METHOD DEVELOPMENT TOWARD THE ANALYSIS OF THE FUNCTIONALIZATION OF PROTEINS

By
Victoria Leigh Calcote

A thesis submitted to the faculty of The University of Mississippi in partial fulfillment of the requirements of the Sally McDonnell Barksdale Honors College.

Oxford
May 2016

Approved by

Advisor: Professor Susan Pedigo

Reader: Professor Davita Watkins

Reader: Professor Jason Ritchie

© 2016
Victoria Leigh Calcote
ALL RIGHTS RESERVED

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Susan Pedigo, for her invaluable guidance and support throughout the researching and writing process. She willingly welcomed me into her laboratory, taught me how to conduct research, answered my countless questions, and supported and encouraged me along the way. I am very grateful for the time spent together, both in Oxford and in Paris. Merci beaucoup.

I would like to thank my second reader, Dr. Davita Watkins, for her contributions and time spent on developing this project. We would not have anything to analyze without your help.

I would also like to thank my third reader, Dr. Jason Ritchie, for taking the time to read my thesis and for expressing interest in my topic.

I would like to express my gratitude to the graduate students and the rest of the research crew in Dr. Pedigo's lab—especially you, Chris. They readily offered guidance and assistance whenever needed, and they provided much encouragement throughout this process.

I owe much appreciation to the Sally McDonnell Barksdale Honors College for the resources and funding they provided, as well as offering me the opportunity of conducting this exciting research.

ABSTRACT

VICTORIA LEIGH CALCOTE: Method Development Toward the Analysis of the Functionalization of Proteins (Under the direction of Dr. Susan Pedigo)

Calmodulin is a ubiquitous small intracellular protein that functions as a receptor for regulatory calcium signals and thus regulates a multitude of physiological processes in organisms as diverse as yeast, fruit flies, and mammals. Since its structure and function have been extensively studied, our goal is to create a calmodulin biomaterial, with the ultimate goal of *in situ* drug delivery or as a scaffold for stem cell growth and differentiation. We plan to create this biomaterial by using "Click Chemistry," the rapid creation of new heteroatom bonds through a copper-catalyzed azide-alkyne cycloaddition. The bonds are created by treating calmodulin as an organic reagent and derivatizing the hydroxyl group of the amino acid tyrosine for an organic fluorophore. However, since tyrosines appear at residues 99 and 138, functionalization could be occurring at multiple sites. A reliable protocol has to be developed to determine the exact position of functionalization on calmodulin. We will focus on determining if and where functionalization occurs by employing an enzyme to cleave between the two tyrosines. Thrombin, a protease, offers a convenient analytical solution by cleaving the protein between two tyrosines, as it hydrolyzes the protein backbone at the C-terminal side of arginine (residue 106), which is located in between the tyrosines (residues 99 and 138). In finding a mechanism through which we are able to successfully separate the tyrosines on calmodulin, we will determine which, if either, has

been derivatized by then performing gel electrophoresis, mass spectrometry, or chromatography.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

LIST OF ABBREVIATIONS

Ca^{2+}	calcium ion
cAMP	cyclic adenosine monophosphate
DMSO	dimethyl sulfoxide
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KCl	potassium chloride
LB	lysogeny broth
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry
MLCK	myosin light-chain kinase
MW	molecular weight
NaCl	sodium chloride
pH	potentia hydrogenii
R	arginine
Rf	relative mobility
RP-HPLC	Reversed Phase High-Performance Liquid Chromatography
STD	standard

Tricine-SDS-PAGE	tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis
UV-Vis	Ultraviolet-visible spectroscopy
Y	tyrosine

INTRODUCTION

Life on this planet evolved in an aqueous environment. The structure and function of the molecules of life are integrally dependent upon the aqueous environment from which they emerged. Globular proteins assume complex tertiary folds that are essential for their function. This property of proteins is dependent upon their interaction with the solvent, water. Since proteins are carbon-based, they are organic molecules. Therefore, one can consider proteins as large, linear organic polymers with regularly spaced organic functional groups. Can proteins withstand the rigors of a typical organic reaction?

Yes, proteins are organic, but as organic reagents, problems arise in the harsh conditions involved that could potentially irreversibly denature the protein. However, calmodulin has been found to be a remarkable protein that reversibly denatures under diverse insults such as denaturants [1] and temperature [2]. It is found primarily in the brain and testes, and in the past before recombinant protein technology was developed, boiling was required to acquire pure calmodulin from tissue homogenates. Unlike most proteins, calmodulin is able to recover and refold to its native, functional form after being subjected to high temperatures and denaturation. This remarkable property prompted the proposal of using calmodulin as a polyfunctional organic reagent. Hence, calmodulin could be considered a polymer with 148 functional groups with 20 possible organic moieties. Functionalization of calmodulin can be achieved through “Click Chemistry,” a

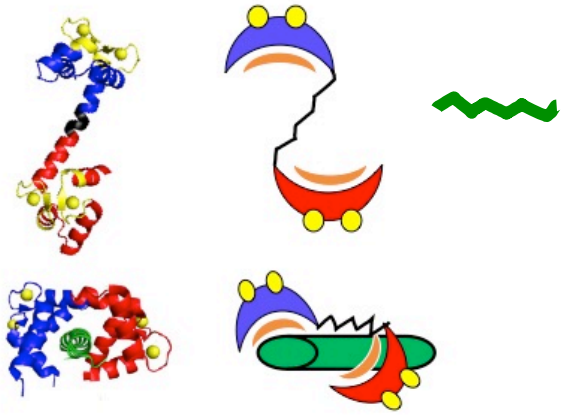
method of creating new heteroatom linkages, rather than energetically expensive carbon-carbon bonds, for fast, reproducible synthesis of new compounds [3]. In the current study, the new heteroatom linkages were directed at the hydroxyl hydrogen on tyrosines, residues 99 and 138, the only tyrosines in calmodulin. We use copper-catalyzed azide-alkyne Click Chemistry to functionalize calmodulin by completing a cycloaddition reaction, essentially replacing the hydroxyl of tyrosine with a triazole group. However, functionalization could be occurring at multiple sites on calmodulin. Not only does calmodulin have two tyrosines, but also it has serine and threonine residues that have hydroxyl groups, which could be functionalized.

1. Calmodulin

Calmodulin is an acidic ubiquitous protein that acts as the primary intracellular receptor of calcium (Ca^{2+}) in eukaryotic cells [4]. It functions as a regulator in a multitude of calcium signal transduction pathways, playing a role in the activation of more than 20 enzymes [5]. These enzymes, including cAMP regulated enzymes and smooth muscle MLCK, are involved in a wide variety of physiological processes, such as muscle contraction, cell proliferation regulation, and regulation of the cell cycle [5]. Furthermore, calmodulin is involved in the regulation of 2 Ca^{2+}-transporting ATPases, which bind Ca^{2+} with high affinity, found in the plasma membrane and the heart sarcoplasmic reticulum [4]. Calmodulin binding relieves the autoinhibition by the cell of MLCK and changes it into its catalytically active form [5]. Hence, calmodulin's extensive influence throughout physiological systems happens through direct interactions with its targets, thereby controlling cell physiology.

Ca^{2+} enters the cells via voltage-dependent Ca^{2+} channels or is released from stores already inside the cell. The influx of free intracellular Ca^{2+} causes the activation of calmodulin-dependent pathways and the formation of the Ca^{2+} –calmodulin complex, which is necessary for interactions with target proteins. Ca^{2+} binding to calmodulin induces a conformational change, just as when the Ca^{2+}-calmodulin complex binds to its substrate and produces a conformational change in both the complex and the substrate (**Figure 1A**). The structural change in the target protein is necessary for its activation and

allows for expression of a biological activity[5]. The study of calmodulin's structure while in the Ca^{2+}–calmodulin complex allows for more understanding of its function and activity.



