Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Kenneth Wingate, Jr. April 10, 2024

A Study of the Effect of Amino Acid Substitution on Cannula Mimetic Proteins

By

Kenneth Wingate, Jr.

Vincent P. Conticello
Adviser

Department of Chemistry

Vincent P. Conticello
Adviser

Katherine Davis
Committee Member

David Lynn
Committee Member

2024

A Study of the Effect of Amino Acid Substitution on Cannula Mimetic Proteins

By

Kenneth Wingate, Jr.

Vincent P. Conticello
Adviser

An abstract of
a thesis submitted to the Faculty of
Emory College of Arts and Sciences of Emory University
in partial fulfillment of the requirements for the degree of
Bachelor of Science with Honors.

Department of Chemistry

2024

Abstract

A Study of the Effect of Amino Acid Substitution on Cannula Mimetic Proteins
By Kenneth Wingate, Jr.

Protein assemblies and self-assembling peptides are finding new and diverse ways to be implemented in various fields of science. Previous research has presented the discovery of protein assemblies and self-assembling peptides throughout nature and the various roles they have in their environments. As more is understood about how they function and what properties contribute to that level of functioning, the application of these proteins becomes more diverse. We studied the structures of supramolecular assemblies from the archaea extremophile *P. abyssi* to better understand its use of strand donation polymerization, the ability to form Cannula A and Cannula B nanotubes with high heat and the presence of calcium, and how residue substitutions affect the overall process. The role that ligands and amino acid residues play in the lateral association of these cannulae is not yet completely known, but substituting residues of Cannula B to create a higher similarity to that of Cannula A led to greater polymerization under less strict conditions.

A Study of the Effect of Amino Acid Substitution on Cannula Mimetic Proteins

By

Kenneth Wingate, Jr.

Vincent P. Conticello
Adviser

A thesis submitted to the Faculty of the
Emory College of Arts and Sciences of Emory University
in partial fulfillment of the requirements for the degree of
Bachelor of Science with Honors.

Department of Chemistry

2024

Acknowledgments

I would like to thank all of the members of the Conticello lab for supporting me every step of the way. A special thanks to Andres Socorro for always being there no matter what time that I may have needed help troubleshooting, understanding concepts, or completing lab work. I would also like to thank my PI, Dr. Vincent P. Conticello for being extremely flexible and available throughout my time in the lab. An especially big thanks goes to Brian (Zijun) Zhao for helping me with so much; from formatting to offering new perspectives, Brian has been a reliable support throughout this process. Without these people, this work would not have been possible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Self-assembling peptides (SAPs) are pervasive throughout nature and present in all kingdoms of life. Ranging in complexity, the monomeric units of peptide assemblies can be simple peptide sequences of six amino acids up to larger peptides such as actin, which has a molecular weight of 42kDa and a primary sequence that is 375 amino acids long.[1] The variety in the monomeric units and the methods of assembly gives rise to a multitude of assemblies, all with different properties. Common examples of higher-order structures formed from SAPs include fibrils, colloidal structures, amorphous aggregates, and amyloid fibrils.[2,3] From such structures, there are creative and varied implementations of SAPs in nature including spider silk, bacterial pili, and keratin.[4,5] Outside of a purely natural context, mankind has started to utilize SAPs to their advantage across multiple disciplines. Integral to medicinal chemistry and material science, SAPs are used for creating hydrogel scaffolds and engineering nanomaterials.[6,7] Some natural SAPS are also known for being antimicrobial, and even have a close link to fibril-forming amyloid-like self-assemblies.[8,9] With such a broad impact, there is a push to understand more about SAPs as a whole.

To understand SAPs, scientists have looked at various natural sources of peptides that form high-order, supramolecular structures. Pili formation on gram-negative

bacteria is an abundant example of protein assemblies that are readily researchable.[10] Pili function ranges from bacterial conjugation and adhesion to twitching motility and promotion of biofilms.[11,12,13] The structures referred to as pili are filamentous assemblies that are often found across many prokaryotic cells.[14] The formation of these pili is dependent upon many factors such as pH, presence of metal cations, and temperature. The numerous roles that pili can fill make studying them an ideal candidate. Although they perform intricate tasks and have complex structures, the greater structure of the protein assembly is dependent upon the primary sequence and the environmental factors around the protein itself. Using pili as a basis for studying protein assemblies has resulted in major contributions to the scientific knowledge of protein assemblies, but there is still a lack of understanding the structure-to-function relationship of protein assemblies. More information stands to be learned about the driving forces involved in the transition from the primary sequence of the peptide to the higher-order structures.

While bacteria have functioned as a model organism for protein assembly studies, there is a limitation to what can be learned from them. The proteins related to an organism are a consequence of the organism's environment and general needs. By expanding these studies to organisms that require different functional roles than that of bacteria, more in-depth studies can be performed to elucidate new knowledge of assemblies. To better understand the structure-to-function relationship, the archaeon *Pyrodictium abyssi*, an extremophile, was studied. It is a part of the genus *Pyrodictium*, which is known to withstand temperatures up to 121°C.[15] As an organism that lives in extreme temperatures, the proteins created by this organism must also be able to withstand the extreme environment of the archaeon. A process that occurs with *P. abyssi* that involves protein assemblies is asexual reproduction. *P. abyssi* forms networks of tubular connections upon undergoing binary fission. These connections are cannulae, protein filaments with large diameters, that are formed from SAPs from

the organism. The proteins believed to be larger involved in the formation of these cannulae are three homologous glycoproteins: Cannula A (CanA), Cannula B, and Cannula C (CanC).[16]

Previous work with assembling and characterizing these same nanotubes, but expressed in bacteria, functioned as the basis for this work. CanA is a 21 kDa hyperthermostable protein that was recombinantly expressed in *E. coli* at 46°C. Unlike the native cannulae, the cannulae produced from *E. coli* are heavily glycosylated due to bacteria not having the same post-translational modification capabilities as archaea.

CanA samples formed filamentous assembles via strand donation polymerization with heating and binding of calcium ions. Apart from the deletion of the signal sequence, the last 20 amino acids form the donor strand. CanB is a 15.6 kDa hyperthermostable that also forms filamentous assemblies. Despite CanA and CanB having high similarities, CanB does not form nanotubes as easily as CanA does. To promote nanotube formation, sequence alignment of the two peptides was performed in order to change some residues from CanB to match those of CanA. This new 16.31 kDa peptide is referred to as CanB mutant (CanB_mut), and its behavior with assemblies can give insight into the relationship between primary structure and supramolecular structure formation; more specifically, it can be seen what amino acids play an active role in calcium binding, strand donation, and lateral association. It is hypothesized that mutating CanB to be more similar to CanA will result in a higher propensity for assembly formation.

Chapter 2

Materials and Methods

2.1 Plasmid DNA isolation and purification

Amplification of the recombinant DNA plasmid was achieved by transforming (see "Transformation protocol" below) commercially competent cells. QIAprep® Spin Miniprep kit plasmid DNA purification protocol was used with 2 mL of starter culture of CanB and CanB_mut. Nanodrop (Thermo Scientific NanoDrop Onec) measurements yielded concentrations of 44.9 ng/μL and 60.8 ng/μL with 260/280 of 1.88 and 1.82 for CanB and CanB_mut, respectively. Excess plasmid DNA was stored at -30°C and CanB and CanB_mut Top10 cells were stored at -80℃ in a solution of 500 μL culture and 500 μL 50% glycerol solution.

2.2 Protein expression

The transformation protocol was followed using BL21gold competent cells with CanB and CanB_mut DNA from plasmid DNA isolation and purification. Primary cultures were incubated at 37°C for 16-24 hours at 200 RPM. Induction of 50 mL secondary cultures was performed by adding 1 M isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 10 Mm then allowing incubation until reaching an optical

density (OD) of 0.4. A 1 mL aliquot was taken before and after IPTG induction. After a 4-hour incubation at 42°C and 200 rpm, the 50 mL solutions were centrifuged for 15 minutes at 10,000 RPM (Sorvall® RC 5C centrifuge) and the 1 mL samples were centrifuged at max speed for 20 minutes in a tabletop centrifuge (Fisher Scientific accuSpin Micro17 centrifuge). Supernatants were discarded and cell pellets were stored at -20℃.

