Distribution Agreement

In presenting this thesis or dissertation 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 or dissertation 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 or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature

_____ _____

Ayda Gonzalez de la Nuez Date

Designing and Engineering Self-Assembling Systems to Control Quaternary Structures

By

Ayda Gonzalez de la Nuez

Master of Science

Chemistry

_____

Vincent Conticello, Ph.D.

Advisor

_____

Khalid Salaita, Ph.D.

Committee Member

_____

Monika Raj, Ph.D.

Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.

Dean of the James T. Laney School of Graduate Studies

_____

Date

Designing and Engineering Self-Assembling Systems to Control Quaternary Structures

By

Ayda Gonzalez de la Nuez

B.A., Bard College, 2016

Advisor: Vincent Conticello, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science
in Chemistry
2021

Abstract
Designing and Engineering Self-Assembling Systems to Control Quaternary Structures
By
Ayda Gonzalez de la Nuez

Supramolecular assemblies that form high-ordered structures from monomers are ubiquitous. Protein and peptide-based nanomaterials are emerging as a promising array of tools in biotechnology due to their biocompatibility, stability, and ability to spontaneously self-assemble into a set of monodispersed structures or experience polymorphism. Self-assembly of these nanomaterials have unique physical and mechanical properties that often dictates the size, morphologies, and functionality. These nanomaterials are particularly attractive due to their critical roles associated with life processes and function. However, the properties that dictate how monomers assemble into specific high-ordered structures are still unknown. The ability to control and rationally design materials that self-assemble is helpful when trying to control the morphology, which has potential uses in biotechnological applications.

Hence, this thesis focuses on understanding the design rules and properties that dictate the self-assembly of two different systems: Encapsulins and peptide-based assemblies. The first chapter focuses on protein-based nanocompartments, encapsulins, which typically support the cell during metabolic processes. We aim to elucidate the principles that dictate the size and symmetry which encapsulins adopt when assembled. Structural analysis suggests that the E-loop, one of the motifs found within the monomer of the encapsulin, is responsible for conformational changes, which give rise to distinct quaternary structures. To probe the E-loop's structural role, we employ protein engineering to modify a native T-1 symmetry encapsulin to adopt the quaternary structure of a T-3 symmetry encapsulin.

Moreover, the second chapter centers on rationally designing *de novo* peptides that assemble into high-ordered structures in order to understand them and precisely control their architecture. Here, we design and characterize the supramolecular assembly of β-sheet peptides, which have been previously used to understand the pathological formation of amyloids. To study the effects of sequence to structure relationship of β-sheet forming proteins, we aimed to survey the landscape of how amino acid identity affects the supramolecular assembly of β-sheet peptides. The results of these projects will lay the initial groundwork to improve our comprehension of the roles that are critical in peptide-based self-assembly. Understanding these properties will offer us a manual for the rational design and control of self-assembling nanomaterials, improving the usefulness of biotechnological applications, including controlled drug delivery and release, nanoreactors, conducting polymers, and potentially shed light into amyloid formation.

Designing and Engineering Self-Assembling Systems to Control Quaternary Structures

By

Ayda Gonzalez de la Nuez

B.A., Bard College, 2016

Advisor: Vincent Conticello, Ph.D.

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science
in Chemistry
2021

Acknowledgements

I want to acknowledge the people that helped me along the way. I would first like to thank Dr. Vincent Conticello, whose expertise was invaluable in formulating the research questions and methodology. Also, I am grateful for accepting me into your lab and giving me the opportunity to explore a different field. Additionally, I would like to thank my former PI, Dr. Stefan Lutz. I would also like to acknowledge my committee members—Dr. Khalid Salaita for being a great teacher and always asking insightful yet challenging questions, and Dr. Monika Raj for offering support as a committee member.

I want to extend my sincere thanks to Dr. Antonio Brathwaite, for offering unique mentorship and support to us, and Dr. Kira Walsh for her sympathy, kindness, and valuable guidance throughout my time at Emory. Thank you for trying your best to make sure our voices were heard.

I am grateful for the Conticello lab: Ordy Gnewou, Duong Nguyen, Jessalyn Rogers, and Shengyuan Wang; thank you for taking me in and helping me learn and adapt to this new lab. Also, fortunate to have met the Lutz lab members: Dr. Matthew Jenkins, Tamra Blue, David White, Evy Kimbrough, Kendra Ireland, Dr. Samantha Iamurri, and Dr. Elsie Williams. I appreciate the mentorship, patience, sympathy, laughs, encouragement, help, during the good times, bad times, and struggles.

Thank you to all the other friends I made at Emory who supported and cared for me during the last few years. Thank you for your company and solidarity.

Eternally grateful to my closest friends who, despite the distance, have been present and for always keeping me grounded.

Lastly and most importantly, I want to thank my family for their unwavering support and love. Everything that I accomplished has been for you and thanks to you. Thank you for being my rock through difficult times and supporting all of the decisions I have made. Also thankful to my nephew, Noah, for bringing joy into my life.

Por último y lo más importante, quiero agradecer a mis padres, Angel y Aida, por su apoyo y amor inquebrantable. Todo lo que logré ha sido para ustedes y gracias a ustedes. Gracias por ser mi roca en tiempos difíciles y por apoyar todas las decisiones que he tomado. También estoy agradecida con mis hermanos Angel y Marlon, y mi sobrino Noah, gracias por traer alegría a mi vida.

It truly takes a village.

Table of Contents

List of Figures

List of Tables

List of Commonly Used Abbreviations

MCP	Prokaryotic microcompartments
BMC	Bacterial microcompartments
TmE	*Thermatoga maritima* encapsulin
MxE	*Myxococcus xanthus* encapsulin
Qte	*Quasibacillus thermotolerans* encapsulin
FLP	Ferritin-like proteins
T-number	Triangulation numbers
A-domain	Axial domain
P-domain	Peripheral domain
E-domain	Elongated loop
SEC	Size exclusion chromatography
TEM	Transmission electron microscopy
PSMα3	Phenol-soluble modulins α3
Cryo-EM	Cryo-electron microscopy
FKFE	Ac-FKFEFKFE-NH_2
MALDI-TOF	Matrix-assisted laser desorption/ionization
CD	Circular dichroism
VLP	Virus-like Particle
FPLC	Fast protein liquid chromatography

Introduction

Biological self-assembly is a widespread phenomenon naturally occurring in nature and in all aspects of life.[1, 2] Self-assembly is defined as a spontaneous process in which a disordered system forms organized structures as a consequence of significant interactions between their building units. In biology, nature can utilize self-assembly to convert simple building blocks into sophisticated architectures that permit for diverse machinery and functional materials at the nanoscale.[3] It plays various important roles and underlines the formation of a wide variety of complex biological structures, including determining the organization of self-assembling compartments and peptides, which dictates a protein's architecture.[4]

Protein-based assemblies are emerging as attractive nanomaterial due to their biocompatibility, plasticity, and ability to spontaneously self-assemble in diverse architectures. The ability to control different types of self-assembling systems can potentially enhance their native function and exploit physical attributes in all fields, including drug delivery, materials chemistry, nanotechnology, and even understanding specific etiology.[5] Therefore, we aim to understand and elucidate the design rubric and principles that dictate the distinctive quaternary structures of two different types of self-assembling systems: encapsulins and peptide-based assemblies.

References:

1. Bowerman, C. J.; Ryan, D. M.; Nissan, D. A.; Nilsson, B. L., The effect of increasing hydrophobicity on the self-assembly of amphipathic beta-sheet peptides. *Mol Biosyst* **2009,** *5* (9), 1058-69.

2. Nilsson, B. L.; Doran, T. M., *Peptide Self-Assembly: Methods and Protocols*. Springer: 2018.

3. Levin, A.; Hakala, T. A.; Schnaider, L.; Bernardes, G. J. L.; Gazit, E.; Knowles, T. P. J., Biomimetic peptide self-assembly for functional materials. *Nature Reviews Chemistry* **2020,** *4* (11), 615-634.

4. Mendes, A. C.; Baran, E. T.; Reis, R. L.; Azevedo, H. S., Self-assembly in nature: using the principles of nature to create complex nanobiomaterials. *Wiley Interdiscip Rev Nanomed Nanobiotechnol* **2013,** *5* (6), 582-612.

5. Colletier, J. P.; Laganowsky, A.; Landau, M.; Zhao, M.; Soriaga, A. B.; Goldschmidt, L.; Flot, D.; Cascio, D.; Sawaya, M. R.; Eisenberg, D., Molecular basis for amyloid-beta polymorphism. *Proc Natl Acad Sci U S A* **2011,** *108* (41), 16938-43.

Chapter 1 Introduction to Encapsulins

In cellular biology, compartmentalization is one of the fundamental features of eukaryote cells. The ability of an organism to compartmentalize endogenous processes is essential in sequestering reactions and molecules to prevent off-target interactions, activation of reactions that are thermodynamically unfavorable, and isolating pathways that contain toxic intermediates.[1] Similar to membrane-based compartmentalization in eukaryotic cells, prokaryotes have evolved to establish physical boundaries within their cellular environment. This is accomplished with the help of proteinaceous compartments, also known as prokaryotic microcompartments (MCPs) or bacterial microcompartments (BMCs), which have been identified in various bacteria.[2] BMCs are made up of various individual shell protein, that self-assemble to form hollow, semipermeable capsids ranging from 40-600 nm in diameter.[3] BMCs putatively help the cells overcome metabolic and physiological challenges and protect the cell from toxic intermediates and undesired side reactions.[3]

Within the last 20 years, a new class of proteinaceous cage structures have been reported which display unique structural and sequence characteristics compared to traditional BMCs.[2] This new class of BMCs protein cages, named encapsulins, have icosahedral geometry and are made up of at least 60 monomers which spontaneously self-assemble to first form dimers, followed by the assembly of homopentamers and/or homohexamers to produce the final nanocompartment.[2] Encapsulins have hollow interiors suitable for encapsulating cargo that helps realize their native function.[2]

Cellular encapsulation of different enzymes in nanocompartments offers a way to regulate and protect the different biochemical reactions that may compromise the cellular metabolism.[4] It

has been previously reported that encapsulation of enzymes has many benefits including increasing the local concentration of enzymes, avoiding undesirable side reactions, and creating the appropriate environments for specialized reactivities.[5] Thus, co-encapsulation of enzymes that execute a series of cascade reactions may have limitless benefits.