A

```
1                  20        Site 1     31     39
ADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSL
  40               56    Site 2         67           75
   GQNPTEAELQDMINEVDADGNGTIDFPEFLTMMARK
 76   Central Helix  93  Site 3        104          112
  MKDTDSEEEIREAFRVFDKDGNGYISAAELRHVMTNL
   113                 129    Site 4       140     148
   GEKLTDEEVDEMIREADIDGDGQVNYEEFVQMMTAK
```

B

Figure 1: *Calmodulin's structure and properties.*
Ribbon and cartoon drawings calmodulin: Ribbon drawing of 3cln.pdb[6] with calcium binding sites shown in yellow[7]. Helices lead into and out of each site. The "central helix" separates two independently folded domains. When calcium binds, hydrophobic surface area is exposed at the base of each pair of sites (orange crescents in cartoon). These hydrophobic surfaces interact with various calmodulin binding agents such as trifluoperazine[8] and peptides from physiological partners (e.g. myosin light chain kinase[9] and ryanodine receptor[10]). In the ribbon[11] and cartoon drawings, the helical peptide from physiological partner is represented as a green cylinder. The unbound peptide is disordered before binding to calcium-calmodulin complex.
Alignment of Calcium Binding Sites: EF-hand sites contain 12 residues, 6 of which directly chelate calcium: five (X, Y, Z, -X, and –Y) offer one carbonyl oxygen to

(**Figure 1** continued) chelate calcium; one (-Z) is a bidentate chelator of calcium. Sequences comprising sites 1 and 2 (N-terminal domain; blue) and sites 3 and 4 (C-terminal domain; red) are aligned to highlight the conserved residues (consensus: $DXDG^{N}/_{D}GX_5E$). There are only 2 tyrosines (Y99, site 3 and Y138, site 4) and no cysteine or tryptophan residues.

Calmodulin is composed of one hundred and forty-eight amino acid residues. In the unbound conformation (**Figure 1A**), it has a length of sixty-five angstroms and a molecular weight of 16,706 Da. The dumbbell-shaped protein is composed of two globular domains or lobes in its folded state. The N-domain comprises residues 1—75 (blue segment in **Figure 1A**), and the C-domain comprises residues 76—148 (red segment in **Figure 1A**). Each domain contains two EF-hand motifs, which are in a helix-loop-helix configuration and act as the binding sites for Ca^{2+}. The N- and C-domains possess homologous sequences, but have different affinities for Ca^{2+}[7] and therefore distinct roles in enzyme activation (Sorenson 4244). The N-domain contains calcium-binding sites I and II and the C-domain contains calcium-binding sites III and IV[12]. The two domains are joined by a "central helix" (black segment and surrounding red and blue in **Figure 1A**): a seven-turn alpha-helix that acts as a "flexible tether" between them, which does not participate in any Calmodulin-peptide interactions[6]. The tether encompasses residues 65 to 92, with the central "solvent-exposed"[12] portion including charged amino acids. At the base of each pair of sites is a collection of hydrophobic amino acids (orange crescents in **Figure 1A**) that are important for interactions with physiological targets, as will be discussed below[6].

Nine aromatic amino acids are found in both the C- and the N- domains[6]. These aromatic residues, specifically tyrosines 99 and 138 (**Figure 1B**), are of particular importance to us, as these provide the potential for functionalization sites with organic

fluorophores and, hence, the focus of our study. Tyrosines 99 and 138 are located in the C-domain and occur in calcium-binding sites III and IV, respectively[6]. The locations and spectroscopic properties of each tyrosine provide us with our method of detecting successful functionalization. Our technique of detecting through spectroscopy the site(s) at which the functionalization of calmodulin took place depends on cleaving the protein between them, thereby separating these aromatic residues into peptides with one tyrosine in each, and performing RP-HPLC with diode array UV detection.

When calmodulin binds Ca^{2+}, it does so cooperatively and distinct conformational changes occur. Since there are four possible binding sites of calcium to calmodulin, "both the number and position of calcium ions bound to CaM may regulate its interactions with targets"[13]. As stated above, the N-terminal and C-terminal lobes have different affinities for calcium and take different roles in controlling the physiology of cells[14].

The conformational changes resulting from calmodulin binding Ca^{2+} have a large role in calmodulin-peptide interactions. The changes produce access to hydrophobic clefts in calmodulin, which are important in binding target proteins, most containing complementary hydrophobic regions[6]. Calmodulin has a unique structure optimized for binding Ca^{2+}. The loops in four EF-hand motifs are made up of sections twelve residues in size[6]: 20 to 31 in Site I, 56 to 67 in Site II, 93 to 104 in Site III, and 129 to 140 in Site IV (**Figure 1B**). As expected, these loops are hydrophilic, as they bind with the charged calcium ion. Glutamate, which always occupies the twelfth residue position in every loop, is essential in its role as a bidentate chelator, binding the calcium ion through its two side chain oxygens[6]. The binding of Ca^{2+} to calmodulin exposes hydrophobic regions (orange crescents in **Figure 1A**), which are created from the side chains of

fourteen specific residues in both the C-terminal and the N-terminal domains[6]. These clefts are 70 percent homologous[6] and each cleft is approximately 10.0 Å x 12.5 Å with a depth of 9.5 Å. These hydrophobic clefts do not appear without this calmodulin-calcium interaction, inhibiting potential target peptide-calmodulin interactions[6].

A long-term interest of this project is to determine whether the functionalization of calmodulin will affect its interactions with target peptides. The main goal of the project is to create a biomaterial, whose three dimensional structure depends on calmodulin peptide-interactions. The target peptides are short segments from natural calmodulin-binding ligands, such as cAMP and skeletal MLCK[5]. The target peptides of calmodulin are not necessarily homologous in sequences, but are similar in secondary structures. They exhibit a tendency for alpha-helix formation and contain either aromatic or long-chain hydrophobic residues twelve amino acids apart[12]. When free of calmodulin, they are disordered[12]. Therefore, peptide binding to calmodulin can be monitored spectrally due to formation of helix in peptide[9]. Additionally, they possess no acidic residues (negatively-charged amino acids) that would repel the glutamate on calmodulin's surface.

Calmodulin and calcium form a complex that is then able to bind target peptides (**Figure 1A**), which occurs primarily with the extension of the centrally located flexible tether. It elongates into a loop, joining the two domains and creating a hydrophobic channel complementary to the secondary structure of the target peptide. Binding of calmodulin to the target peptide induces the formation of a helical structure in the peptide, such that the peptide-calmodulin interaction resembles a hotdog in a bun (**Figure 1A**). The calcium-calmodulin complex binds the target peptide very tightly ($K_D = 0.1 \rightarrow$

10 nm), because as it folds in on itself, it forms a condensed globular shape (Sorenson 4246). This shape results in less of the surface exposed to the solvent, leading to a decreased solvation free energy that supports tight binding[12]. Multiple interactions, mostly hydrophobic but including electrostatic, stabilize the new complex. Methionines in the hydrophobic cleft play a role in forming the hydrophobic channel complementary to the target peptides. Small changes in its side-chain configurations provide a mechanism for various peptides to bind calmodulin. The electrostatic interactions are a result of the abundance of glutamate and aspartate in calmodulin that interact with basic residues in the peptide. For example, in the twenty-six-residue peptide M13, found in Myosin Light Chain Kinase (MLCK), there are 14 lysine and arginine residues that interact with the acidic residues of calmodulin and stabilize the calcium-calmodulin-M13 complex [12, 15].