2.3 Transformation protocol

Cells were thawed on ice for 10-20 minutes. Once thawed, 2 μL DNA were added to the cells while remaining on ice for an additional 10-20 minutes before a 45-second heat shock at 42 ± 1°C. Following the heat shock, cells were returned to ice for 2 minutes. 300 μL LB Miller broth (IBI Scientific) or terrific broth (TB) (IBI Scientific) were added to the cells and incubated for 1 hour at 37°C before plating on an agar plate. Agar plates used in the experiments were LB-kanamycin.

2.4 Protein isolation and purification

Thawed cells were lysed by resuspension in 2 mL non-denaturing lysis buffer with $MgCl_2$ added to a final concentration of 1mM and the addition of 50 μL lysozyme and 1 μL benzonase (EMD Millipore Corp, 71205-3). The samples were then incubated for 1 hour in a shaker at a low RPM. Cell samples were centrifuged at 10,000 RPM for 20 minutes. Pellets were frozen and stored at -20°C and supernatants were collected. To purify the protein, the supernatant was subjected to multiple rounds of dialysis with low-salt buffer and decreasing concentrations of EDTA. The supernatant was first dialyzed against 5mM EDTA low-salt buffer twice for 2 hours each round of dialysis. Then, the supernatant was dialyzed twice against 0.01 mM EDTA low-salt buffer for 2 hours for each dialysis. The solutions were then heated at 80°C for 20 minutes and

then spun down at max speed in a tabletop centrifuge for 20 minutes. After heating and collecting the supernatant, the last round of dialysis was performed against a low-salt buffer with no EDTA overnight. All dialysis was performed with SnakeSkin dialysis tubing (Thermo Scientific, 68700) with a molecular weight cut-off (MWCO) of approximately one-half the protein size. All protein samples were stored at -20°C.

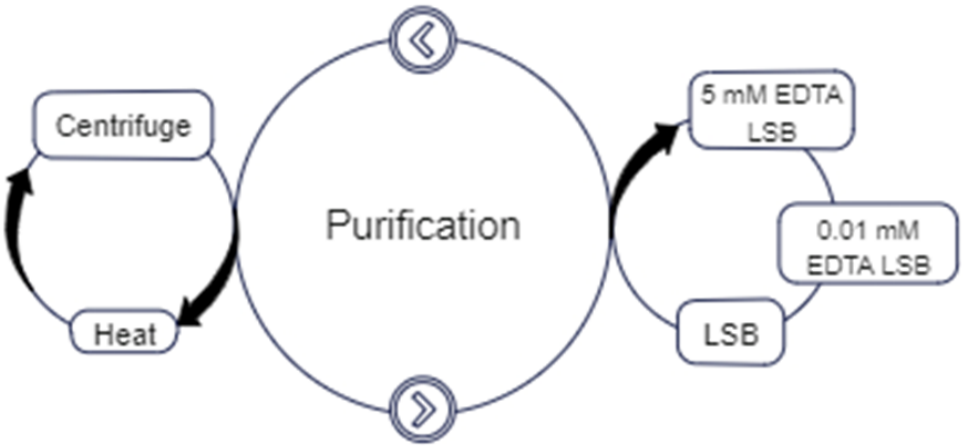

Figure 2.1: A representation of the purification cycle performed on the protein samples. Cycles could be repeated to maximize protein purity.

2.5 Assembly conditions

CanA assembly conditions were used as a basis for optimizing assembly conditions for CanB and CanB_mut. This partial optimization only took into consideration pH, Presence of $CaCl_2$ and $MgCl_2$, temperature, and incubation time while annealing.

2.6 Sample assembly preparation

1 μL of $CaCl_2$ was added 49 μL of purified protein sample and then heated at 60-80°C for a range of 1 to 48 hours. Samples of CanA were prepared to have a final calcium ion

Table 2.1: A table of the properties seen to affect the assemblies the most. The initial guess for the values of the assembly parameters was based on assembly conditions for CanA.

Optimization Parameters and Values				
pH	5		7	
Temperature	20°C		80°C	
Presence of salts	$MgCl_2$		$CaCl_2$	
Incubation time	1 hour	2 hours	1 day	5 days

concentration of 10 mM and samples of CanB and CanB_mut were prepared to have a final calcium ion concentration of 20 mM. CanA requires as little as 15 minutes but is incubated for 1 hour for full assemblies. CanB and CanB_mut were heated for 2 hours for samples of high concentration and upwards to 48 hours for lower concentrations (> 100 μM). After annealing, samples were left at room temperature to cool before preparing TEM grids. For seeding, CanA was incubated at 80°C for 15 minutes to form small assemblies that were confirmed by TEM. The resulting solution was then passed through centrifuge filters to filter out the monomers and leave behind only small assemblies. Seeding assembly samples were then prepared with 5 μL of CanA seeds added to 44 μL CanB or CanB_mut and 1 μL $CaCl_2$.

2.7 Circular dichroism (CD) spectroscopy

To characterize the secondary structures of the proteins, a minimum of 25 mL of the assembled protein sample was prepared and used at 25°C. The CD spectra acquisition was performed using a Jasco J-1500 CD spectrophotometer with a 0.1 mm path length cuvette. The images were doubly processed by subtracting the background, and the second processing was curve smoothing with 10 data point interval averaging.

2.8 Transmission electron microscopy (TEM)

TEM was used to view the nanoscale assembly that formed from CanA, CanB, CanB_mut, and the seeding experiments. Grids for all samples were prepared via negative staining on formvar/carbon-supported copper grids. Each protein sample was placed on a grid and remained for one minute before wicking the excess buffer. A solution of 1:1 NanoVan:NanoW (Ted Pella, INC. Prod No. 14702) (Nanoprobes 2011-5mL) was prepared and stored at 2°C. Each grid was negatively stained for one minute in the NanoVan/NanoW solution. Excess stain was wicked off the grids, and grids were placed in a desiccator until imaged at 80-120 kV with JOEL JEM-1400 or Hitachi HT-7700.

Chapter 3

Results and Discussion

3.1 Secondary Structure Characterization

The CD of CanA assemblies at 25°C presented peaks around 210 cm^{-1} and 230 cm^{-1}. The CD curve reflects the β-sheet structures found throughout the jelly roll fold that is present in CanA.

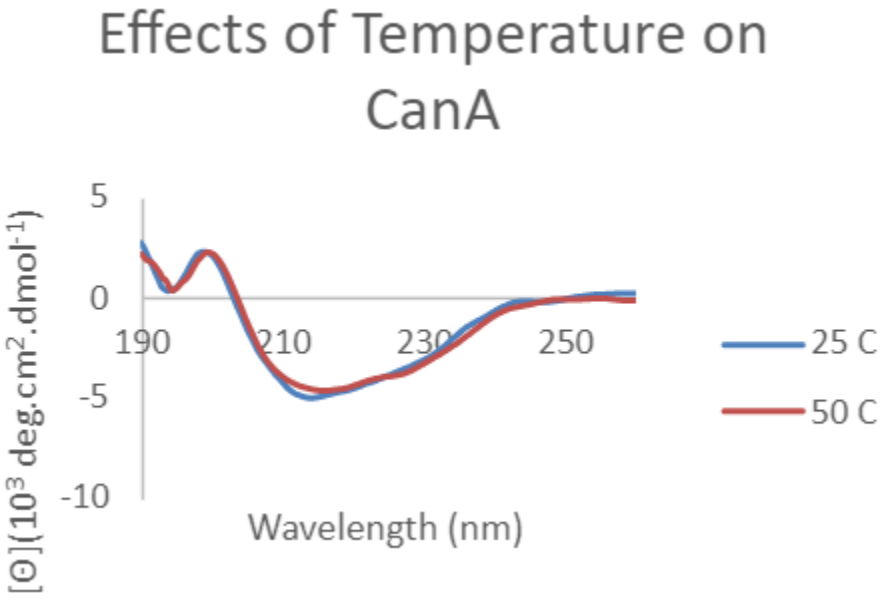

Figure 3.1: CD curve of CanA at 25°C and 50°C with wavelengths ranging from 190-280nm.