The size of the encapsulins correlates with the type and amount of cargo encapsulated.[6] Encapsulins specifically encapsulate cargo proteins via small C-terminal targeting peptides located on the cargo enzyme.[2, 7] Encapsulated cargo proteins such as ferritin-like proteins (FLPs) or dye-decolorizing peroxidases help encapsulins achieve their native function, which is believed to sequester and remediate toxic species generated during metabolic stress.[6, 8] Concurring with the known biological functions, encapsulins contain surface pores that are 3-5 Å in diameter at the 2, 3, and 5-fold symmetry axes.[9] The pores regulate the flux of small molecules and ions from the cytosol to the interior of the encapsulin while also serving to keep larger molecules out of the cavity and shielding the lumen from the exterior environment.

We were specifically interested in the encapsulin from the hyperthermophilic organism, *Thermatoga maritima* depicted in Figure 1.1A. The *T. maritima* encapsulin (TmE) is constructed from a 31 kDa monomer. Interestingly, the TmE monomer shows considerable structural similarity to viral capsid proteins, specifically the capsid protein gp5 of the HK97 virus.[2] Thus, encapsulins are believed to have originated from the same common ancestor as viruses,[10] however, they differ in sequence homology and biological function. Based on the crystal structure of the TmE, the quaternary structure is made up of 60 homomeric monomers that spontaneously self-assemble into a spherical structure with icosahedral symmetry. The monomers form vertices that make up three major pores: the 2-fold pore, 3-fold pore, and 5-fold pore.[11] These pores are lined with negatively

charged residues that are believed to sequester metal cations, specifically Fe (II), which is then mineralized into Fe(III) within the encapsulated FLP.[11]

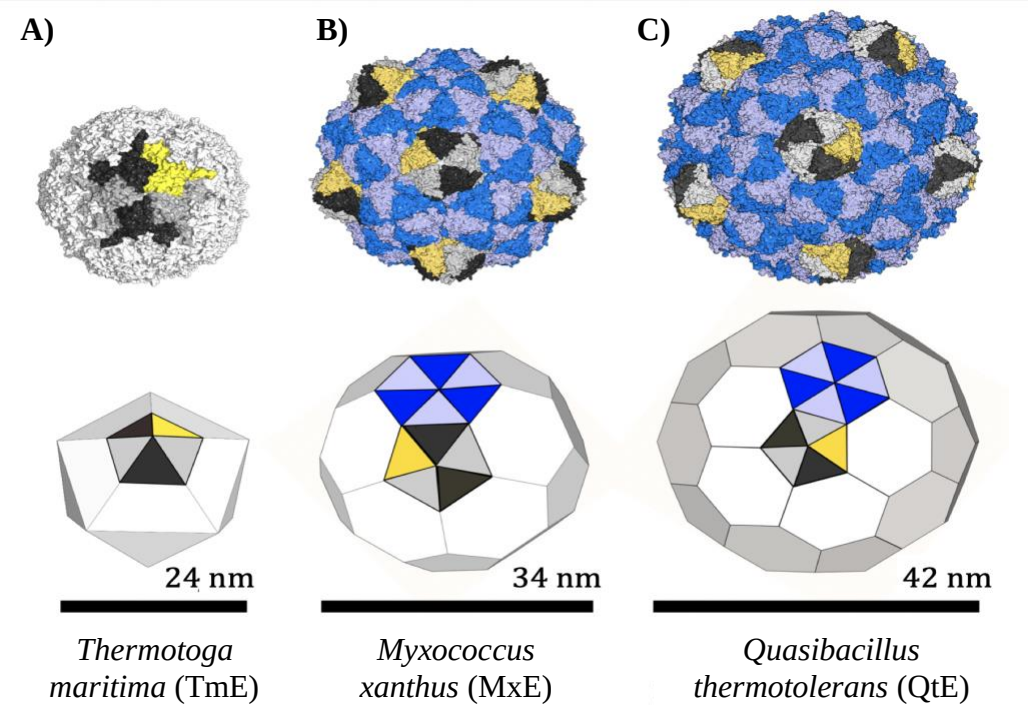

Figure 1.1 (A) schematic representation of T-1 encapsulin with crystal structure of TmE (PDB 3dkt). (B) schematic representation of T-3 encapsulin with crystal structure of MxE (PDB 4pt2). (C) schematic representation of T-4 encapsulin with crystal structure of QtE (PDB 6nj8).

Once assembled, the TmE are extremely stable proteins that can withstand a variety of environmental conditions. For example, they are resistant to degradation by nonspecific proteases, stable over a range of pHs (3-10) and in the presence of several chaotropic salts, and they are thermostable in temperatures up to 80°C.[11] Remarkably, the packaged cargo proteins are stabilized and allow for catalysis to occur in the encapsulin lumen.[12] As such, the robustness, stability, and evolvability of the encapsulins makes them extremely promising platforms for a diverse array of biotechnology applications such as the development of programmable nanoreactors in biocatalytic cascades and targeted drug delivery vehicles.[13]

Despite the abundance of capsid structures in nature, the physical driving forces governing the spontaneous self-assembly of encapsulins into different quaternary structures is currently unknown and represents an unmet challenge in chemical biology.[14] Thus, it is imperative to understand the principles that dictate the size and symmetry that encapsulins adopt when assembled in order to rationally evolve them with novel properties and functionalities in a predictable manner.

Once we accurately understand what directs the self-assembly of encapsulins, we may be able to exploit the different structures we create by encapsulating enzymes of varying sizes. It is essential that foreign cargo can be easily expressed, and the extent of encapsulation can be controlled. The ability to effectively tune the expression of exogeneous cargo would greatly benefit industrial biocatalytic systems by maximizing encapsulated catalyst loading for faster substrate turnover without overcrowding the encapsulins to an extent that reduces cargo enzyme activity. Moreover, in order to make selective nanoreactors, the Lutz lab recently redesigned the 5-fold pore of the TmE encapsulin, enlarging its diameter (~18 Å) for enhanced substrate exchange into and out of the lumen. Another application of the TmE is to exploit this larger pore and use protein engineering to create selectivity filters. This system, coupled with the appropriate cargo enzyme, and controlled assembly, can create extremely useful nanoreactors that show selectivity towards specific substrates.

Our goal was to develop the next generation of nanocontainers that are fully tailored to specific applications in order to maximize the benefits encapsulins offer for nano-bioreactor applications. We aimed to tailor encapsulins to particular needs, such as, understanding and modulating self-assembly of encapsulins and in turn modifying size. We hypothesized that the robustness, spatial organization capabilities and malleability of encapsulins makes them highly

suitable scaffolds for creating tunable nanoreactors with quasi-equivalent structures and pore selectively.

1.1 Investigating the principles which drive the self-assembly of TmE

Symmetry of encapsulins

The diversity of encapsulins is still being investigated, however they have been observed in different sizes and polyhedral symmetries.[6] Similar to viruses, the symmetry of encapsulins is governed by what Casper and Klug termed "triangulation (T) numbers." The T numbering is a system used to describe the ratio of pentameric to hexameric faces that form the icosahedral structures.[11] It is an integer that describes the amount of times an equilateral triangle can be divided in order to maintain the shape of the icosahedron.[15] Variations in the triangulation number changes the ratio of pentameric to hexameric faces on the exterior of the compartments, which, in turn, dictates the size and structural organization of the icosahedron.[15]

Encapsulins of different sizes

The encapsulins from *Myxococcus xanthus* (MxE) form larger structures than TmE and assemble from 180 monomers that form 12 pentameric vertices and 20 hexameric faces, yielding a 34 nm diameter encapsulin adopting T-3 symmetry.[8] Furthermore, the encapsulins from *Quasibacillus thermotolerans* (QtE) are 42nm in diameter with T-4 symmetry assembled from 240 monomers that form 12 pentameric and 30 hexameric capsomers[16] as depicted in Figure 1.1B & C.

We currently do not know what dictates the spontaneous self-assembly of encapsulins into different quaternary structures. The study of oligomer self-assembly is relevant due to its ubiquity

in natural systems and its potential in synthetic biology to develop novel protein assemblies. As protein engineers, it is essential to understand what principles in the monomeric structure dictates the size and symmetry encapsulins assume when assembled, in order to rationally evolve them with novel properties and functionalities. As a proof of concept, we aimed to convert a native T-1 symmetry encapsulin, TmE, into a T-3 nanocontainer. We strived to understand the physical driving forces governing the spontaneous self-assembly of encapsulins, but also be able to dictate monomer arrangement at a nanoscale.

Significance and applications

Although encapsulins with larger symmetries like MxE and QtE exist, not all encapsulins are able to withstand different types of hash environments. Gaining the ability to create TmE that are larger in size, have many benefits. First, it will allow us to build a larger thermophilic encapsulin that is robust and stable, and to package either higher quantities or larger cargoes. Additionally, we may be able to construct an array of TmEs that vary in size to serve as nanoreactors, while also creating encapsulins that are pleomorphic, capable of forming quasi-equivalent structures, similar to some viruses, that can alter morphology or function in response to environmental stimuli.[17,18] More importantly, if we are able to control symmetry, we will be able to control when the TmE changes its morphology, which has potential value for drug delivery applications.

1.2 The encapsulin monomer

In order to begin understanding why encapsulin monomers form different quaternary structures, we examined the structural differences in the monomers TmE, MxE, and QtE. The encapsulin monomers adopt a stereotypical DNA bacteriophage HK97 fold consisting of three distinct structural domains: an axial (A) domain, a peripheral (P) domain, and an elongated (E) loop[11], as depicted in Figure 1.2. The A-domain forms the 5-fold pore, the P-domain forms the 3-fold pore, and the E-loop makes up the 2-fold pore.[11]

At the monomeric level, the TmE and MxE proteins are structurally conserved with 23% sequence identity and 44% sequence similarity. Likewise, TmE and QtE have a 22% sequence identity and 44% sequence similarity. TmE and MxE are roughly the same size, with monomers that have a molecular weight of 31 kDa and 32.6 kDa, respectively.[2, 8] However, subtle differences in the proteins' sequences result in either T-1 or T-3 capsids. When we analyzed the superimposed crystal structures of the TmE, MxE, and QtE monomers, we observed that the A-domains and P-domains align well with each other, however the largest structural deviation observed between the monomers is in the E-loop.