2. Click Chemistry

As stated earlier, our principal goal is to create a calmodulin biomaterial, utilizing calmodulin's unique Ca^{2+}- and peptide-binding characteristics that can be functionalized through the use of a Click Chemistry reaction. Specifically, we will use copper-catalyzed azide-alkyne Click Chemistry, an important component in bioconjugation, which can be defined as the creation of covalent bonds to biological molecules. In general, Click Chemistry involves the process of creating a larger molecule from the assembly of smaller units, linking them together by heteroatom bonds, at a lower energy cost than employing new carbon-carbon bonds [3]. This method mostly avoids having to form new carbon-carbon bonds, which are less energetically favorable and more complex to make. Overall, Click Chemistry reactions target specific protein functional groups, are much more energetically favorable, and should produce a single functionalized product. For example, when comparing equilibrium aldol reactions to a Click Chemistry reaction, the new carbon-carbon bonds formed in equilibrium aldol reactions are driven by thermodynamic energy of 3 kcal mol^{-1} , whereas Click reactions are driven by more than 20 kcal mol^{-1} [16]. The energy difference demonstrates the fact that Click Chemistry provides a good leaving group for propargyl addition and stabilization of an azide reactant into a heterocycle. Moreover, the reactions can take place in solvents with high vapor pressure, making the reaction easy to separate from the solvent.

Click chemistry does not generate novel molecules, but adds new functional groups to existing molecules. As George S. Hammond said, "The most fundamental and lasting objective of synthesis is not the production of new compounds, but production of properties" [3]. The Click Chemistry reaction we will use on calmodulin is a copper-catalyzed azide-alkyne cycloaddition, which adds a 1,2,3-triazole group to tyrosines at positions Y99, Y138, or both. In future studies, the 1, 2, 3-triazole group will be used to attach a chromophore or fluorescent probe and, eventually, a drug.

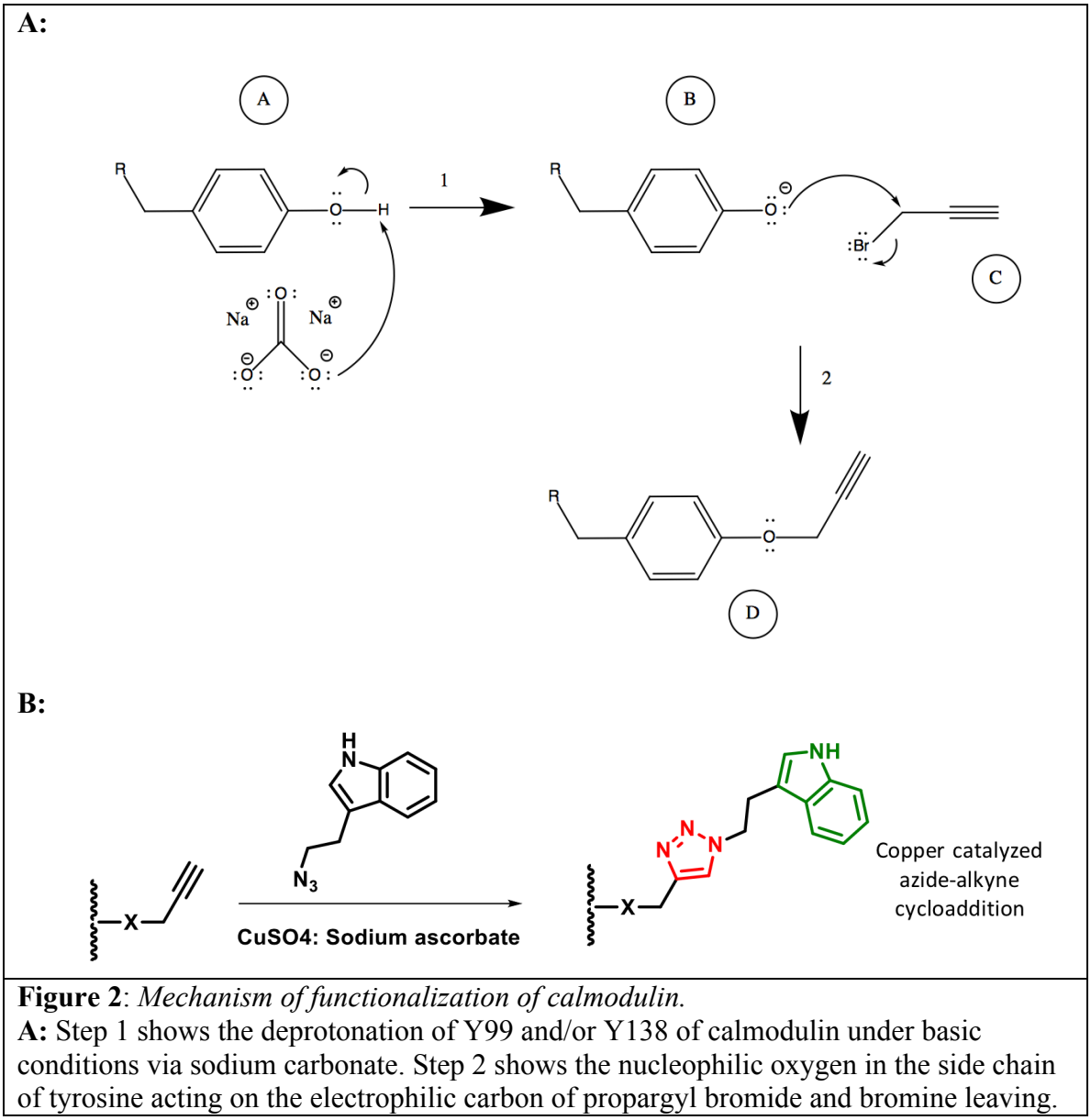


Figure 2: *Mechanism of functionalization of calmodulin.*
A: Step 1 shows the deprotonation of Y99 and/or Y138 of calmodulin under basic conditions via sodium carbonate. Step 2 shows the nucleophilic oxygen in the side chain of tyrosine acting on the electrophilic carbon of propargyl bromide and bromine leaving.

(**Figure 2** continued) Part **B** demonstrates the "Click Chemistry" reaction: the coupling of the propargyl group to an azide, creating a 1,2,3-triazole group. The copper catalyzed azide-alkyne cycloaddition was used in order to produce only the 1,4-disubstituted regioisomer. Without copper, the reaction would produce a mixture of 1,5- and 1,4-disubstituted regioisomers.

Our technique involves functionalizing the hydroxyl group on the tyrosine with a propargyl group, which can then be coupled with an azide through a copper catalyzed azide-alkyne cycloaddition reaction to create a triazole group (**Figure 2**). The result is a functional group of our choosing attached to tyrosine on calmodulin. The advantage of the Click Chemistry reaction is the flexibility of the azide coupling to the propargyl group. We can attach fluorescent probes and track the location of the derivatized calmodulin in vivo. If the fluorescent signal changes upon binding Ca^{2+} and calmodulin's physiological targets, we can track spatially calmodulin's interactions in vivo. Calmodulin can be used as an exogenous calcium sensor or sink inside or outside cells. Fluorescent labeling through the Click Chemistry reaction thus enables several major technologies.

In order to prepare a protein for organic solvents, one must dialyze the protein against water to remove salts. Then, one must lyophilize the protein to remove the water from the protein, creating a protein powder. However, even "dry" protein will have some associated water and counterions. We use the established techniques of proteins in organic solvents, but most of the work to date involves studies of enzyme activity in organic solvents. Hence, the majority of the information available on proteins in organic solvents is in regards to enzyme activity [17] and not directly relevant to our studies.

Proteins have evolved to function in aqueous environments and usually do not interact well with organic solvents. Surprisingly, they can achieve stability and maintain

catalytic ability in certain organic solvents [18]. In order to complete the copper-catalyzed azide-alkyne cycloaddition Click Chemistry we require for the modified calmodulin, we will have to use nonaqueous solvents. Due to kinetic trapping, proteins in hydrophobic organic solvents generally keep their native structure. Furthermore, it was found that proteins have a "pH memory" from the solvent in which it was prepared. However, proteins in hydrophilic or polar organic solvents may denature, as the solvent molecules disrupt bound water molecules at the surface and the hydrogen bonds with them[12]. Hence, for a protein to maintain its structure and be flexible enough to function, there needs to be a combination of polar and nonpolar organic solvents [16].

