt, a 20% SDS-PAGE gel with 10 mM EGTA is sometimes run to ensure that CaM was purified with little contamination.

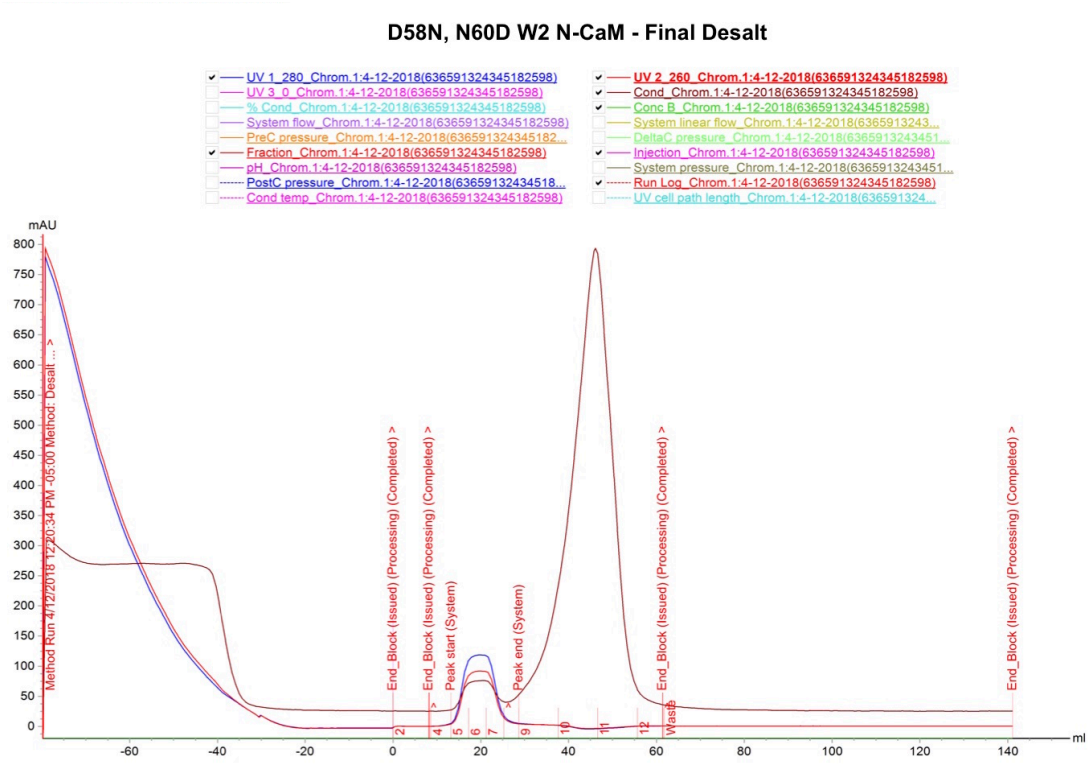

Figure 3.2.6: Final desalting step for CaM construct D58N, N60D W2 N-CaM, in which fractions 5-7 were collected. The blue trace corresponds to the absorbance at a wavelength of 280nm, representing the purified CaM construct.

Analyzing the absorbance of the solutions at 260 nm and 280 nm allowed for the assessment of whether CaM was purified and extracted. The maximum absorption for nucleic acids is 260nm. Nucleic acids are able to absorb UV light because they contain aromatic pyrimidine and purine nitrogenous bases, composed of a single ring structure and a double ring structure respectively. Proteins absorb UV light at 280nm due to the aromatic structure of certain amino acid side chains. Proteins have different extinction coefficients based on the number of

aromatic amino acids they contain; if a protein contains high numbers of Tryptophan, Tyrosine, Cysteine (due to double-bonding conjugated systems), and Phenylalanine residues, the protein will have a larger extinction coefficient. The N-lobe CaM with the tryptophan reporter has an extinction coefficient of 5500 M^{-1} cm^{-1} and a molecular weight of 9036 g/mol, whereas wild-type CaM has an extinction coefficient of 3020 M^{-1} cm^{-1} and a molecular weight of 16709 g/mol. The absorbance at a range of 250 nm to 350 nm was measured using a spectrophotometer. The Beer-Lambert Law describes the linear relationship between the absorbance of a sample to the concentration of the protein absorbing light. The Beer-Lambert law is $A = \varepsilon lc$, where ε is the extinction coefficient, which is dependent on the wavelength of light, l is the path length of light through the sample and c is the concentration. The concentration of CaM was calculated using the Beer-Lambert law since the absorbance is linearly proportional to the concentration (Figure 3.2.7 below).