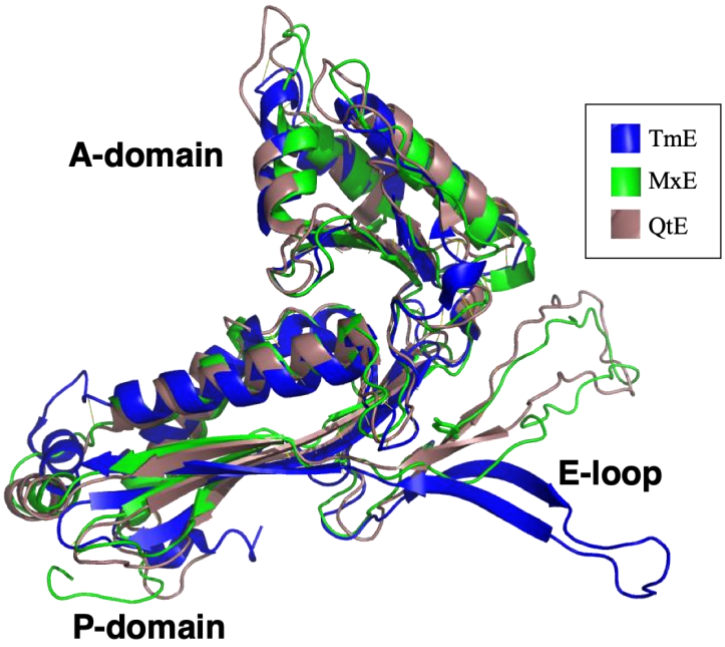

Figure 1.2 Superimposed crystal structures of two monomers form the encapsulins. One monomer from the TmE (blue, pdb: 3DKT) and a monomer from MxE (green, pdb: 4PT2).

The E-loop

The E-loop of TmE is shorter and rotated distally outward relative to the MxE E-loop (Figure 1.2), allowing it to form a strong extended beta sheet interaction with its neighboring monomer.[11] Interestingly, this feature is not present in the MxE or QtE.[2, 11, 16] TmE does not display quasi-equivalence, in other words, all of its monomers adopt the exact same structure in the assembled encapsulin, and we believe that the structural differences in its E-loop relative to those of MxE and QtE are what lock it into a single conformation. Key differences in the amino acid composition, rigidity, and overall length are believed to influence the flexibility of the E-loop. It has been theorized that E-loop flexibility leads to larger symmetry and final assembled size of the encapsulins, because of an abundance of hexamer structures.[11] Based on these observations, we hypothesized that the E-loop of the TmE monomer lacks the flexibility and the conformational

mutability due to the strong inter- and intra-subunit interactions in the encapsulin quaternary structure to form T-3 symmetry of quasi-equivalent structures.

From an engineering point of view, the E-loop presents the most promising region for genetic modification to modulate the TmE quaternary structure. Previous work in the Lutz lab aimed to genetically grafting the E-loop from MxE into the TmE monomer in place of the native E-loop, depicted in Figure 1.3. However, the resulting E-loop chimeras were found to have lost their ability to assemble into functional nanocontainers, Figure 1.3 (unpublished work). Thus, we decided to examine the individual residues within the E-loop and their potential contributions to the overall E-loop structure in more detail.

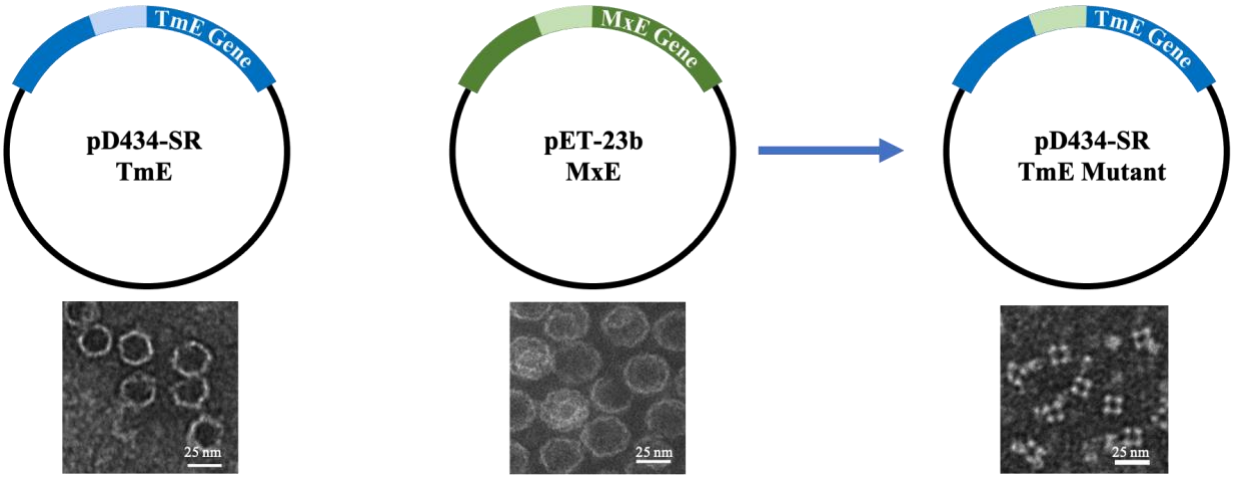

Figure 1.3 A) The pD434-SR plasmid containing the TmE gene (dark blue) and E-loop (light blue); B) The pET-23b plasmid containing the MxE gene (dark green) and E-loop (light green); C) The variant, pD434-SR plasmid containing the TmE gene (dark blue) and the E-loop (light green) from the MxE gene. Below are the corresponding TEM images for the assemble gene and variant. All scales = 25 nm. (*Adapted from Angel Gonzalez-Valero).*

We attempted to destabilize the inter- and intra-subunit sidechain interactions which could be locking the TmE E-loop in a single conformation, thus restricting it to only form T-1 nanocontainers.

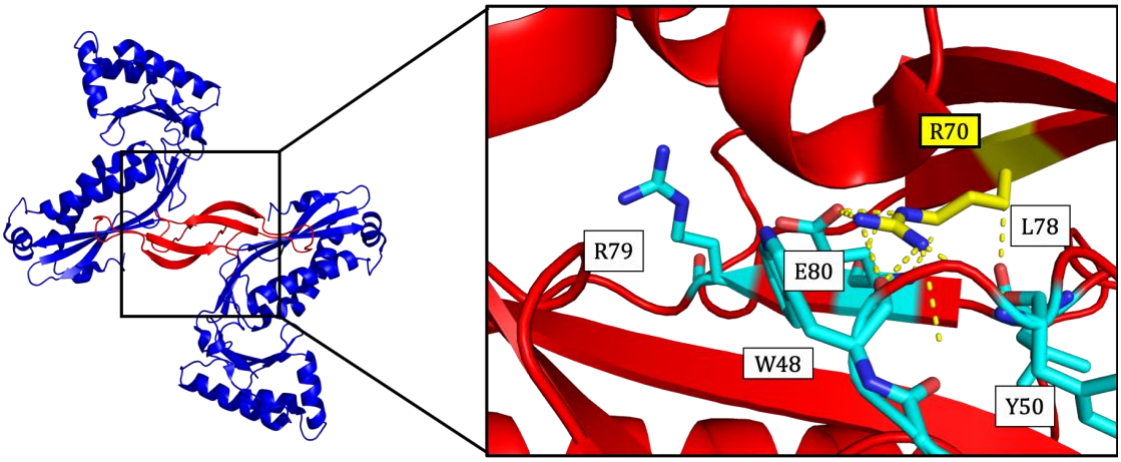

Figure 1.4 A close up of the interface between two TmE monomers that have inter-subunit contacts with neighboring residues. The E-loops are depicted in red.

We selected arginine 70 (R70) of TmE for mutation due to its numerous hydrogen bonding contacts with neighboring monomers at the 2-fold symmetry axis, as depicted in Figure 1.4. We reasoned that by mutating the arginine to a hydrophobic residue such as isoleucine, which has similar length and structure, we will eliminate the hydrogen bonding contacts and thereby increase the flexibility of the E-loop.

1.3 Results and Discussion

The TmE R70I mutant was created via site-directed mutagenesis and was subsequently expressed in *E. coli* BL21(DE3) cells. Initial expression of the R70I variant was confirmed via SDS-PAGE and the samples were purified by HiTrap Q-FF anion exchange chromatography followed by size exclusion chromatography (SEC). The purified variant was then analyzed by transmission electron microscopy (TEM).

Based on previous experiments, wild-type T-1 encapsulins elute at a volume of 80-84 mL during SEC separation, while the wild-type T-3 encapsulins elute at a volume of ~76 mL. The initial SEC trace of the TmE R70I mutant showed an interesting shoulder/peak at an elution volume between 71-76 mL (indicated by red arrow in Figure 1.5), which may indicate formation of bigger or incomplete encapsulins. We visualized several collected fractions of the variant from the SEC by TEM. We also observed a major peak at ~82 mL, which indicates that there are likely wild-type T-1 nanocontainers being formed as well.

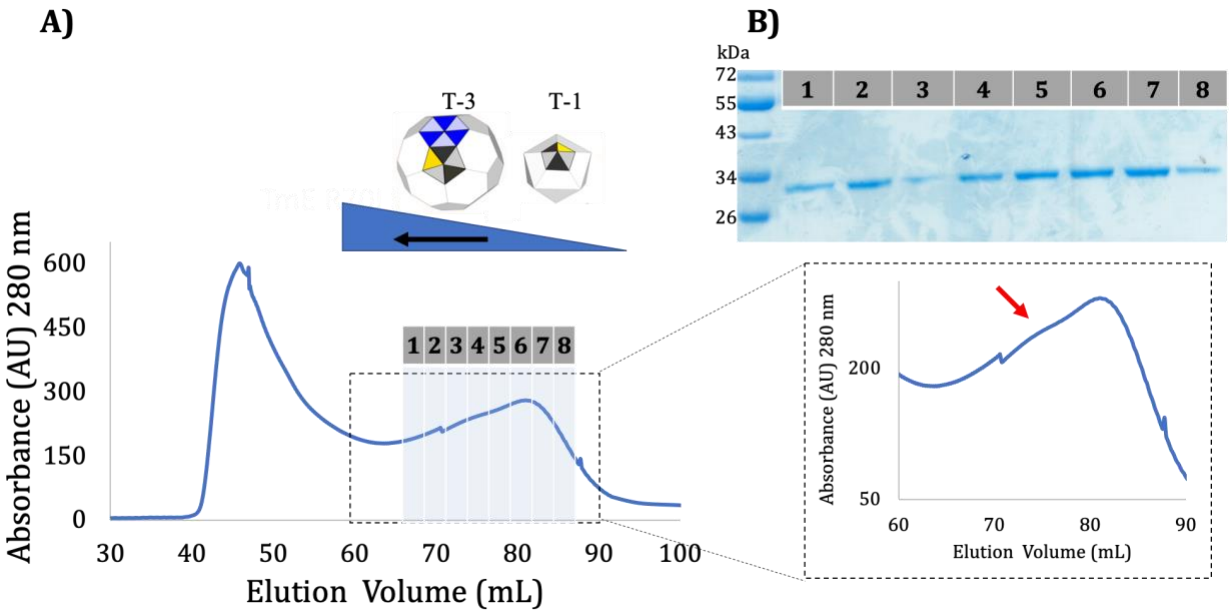

Figure 1.5 A) SEC of TmE R70I variant. B) SDS-PAGE gel of fractions obtained from SEC. The gel confirms the presence of the TmE monomer (31 kDa).