3. New Analytical Methods for Assaying Functionalized Products

Figure 3 below illustrates the logic of ongoing experiments in the collaboration between the Pedigo and Watkins laboratories.

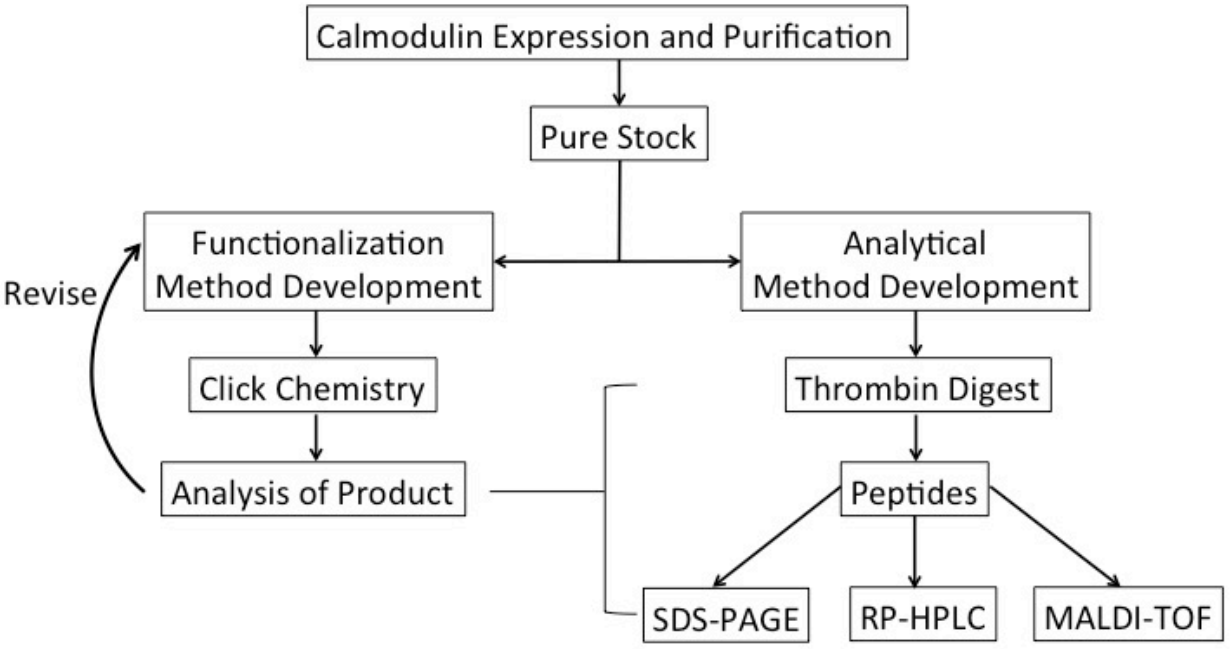


Figure 3: *Overview of experimental procedures.*
In the analytical portion of the experiment, various chromatography can be used to evaluate the peptides. SDS-PAGE will give an estimation of molecular weights of each peptide fragment, RP-HPLC will assess which fragment is derivitized, and MALDI-TOF is an alternative method which would tell how many propargyl groups, if any, the functionalized calmodulin contains.

In order to obtain the pure calmodulin stock necessary for functionalization and analysis, we transformed *E. coli* with a plasmid that expresses rat calmodulin and used it to create a pure stock. We divided the pure stock between Dr. Watkins' lab for functionalization and our lab for analytical method development.

Dr. Watkins has developed a method of functionalization and performed Click Chemistry via a copper-catalyzed azide-alkyne cycloaddition on calmodulin. The functionalized calmodulin was then analyzed by RP-HPLC, which revealed that the untreated and treated calmodulin have the same retention time, but different spectra (**Figure 4**). The spectra offer proof that tyrosines are indeed modified, as the UV spectra changes at the absorbance region of tyrosine (maximum at 280 nm). A new large absorbance can be observed at around 300 nm, which is indicative of successful modification of tyrosine in calmodulin.

Before use on the functionalized calmodulin, we used the unmodified stock to develop methods to analyze the functionalization of calmodulin. In order to analyze the functionalization at positions Y99 and Y138, we needed a specific protease to cleave between them. Thrombin was chosen, as it cleaves specifically after R106, effectively separating both tyrosines. The peptides obtained from cleavage were then assessed with SDS-PAGE, which gave approximate molecular weights for the peptide fragments. MALDI-TOF will be used to analyze fragment size and RP-HPLC with diode array detection will be used to determine retention time of the fragments and their spectrum. Once the methods have been developed and refined, we can perform an assay on the functionalized calmodulin, which could then be used to revise the method of functionalization.

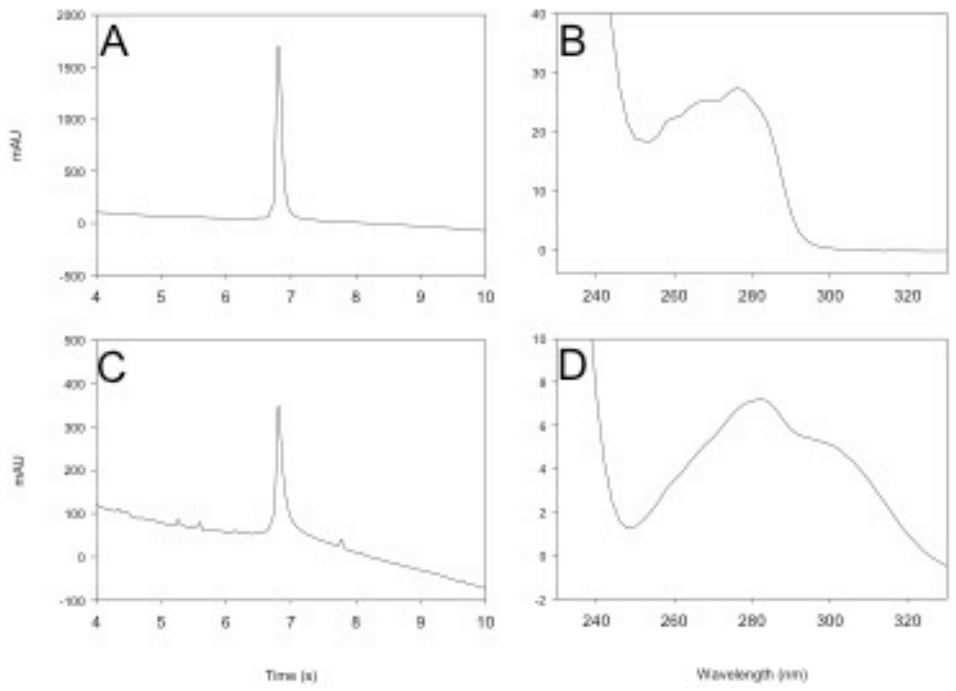


Figure 4: *Evidence for functionalization of tyrosines in calmodulin*

Conditions: Propargyl Bromide + $CsCO_3$ in DMF at room temperature for 3 hours. Precipitated in ether. Precipitate resuspended in DMSO and water. Dialyzed against water, the analyzed by RP-HPLC. Untreated calmodulin (A) elutes at 6.8 min. The spectrum of the peak (B) is identical to the spectrum of calmodulin in water, and typical for a protein that contains only phenylalanine and tyrosine (no tryptophan). The functionalized calmodulin (C) has the same retention time as unmodified calmodulin. Its spectrum is dramatically different with a large shoulder at ~ 300 nm. We confirmed that calmodulin was intact with SDS-PAGE. Data obtained by S. Pedigo.

4. Thrombin Digest

We chose thrombin for the digestion of calmodulin, as we know it to have two potential cleavage sites on calmodulin from previous experimentation[19]. Thrombin was chosen out of many serine proteases in Verhoeven's experiment because of its sensitivity to the tertiary structure of proteins and its highly selective cleavage. The high degree of specificity it exhibits confers a degree of confidence, reproducibility and ease to the experiment, which is why we used it in our proteolysis, also. Calmodulin has six arginines (**Figure 1B**), and one at 106, which is of particular interest to us, as it is positioned between the two tyrosines at positions 99 and 138. Thrombin's highly specific nature makes it ideal for our use in separating tyrosines on calmodulin, as there is large literary evidence for cleavage at only R106[19].