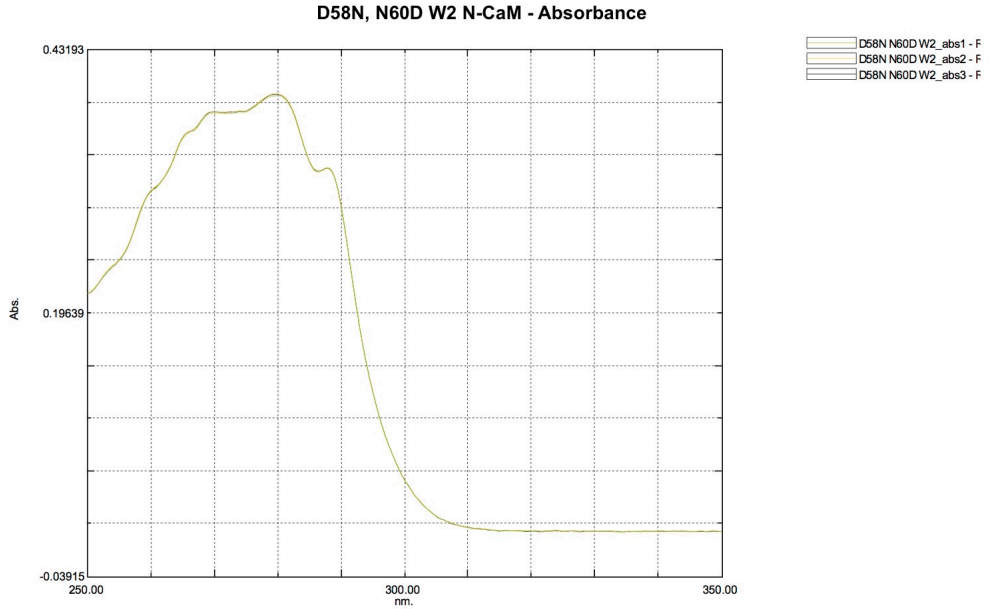

Figure 3.2.7: Absorbance data for construct D58N, N60D W2 N-CaM. The concentration was determined based on the extinction coefficient 5500 M^{-1} cm^{-1} for N-lobe CaM with a tryptophan reporter. The concentration was determined by averaging the three absorbance measures at 280 nm and 350 nm and applying the Beer-Lambert law to determine the concentration of protein in the sample. The concentration of the construct D58N, N60D W2 N-CaM purified was 364.818 μM.

Next, we measured the site-specific CaM binding curves and luminescent decays using a spectrofluorometer from Photon Technologies International. The mutated CaM proteins were analyzed by testing the binding capabilities of the CaM constructs when exposed to terbium ions (Tb^{3+}). Tb^{3+} ions bind similarly to calmodulin, thus these lanthanide ions are used as an alternative binding species for the analyzing the protein's affinity for calcium (Martin and Richardson, 1979). The reporter tryptophan amino acid was incorporated into CaM in order to measure the excitation of terbium ions that were bound to CaM. Tb^{3+} is able to luminesce when excited, which allows measurements of the relative concentration of terbium bound to the protein. A flash lamp of the spectrofluorometer emitted light at a frequency of 100 Hz with an average of 10,000 light pulses per experiment. The 292 nm light emitted excites the reporter tryptophan amino acid, which directly transfers its energy to the Tb^{3+} bound to the particular EF-hand of interest in CaM (Figure 3.2.8 below). When the tryptophan residue is excited and its energy is transferred to the neighboring Tb^{3+}, the Tb^{3+} will in turn emit a wavelength of 525nm. The terbium luminescence decay curve was measured by the spectrofluorometer for each of the different CaM constructs. The large range of 1μM to 10mM Tb^{3+} solutions were stored in Teflon vials and MES solution was used as the base solution.