We collected 8 fractions from the SEC purification, as depicted Figure 1.5A. We first confirmed the existence of the TmE via SDS-PAGE analysis. We were interested in further characterizing fractions at 1, 3, 4, and 6 via Transmission Electron Microscopy (TEM). Figure 1.6 depicts the TEM images of the four fractions of interest. We observed that encapsulins were present in all four of the different fractions collected; however, they all contained fully formed, natively sized (24 nm) TmE. We also observed some incomplete assemblies, which made the encapsulins appear larger, but upon closer inspection, they were ~ 24 nm in diameter.

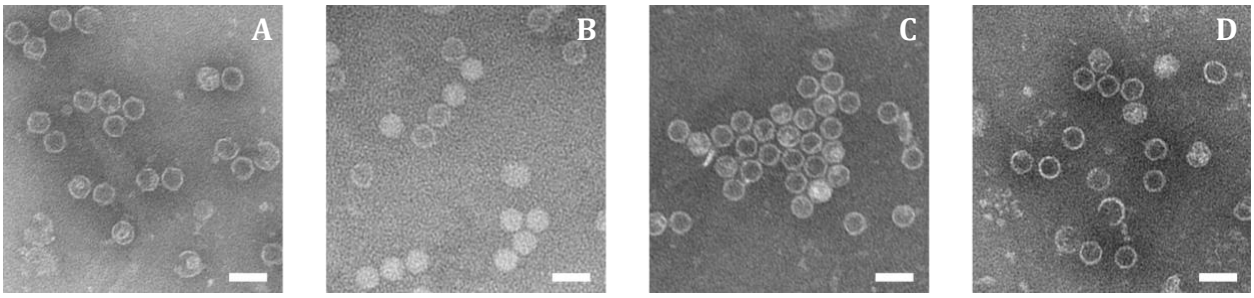

Figure 1.6 TEM images of the fractions collected from SEC at (a) fraction 1, (b) fraction 3, (c) fraction 4, (d) fraction 6. Scale bar represents 20 nm.

TmE G_{41}PYGWEYAAHPLGEVEV--------LSDENE--VKWGLRKSLPLIEL$_{78}$
MxE G_{41}PLGAGVQTVPYDEFQGVSPGAVDIVGEQETAMVFTDARKFKTIPIIY$_{86}$

Figure 1.7 Sequence alignment of the TmE monomer and the MxE monomer.

Moving forward, more mutations need to be made on the E-loop of the TmE monomer, specifically at residues L78, E57, Y48, G80, and E71 because these amino acids make various hydrogen bonding contacts intramolecularly and between the beta-sheets of the neighboring monomers. Additionally, when we align the sequences of TmE and MxE, depicted in Figure 1.7, we observe that the E-loop of the MxE is longer and the TmE is shorter. Thus, in the future more residue insertions should be made, specifically, 8 residues of the MxE E-loop into the E-loop of

the TmE. We believe that the additions of these residues, will make the E-loop longer and more flexible, while interrupting the beta-sheet interactions currently present. Giving the E-loop more flexibility and disrupting the beta-sheets interactions may give the TmE monomer the ability to form quasi-equivalent structures.

1.4 Conclusion

The main goal of this study was to investigate and understand the governing principles which drive the self-assembly of the *T. maritima* encapsulin, to customize different quaternary structures by interrupting inter- and intra-subunit interactions. We successfully created a TmE variant via rational design, that disrupted critical hydrogen bonding interactions caused by an arginine amino acid between the neighboring monomers. However, subsequent purification via FPLC and morphological characterizations using TEM showed that although the TmE assembled, there was no change in size, indicating that perhaps more hydrogen-bonding disruptions were needed. Thus, there are other residues and regions that might be important to continue to study in order to start answering the proposed question of what drives self-assembly. Unfortunately, this project was not completed due to unforeseen circumstances. Thus, this project is in its initial stages and there is still a lot of work to be done to bring this work to completion.

1.5 Materials and Methods

All electrocompetent *E. coli* DH5α cells were purchased from ThermoFisher Scientific (Waltham, MA). The chemically competent *E. coli* BL21(DE3) cells were purchased from New England Biolabs (Ipswich, MA). All DNA primers were ordered from Integrated DNA Technologies (Coralville, IA). All chemicals and reagents were obtained from Sigma Aldrich (St. Louis, MO). Carbon film 200 mesh copper electron microscopy grids were purchased from Electron Microscopy Sciences (Hatfield, PA). The chromatography columns were purchased from GE Healthcare (Marlborough, MA). The pD434-SR vector containing the *T. maritima* encapsulin gene that was been codon-optimized was purchased from ATUM (Newark, CA). Cells were cultured in lysogeny broth (LB) supplemented with either 100 µg/mL ampicillin or 50 µg/mL streptomycin or both.

Protein overexpression: The pD434-TmE variants were overexpressed and purified following the same protocol. The electrocompetent *E. coli* BL21(DE3) cells were transformed with the pD434-TmE vector containing the new TmE variant. These transformed cells were subsequently grown overnight at 37°C on LB agar plates that contained 0.1 mg/mL ampicillin. Single colonies from the agar plates were used to inoculate 5 mL of liquid LB with 0.1 mg/mL ampicillin. This was incubated overnight at 37°C. After about 18 hours of overnight incubation, 1 mL from the starter culture was used to inoculate 1000 mL of new LB media also with 0.1 mg/mL ampicillin. The 1000 mL culture was then incubated at 37°C in shaker until it reached an OD_{600} value between 0.5 – 0.8. When it reached the appropriate OD_{600}, 0.1 mM IPTG was used to induce protein expression. The protein expression was done at 37°C for 18-24 hours. After, the cell pellets were centrifuged at 4000 rpms for 40 minutes at 4°C, collected and stored at -20°C.

Protein purification: All of the TmE and TmE variant proteins were purified using the same procedure adapted from Williams *et al.*[19] The cell pellets that were stored at -20°C were left in ice to thaw. Once thawed, the TmE containing cells were resuspended in 10 mL of 50 mM Tris-HCl at a pH 7.5 and a Protease Inhibitor Cocktail and 5 µg/mL DNase I were added to the mix. Subsequently, the cells were lysed, 50 mL at a time, using sonication for a total of 3 minutes and 10 seconds for 7 cycles of 10 seconds of sonification and 20 seconds of rest. The various falcon tubes of 50 mL of protein were centrifuged and the clarified lysate supernatant was collected for the next step of the purification process. The clarified lysate was first purified with a HiTrap Q-FF anion exchange column. Then, the collected flowthrough was mixed with, 1 g of PEG-800 and 0.2 g of NaCl until dissolved and mixed with biomixer rocking table for 1 hour. The mixture was centrifuged at 4000 rpms for 30 minutes at 4°C, the white pellet was collected and stored at 4°C. The following day, the pellet was prepared for the final purification step, it was resuspended in 7 mL of 50 mM HEPES-KOH with pH 7.5 and 100 mM NaCl. Bovine trypsin was added to achieve a final concentration of 0.05 mg/mL, then the samples were kept for 10 mins at 37˚C, filtered through a 0.2 µm nylon syringe filter. A NGC Chromatography Fast Protein Liquid Chromatography (FPLC) system (Bio-Rad, Hercules, CA) was used, to purify the sample with a HiPrep 16/60 Sephacryl S-500HR size exclusion chromatography column. The resulting fractions with the TmE proteins were collected and store at 4°C until further characterization.

Transmission electron microscopy: For TEM analysis, about 0.1 mg/mL 5 µL of purified encapsulin in 50 mM HEPES-KOH pH 7.5 was added to a Carbon Film 200 Mesh Copper grid (CF200-CU, Electron Microscopy Sciences, Hatfield, PA). The sample drop was left for 2.5 minutes then rinsed by dipping grids in two drops of double distilled water. The excess liquid was removed by blotting with a filter paper, then 5 µL of 1% phosphotungstic acid stain diluted in

double distilled water, adjusted to pH 6.5 using KOH was immediately added and left for 20 secs. The stain was quickly removed by gently blotting with filter paper. The TEM grids were air dried for 5 mins, and then left in vacuum for about 10 minutes. Samples were then stored or examined on a Hitachi HT-7700 Transmission Electron Microscope at 80 keV (Hitachi, Tokyo, Japan).

1.6 References

1. Diekmann, Y.; Pereira-Leal, J. B., Evolution of intracellular compartmentalization. *Biochem J* **2013,** *449* (2), 319-31.

2. Sutter, M.; Boehringer, D.; Gutmann, S.; Gunther, S.; Prangishvili, D.; Loessner, M. J.; Stetter, K. O.; Weber-Ban, E.; Ban, N., Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nat Struct Mol Biol* **2008,** *15* (9), 939-47.

3. Kerfeld, C. A.; Aussignargues, C.; Zarzycki, J.; Cai, F.; Sutter, M., Bacterial microcompartments. *Nat Rev Microbiol* **2018,** *16* (5), 277-290.

4. Giessen, T. W.; Silver, P. A., A Catalytic Nanoreactor Based on in Vivo Encapsulation of Multiple Enzymes in an Engineered Protein Nanocompartment. *Chembiochem* **2016,** *17* (20), 1931-1935.