Thrombin is a glycoprotein that acts as a serine protease, which has functions in various biological processes. Primarily, it is known for its role in coagulation, or blood clotting. Once activated, it begins as prothrombin, a zymogen, which is then proteolytically activated in the penultimate step of the coagulation cascade[20]. It cleaves fibrinogen to convert it into its active form fibrin, which stems blood loss in damaged blood vessels by creating a protein meshwork as a new temporary physical barrier. In addition to fibrinogen, thrombin activates factors V, VII, XI, and XIII in blood coagulation[21]. Thrombin's role as a multifunctional enzyme allows it to interact with a number protein substrates, making it ideal for our use of calmodulin proteolysis, as will

be discussed. We chose thrombin as it cleaves on the C-terminal side of R106 (**Figure 4**), which lies in between the two potentially derivatized tyrosines of calmodulin, at positions 99 and 138 (**Figure 5**).

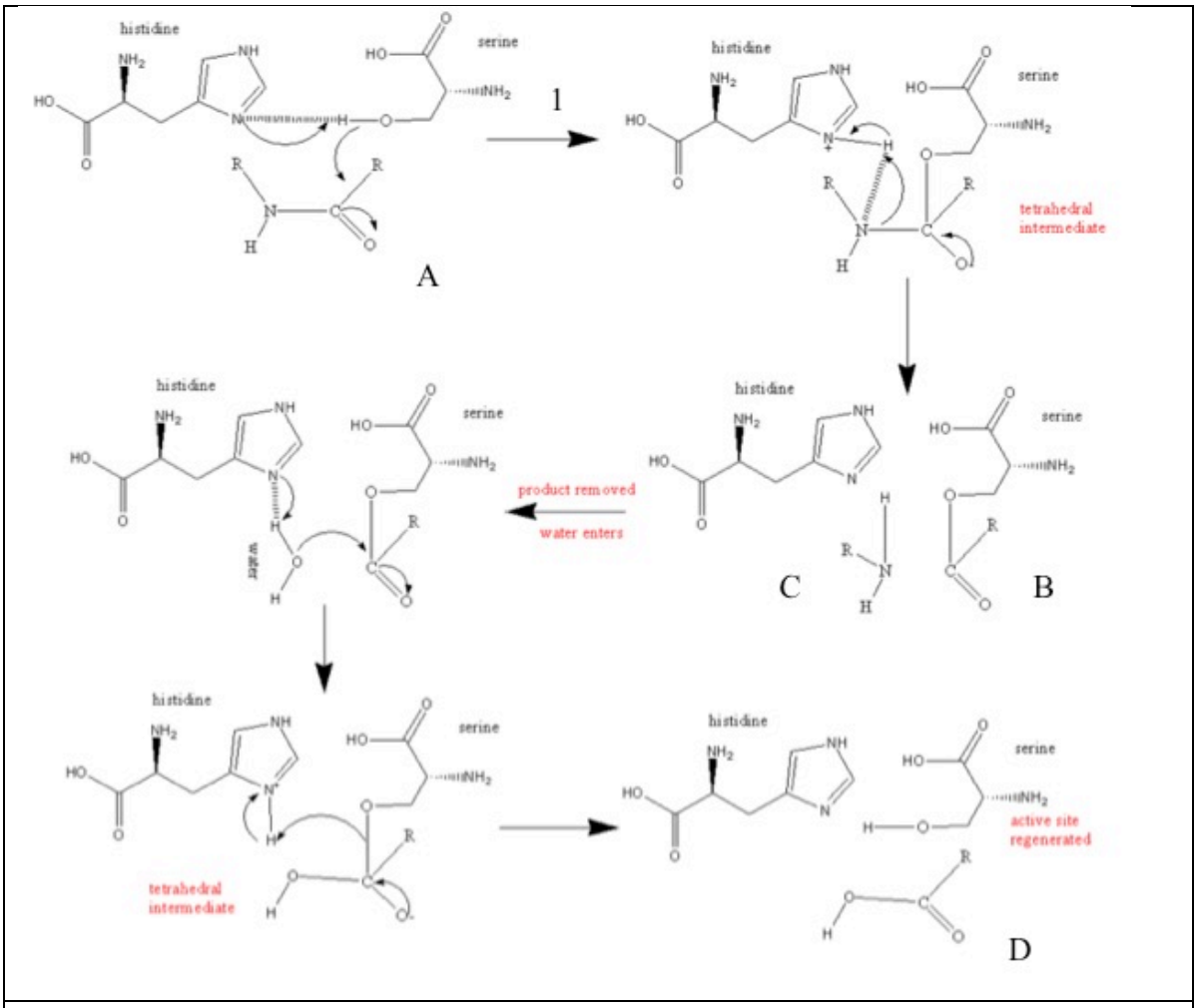


Figure 5: *Thrombin cleavage mechanism* [26]
A is the polypeptide substrate of thrombin; in our case, calmodulin. **B** is the covalent intermediate esterified by serine hydroxyl in active site. **C** is the new N-terminus formed on the peptide after hydrolysis of the peptide bond. **D** is the new C-terminus formed upon hydrolysis of the ester-linked peptide. In Step 1, thrombin, a serine protease, acts on the polypeptide substrate and forms a esterified intermediate. This happens when the nitrogen in the side chain of histidine deprotonates the oxygen in the side chain of serine. The serine oxygen then acts as a nucleophile and attacks the electrophilic carbonyl carbon in the substrate.

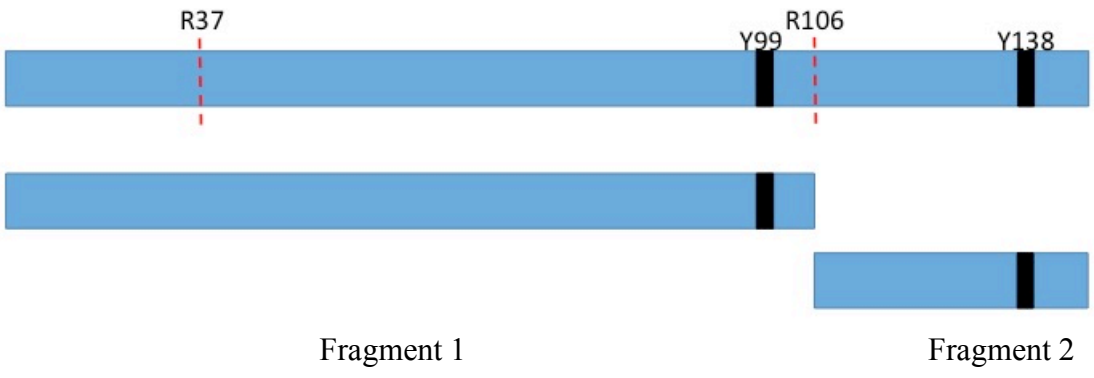


Figure 6: *Overview of thrombin cleavage of R106 in calmodulin*
Thrombin was specifically chosen for its ability to differentiate at position 106, in between tyrosines 99 and 138. Molecular weight of calmodulin is 16706 Da. The expected molecular weights of Fragments 1 and 2 are 11881 Da 4823 Da, respectively. Fragment 1 should resolve on Tris-Tricene gel, but Fragment 2 is too small to resolve.