Secondary structures of CanA, CanB, and CanB_mut were also supported by AlphaFold predictions. The AlphaFold predictions were found to be high confidence for all three samples. CanA, CanB, and Can_mut consist of a jelly roll fold, which is considered to be a form of a beta-barrel. The jelly roll fold consists of 8 antiparallel beta strands, but the jelly roll fold for these 3 structures is missing a beta-strand. This missing strand is filled by the proposed strand donation motif for polymerization. That is, the N-terminus tail that is formed after the protein folds is inserted into the space where the missing beta strand would be.

Although there is a structure of CanA that resulted from studies with Cryo-EM, no information can be inferred about the conformational state of the protein when it is not assembled and there is no calcium bound. To investigate the unassembled structure of CanA, X-ray crystallography was used in addition to AlphaFold. The structure of the CanA protomer solved from Cry-EM and from X-ray crystallography were aligned. The structural alignment showed a change in the conformation of two loops involved in the binding of calcium. This suggests a change in the configuration of the loops when undergoing assembly.

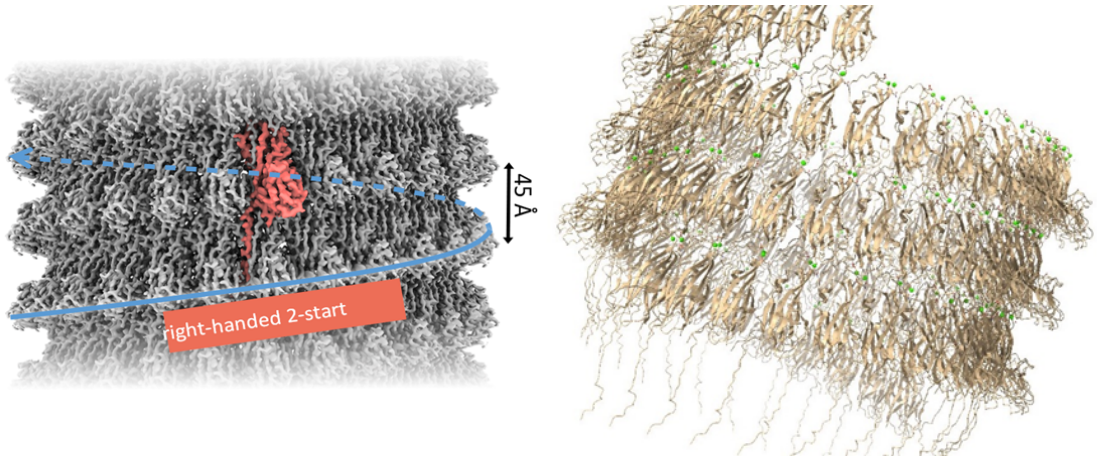

Figure 3.2: Solved supramolecular assembly of CanA, a right-handed helical nanotube (left). The blue arrow indicates the 2-start aspect of the assembly. ChimeraX depiction of the CanA nanotube with calcium ions (green) in their respective binding sites.[17]

These structures are consistent with the resolved Cryo-EM structures that reflect

calcium binding and X-ray crystallography structures without calcium binding. Two main differences in the CanA structure were found and may have a significant impact on the polymerization process. One loop (loop 34-40) does not appear in the X-ray structure. This suggests that the loop is highly flexible or disordered. The ability of this loop to take on different conformations may aid in binding one calcium ion and forming favorable, stable interactions between the ions and respective residues. The second loop, loop 163-172, is found in a position that prohibits strand donation. It is possible that heating the protein introduces the required energy for loop 163-172 to overcome stabilizing interactions in its original folded state and find more favorable, lower-energy states involving ion-dipole interactions upon binding calcium. CanA has more than one critical calcium ion binding pockets that influence the structure of the protomer. The previously mentioned binding pocket is internal, but there is also an intermolecular calcium ion binding site. The residues that are responsible for this ion binding are Glu-161, Glu-164, and Gly-37. The importance of these residues can be assessed by replacing them with various amino acids and assessing assembly. Without further study of mutations to this loop and binding site, the position of the binding site suggests that this calcium-binding may aid in the lateral association of the polymerization process. Assuming that the binding pockets are critical for both strand donation and lateral association, and subsequently overall polymerization, it is mostly conserved across CanA, CanB, and CanB_mut. The changes to CanB to form CanB_mut can be found in the two strands that interact with the donor strand and in the last amino acid of the sequence. In the folded structure of CanA, CanB, and CanB_mut, the final amino acid is found adjacent to the calcium ion that binds in loop 163-172 to form interactions with Gln-169. The change from Ser-146 to Gly-146 in CanB to Can_mut creates a larger binding pocket because glycine is more flexible than serine and has a smaller pendant group that causes less steric issues with space occupancy.

Figure 3.3: Depiction of AlphaFold generated CanB protomer (left) with the calcium ion in green and CanB_mut protomer (right).[18]

Figure 3.4: Depiction of AlphaFold generated CanA protomer with two calcium ions (gray).

3.2 Supramolecular Assemblies

Negative stain TEM was used to analyze the assemblies that formed from CanA, CanB, and CanB_mut. The TEM images of CanA showed dense homogenous nanotubes with repeating ridges on the surface resulting from an external loop in the CanA protomer. CanA exhibited a clear increase in assembly abundance with increased incubation time. Assembly for CanA only occurred in the presence of calcium and heat. Due to the lack of assembly in the absence of heat, it can be concluded that heat is an essential energetic input for polymerization to occur and that this process also requires

enthalpic gain conferred by the presence of calcium ions.

CanB assemblies were found to be essentially absent at concentrations below 80 μM. At higher concentrations ($\sim$400 μM), CanB exhibited moderate assembly formation. CanB and CanB_mut Samples with no $CaCl_2$ and CanB and CanB_mut samples incubated at 20℃ did not form assemblies. As with CanA, the use of heat and calcium ions is necessary for polymerization and most likely for the same reasons. Where CanB and CanB_mut differ from CanA is the concentration at which assembly occurs. CanB and CanB_mut exhibit higher minimum monomer concentration for favorable polymerization. The value of the minimum monomer concentration is correlated to the propensity to form more complex and ordered structures. In this case, that is the formation of nanotubes. The lack of nanotube formation could be due to a need for the protein to form highly concentrated regions or coacervates in the solvent, and this may not be occurring because of less favorable interactions in the periphery of the monomers where lateral association occurs or where the donor strand inserts into the acceptor groove.

Little difference was found in incubating CanB and CanB_mut samples for 2 days to 5 days. The abundance of assemblies was similar and both assemblies and aggregates were present. In some of the protein samples, there was a noticeable amount of incomplete nanotubes. These were often found in clusters with other incomplete nanotubes and could be due to competition between controlled polymerization and aggregation. The incomplete nanotubes could also still be in the process of polymerizing. It is doubtful that the nanotubes were fully formed first and then broken to create these partial nanotubes due to the lack of complimentary shape and characteristics of nanotube fragments to those nearby.

Another effect that resulted in differences in assembly formation of CanB and CanB_mut was pH. More nanotube formation was exhibited at pH 5 than at pH 7. A major suspect in this result is the isoelectric points (pI) of the proteins, which were

calculated to be 4.47, 4.73, and 4.77 for CanA, CanB, and CanB_mut, respectively. It should be noted that CanA has the lowest calculated pI. At pH 5 and pH 7, CanA formed more nanotubes than both CanB and CanB_mut. This is also consistent with CanA having the greatest difference between its pI and both pH values. When molecules are in a solution whose pH is greater than the molecule's pI, the molecule will have a net negative charge. It is suggested that CanA has an overall more negative charge because the pI of CanA is less than that of CanB and CanB_mut. This would impact calcium ion binding and aid in Ca^{2+}-mediated associations during polymerization. pI alone is insufficient for explaining the polymerization trends exhibited across the three samples as it does not hold for CanB and CanB_mut. To understand the effect of pI on the CanB and CanB_mut, the individually substituted residues should be analyzed based on the interactions that would result in the higher-order structure. Theoretically, the functional group of each amino acid would be deprotonated at a pH above the pI. With the substitution of Ser-146 to Gly-146, there is no significant change relative to pI and charges as both residues remain relatively neutral.