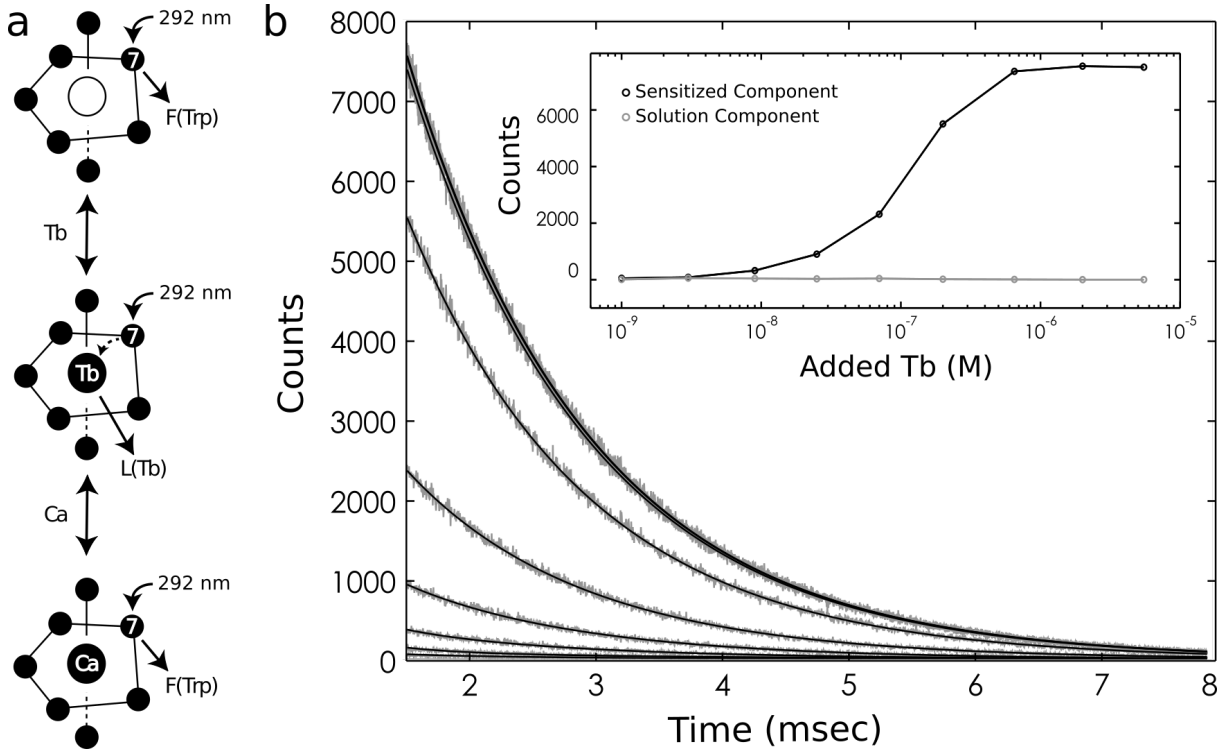

Figure 3.2.8: Site-specific CaM binding curves and luminescent decays measured by a spectrofluorometer utilizing the luminescent properties of the reporter tryptophan amino acid in the EF-hand binding site and Tb^{3+} bound to EF-hand binding site of interest in CaM (Greeson-Bernier et al., 2013).

Multiple exponential luminescent decay curves are generated by each Tb^{3+} titration experiment, in which the amplitudes of Tb^{3+} decay are plotted. The raw data luminescent decays are fitted to a double exponential fitting curve for binding analysis. In the double exponential decay equation, $y = A_1 e^{-t/\tau_1} + A_2 e^{-t/\tau_2}$, the $A_1 e^{-t/\tau_1}$ part of the equation represents the fast component of Tb^{3+} in solution and the $A_2 e^{-t/\tau_2}$ part of the equation represents the slower component of CaM bound to Tb^{3+}. The fast component represents the Tb^{3+} that is free in solution and therefore the signal decay extinguishes quickly. The slow component represents the Tb^{3+} that is bound to the CaM construct and the signal decays more slowly, indicating binding. The double exponential algorithm is used to generate sigmoidal binding curves, which indicates cooperative binding of CaM to its substrates, Tb^{3+} and Ca^{2+}. We infer that CaM experiences positive

cooperativity evident in the sigmoidal curves of the binding sites, however the extent or value for positive cooperativity cannot be assumed or determined.

4. Results and Discussion

The raw data luminescent decays are fitted to a double exponential curve for the binding analysis of the N-lobe EF-hand binding domains of CaM as discussed in the methods. The double exponential curve fitting consists of an algorithm, called the Levenberg-Marquardt algorithm, which determines the coefficients for the double exponential equation $y = A_1 e^{-t/\tau_1} + A_2 e^{-t/\tau_2}$, based on the raw data values. The algorithm uses iterative Chi-squared analyses to estimate the best values for the coefficients based on the data. The Chi-squared equation used is $\sum(\frac{y-y_i}{\sigma_i})$, where y is the fitted value for a given point, y_i is the measured raw data value, and σ_i is the estimated standard deviation for the data set. The algorithm is iterative meaning it tries different coefficients until it finds the coefficient that minimizes the value of the chi-squared value, maximizing the accuracy of the binding curve and delivering a nonlinear, least squares fitting curve.

The binding curves are dependent on three factors: the ability of the tryptophan amino acid to absorb and experience excitation from the 292 nm wavelength light pulse, the extent of energy transfer to the Tb^{3+} bound to the particular EF-hand of interest in CaM, and the emission of light with a wavelength of 525 nm by the neighboring Tb^{3+}. As a result, these factors are slightly affected by the different chemical environments of the distinct CaM constructs with different mutations. This results in the different non-normalized raw amplitudes on the binding curves for the data in Figures 4.1 and 4.2. It is most likely that the Tb^{3+} emitting light at 525 nm wavelength light mostly affects the non-normalized raw amplitude of the data because the different mutations in the EF-hand binding domain are close to and impact binding of the bound Tb^{3+}. However, the constructs with the sigmoidal shape and resemblance to the wild-type binding curve suggest that binding has been restored.