Thrombin is a highly charged molecule with overall dimensions of 45 Å x 45 Å x 50 Å and two chains: A and B. These chains form a "single contiguous body" and are not segregated to different domains[22]. In human alpha-thrombin, the A- and B- chains possess 36 and 259 residues, respectively[22]. The A- and B-side chains interact through a disulfide bridge between cysteines at positions 1 and 122 and side chain interactions, also. Of these side chain interactions, the majority that contribute to the overall molecule stability are salt bridges that form between acidic and basic residues. These chains form two domains, with the catalytic residues between them. The catalytic residues include serine, histidine, and asparagine, and contribute to the active site, which lies perpendicular to them. The active site forms a deep well structure so that steric hindrance contributes to its great specificity. A high number of negative residues are concentrated near the active site, nestled within a molecule with an already high proportion of charged residues. The side chains of seven aspartic and glutamic acid residues contribute to the

charge here and aid in binding by positioning the positive side chain of arginine of the substrate into the active site.

The specificity of thrombin is similar to trypsin, as they are both serine proteases that cleave basic residues; however, they have profoundly different specificities. Trypsin is used in the digestive system to break down all proteins, but thrombin has a much more specific substrate in blood clotting, necessitating the increased specificity of cleavage sites. Whereas trypsin cleaves at all lysines at arginines, thrombin cleaves only after specific arginines (**Figure 6**). Thrombin is much more specific because it responds to the tertiary structure of the protein substrate due to its canyon-like active-site and fibrin(ogen)-recognizing exosite [22].

There are six arginines in calmodulin, located at positions 37, 74, 86, 90, 106, and 126 (**Figure 1B**), most of which have already been observed as substrates of trypsin[23]. There is a "calcium-induced change in the relative susceptibility of calmodulin to thrombin"[19], meaning that thrombin cleaves at a certain position depending on the presence of calcium. Thrombin cleaves the bond at R37-S38 in the moderate presence of calcium and at R106-H107 in the apo-state (absence of calcium) (**Figure 6**). No secondary cleavage is exhibited once the first bond has been cleaved. This absence of secondary cleavage could be due to the sensitivity of thrombin to a substrate's tertiary structure. After calmodulin has been cleaved, its tertiary structure is lost, therefore losing its ability to bind thrombin. In order to bind thrombin at one site, the other potential cleavage site would have to remain uncleaved; the absence of secondary cleavage could be tied in with thrombin's exosite, previously seen in fibrin(ogen) binding. This entails that one domain would serve as an exosite and the other as an active site for thrombin.

Once cleavage fragments are obtained from the digestion of thrombin, they are analyzed with SDS-PAGE, which gives the sizes of the fragments for comparison to the expected fragment sizes. My research focal point was to find an enzyme and protocol that would cleave stock calmodulin reproducibly and a technique of resolving the protein fragments acquired, so that further testing could be performed on the derivatized calmodulin. SDS-PAGE is a commonly used practice to separate denatured proteins based on mass-to-charge ratios. However, proteins that fall below 20 kDa do not resolve well in the Laemmli system, which uses glycine as a trailing ion[24], and streak when used with higher acrylamide concentrations in the separating gel. Since calmodulin weighs 16.7 kDa and the expected cleavage fragments of 1-106 and 107-148 would weigh 11881 kDa and 4823 kDa, respectively, the Laemmli system poses a problem. Thus, we selected tricine SDS-PAGE as the method of protein fragment resolution following proteolysis with thrombin. This particular approach combines components that allow for resolution of low molecular weight proteins, in range of 1-100 kDa, with greater resolution of proteins in the 5 and 20 kDa range[24].

EXPERIMENTAL PROCEDURES

Materials

Thrombin (EC 3.4.21.5) was purchased from Akron Biotechnology (Boca Raton, FL). Spectra Multicolor Low Range Protein Ladder (26628) was obtained from ThermoFisher Scientific (Waltham, MA). Acrylamide/bis-acrylamide, 40% solution, was obtained from Sigma (St. Louis, MO). Recombinant rat CaM was over-expressed bacterially, purified, characterized, and stored as described previously (Pedigo & Shea, 1995).

Methods

Overexpression and Purification of Calmodulin from E. coli. In order to obtain a pure stock of calmodulin for use in both functionalization and analytical methods, calmodulin was expressed and purified from *E. coli* in accordance with **Figure 7**. In order to achieve this, on the first day, Agar plates were prepared and inoculated with pre-transformed cells. On the second day, 100 mg/mL ampicillin stock was added to one LB culture. Solutions of 20% glucose, 1 M potassium phosphate, overnight culture and 100 μg/mL ampicillin stock were then added to each 1 L LB culture. As the bacteria grew, absorbance at 600 nm was then checked to 0.6 to 1.0 AU. To induce, 0.4 M IPTG was

then added to each culture and centrifuged. The cells were later harvested by centrifugation. They were then resuspended in 20 mM HEPES, 100 mM KCl, and pH 7.4 buffer. The cells are lysed through sonication and centrifugation, then the supernatant and pellet were separated. To the supernatant 1 M calcium was added, heated to 80 °C to precipitate cellular proteins, placed on ice, and centrifuged. The supernatant and pellet were again separated. The supernatant was then treated with DNase and dialyzed against 1 M NaCl and 10 mM Tris at a pH of 7.5. Chromatography was then performed to obtain the purified protein. Once the protein eluted from the column, it was dialyzed against water and lyophilized.

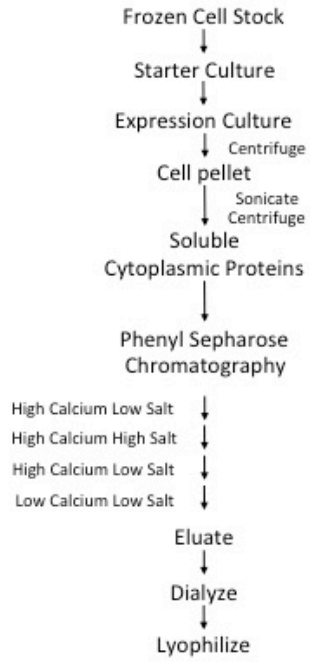


Figure 7: *Expression and purification of calmodulin from* E. coli.
This flow chart illustrates the procedural steps involved in the expression and purification of calmodulin from *E. coli*. The gene for rat calmodulin was cloned into a PT7 vector that conferred Ampicillin resistance to BL21 (DE3) *E. coli*. Transformed cells were frozen as a glycerol stock.

Thrombin Digest. In order to obtain a reproducible technique for the cleavage of calmodulin, a thrombin digest was designed to completely cleave calmodulin and provide correct products. Optimizing the method involved finding the optimal concentrations of

calmodulin, the optimal ratio of protein to enzyme, and the optimal time of digestion. The optimal concentration for the calmodulin was found to be 100 μM. This level of protein allowed for adequate band density on a Coomassie stained SDS-PAGE gel. The thrombin stock was 6.8 μg/μL. The optimal ratio of protein to enzyme was eventually found to be 0.7 u for every 14 μg calmodulin, based off of the protocol of Verhoeven [19]. The time that produced the highest yield of digested calmodulin was found to be overnight. Hence, thrombin was added to calmodulin at the optimal concentrations in a reaction microfuge tube and let run overnight, collected, and frozen.

An additional 20 μL calmodulin sample was digested with thrombin for mass spectrometry analysis of fragment size and RP-HPLC analysis of fragments to determine retention time.

SDS PAGE. To test the successfulness of the calmodulin digest by thrombin and to ascertain thrombin did in fact cleave the R106-H107 bond, protein digests were assessed by SDS-PAGE on Tris-Tricene gels. Acrylamide concentrations of 20% Tris-Tricene gels were created according to the protocol of Schagger et al. Samples were added to equal volumes of 2X reducing and denaturing loading buffer and boiled for one minute. Samples were then loaded onto gels and run at 100 V for one and a half to two hours. The gel was then removed and placed in Coomassie blue stain for twenty minutes and then destained overnight.

RESULTS AND DISCUSSION

UV-VIS FIGURE SHOWING DERIVITIZATION

Ultraviolet-visible spectroscopy was performed on the modified calmodulin, and an additional peak was found (**Figure 4**). This peak confirmed the successful functionalization of calmodulin. However, the tyrosine on which it occurred, at either position 99 or 138, is unknown. Therefore, we aspired to establish a reproducible technique separating the two tyrosines in order to subject them to future study to confirm the existence and location of functionalization.