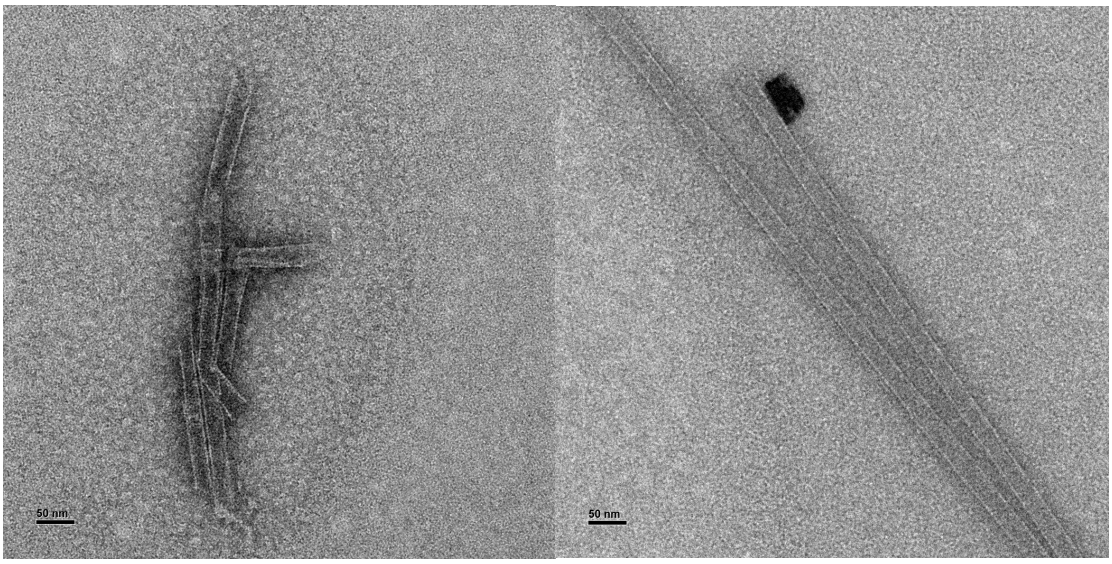

Figure 3.5: TEM images of CanB with 20 mM $CaCl_2$ incubated for 2 days at 80°C. Stained with 1:1 nanoVan/nanoW.

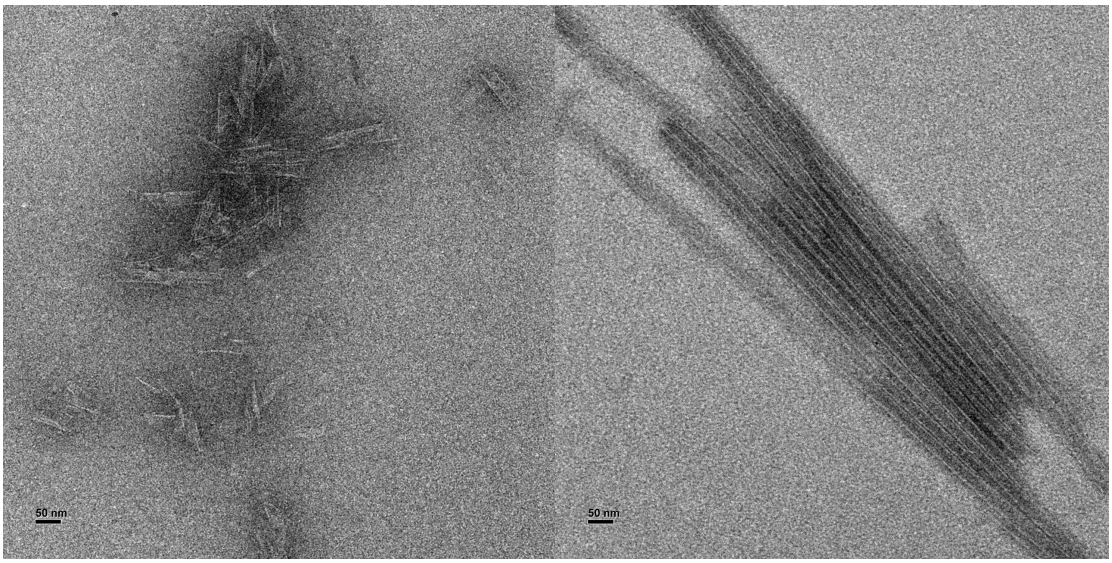

Figure 3.6: TEM images of CanB_mut with 20 mM $CaCl_2$ incubated for 2 days at 80°C. Stained with 1:1 nanoVan/nanoW.

While docking predictions were attempted, no useful results came from the predicted structures. Some pdb files were unusable and the usable files did not predict strand-donation, but rather various types of lateral association. The docking predictions from ZDock and HDock were performed with pdb files of CanB as both the receptor protein and ligand.[19,20] The result of no strand donation appearing in the predicted structures may be due to a lack of defined contacts and computed favorable interactions within the strand-accepting groove.

Chapter 4

Conclusion and Future Directions

It has been demonstrated that CanB_mut has a higher capability of forming assemblies than CanB. In separate samples of CanB and CanB_mut with identical conditions, CanB_mut yielded a higher abundance of nanotubes. While CanB and CanB_mut are both able to form assemblies to varying degrees, the assemblies were not uniform in length or diameter. Preliminary work of seeding CanB and CanB_mut with CanA was inconclusive. This research gives insight into how small variations within primary sequences of SAPs can promote supramolecular assembly formation.

While this research has yielded information about editing primary sequences of proteins to promote supramolecular assembly formation, such as the substitutions of Ser-106 and Gln-107 to Lys-106 and Glu-107 to promote lateral association, the true impact of these edits cannot be fully understood without analyzing solved cryo-EM or X-ray crystallography structures. Having solved structures allows for the analysis of binding pockets associated with the protomers, including both external and internal binding loops, and the specific interactions involved in both lateral associations and strand donation. This can give further insight into the large knowledge gap that still exists: The role that the critical calcium ion binding pockets play in influencing the structure of the polymer is still unknown. To achieve the cryo-EM structures, further

optimization of the assembly conditions must be performed. Additionally, staining techniques need to be refined to improve TEM visualization, specifically for imaging CanB, CanB_mut, and the seeding experiments. Based on observations with previous overstained samples of CanB and CanB_mut, the negative stain should be left on the grid for less than one minute and the time that the sample remains on the grid should be increased to greater than one minute.

By crystallizing the proteins and using an X-ray structure, we can see the difference in the conformation of the protein monomers and the protomers once assembly has occurred. Discrepancies between cryo-EM and X-ray structures could attest to conformational changes that occur upon calcium binding, heating, or strand donation. More information about disordered monomer regions can be understood by studying the X-ray structures of these proteins. To further interpret the contribution of various residues in this assembly process, mutagenic studies that exchange residues involved in strand donation, loops relevant to strand donation, and calcium binding pockets can be performed to assess polymerization and gain more knowledge of the higher-order structures. By comparing the resulting structures, the importance and role of various amino acids can be determined by qualitative observations and matchmaker alignments.

To gain an in-depth understanding of calcium binding in CanA, CanB, and CanB_mut, Isothermal Titration Calorimetry can be performed to obtain data relevant to the thermodynamics of Ca^{2+} binding. ITC can be beneficial in determining the number of calcium ions that bind in each CanA, CanB, and CanB_mut, and the binding affinity of the ligand can also be calculated. A major limitation of this experiment is the determination of protein concentration. Enough protein sample must be used to produce sufficient heat for interpretable data. There is a possibility that high protein concentrations may trigger aggregation or polymerization, but that is unlikely for CanB and CanB_mut due to lower propensities to form supramolecular structures.