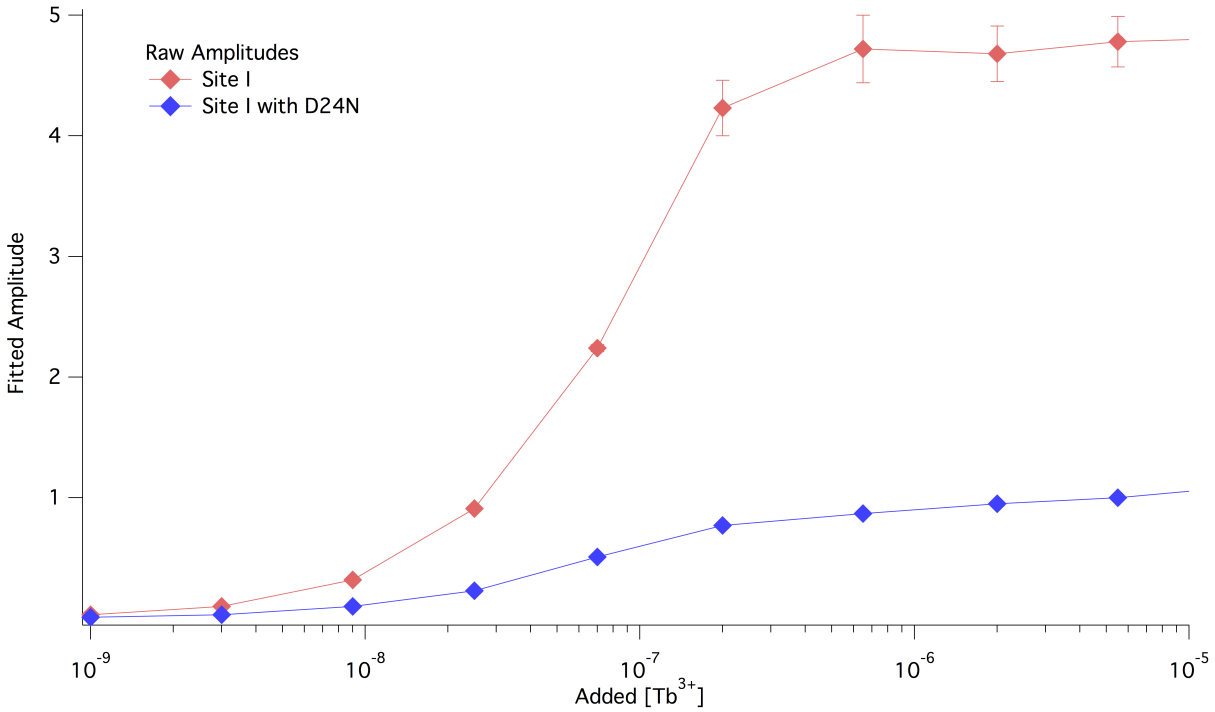

Figure 4.1: Non-normalized binding curve for covariant position 24 compared to the wild type in EF-hand binding site 1 in the N-lobe of CaM. The D24N mutation removes the necessary aspartate amino acids in site 1. The raw amplitude is plotted versus the increasing Tb^{3+} concentrations and the data is fitted using the Levenberg-Marquardt algorithm described above. Binding curve data collected by Suzanna Bennett.

Figure 4.1 depicts the non-normalized binding curve for the covariant position 24 in the EF-hand binding site 1 of the N-lobe of CaM compared to the non-normalized binding curve for wild-type EF-hand binding site 1 of the N-lobe of CaM. The raw amplitude is plotted versus the increasing Tb^{3+} concentrations and the data indicates that the D24N mutation disrupts binding in site 1. The binding curve for the wild-type EF-hand binding site 1 is sigmoidal and the binding curve for the EF-hand site 1 mutation D24N reaches a very low raw amplitude and loses the sigmoidal curve, indicating that the D24N mutation disrupts binding in site 1 of the N-lobe of CaM. The small raw amplitude of the CaM construct with the D24N mutation is probably due to the artifacts of the fitting algorithm or is attributed to the fact that the different chemical environment of the mutant CaM slightly impacts the ability of the bound Tb^{3+} to emit light at a

wavelength of 525 nm. The site 1 with D24N mutation binding curve is evidently different from the sigmoidal site 1 N-CaM wild-type binding curve, which indicates that binding has been negatively impacted.