A reliable method in which to cleave in between tyrosines 99 and 138 of calmodulin was found by using 6.8 μg/μL Thrombin and 100 μM calmodulin. After the cleaved peptides were subjected to SDS-PAGE, bands of the expected molecular weights appeared on the gels, confirming the successful cleavage (**Figure 8**). Cleavage of calmodulin after R106 should only result in two bands on the SDS-PAGE: the large (107-148) and small (1-106) peptides. However, three discrete bands are observed. The third and last band seen in lanes 3 and 4 could be the small fragment (1-37) resulting from thrombin cleavage after R37. If this happened, the large fragment (38-148) could have co-eluted with the large fragment (107-148) from cleavage after R106, creating one large band on the gel.

Cleavage after R37 is possible, but there is strong data supporting selective cleavage after R106 in apo conditions[19]. We used EDTA to ensure the apo state for calmodulin. In the Verhoeven study, the bands from cleavages at R37 and R106 had distinct migrations[19]; however, a gradient gel was used, whereas we used 20% Tris-Tricene SDS-PAGE. It should be noted that if there is cleavage after R37, both tyrosines would be located in the large fragment, rather than after R106, which separates the tyrosines into the large and small fragments. We can conclude that thrombin was successful in cleaving calmodulin at the R106-H107 bond, but cleavage at another site may have been observed. Future spectroscopic study of calmodulin cleaved by thrombin could conclude at which site was calmodulin cleaved and the percentages of cleavage at each site. However, we can confidently say that thrombin effectively separated the two tyrosines and rendered them to future spectroscopy.

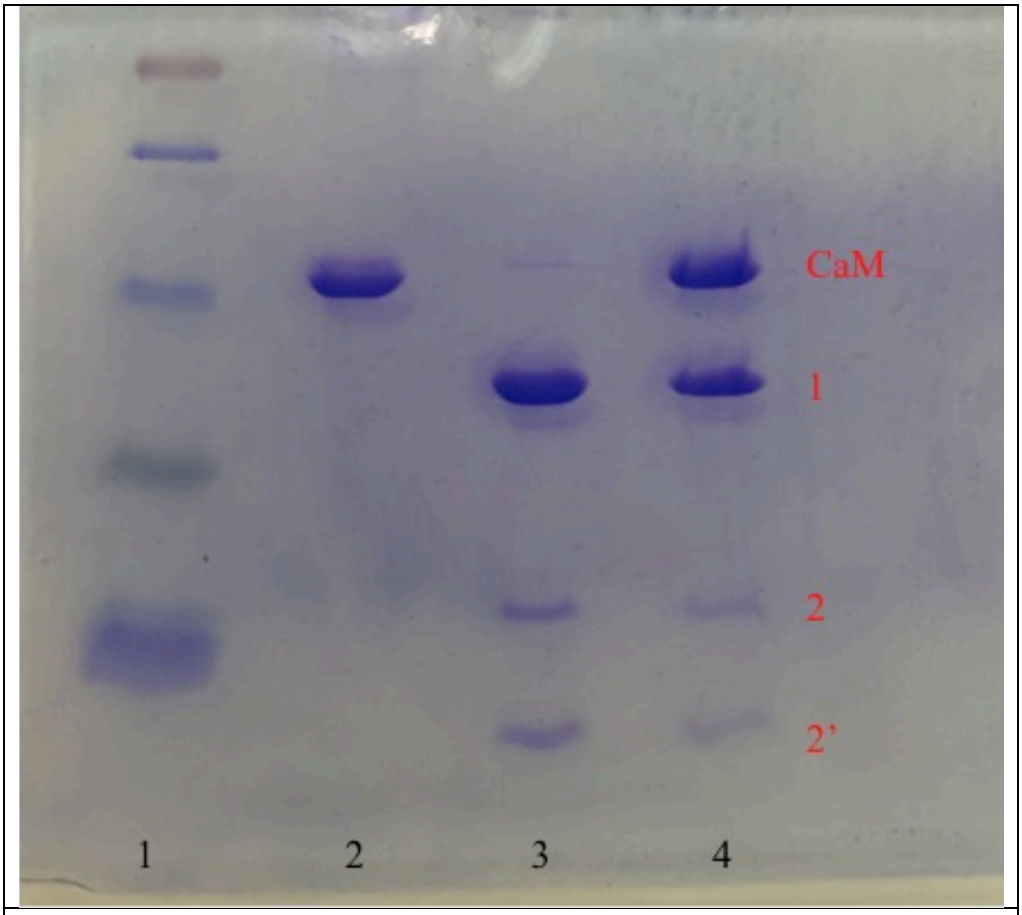


Figure 8: *SDS-PAGE analysis of cleavage fragments.*
20% Tris-Tricene SDS-PAGE. Lane 1: 5 µL Low molecular weight standard. Lane 2: 3 µL calmodulin/EDTA. Lane 3: 6 µL Digested calmodulin. Lane 4: 3 µL calmodulin/EDTA added to 6 µL Digested calmodulin. Lanes 3 and 4 contain peptide bands resulted from the R106-H107 thrombin cleavage of calmodulin.

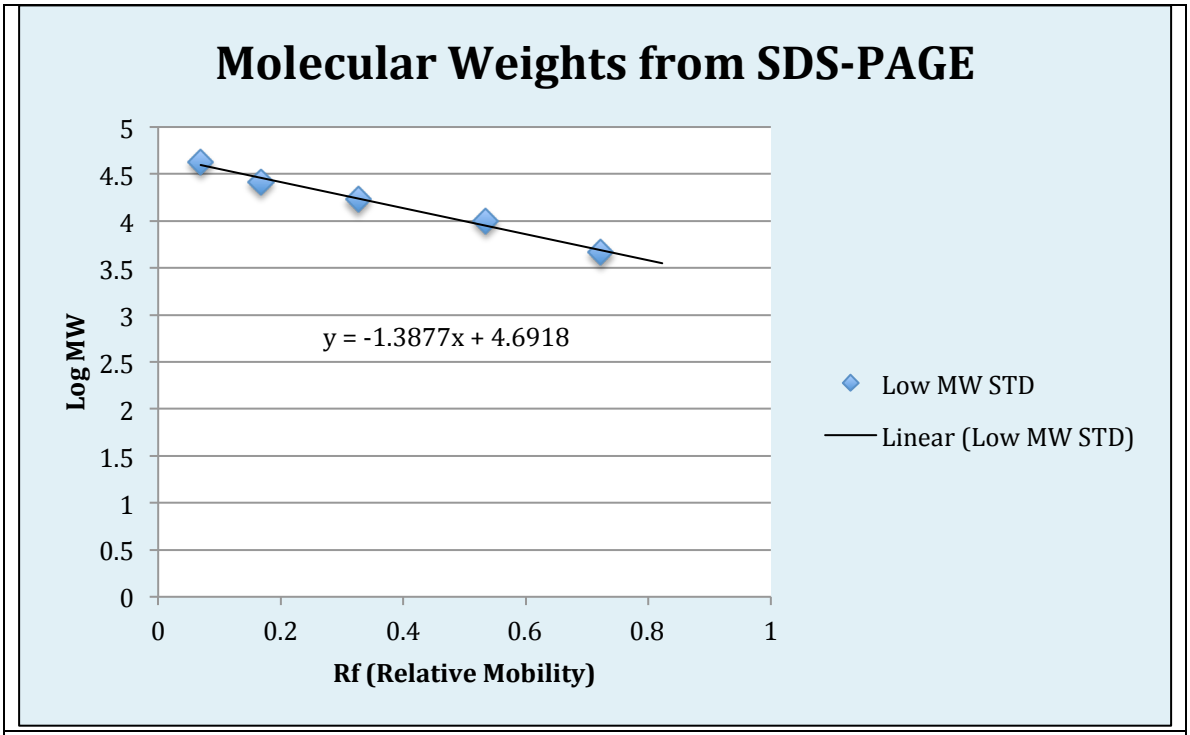


Figure 9: *Measuring molecular weights from SDS-PAGE.*
R_f vs. Log MW of the low molecular weight standard. R_f = migration distance of protein/migration distance of dye front. The values were obtained by measuring distances between bands and the buffer front of the gel. The R_f was obtained and then plotted against the log of the molecular weight, which are known, of each standard band. The graph then allows us to obtain an equation to use in determination of the molecular weights of peptide fragments. The outlying value was left in, as the removal of the apparent abhorrent point skewed values. It is clear that mass spectrometry is needed to determine molecular weight values.