Further investigation of the protein structures and ligand binding can be carried out with 2D NMR and the optimized buffer used for assembly.

The seeding projects inform us about the compatibility of the three protein samples with one another. Studying the three different combinations of proteins can be taxing time-wise and expensive. Another approach that can supplement the information that would be gained from the seeding projects is a computational prediction. Supposing that the main interactions that promote assembly are captured by the donor strand and the jelly fold roll, docking predictions can be made with CanA, CanB, and CanB_mut to understand preferential binding. These docking interactions should yield higher confidence results than the protein docking discussed earlier. This docking would involve a peptide (the donor strand) docking to a protein smaller than the original monomers. This is less computationally taxing, and there are many resources, such as HDock, ZDock ClusPro, and HADDOCK, that are readily available.

Appendix A

Supplemental Information

Organism: *P. abyssi*

Sequence No: 1

Sequence name: CanA protein

Length: 193

Type: Protein

Amino acid SEQ

TTQSP LNSFY MTTQS PLNSF YATGT AQAVS EPIDV ESHLG SITPA AGAQG SDDIG YAIVW 60

IKDQV NDVKL KVTLA NAEQL KPYFK YLQIQ ITSGY ETNST ALGNF SETKA VISLD NPSAV 120

IVLDK EDIAV LYPDK TGYTN TSIWV PGEPD KIIVY NETKP VAILN FKAFY EAKEG MLFDS 180

LPVIF NFQVL QVG 193

Figure A.1: Protein sequence of CanA. The underlined residues are the secretion signal and were deleted.

Organism: *P. abyssi*

Sequence No: 3

Sequence name: CanB protein

Length: 146

Type: Protein

Amino acid SEQ

MTTQS PLNSF YATGT AAATS EPIDV ESHLS SIAPA AGAQG SQDIG YFNVT AKDQV NVTKI 60

KVTLA NAEQL KPYFK YLQIV LKSEV ADEIK AVISI DKFSA VIILD SQDFD SNNRA KISAT 120

AYYEA KEGML FDSLP LIFNI QVLSV S 146

Figure A.2: Protein sequence of CanB.

Organism: *P. abyssi*

Sequence No: 4

Sequence name: CanB_mut DNA

Length: 441

Type: DNA

Nucleic acid SEQ

ATG ACG ACG CAA TCC CCG CTG AAT AGC TTT TAT GCA ACC
GGT ACC GCA CAA GCC GTT TCG GAG CCG ATT GAC GTT GAG
AGC CAC CTG AGC AGC ATT GCG CCG GCG GCC GGT GCA CAG
GGT TCC CAA GAT ATC GGC TAC TTC AAT GTC ACC GCG AAA
GAT CAA GTG AAC GTG ACG AAG ATT AAA GTT ACT CTG GCG
AAT GCT GAG CAA CTG AAG CCG TAT TTC AAA TAC TTG CAG
ATC GTG TTG AAG TCT GAA GTG GCG GAC GAA ATC AAA GCT
GTC ATT AGC ATC GAC AAA TTC AGC GCA GTC ATC ATT CTG
GAC AAG GAA GAT TTT GAT AGC AAC AAC CGT GCG AAG ATT
AGC GCC ACC GCG TAC TAT GAA GCG AAA GAG GGC ATG CTG
TTC GAC AGC CTG CCA GTT ATC TTT AAC TTT CAG GTC CTG
AGC GTT GGT TAA

Figure A.3: Protein sequence of CanB _mut.

Organism: *P. abyssi*

Sequence No: 4

Sequence name: CanB_mut DNA

Length: 441

Type: DNA

Nucleic acid SEQ

ATG ACG ACG CAA TCC CCG CTG AAT AGC TTT TAT GCA ACC
GGT ACC GCA CAA GCC GTT TCG GAG CCG ATT GAC GTT GAG
AGC CAC CTG AGC AGC ATT GCG CCG GCG GCC GGT GCA CAG
GGT TCC CAA GAT ATC GGC TAC TTC AAT GTC ACC GCG AAA
GAT CAA GTG AAC GTG ACG AAG ATT AAA GTT ACT CTG GCG
AAT GCT GAG CAA CTG AAG CCG TAT TTC AAA TAC TTG CAG
ATC GTG TTG AAG TCT GAA GTG GCG GAC GAA ATC AAA GCT
GTC ATT AGC ATC GAC AAA TTC AGC GCA GTC ATC ATT CTG
GAC AAG GAA GAT TTT GAT AGC AAC AAC CGT GCG AAG ATT
AGC GCC ACC GCG TAC TAT GAA GCG AAA GAG GGC ATG CTG
TTC GAC AGC CTG CCA GTT ATC TTT AAC TTT CAG GTC CTG
AGC GTT GGT TAA

Figure A.4: DNA sequence of CanB _mut.

```
CLUSTAL O(1.2.4) multiple sequence alignment

CanA          -TTQSPLNSFYATGTAQAVSEPIDVESHLGSITPAAGAQGSDDIGYAIVWIKDQVNDVKL   59
CanB          MTTQSPLNSFYATGTAAATSEPIDVESHLSSIAPAAGAQGSQDIGYFNVTAKDQVNVTKI   60
CanB_Mut      MTTQSPLNSFYATGTAQAVSEPIDVESHLSSIAPAAGAQGSQDIGYFNVTAKDQVNVTKI   60
               *************** *.**********.**:********:****  *  ***** .*:

CanA          KVTLANAEQLKPYFKYLQIQITSGYETNSTALGNFSETKAVISLDNPSAVIVLDKEDIAV   119
CanB          KVTLANAEQLKPYFKYLQIVLKSEVA---------DEIKAVISIDKFSAVIILDSQDFDS   111
CanB_Mut      KVTLANAEQLKPYFKYLQIVLKSEVA---------DEIKAVISIDKFSAVIILDKEDFDS   111
               ******************* :.*            .* *****:*: ****:**.:*:

CanA          LYPDKTGYTNTSIWVPGEPDKIIVYNETKPVAILNFKAFYEAKEGMLFDSLPVIFNFQVL   179
CanB          N---------------------------NRAKISATAYYEAKEGMLFDSLPLIFNIQVL   143
CanB_Mut      N---------------------------NRAKISATAYYEAKEGMLFDSLPVIFNFQVL   143
                                         * :. .*:**************:***:***

CanA          QVG     182
CanB          SVS     146
CanB_Mut      SVG     146
               .*.
```

Figure A.5: Sequence alignment of CanA, CanB, and CanB_mut.