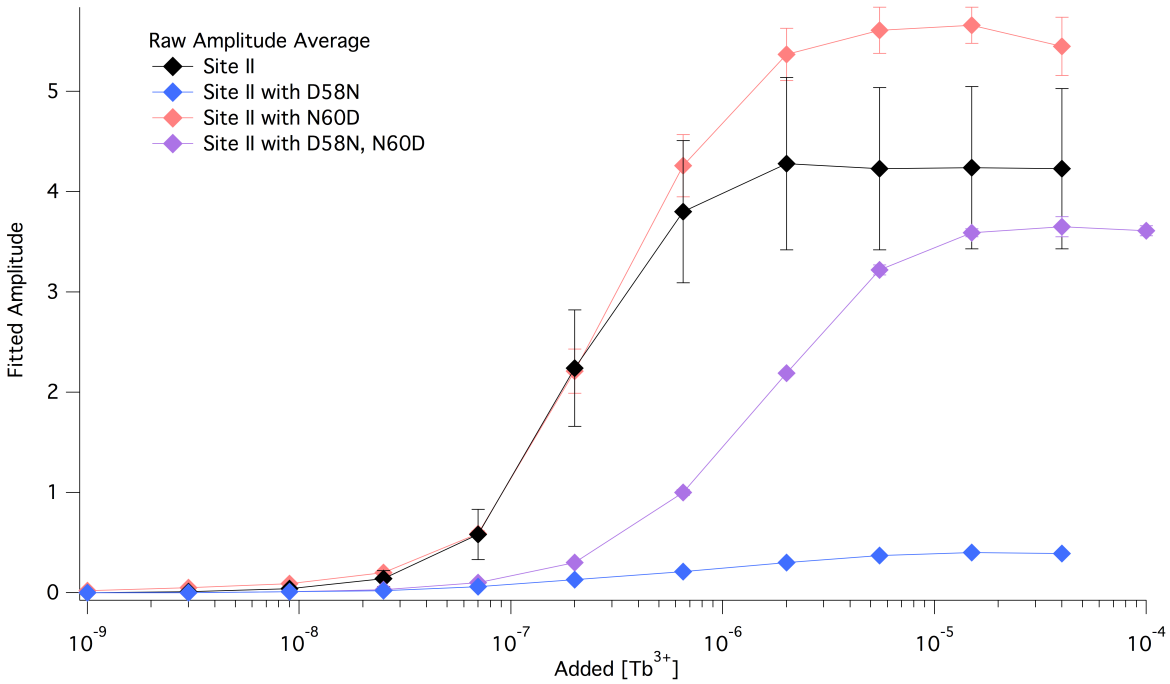

Figure 4.2: Non-normalized binding curve for covariant positions 58 and 60 compared to the wild type in EF-hand binding site 2 in the N-lobe of CaM. The D58N mutation removes the necessary aspartate amino acids from site 2. The N60D mutation adds an aspartate amino acid to site 2. The combined D58N and N60D mutations in site 2 restore binding. The raw amplitude is plotted versus the increasing Tb^{3+} concentrations and the data is fitted using the Levenberg-Marquardt algorithm described above. Binding curve data collected by Suzanna Bennett.

Figure 4.2 depicts the non-normalized binding curve for the covariant positions 58 and 60 in the EF-hand binding site 2 of the N-lobe of CaM. The raw amplitude is plotted versus increasing Tb^{3+} concentration. The D58N mutation in site 2 of the N-lobe of CaM appears to have disrupted binding in site 2 due to the removal of a necessary aspartate amino acid in the binding site; this is evident in the very low amplitude and loss of the sigmoidal shape of the site 2 with D58N binding curve. However, the combined D58N & N60D mutations restore binding in site 2 by providing the necessary aspartate in the covariant position 60. Although the binding

curve does not reach the raw amplitudes like the wild-type site 2 binding curve, binding was rescued with both the D58N & N60D mutations, as the raw amplitude reaches around 3.5 and is sigmoidal. However, the curve appears to be right shifted, which means that the apparent affinity for calcium has decreased with the combined D58N & N60D mutations. The N60D mutation by itself has little impact on binding at site 2 because the binding curve for the N60D mutation is very similar to the wild-type curve. There seems to be selective pressures to conserve at least one aspartate in the covariant positions of the EF-hand binding site 2 and the presence of two aspartates in the covariant positions of the binding site has little affect on calcium binding.

5. Conclusions

The central goal of our research is to examine whether the basic principles of covariance apply to CaM EF-hand binding domains. Our research is unique because it applies the concept of covariance, which is usually used to predict protein structure, on a highly conserved protein with a known crystallized structure. Covariance analyses on CaM EF-hands indicate which amino acid contacts are under evolutionary selective pressure and are most important for Ca^{2+} binding and coordination in the N-lobe EF-hand domains.