Table 1. Expected v. Actual Molecular Weights of Peptide Fragments.

	MW (Da)	Experimental MW (Da)
Calmodulin (CaM)	16706	16578
Large Fragment (1)	11892	10364
Small Fragment (2)	4832.0	3745.1
Small Fragment (2')	?	2164.9
The molecular weight of the smallest bands are not well-predicted because they are lighter than the smallest standard.		

As a reproducible technique was found to cleave calmodulin, it is now possible to test the potentially functionalized calmodulin using RP-HPLC and MALDI-TOF.

CONCLUSION

In conclusion, informal strategies for functionalizing calmodulin require thorough development of analytical methods. These analytical methods seek to provide information on the location and nature of functionalization of calmodulin. Enzyme to substrate ratio and substrate amounts were established, as well as the digestion time by thrombin. The cleavage of calmodulin was achieved by a specific and carefully chosen protease, thrombin, which was chosen for its specificity towards cleavage of the R106 peptide bond. Cleavage of calmodulin after position 106 successfully separates the two tyrosines at positions 99 and 138, which are sites of potential functionalization. In order to ascertain at which site functionalization has taken place, the tyrosines were separated by thrombin, creating two cleavage fragments. These fragments were then subjected to SDS-PAGE testing, which provided an estimation of the fragment molecular weights. The expected molecular weights of the fragments were then compared to the estimated molecular weights. We concluded that thrombin achieved cleavage of calmodulin into one large fragment, likely corresponding to residues 1-106. Two small fragments were found; presumably one is the 37-148 fragment. Further testing in regards to mass spectrometry has been arranged in order to determine the exact number and location of functionalized groups on calmodulin.

In conclusion, we established protocols for calmodulin cleavage by thrombin and cleavage fragment evaluation by SDS-PAGE. The establishment of analytical methods

towards the functionalization of proteins allows for further research in the area of calmodulin as a biomaterial. Once a reliable and reproducible analytical method has been established, more functional groups can be explored in the functionalization of calmodulin.

LIST OF REFERENCES

1. Hobson, K.F., N.A. Housley, and S. Pedigo, *Ligand-linked stability of mutants of the C-domain of calmodulin.* Biophys Chem, 2005. **114**(1): p. 43-52.
2. Tsalkova, T.N. and P.L. Privalov, *Thermodynamic study of domain organization in troponin C and calmodulin.* J Mol Biol, 1985. **181**(4): p. 533-44.
3. Kolb, H.C., M.G. Finn, and K.B. Sharpless, *Click Chemistry: Diverse Chemical Function from a Few Good Reactions.* Angew Chem Int Ed Engl, 2001. **40**(11): p. 2004-2021.
4. Cohen, P. and C.B. Klee, eds. *Calmodulin*. Molecular Aspects of Cellular Recodnition, ed. P. Cohen. Vol. 5. 1988, Elsevier: Amsterdam. 371.
5. Means, A.R., et al., *Regulatory functions of calmodulin.* Pharmacol Ther, 1991. **50**(2): p. 255-70.
6. Babu, Y.S., C.E. Bugg, and W.J. Cook, *Structure of calmodulin refined at 2.2 A resolution.* J Mol Biol, 1988. **204**(1): p. 191-204.
7. Evans, T.I.A., J.W. Hell, and M.A. Shea, *Thermodynamic linkage between calmodulin domains binding calcium and contiguous sites in the C-terminal tail of CaV1.2.* Biophys Chem, 2011. **159**: p. 172-187.
8. Cook, W.J., L.J. Walter, and M.R. Walter, *Drug binding by calmodulin: crystal structure of a calmodulin- trifluoperazine complex.* Biochemistry, 1994. **33**(51): p. 15259-65.
9. Blumenthal, D.K., et al., *Identification of the calmodulin-binding domain of skeletal muscle myosin light chain kinase.* Proc Natl Acad Sci U S A, 1985. **82**(10): p. 3187-91.
10. Maximciuc, A.A., et al., *Complex of calmodulin with a ryanodine receptor target reveals a novel, flexible binding mode.* Structure, 2006. **14**(10): p. 1547-56.
11. Ikura, M., et al., *Solution structure of a calmodulin-target peptide complex by multidimensional NMR.* Science, 1992. **256**(5057): p. 632-8.
12. Clore, G.M., et al., *Structure of calmodulin-target peptide complexes.* Current Opinion in Structural Biology, 1993. **3**: p. 838-845.
13. Sorensen, B.R. and M.A. Shea, *Interactions between domains of apo calmodulin alter calcium binding and stability.* Biochemistry, 1998. **37**(12): p. 4244-53.
14. Yagi, K., et al., *Interaction between calmodulin and target proteins.* Adv Exp Med Biol, 1989. **255**: p. 147-54.
15. Meador, W.E., A.R. Means, and F.A. Quiocho, *Target enzyme recognition by calmodulin: 2.4 A structure of a calmodulin-peptide complex.* Science, 1992. **257**(5074): p. 1251-5.

16. Soares, C.M., V.H. Teixeira, and A.M. Baptista, *Protein structure and dynamics in nonaqueous solvents: insights from molecular dynamics simulation studies.* Biophys J, 2003. **84**(3): p. 1628-41.
17. Mattos, C. and D. Ringe, *Proteins in organic solvents.* Curr Opin Struct Biol, 2001. **11**(6): p. 761-4.
18. Oppenheim, S.F., et al., *Aromatic hydroxylation catalyzed by toluene 4-monooxygenase in organic solvent/aqueous buffer mixtures.* Appl Biochem Biotechnol, 2001. **90**(3): p. 187-97.
19. Shea, M.A., A.S. Verhoeven, and S. Pedigo, *Calcium-induced interactions of calmodulin domains revealed by quantitative thrombin footprinting of Arg37 and Arg106.* Biochemistry, 1996. **35**(9): p. 2943-57.
20. Pozzi, N., et al., *Crystal structure of prothrombin reveals conformational flexibility and mechanism of activation.* J Biol Chem, 2013. **288**(31): p. 22734-44.
21. Pechik, I., et al., *Crystal structure of the complex between thrombin and the central "E" region of fibrin.* Proc Natl Acad Sci U S A, 2004. **101**(9): p. 2718-23.
22. Bode, W., D. Turk, and A. Karshikov, *The refined 1.9-A X-ray crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human alpha-thrombin: structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships.* Protein Sci, 1992. **1**(4): p. 426-71.
23. Mackall, J. and C.B. Klee, *Calcium-induced sensitization of the central helix of calmodulin to proteolysis.* Biochemistry, 1991. **30**(29): p. 7242-7.
24. Schagger, H. and G. von Jagow, *Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa.* Anal Biochem, 1987. **166**(2): p. 368-79.
25. Pedigo, S., & Shea, M. A., *Quantitative endoproteinase GluC footprinting of cooperative Ca2+ binding to calmodulin: proteolytic susceptibility of e31 and e87 indicates interdomain interactions*. Biochemistry, 1995. **34**(4): p. 1179-1196.
26. Hedstrom, L. Serine protease mechanism and specificity. Chem Rev, 2002. **102** (12): 4501-24.