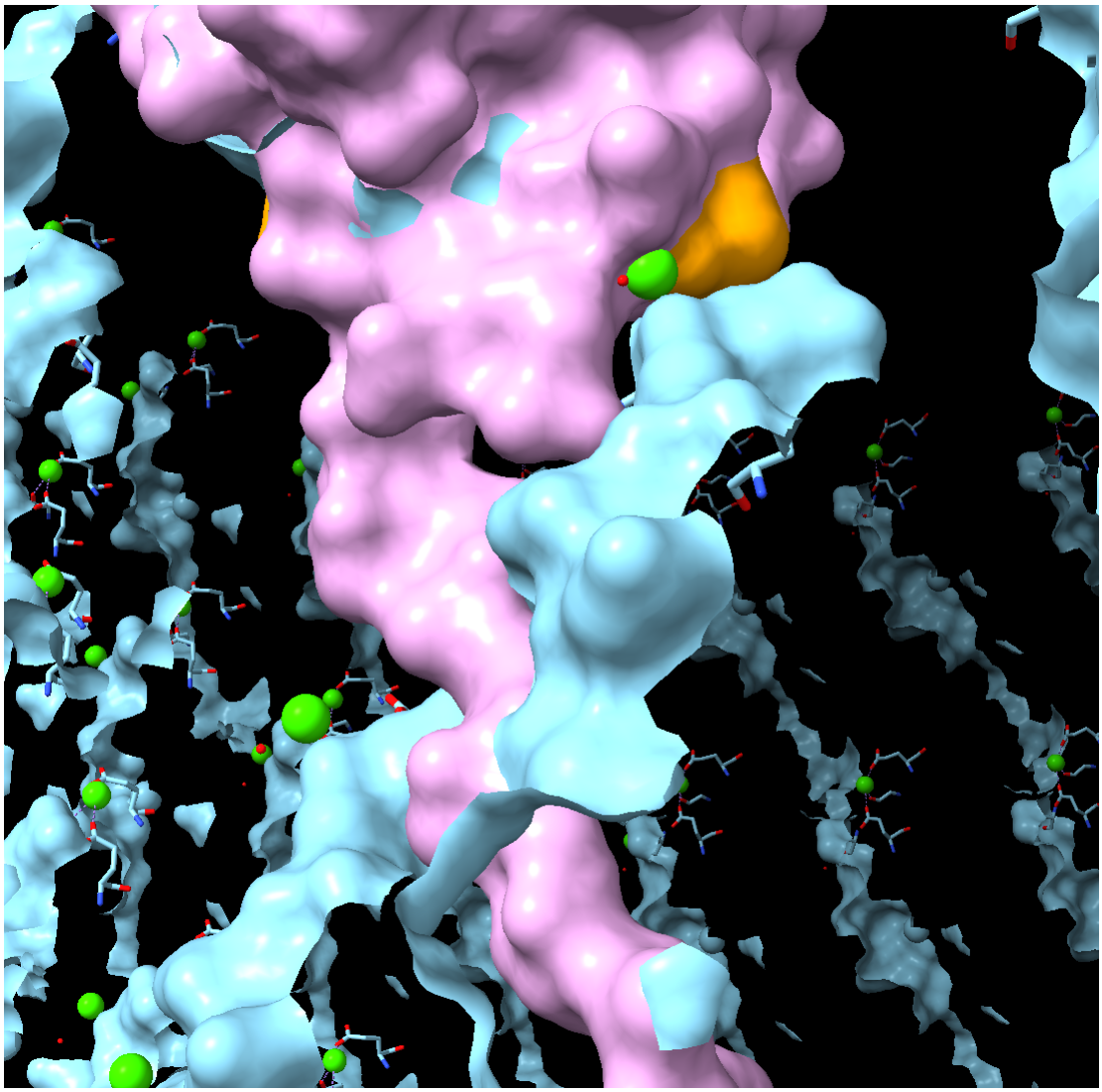

Figure A.6: Space filling model of Ser-106 residue (orange) from CanB next to bound calcium ion (green) in the partially-hidden supramolecular assembly of CanA (blue).

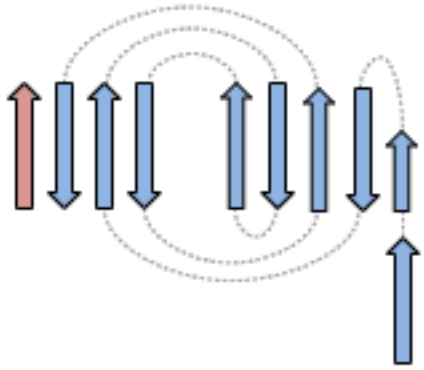

Figure A.7: Jelly fold roll of CanA. The red strand is a donated strand from another protomer and the bottom right blue strand is the donor strand.

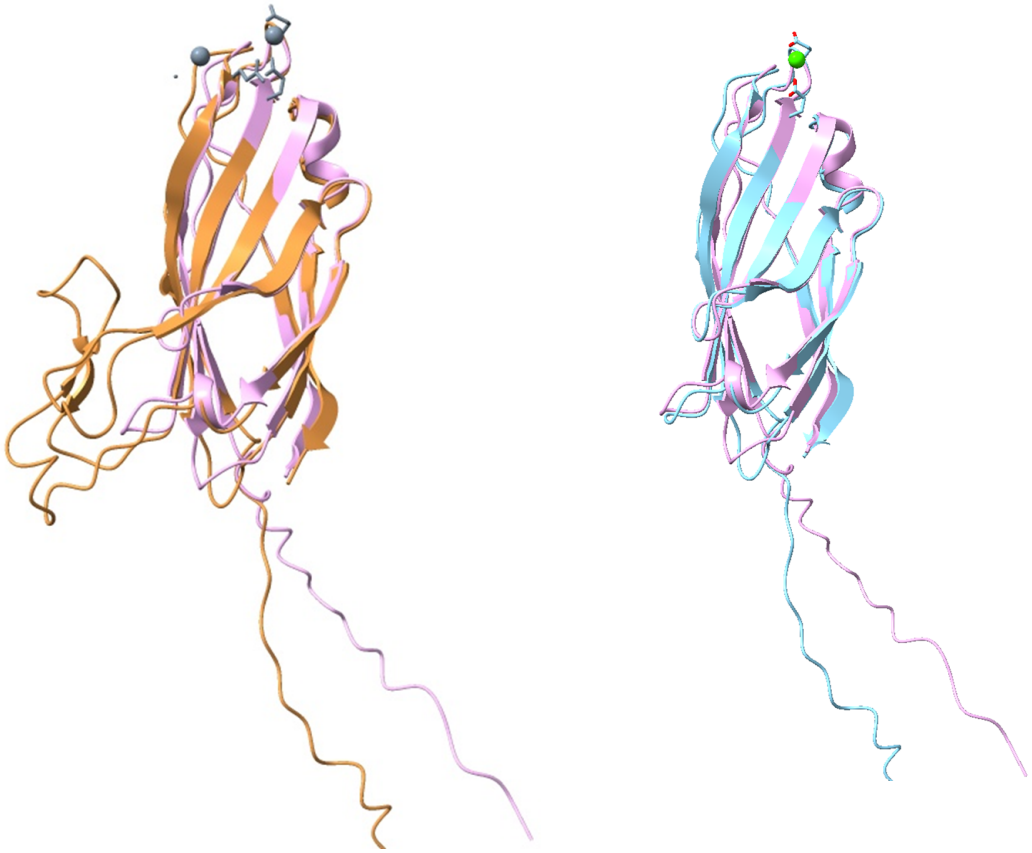

Figure A.8: ChimeraX Matchmaker alignment of CanA and CanB_mut (left) and CanB with CanB_mut (right).

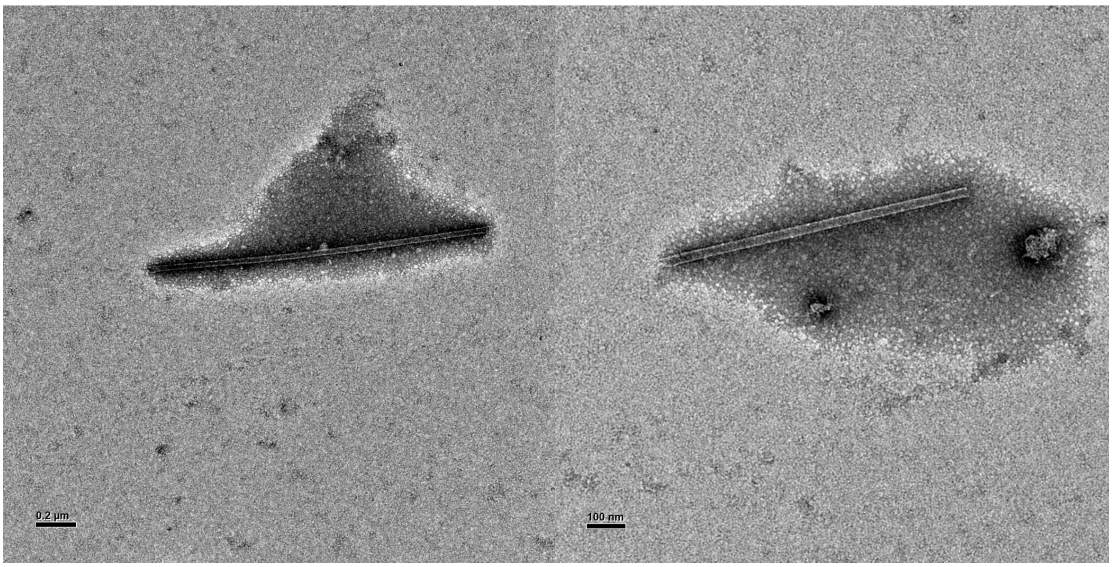

Figure A.9: TEM images of CanB with 20 mM $CaCl_2$ incubated for 2 days at 80°C. Stained with 1:1 nanoVan/nanoW. The addition of salt occurred before annealing. No tubes are present for salt added after annealing (not depicted).