Positions 22 and 24 in EF-hand site 1 and positions 58 and 60 in EF-hand site 2 were identified as the two high scoring amino acid pairs in the N-lobe EF-hands based on covariance algorithms and multiple sequence analyses. We created CaM constructs by mutating the amino acids in these locations and performed accompanied binding analyses to understand the effects of the mutations on calcium binding. We have found that both the D24N mutation in site 1 and the D58N mutation in site 2 disrupt binding likely due to the removal of a necessary aspartate in the binding site. However, the combined D58N and N60D mutations restore binding in site 2 by providing the necessary aspartate in the covariant location. The N60D mutation by itself has little impact on calcium binding in site 2. Therefore, it is evident that evolution conserves at least one aspartate in the covariant positions of the binding site and the presence of two aspartates in the covariant positions of the binding site has little affect on calcium binding. This study indicates that positions 58 and 60 in the EF-hand site 2 of CaM covary and are probably amino acid contacts that are most important for Ca^{2+} coordination in the site 2 EF-hand binding domain. Interrupting the covariance pattern disrupted Ca^{2+} binding, however the binding was rescued by restoring the covariance pattern in site 2 of the N-lobe of CaM.

6. Future Directions

The overall aim of our research is to examine the effects of mutations in covariant positions in the EF-hand binding domains on the structure and function of calmodulin. This study aims to contribute to the understanding of the biophysical interactions between Ca^{2+} and the highly conserved four distinct EF-hand binding domains of CaM with a long term goal of detailing calcium-mediated cellular processes and how improper binding of calcium can lead to various human pathologies. We are currently studying the covariant positions in the EF-hand site 1, specifically the D22N mutation and the combined mutations D24N & T28D as well as D22N & T28D. With our findings described about the effects of the mutations in the EF-hand site 2, we are paying special attention to the selective pressures affecting the net number of aspartates coordinating Ca^{2+}. Future directions also include structurally analyzing the covariant positions in the C-lobe of CaM, specifically the high scoring positions 138, 140, 159 and 160 determined by the covariance algorithm analyses. Additionally, studying covariance patterns of other calcium-binding proteins with EF-hand binding domains would contribute to our understanding of how evolutionary selective pressures influence structure and binding in this family of proteins. Hopefully our findings and unique application of covariance patterns to a protein with a known structure contribute to future studies of calcium-binding proteins.

References:

Adelman, J. P., Maylie, J., & Sah, P. (2012). Small-conductance Ca2+-activated K+ channels: form and function. *Annual review of physiology*, *74*, 245-269. doi: 10.1146/annurev-physiol-020911-153336

Anishchenko, I., Ovchinnikov, S., Kamisetty, H., & Baker, D. (2017). Origins of coevolution between residues distant in protein 3D structures. *Proceedings of the National Academy of Sciences*, *114*(34), 9122-9127. doi: 10.1073/pnas.1702664114

Ashenberg, O., & Laub, M. T. (2013). Using analyses of amino acid coevolution to understand protein structure and function. In *Methods in enzymology* (Vol. 523, pp. 191-212). Academic Press. doi: 10.1016/B978-0-12-394292-0.00009-6

Babu, Y. S., Sack, J. S., Greenhough, T. J., Bugg, C. E., Means, A. R., & Cook, W. J. (1985). Three-dimensional structure of calmodulin. *Nature*, *315*(6014), 37.

Carafoli, E., Santella, L., Branca, D., & Brini, M. (2001). Generation, control, and processing of cellular calcium signals. *Critical reviews in biochemistry and molecular biology*, *36*(2), 107-260. doi: 10.1080/20014091074183

Crotti, L., Johnson, C. N., Graf, E., De Ferrari, G. M., Cuneo, B. F., Ovadia, M., ... & Pedrazzini, M. (2013). Calmodulin mutations associated with recurrent cardiac arrest in infants. *Circulation*, 127(9):1009-17. doi: 10.1161/CIRCULATIONAHA.112.001216.

Edington, S. C., Gonzalez, A., Middendorf, T. R., Halling, D. B., Aldrich, R. W., & Baiz, C. R. (2018). Coordination to lanthanide ions distorts binding site conformation in calmodulin. *Proceedings of the National Academy of Sciences*, *115*(14), E3126-E3134. doi: 10.1073/pnas.1722042115

Ekeberg, M., Lövkvist, C., Lan, Y., Weigt, M., & Aurell, E. (2013). Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models. *Physical Review E*, *87*(1), 012707. doi: 10.1103/PhysRevE.87.012707

Fodor, A. A., & Aldrich, R. W. (2004). Influence of conservation on calculations of amino acid covariance in multiple sequence alignments. *Proteins: Structure, Function, and Bioinformatics*, *56*(2), 211-221. doi: 10.1002/prot.20098

Giese, K. P., & Mizuno, K. (2013). The roles of protein kinases in learning and memory. *Learning & memory*, *20*(10), 540-552. doi: 10.1101/lm.028449.112.