5. Giessen, T. W.; Silver, P. A., Encapsulation as a Strategy for the Design of Biological Compartmentalization. *J Mol Biol* **2016,** *428* (5 Pt B), 916-27.

6. Giessen, T. W.; Silver, P. A., Widespread distribution of encapsulin nanocompartments reveals functional diversity. *Nat Microbiol* **2017,** *2*, 17029.

7. Cassidy-Amstutz, C.; Oltrogge, L.; Going, C. C.; Lee, A.; Teng, P.; Quintanilla, D.; East-Seletsky, A.; Williams, E. R.; Savage, D. F., Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry* **2016,** *55* (24), 3461-8.

8. McHugh, C. A.; Fontana, J.; Nemecek, D.; Cheng, N.; Aksyuk, A. A.; Heymann, J. B.; Winkler, D. C.; Lam, A. S.; Wall, J. S.; Steven, A. C.; Hoiczyk, E., A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *EMBO J* **2014,** *33* (17), 1896-911.

9. Tamura, A.; Fukutani, Y.; Takami, T.; Fujii, M.; Nakaguchi, Y.; Murakami, Y.; Noguchi, K.; Yohda, M.; Odaka, M., Packaging guest proteins into the encapsulin nanocompartment from Rhodococcus erythropolis N771. *Biotechnol Bioeng* **2015,** *112* (1), 13-20.

10. Akita, F.; Chong, K. T.; Tanaka, H.; Yamashita, E.; Miyazaki, N.; Nakaishi, Y.; Suzuki, M.; Namba, K.; Ono, Y.; Tsukihara, T.; Nakagawa, A., The crystal structure of a virus-like particle from the hyperthermophilic archaeon Pyrococcus furiosus provides insight into the evolution of viruses. *J Mol Biol* **2007,** *368* (5), 1469-83.

11. Nichols, R. J.; Cassidy-Amstutz, C.; Chaijarasphong, T.; Savage, D. F., Encapsulins: molecular biology of the shell. *Crit Rev Biochem Mol Biol* **2017,** *52* (5), 583-594.

12. Lau, Y. H.; Giessen, T. W.; Altenburg, W. J.; Silver, P. A., Prokaryotic nanocompartments form synthetic organelles in a eukaryote. *Nat Commun* **2018,** *9* (1), 1311.

13. Patterson, D. P.; Schwarz, B.; Waters, R. S.; Gedeon, T.; Douglas, T., Encapsulation of an enzyme cascade within the bacteriophage P22 virus-like particle. *ACS Chem Biol* **2014,** *9* (2), 359-65.

14. Panahandeh, S.; Li, S.; Zandi, R., The equilibrium structure of self-assembled protein nano-cages. *Nanoscale* **2018,** *10* (48), 22802-22809.

15. Prasad, B. V.; Schmid, M. F., Principles of virus structural organization. *Adv Exp Med Biol* **2012,** *726*, 17-47.

16. Giessen, T. W.; Orlando, B. J.; Verdegaal, A. A.; Chambers, M. G.; Gardener, J.; Bell, D. C.; Birrane, G.; Liao, M.; Silver, P. A., Large protein organelles form a new iron sequestration system with high storage capacity. *Elife* **2019,** *8*.

17. Atanasova, N. S.; Demina, T. A.; Krishnam Rajan Shanthi, S. N. V.; Oksanen, H. M.; Bamford, D. H., Extremely halophilic pleomorphic archaeal virus HRPV9 extends the diversity of pleolipoviruses with integrases. *Res Microbiol* **2018,** *169* (9), 500-504.

18. Joshi, H. M.; Toleti, R. S., Nutrition induced pleomorphism and budding mode of reproduction in Deinococcus radiodurans. *BMC Res Notes* **2009,** *2*, 123.

19. Williams, E. M.; Jung, S. M.; Coffman, J. L.; Lutz, S., Pore Engineering for Enhanced Mass Transport in Encapsulin Nanocompartments. *ACS Synth Biol* **2018,** *7* (11), 2514-2517.

Chapter 2 Introduction to Nanotubes

Biological peptide self-assemblies are ubiquitous in nature.[1, 2] Dictated by their physiochemical characteristics, peptide assemblies can form an array of structures with greater diversity than usual non-biological materials.[3] The structural arrays of self-assembling peptides give rise to numerous functional possibilities. For example, natural peptide-based assemblies can be found in the form of actin filament, microtubules binding protein, ribosomes, and filamentous bacteriophage, and much more.[4, 5] Due to their functional versatility, peptide assemblies have potential as scaffolds for cell and tissue regeneration, locomotion, drug delivery vehicles with greater targeting ability, stability and controlled drug release.[3] Moreover, self-assembling peptides are also used as drugs themselves as certain peptides have been reported to demonstrate antimicrobial activity and as polymers.[3, 6]

Peptide-based assembly is defined as a collection of peptides that spontaneously self-organize into an ordered structure that are composed of α-helices and β-sheets, or a mixture of both. As such, α-helices and β-sheets are the building blocks of proteins, and this makes them a good starting point to produce synthetic functional material.[7] Due to their structural diversity and simple synthesis,[8] short peptides that form α-helices or β-sheets can be manipulated based on their sequence and environmental conditions for controlled *de novo* design of biological materials. The *de novo* designs are particularly advantageous in the development of synthetic peptides because we can classify appropriate peptides in correlation to their high-ordered structures, which is valuable when synthesizing structurally specific peptides.

The cross β-sheets self-assembly are the most notable ones due to their prevalence in the study of amyloid diseases.[1] Unregulated self-assembly of β-sheet forming proteins can lead to

abnormal misfolding and aggregation of peptides, which is a feature of many amyloid diseases,[1] including Alzheimer's disease, Parkinson's disease, and type II diabetes.[9, 10] Thus, there is significant effort devoted to understanding and exploiting peptide self-assembly to generate materials for biotechnology to better understand and combat these diseases.

On the other hand, cross-α assemblies can be found in various biological systems, including phenol-soluble modulins α3 (PSMα3) peptide, secreted by the *Staphylococcus aureus* bacterium.[11] PSMs peptides incite inflammatory responses and are capable of lysing human cells by the self-assembly of cross-α fibrils.[12] Moreover, the amphiphilic PSMα3 cross-α architectures are linked to various virulence activities, which are reported to be caused by the formation of amyloid fibrils in their cross-α architectures.[12] This cross-α helical structure in amphiphilic peptides is linked to antimicrobial activity, but it is also unique in architecture due to its resemblance to cross-β amyloids.[13]

Due to their relevance in health and uniquely powerful nanotechnology capabilities, such as studies of amyloid disease[14] and tissue regeneration[15], self-assembling peptides have received growing attention, and a significant amount of work has been accomplished to synthetically mimic their native counterparts, using simpler sequence motifs that display similar structures.[16] Generally, supramolecular self-assemblies structures in the nanoscale have been difficult to rationally design and analyze.[17] However, biomolecules such as proteins and peptides, incorporating either natural and unnatural amino acids, present the advantage of sequence specificity when trying to synthetically mimic them. Correlations can be made from biomolecules sequence specificity and their quaternary structures[4], making biomolecules, such as peptides, potentially useful design elements for constructing nanomaterials as they give access to manipulating the sub-unit interactions.

Protein assembly formation is dictated based on interactions between sub-units, where the final structure depends on many physicochemical properties such as length, polarity, hydrophobicity and other non-covalent interactions.[7] The key non-covalent binding interactions that mainly govern self-assembly are weak intermolecular or colloidal forces, such as hydrogen bonds, van der Waals interactions, electrostatic interactions, hydrophobic forces, π-π interactions, and other steric interactions.[18, 19] Therefore, in order to control the structural assembly of synthetic peptides we need to use these interactions to our advantage.

Enhancing the understanding that scientists have of the methods and formation rules employed by biological systems allows researchers the ability to construct *de novo* designed sequences that can be used in a diverse array of applications, such as peptide-based nanostructured assemblies.[20] In other words, if we can understand the complexities of peptide self-assembly, then we can reliably and accurately predict quaternary structures of these assemblies and effectively control them to our advantage. Thus, it is imperative that we get a better understanding of what residues are critical and why, in different self-assembly architectures. Even with the advances in computational structure predications currently available, defining the relationship between sequence and supramolecular structure remains a significant challenge for the *de novo* design of protein structures, as the predictability of self-assembled structure is challenging to achieve due to the atypical assembly arrangements of some peptides.[20, 21]

2.1 Insight into the peptide motifs

As early as the 1950s, researchers have investigated synthetic peptides. During this period, scientists discovered the most fundamental structure in structural biology: the α-helix and the β-sheet.[22] However, they did not have the equipment available to characterize different peptide

sequences and determine structural information at high resolution.[18] Since then, our technological advancements have allowed scientists to discover that proteins are complex structures are comprised of a small number of secondary folds and up to an atomic level resolution of high-ordered structures. As such, α-helices and β-sheets have become one of the most extensively used motifs in the development of new nanomaterials.