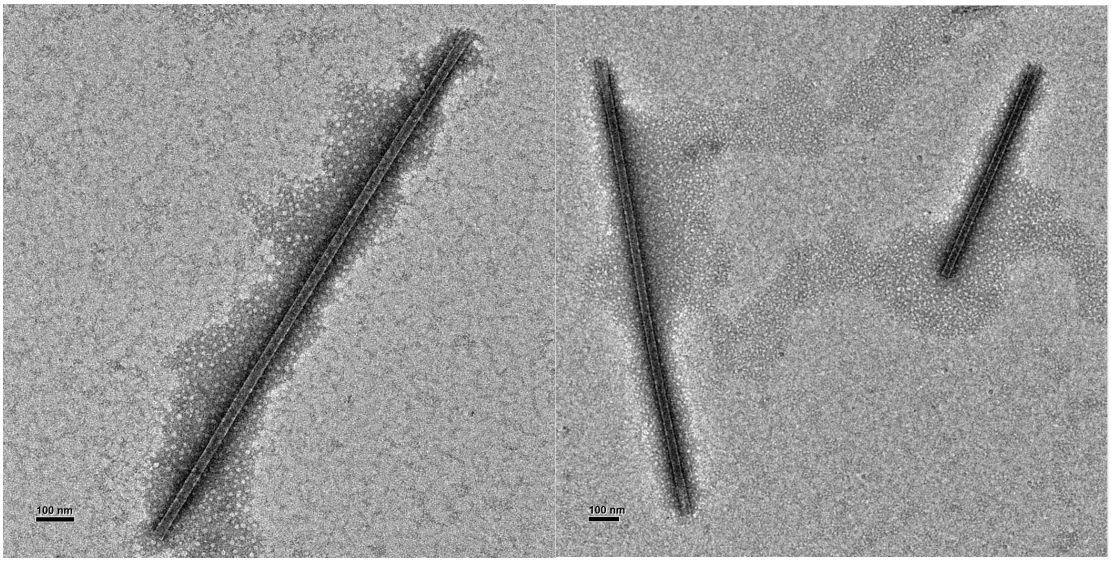

Figure A.10: TEM images of CanB_mut with 20 mM $CaCl_2$ incubated for 2 days at 80°C. Stained with 1:1 nanoVan/nanoW. The addition of salt occurred before annealing. No tubes are present for salt added after annealing (not depicted).

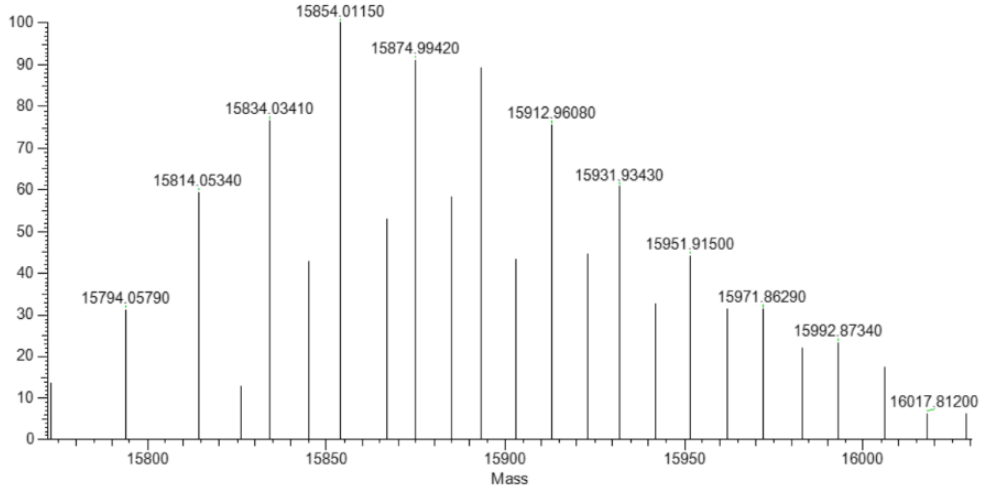

Figure A.11: Deconvoluted Electron Spray Ionization Mass spectrometry (ESI-MS) of CanB_mut. The average calculated mass was 15875, and the monoisotopic mass was 15865. These values were calculated with N-terminal methionine.

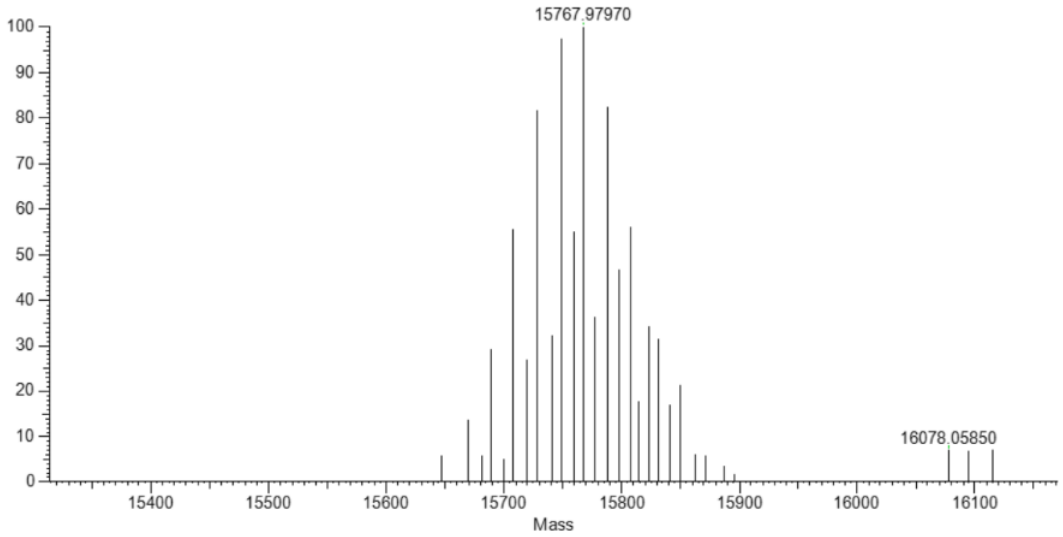

Figure A.12: Deconvoluted Electron Spray Ionization Mass spectrometry (ESI-MS) of CanB. The average calculated mass was 15788, and the monoisotopic mass was 15778. These values were calculated with N-terminal methionine.

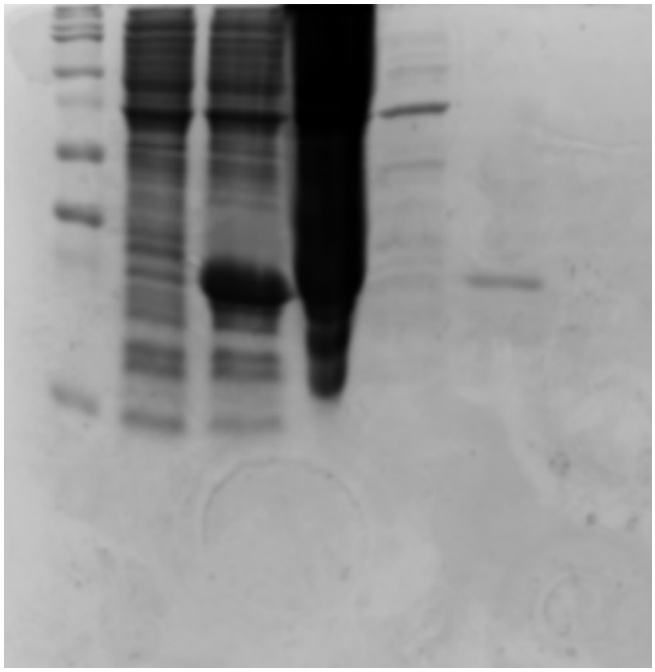

Figure A.13: SDS-PAGE of CanB. The lanes are whole cells, IPTG-induced cells, lysis pellet, heat pellet, and CanB (from left to right).