Greeson-Bernier, J. N., Middendorf, T. R., & Aldrich, R. W. (2013). Site-Specific Measurement of Terbium and Calcium Binding to Calmodulin and Calmodulin Fragments. *Biophysical Journal*, *104*(2), 98a. doi: 10.1016/j.bpj.2012.11.583

Haeseleer, F., Imanishi, Y., Sokal, I., Filipek, S., & Palczewski, K. (2002). Calcium-binding proteins: intracellular sensors from the calmodulin superfamily. *Biochemical and biophysical research communications*, 290(2), 615-623. https://doi.org/10.1006/bbrc.2001.6228

Halling, D. B., Liebeskind, B. J., Hall, A. W., & Aldrich, R. W. (2016). Conserved properties of individual Ca2+-binding sites in calmodulin. *Proceedings of the National Academy of Sciences*, *113*(9), E1216-E1225. doi: 10.1073/pnas.1600385113.

Halling, D. B., Kenrick, S. A., Riggs, A. F., & Aldrich, R. W. (2014). Calcium-dependent stoichiometries of the KCa 2.2 (SK) intracellular domain/calmodulin complex in solution. *The Journal of General Physiology, 143 (2), 231-252*. doi: 10.1085/jgp.201311007

Hoeflich, K. P., & Ikura, M. (2002). Calmodulin in action: diversity in target recognition and activation mechanisms. *Cell*, *108*(6), 739-742. https://doi.org/10.1016/S0092-8674(02)00682-7

Kawasaki, H., Nakayama, S., & Kretsinger, R. H. (1998). Classification and evolution of EF-hand proteins. Biometals, 11(4), 277-295.

Keen, J. E., Khawaled, R., Farrens, D. L., Neelands, T., Rivard, A., Bond, C. T., ... & Maylie, J. (1999). Domains responsible for constitutive and Ca2+-dependent interactions between calmodulin and small conductance Ca2+-activated potassium channels. *Journal of Neuroscience*, *19*(20), 8830-8838.

Kahl, C. R., & Means, A. R. (2003). Regulation of cell cycle progression by calcium/calmodulin-dependent pathways. *Endocrine reviews*, *24*(6), 719-736. doi: 10.1210/er.2003-0008

Li, W., & Aldrich, R. W. (2009). Activation of the SK potassium channel-calmodulin complex by nanomolar concentrations of terbium. *Proceedings of the National Academy of Sciences*, *106*(4), 1075-1080. doi: 10.1073/pnas.0812008106

Li, W., Halling, D. B., Hall, A. W., & Aldrich, R. W. (2009). EF hands at the N-lobe of calmodulin are required for both SK channel gating and stable SK–calmodulin interaction. *The Journal of general physiology*, *134*(4), 281-293.

Lisman, J., Schulman, H., & Cline, H. (2002). The molecular basis of CaMKII function in synaptic and behavioural memory. *Nature Reviews Neuroscience*, 3(3), 175-190. doi:10.1038/nrn753

Makita, N., Yagihara, N., Crotti, L., Johnson, C. N., Beckmann, B. M., Roh, M. S., ... & Homfray, T. (2014). Novel Calmodulin Mutations Associated With Congenital Arrhythmia Susceptibility Clinical Perspective. *Circulation: Genomic and Precision Medicine*, *7*(4), 466-474. doi: 10.1161/CIRCGENETICS.113.000459

Marsman, R. F., Barc, J., Beekman, L., Alders, M., Dooijes, D., van den Wijngaard, A., ... & Bezzina, C. R. (2014). A mutation in CALM1 encoding calmodulin in familial idiopathic ventricular fibrillation in childhood and adolescence. *Journal of the American College of Cardiology*, *63*(3), 259-266. doi: 10.1016/j.jacc.2013.07.091

Martin, B., & Richardson, F. S. (1979). Lanthanides as probes for calcium in biological systems. *Quarterly reviews of biophysics*, 12(2), 181-209. https://doi.org/10.1017/S0033583500002754

Moncrief, N. D., Kretsinger, R. H., & Goodman, M. (1990). Evolution of EF-hand calcium-modulated proteins. I. Relationships based on amino acid sequences. *Journal of Molecular Evolution*, *30*(6), 522-562.