The α-Helical motif

α-helices are the most abundant and stable motifs found in protein secondary structure.[22,23] This secondary structure is produced when amino acid polymers form a right-handed or left-handed helices in which the sidechains point outward from the central hydrogen-bonded coil, towards the outer helical surface.[24] α-helices have 3.6 amino acid residues per turn and are stabilized by non-covalent interactions, mainly hydrogen bonds between the nitrogen and carbonyl carbon of structurally adjacent residues, as depicted in Figure 2.1A & B.[25] Sometimes, during self-assembly, two or more α-helices intertwine with one another to produce a coiled coil conformation seen in Figure 2.1C.[19] One of the defining characteristics of a coiled coil is the specific packaging of amino acids, where the hydrophobic sidechains are packed inside the helix and excluded from the aqueous environment.[26]

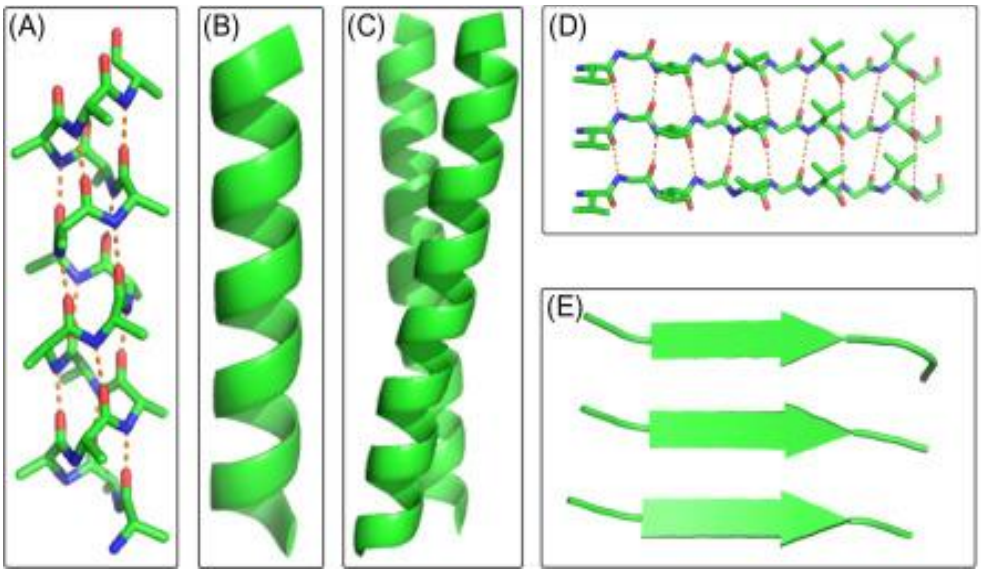

Figure 2.1 Structures of common peptide motifs. (A) α-helix structure with intramolecular hydrogen bonding; (B) cartoon representation of an α-helix; (C) cartoon representation of a coiled-coil; (D) β-sheet structure with intramolecular hydrogen bonding; (E) cartoon representation of β-strands. Source: Image adapted from Boyle *et al*.[24] (PDB 4DZM).

β-Sheets

Another common secondary fold found in proteins are called β-strands. Commonly, the primary structure of β-strands contains alternating hydrophobic and polar amino acids, which lead to an extended backbone structure. The hydrogen bonding between two or more parallel β-strand regions creates β-sheets.[19] Two types of alignments are possible within β-sheets, parallel and antiparallel. The distances between the amino acids that form hydrogen bonds in a parallel β-sheets differ, and this forces the hydrogen bonds to be at an angle relative to the protein backbone within a strand (Figure 2.1 D & E). In turn, the amino acids involved in hydrogen bonds in the antiparallel β-sheets are the same distance, and the hydrogen bonds are at a 90° angle relative to the protein strand. Thus, the antiparallel β-sheets alignments are energetically favorable due to the alignment of hydrogen bonds.[27] The typical cross-β structure found in many amyloid assemblies entails multiple β-sheet peptides arranged in either parallel or antiparallel orientation, stabilized

by hydrogen-bonding along the fibril axis with noncovalent ionic and hydrophobic interactions between amino acid sidechains.[1, 28]

Cryo-EM

The development of the high-resolution cryo-electron microscopy (cryo-EM) methods are invaluable in determining the quaternary structures of peptide and protein assemblies. This method is critical for determining the near-atomic resolution structure of helical peptide filaments.[21] Cryo-EM is a powerful tool when trying to manipulate quaternary structures of peptide assemblies by investigating the correlation between *de novo* peptide designs and quaternary structures. In other words, information obtained from cryo-EM analysis is valuable when controlling the function of peptide assemblies.

Recently, the Conticello lab verified the importance of these methods. They reported using Rosetta modeling and EM density maps obtained from cryo-EM micrographs to reconstruct a peptide nanotube after it was rationally designed. This study provides insight on how rational design of peptides can inform us about the critical non-native structural interactions that exist and that can be modified based on a small amino acid change, ultimately affecting the quaternary structure of a peptide assembly.[21] More specifically, they demonstrate that changing four arginine amino acids in a 29 residue sequence called, Form I, could completely change the structure and helical symmetry of the assembly derived from the resultant peptide, Form II, as depicted in Figure 2.2. The cryo-EM reconstructions indicate that the promoters in Form I assembled into 4 cross-α helical stacks that were single-walled and Form II assembled into 3 cross-α helical stacks that were double walled. Consequently, the mutagenesis of important amino acid residues caused the adaption of two alternative helical structures. These findings indicate that this approach is

beneficial when trying to establish rational designs from *de novo* synthetic peptides.[20, 21] We use the acquired information from these tools to understand and predict the structure of specific types of cross-α assemblies.

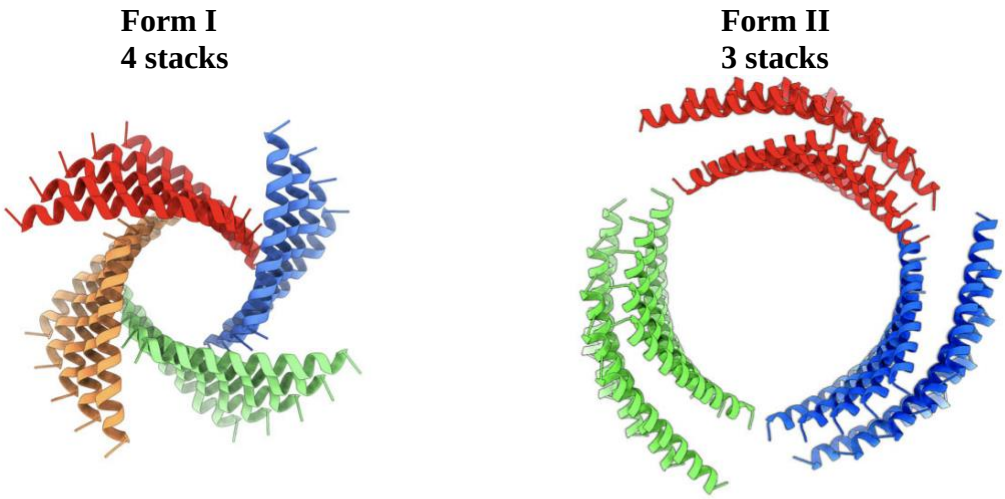

Figure 2.2 Cryo-EM structure of Form I and Form II filaments. Adapted from Conticello *et al.*[21]

2.2 Sequence Design

Furthermore, the Conticello lab expanded these initial studies into exploring the principles of β-sheet self-assembly based on sequence design. While many synthetic β-sheet assemblies have been described, few high-resolution structural analyses are available for designed sequences. In order to accomplish this goal, we focused on the extensively studied β-sheet forming synthetic octapeptide sequence that is amphiphilic, Ac-FKFEFKFE-NH_2 (FKFE).[29, 30] This synthetic peptide was initially reported in 2000[29], and is derived from a sequence discovered in the protein Zuotin, a DNA binding protein, containing alternating hydrophobic sidechains. Since its initial use, the FKFE peptide is reported to have left-handed helical ribbon intermediates during the assembly of this β-sheet peptide.[31] More recently, in 2012, the FKFE peptide was investigated for

a variety of reasons such as, potential immunotherapy uses,[32] and to understand the driving force for the assembly of β-amyloid protein in order to predict conditions that will prevent plaque formation that is essential amyloid disease.[29] Additionally, amphiphilic peptides have demonstrated a higher propensity to self-assemble into amyloid-like β-sheet fibrils when their sequence patterns consist of alternating hydrophobicity,[33] and have been reported to demonstrate structural polymorphism.[20] Thus, the synthetic amphiphilic FKFE peptide is an attractive peptide to investigate. Recently, the Conticello lab was able to determine the atomic structure for the FKFE peptide using cryo-EM analysis.

The cryo-EM reconstruction of the FKFE peptide demonstrated two populations of nanotubes, a thin and a thicker one. In Figure 2.3, we see the cryo-EM micrograph of the different FKFE tubes, the top view or 3D reconstruction of the thinner tube (black arrow), which is made up four protofilaments and of the thicker tube (white arrow), comprised of five protofilaments. This indicates the structural polymorphism found in one sample. They also studied the dodecapeptide of FKFE and found that this peptide assembles into twisted fibrils and ribbons as opposed to a nanotube. This demonstrates that quaternary structure of peptide assemblies is dependent on length of peptide sequence and amino acid composition.[7]

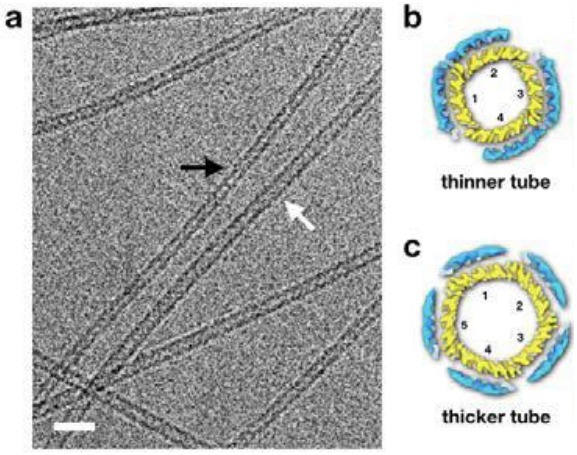

Figure 2.3 A) Cryo-EM micrograph of FKFE tubes. Scale bar = 20nm; B) Top view of the 3D reconstruction of thinner tube (black arrow); C) Top view or 3D reconstruction of the thicker tube (white arrow). Adapted from Conticello *et al.* (*unpublished work*).

The culmination of these results leads us to investigate the effect of substitutions within a sequence and the impact that may have on higher-ordered assembly. This section will focus on

investigating the octapeptide, FKFE, that forms β-sheet assemblies in order to improve the *de novo* design and guided control of peptide assemblies. Therefore, we propose the use of rational design to investigate the self-assembly of newly designed FKFE derivatives to study the effect of the amino acid sequence pattern of amphipathic peptides as β-sheet structures. The following discussion presents the initial work for these β-sheet assembly studies.

2.3 Results and Discussion

We aimed to survey the landscape of how amino acid identity affects the supramolecular assembly of β-sheet peptides, to study the effects of sequence to structure relationship of these β-sheet forming proteins. In this pursuit, we designed six analogs for the FKFE peptide containing unnatural amino acids listed in Table 1 and depicted in Figure 2.4. Overall, we wanted the analogs to have similar sizes and physical properties as the phenylalanine residues within the original FKFE peptide. Thus, we only changed the identity of the aromatic phenylalanine group to understand how it will affect the resulting assemblies. The rationale for choosing 3-benzothienyl-L-alanine, β-2-L-thienylalanine, β-3-L-thienylalanine, is based on their similarity to alkythiophenes, which is among the most useful for creating conducting polymers and in the creation of enzyme-based sensory elements.[34] Moreover, L-homophenylalanine was designed to examine the effects of extending the side chain length in assembly. The 4-bromo-L-phenylalanine and 4-cyano-L-phenylalanine, were selected to evaluate the tolerance for unnatural acids with bulkier groups during self-assembly, with the former containing a typical halogen. We demonstrate

that these short synthetic peptides can self-assemble into different types of high-ordered structures with a diverse array of architectures.