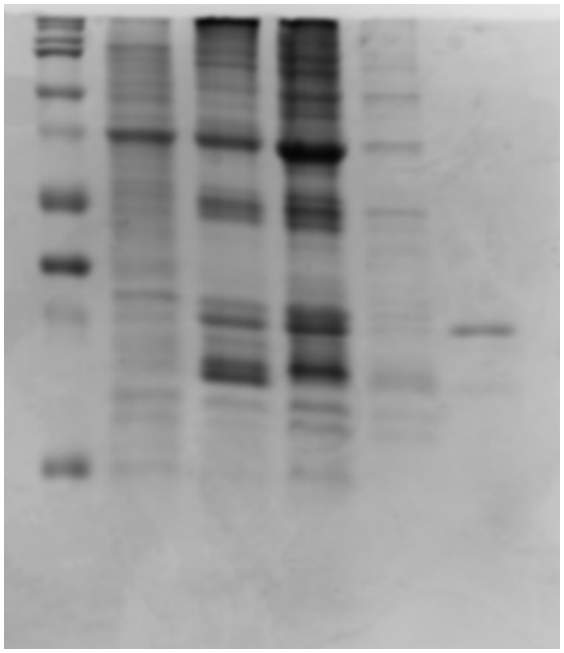

Figure A.14: SDS-PAGE of CanB_mut. The lanes are whole cells, IPTG-induced cells, lysis pellet, heat pellet, and CanB_mut (from left to right).

Figure A.15: First HDock prediction of CanB_mut docking.

Figure A.16: Second HDock prediction of CanB_mut docking.

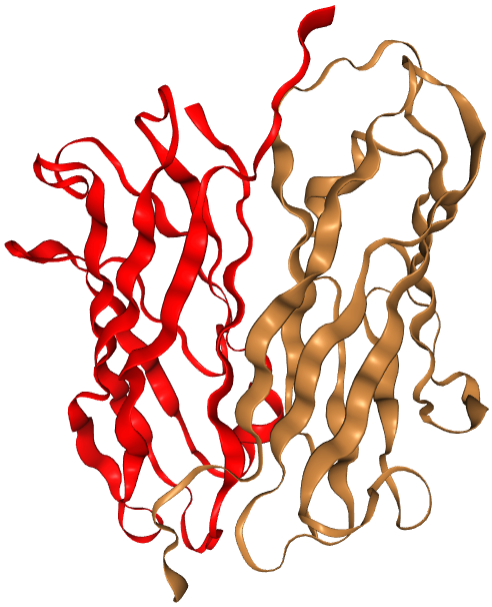

Figure A.17: Third HDock prediction of CanB_mut docking. HDock predictions from 4-10 are not depicted.

References

1. Glyakina AV, Galzitskaya OV. *Bioinformatics Analysis of Actin Molecules: Why Quantity Does Not Translate Into Quality?* Front Genet. 2020 Dec 10;11:617763. doi: 10.3389/fgene.2020.617763. PMID: 33362870; PMCID: PMC7758494.

2. Gazit E.*Self assembly of short aromatic peptides into amyloid fibrils and related nanostructures.* Prion. 2007 Jan-Mar;1(1):32-5. doi: 10.4161/pri.1.1.4095. Epub 2007 Jan 2. PMID: 19164892; PMCID: PMC2633705.

3. Hamley, I. W. *Protein Assemblies: Nature-Inspired and Designed Nanostructures.* Biomacromolecules 2019, 20 (5), 1829-1848

4. Seroski, D. T., & Hudalla, G. A. (2018). *Self-Assembled Peptide and Protein Nanofibers for Biomedical Applications. Biomedical Applications of Functionalized Nanomaterials*, 569-598. https://doi.org/10.1016/B978-0-323-50878-0.00019-7

5. Lazarus B., Luu R., Ruiz-Perez., Almeida Bezerra W., Becerra-Santamaria K., Leung V., Lopez Durazo V., Jasiuk I., Barbosa J., Meyers M.*Equine hoofwall: Structure, properties, and bioinspired designs*, Acta Biomaterialia, Volume 151, 2022, 426-445

6. Gelain, F., Luo, Z., Rioult, M., & Zhang, S. (2021). *Self-assembling peptide scaffolds in the clinic.* Npj Regenerative Medicine, 6(1), 1-8. https://doi.org/10.1038/s41536-020-00116-w

7. Römer, L.; Scheibel, T. *The Elaborate Structure of Spider Silk.* Prion 2008, Vol. 2, Issue 4.

8. Bücker, R., Seuring, C., Cazey, C., Veith, K., Grünewald, K., & Landau, M. (2022). *The Cryo-EM structures of two amphibian antimicrobial cross-β amyloid fibrils.* Nature Communications, 13(1), 1-11. https://doi.org/10.1038/s41467-022-32039-

9. Ou, J. T., & Anderson, T. F. (1970). *Role of Pili in Bacterial Conjugation.* Journal of Bacteriology, 102(3), 648-654. https://doi.org/10.1128/jb.102.3.648-654.1970

10. Proft T, Baker EN. *Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease.* Cell Mol Life Sci. 2009 Feb;66(4):613-35. doi: 10.1007/s00018-008-8477-4. PMID: 18953686.

11. Ou, J. T., & Anderson, T. F. (1970). *Role of Pili in Bacterial Conjugation.* Journal of Bacteriology, 102(3), 648-654. https://doi.org/10.1128/jb.102.3.648-654.1970

12. O'Donoghue, E.; Diego, S.; Frey, G.; Diego, S.; Short, J. M.; Mar, D.; Lafferty, W. M.; Chow, K. (76) Inventors: Nelson R. Barton, San Diego, CA.

13. Zegadło, K., Gieroń, M., Żarnowiec, P., Durlik-Popińska, K., Krecisz, B., Kaca, W., & Czerwonka, G. (2023). *Bacterial Motility and Its Role in Skin and Wound Infections.* International Journal of Molecular Sciences, 24(2). https://doi.org/10.3390/ijms24021707

14. Egelman, E. H. *Cryo-EM of Bacterial Pili and Archaeal Flagellar Filaments.* Current Opinion in Structural Biology 2017, 46, 31–37.

15. Lin, T. J., Sebae, G. E., Jung, H., Jung, H., Park, S., & Holden, J. F. (2016). *Pyrodictium delaneyi sp. Nov., a hyperthermophilic autotrophic archaeon that reduces Fe(III) oxide and nitrate*. International Journal of Systematic and Evolutionary Microbiology, 66(9), 3372. https://doi.org/10.1099/ijsem.0.001201

16. Nickell, S.; Hegerl, R.; Baumeister, W.; Rachel, R. Pyrodictium *Cannulae Enter Periplasmic Space but Do Not Enter the Cytoplasm, as Revealed by Cryo-Electron Tomography*. Journal of Structural Biology 2003, 141 (1), 34–42.

17. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, Ferrin TE. Protein Sci. 2023 Nov;32(11):e4792.

18. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. Protein Sci. 2021 Jan;30(1):70-82.

19. Yan Y, Zhang D, Zhou P, Li B, Huang S-Y. HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. Nucleic Acids Res. 2017;45(W1):W365-W373.

20. Pierce BG, Hourai Y, Weng Z. (2011) Accelerating Protein Docking in ZDOCK Using an Advanced 3D Convolution Library. PLoS One 6(9): e24657.