Nyegaard, M., Overgaard, M. T., Søndergaard, M. T., Vranas, M., Behr, E. R., Hildebrandt, L. L. & Fosdal, I. (2012). Mutations in calmodulin cause ventricular tachycardia and sudden cardiac death. *The American Journal of Human Genetics*, 91(4), 703-712. https://doi.org/10.1016/j.ajhg.2012.08.015

O'Day, Danton H., Kristeen Eshak, and Michael A. Myre. "Calmodulin Binding Proteins and Alzheimer's Disease." *Journal of Alzheimer's Disease* 46.3 (2015): 553-69. Web. doi: 10.3233/JAD-142772

Pipilas, D. C., Johnson, C. N., Webster, G., Schlaepfer, J., Fellmann, F., Sekarski, N., ... & Crotti, L. (2016). Novel calmodulin mutations associated with congenital long QT syndrome affect calcium current in human cardiomyocytes. *Heart rhythm*, 13(10), 2012-2019. https://doi.org/10.1016/j.hrthm.2016.06.038

Reed, G. J., Boczek, N. J., Etheridge, S., & Ackerman, M. J. (2015). CALM3 mutation associated with long QT syndrome. *Heart rhythm: the official journal of the Heart Rhythm Society*, *12*(2), 419. doi: 10.1016/j.hrthm.2014.10.035

Sasagawa, T., Ericsson, L. H., Walsh, K. A., Schreiber, W. E., Fischer, E. H., & Titani, K. (1982). Complete amino acid sequence of human brain calmodulin. *Biochemistry*, *21*(10), 2565-2569.

Soderling, Thomas R. "The Ca2 Calmodulin-dependent Protein Kinase Cascade." *Trends in Biochemical Sciences* 24.6 (1999): 232-36. Web.

Torruella, G., Derelle, R., Paps, J., Lang, B. F., Roger, A. J., Shalchian-Tabrizi, K., & Ruiz-Trillo, I. (2011). Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. *Molecular biology and evolution*, *29*(2), 531-544.

Vigil, D., Gallagher, S. C., Trewhella, J., & García, A. E. (2001). Functional dynamics of the hydrophobic cleft in the N-domain of calmodulin. *Biophysical Journal*, *80*(5), 2082-2092.

Zhang, Xiao-Dong, Deborah K. Lieu, and Nipavan Chiamvimonvat. "Small-conductance Ca2 – activated K Channels and Cardiac Arrhythmias." *Heart Rhythm* 12.8 (2015): 1845-851. Web. doi: 10.1016/j.hrthm.2015.04.046

Acknowledgements:

I would first and foremost like to thank Dr. Aldrich for allowing me to join his lab and for giving me the valuable opportunity of developing a project in his laboratory as an undergraduate researcher. Dr. Aldrich has constantly supported me throughout the past two years and has provided so many opportunities for me to learn from the incredible group of scientists in the ion channel community. I am so appreciative of all of his patience, advice, and guidance that have significantly contributed to my growth as a scientist and as an individual. Thank you for investing the time in teaching me concepts in neuroscience and biochemistry, introducing me to novel discoveries in the field, and for helping me learn and discover through research. Your passions in neuroscience, biophysics and biochemistry have definitely inspired me to continue to pursue research in medical school and beyond.

I would next like to sincerely thank Suzy Bennett for training me in laboratory procedures and for initiating, designing and working with me on the calmodulin covariance project. Suzy is such a brilliant and inspiring scientist and has really been a role model for me. Thank you for always being so open to answering my questions and for teaching me about the theory behind covariance and coevolution and about all the details regarding calmodulin research. I am so grateful to have had the opportunity to work with you on this project. Thank you for being a friend and mentor to me, and I cannot wait to see the incredible work you will do and accomplish at Stanford.

I would also like to thank Dr. Brent Halling for also training me on many of the laboratory procedures we use and for always being so kind and welcoming to any questions I had about the methods or theory involved. Dr. Halling helped me extensively with the protein expression and purification experiments especially when troubleshooting. Thank you for

encouraging me with your genuine passion in conducting research and patience when teaching me about laboratory procedures and concepts. I would also like to thank Dr. Tom Middendorf for also being so open to answering my questions about the site-specific binding curves and mathematical analyses that are involved in the theory behind calmodulin binding measurements. I would also like to thank Ashley, Jonathan, Margaux, Yelena, and the other members and former members of the Aldrich lab for supporting and training me in different laboratory procedures. This lab has really been a family to me and I am going to miss everyone after graduation.

I would also like to thank the directors, advisors, and coordinators in the Health Science Scholars honors program for all of their support in helping me write and develop my thesis. I would also like to thank all of the incredible science professors I have had during my time at UT, especially Dr. O'Halloran, Dr. De Lozanne, Dr. Barrick, Dr. Wilcox, Dr. Machart, and Dr. Stovall for always encouraging me to follow my passions and prompting me to inquire and grow as student. I would also like to thank the Freshman Research Initiative, especially Dr. Stovall and the Aptamer Stream, for allowing me to conduct research, present my work at conferences, and introducing me to a wide range of biochemical research techniques as an underclassman.

Finally, I would like to thank my loving family for their constant support in my pursuits in science and for shaping me into the person I am today.

Uma Vaidyanathan

May 2018