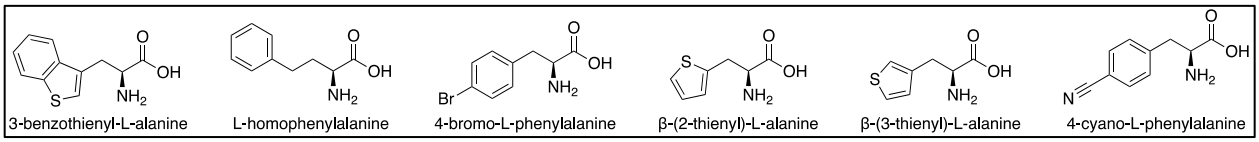

Figure 2.4 Representation of the unnatural amino acids selected for modification of the FKFE peptide.

All of the seven peptides were synthesized via solid-phase peptide synthesis with N-terminal acetylation and C-terminal amidation. The peptides were purified by high pressure liquid chromatography (HPLC). However, we were unable to purify two of the peptides (2K2E, 4K4E), as a result of their extreme hydrophobic character, which could not be solubilized using the protocol developed for FKFE. We purified the four remaining peptides, FKFE, 3K3E, 5K5E, 6K6E, 7K7E via HPLC. Their purity of the sample was confirmed by Matrix-assisted laser desorption/ionization (MALDI-TOF).

Table 1 Nomenclature of amino acid peptide sequences with the modified unnatural amino acid used.

Name	Peptide Sequence	Modification	Characterized by HPLC
FKFE	Ac-FKFEFKFE-NH_2	-	Characterized
2F2K	Ac-2K2E2K2E-NH_2	3-benzothienyl-L-alanine	Not characterized
3K3E	Ac-3K3E3K3E-NH_2	L-homophenylalanine	Characterized
4K4E	Ac-4K4E4K4E-NH_2	4-bromo-L-phenylalanine	Not characterized
5K5E	Ac-5K5E5K5E-NH_2	β-2-L-thienylalanine	Characterized
6K6E	Ac-6K6E6K6E-NH_2	β-3-L-thienylalanine	Characterized
7K7E	Ac-7K7E7K7E-NH_2	4-cyano-L-phenylalanine	Characterized

The self-assembly of a peptide is affected by many factors, including peptide concentration, temperature, pH, and salt concentration. As such, we aimed to observe these new peptides at the same concentration, 3mg/mL, in a variety of conditions. Namely, we screened the four samples in 10 mM buffers: pH 2.0 TFA, pH 4.0 Acetate, pH 7.0 MOPS, and pH 10.0 CAPS

buffers, and they were annealed at 90°C. For the first peptide analog we examined, 3K3E, we observed peptide self-assembly at each of the examined pHs. At pH 4, 7, and 10, we observed peptides assemble into nanotapes. In other words, the self-assembled peptides look like they are flat 2-D nanotubes that look like tapes as depicted in Figure 2.5B-D. The tapes width appears to be distributed from 40 nm to 75 nm. This tape width range was similar when using pH 4, 7, and 10, suggesting the possibility of shared identical structures. Contrastingly, in pH 2 the peptide did not seem to assemble properly all throughout the grid. However, in other places in the grid, we see the formation of other types of materials that we did not expect find. We observed nanosheets that seem to be of square shape and approximately 400 nm in length in Figure 2.5A. It is interesting to find that within one sample we see two different populations of materials that appear to change based on their pH environment. Recently, the Conticello lab reported a shape-shifting peptide material that was pH-dependent and formed collagen tubes and nanosheets,[35] thus these different architectures are possible. Based on these findings, it will be promising to explore this 3K3E peptide as potential polymorphous materials. To learn more about the secondary structures of the synthetic peptides, we used circular dichroism (CD) analysis. In Figure 2.6, the CD spectra for the self-assembled 3K3E peptide at pH 10 shows a maximum at about 195 nm. While at pH 2, 7, and 10 we observe a minimum at about 215 nm, which are in accord with the classical β-sheet formation previously reported.[1]

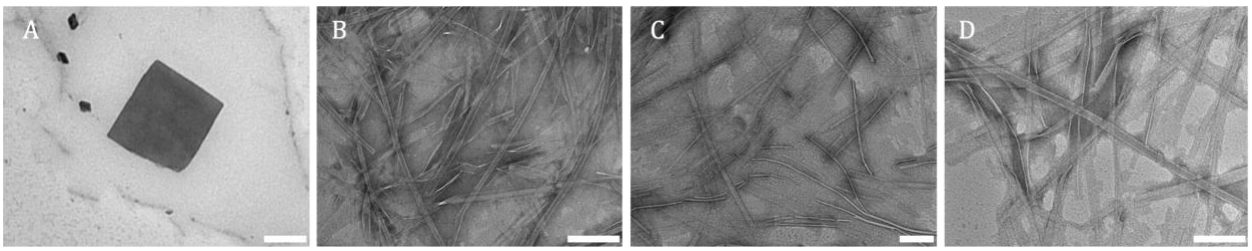

Figure 2.5 Representative negatively stained TEM images of the 3mg/mL 3K3E peptide analogs annealed at 90°C at A, E) pH 2 TFA; B) pH 4 Acetate; C) pH 7 MOPS; D) pH 10 CAPS, annealed at 90°C. All scale bars = 200 nm.

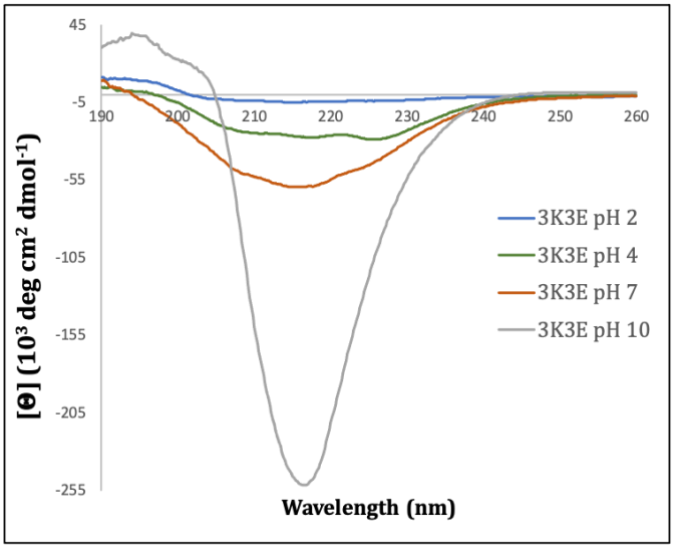

Figure 2.6 Circular dichroism spectra of 3K3E in pH 2.0 TFA (blue), pH 4 acetate (green),pH 7.0 MOPS (orange), pH 10.0 CAPS (grey) annealed at 90°C.

In the second peptide analog we studied, 5K5E, we observed peptide self-assembly formation at each of the examined pHs except for pH 2. At pH 2, the 5K5E peptide did not seem to assemble properly, as depicted in Figure 2.7A, they look more like broken fibrils with a width of approximately 50 nm. At pH 4, 7, and 10, we observed the peptide assemble into long fibrils, with a width range of 25 nm -50 nm. Interestingly, the fibrils observed at pH 4 seem more aggregated in comparison to the other pHs. In Figure 2.7 C and D we see that at pH 7 and pH 10, we still see the assembles peptides, however they are less aggregated.

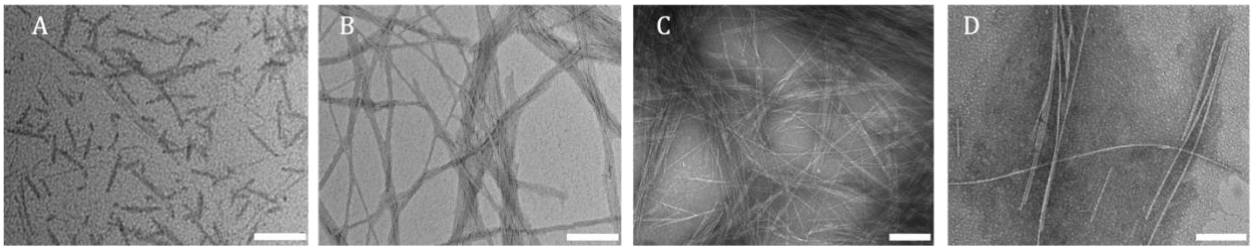

Figure 2.7 Representative negatively stained TEM images of the 3mg/mL 5K5E peptide analogs annealed at 90°C at A) pH 2 TFA; B) pH 4 Acetate; C) pH 7 MOPS; D) pH 10 CAPS. All scale bars = 200 nm.

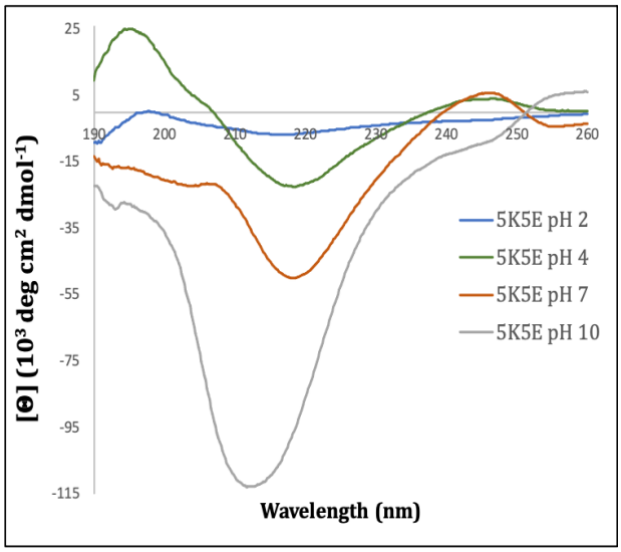

Figure 2.8 Circular dichroism spectra of 5K5E in pH 2.0 TFA (blue), pH 4 acetate (green), pH 7.0 MOPS (orange), pH 10.0 CAPS (grey) annealed at 90°C.

The morphology changes of 5K5E were observed in the CD spectra from Figure 2.8. For the 5K5E, the β-2-L-thienylalanine containing analog, the spectra were slightly similar to that of 3K3E. At pH 2, we observe at small peak at 198 nm and a minimum at 218 nm, both with very low intensity. At pH 4, there is a maximum peak at 195 nm and a minimum at 218 nm, which correspond to that of a classical antiparallel β-sheet. In contrast, we observe an intense minimum peak at about 212 nm and a slight peak at 195 nm at pH 10. There exists another minimum peak at about 218 nm at pH 7, with a maximum peak at 247 nm and a slight peak at 206 nm. At all pHs we observe the classical antiparallel β-pleated sheets[36] at the negative peak between 212-218 nm and a slight maximum present near 195 nm. Additionally, the maximum near 245 nm may be attributed to aromatic group in the protein side chain, but with a weaker absorbance.

We find the most different types of assemblies from the synthesized variants in the 6K6E peptide. At pH 2, 4, 7, and 10, we observed peptide assemble into twisted fibrils (Fig. 4). Specifically, at pH 2, we see these twisted fibrils that are about 50 nm in width and sometimes seem to twist around each other, as demonstrated in Figure 2.9A. Interestingly, at pH 4, the majority of the 6K6E samples are not twisted fibrils, they look like long nanotubes, however, we do observe very few twisted fibrils in some areas of the TEM image in Figure 2.9B.

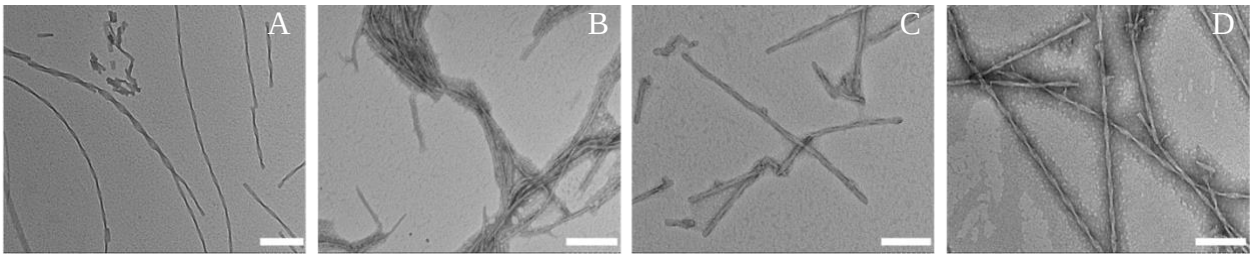

Figure 2.9 Representative negatively stained TEM images of the 3mg/mL 6K6E peptide analogs annealed at 90°C at A) pH 2 TFA; B) pH 4 Acetate; C) pH 7 MOPS; D) pH 10 CAPS. All scale bars = 200 nm.

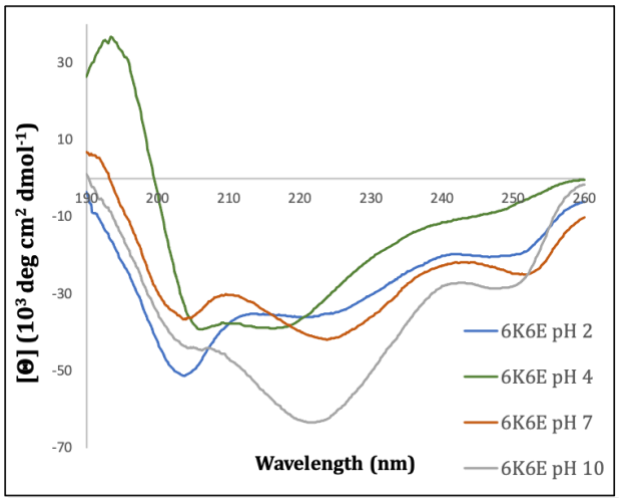

Figure 2.10 Circular dichroism spectra of 6K6E in pH 2.0 TFA (blue), pH 4 acetate (green), pH 7.0 MOPS (orange), pH 10.0 CAPS (grey) annealed at 90°C.

Interestingly, for the β-3-L-thienylalanine containing analog, we do not observe a classical β-sheet signature in CD spectra showed in Figure 2.10. At pH 2 we see five peaks: three local minimums at 203 nm, 222 nm, 252 nm and two local maximums at 209 nm and 240 nm. At pH 4, we see the classical signature for an α-helix, with a maximum peak at 193 nm and two minimum peaks at 205 and 218 nm. At pH 7 there are also three local minimums at 204 nm, 224 nm, 252 nm and two local maximums at 212 nm and 242 nm. Moreover, at pH 10 we observe three minimums at similar locations, 204 nm, 221 nm, 250 nm and two local maximums at 208 nm and 241 nm. The peaks first minimum peaks cannot be attributed to any classical secondary structure. Additionally, the peaks found at the second minimum could be attributed to a β-sheet signature. In a normal CD spectrum, we don't see any evidence for secondary structure beyond about 225 nm. However, here we see absorptions at much longer wavelengths, and these may be credited to the contribution from the aromatic rings because the aromatic groups are highly organized structures.

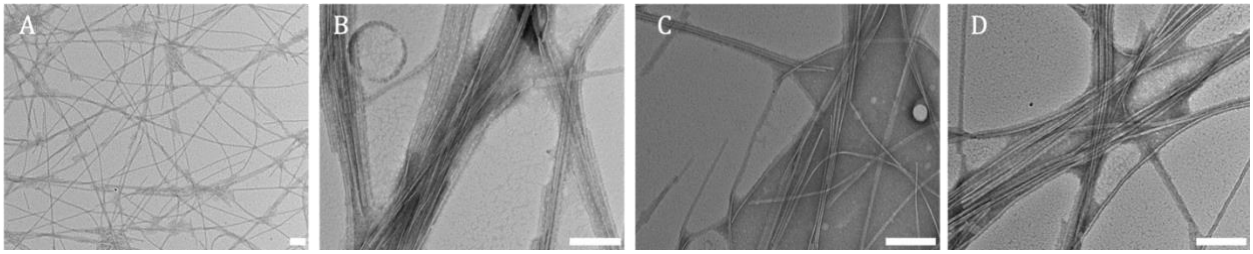

Figure 2.11 Representative negatively stained TEM images of the 3mg/mL 7K7E peptide analogs annealed at 90°C at A) pH 2 TFA; B) pH 4 Acetate; C) pH 7 MOPS; D) pH 10 CAPS. All scale bars = 200 nm.

The fourth peptide analog we examined, 7K7E, we identify peptide self-assembly formation at each of the examined pHs. At pH 2, 4, 7, and 10, we observed peptides assemble into fibrils that range in width from 50 nm to 75nm. In Figure 2.11A, there are very long fibrils formed, and throughout the grid they are aggregated and at times intertwined. In Figure 2.11B & C, we find tapes in most of the grids. Lastly, in Figure 2.11 D, pH 10, we see the formation of mostly fibrils again.

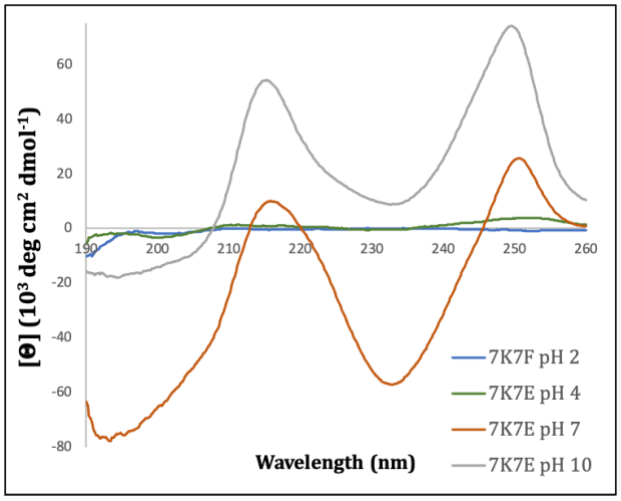

Figure 2.12 Circular dichroism spectra of 7K7E in pH 2.0 TFA (blue), pH 4 acetate (green), pH 7.0 MOPS (orange), pH 10.0 CAPS (grey) annealed at 90°C.

Furthermore, for the 4-cyano-L-phenylalanine containing analog, the CD spectra in Figure 2.12, is similar to the peptides previously discussed because it does not show a classical signature of typical secondary structures. At pH 2 and pH 4 the absorption is essentially flat, which is interesting because the according to the TEM image, there is fibril formation. At pH 7 there are also three local minimums at 204 nm, 224 nm, 252 nm and two local maximums at 212 nm and 242 nm. Moreover, at pH 10 we observe three minimums at similar

locations, 204 nm, 221 nm, 250 nm and two local maximums at 208 nm and 241 nm. The peaks first minimum peaks cannot be attributed to any classical secondary structure. Additionally, the peaks found at the second minimum could be attributed to a β-sheet signature. Any absorption above 225 nm is due to aromaticity of the side chains, since they are highly ordered. In this sample, we once again observe a level of polymorphism where we see the formation of both fibrils and tapes at differing pHs.

2.4 Conclusion

Our primary goal in this project was to understand the structural principles that cause the design of synthetic β-sheet assemblies. We designed six derivates from the FKFE family with the use of unnatural amino acids and characterize four. The FKFE derivatives were confirmed to self-assemble into nanostructures under several pHs: 2.0, 4.0, 7.0, 10.0. Using TEM images, we observe the different morphologies that these assemblies take depending on their environmental conditions. Moreover, two FKFE analogs demonstrated potential polymorphism under differing pH conditions. We observed nanotubes, twisted fibrils, nanosheets, and tapes in the synthetic peptide assemblies. It is promising to see such small changes to the peptide sequence can cause such large differences in the assembled structures. The CD spectra confirmed some of the secondary structures of the FKFE analogs, however, we also observed unusual absorbance peaks at higher wavelengths due to the high-order of the aromatic groups. This was only the initial steps to learning more about the structural-sequence relationship of synthetic peptides. Moving forward, it will be interesting to explore more synthetic analogs to understand the critical residues that dictate specific high-ordered structures. Additionally, the FKFE peptide analogs need to be furthered characterized with atomic force microscopy, small angle x-ray scattering, and cryo-EM to learn more about the atomic structure at a higher resolution.

2.5 Materials and Methods

Peptide Synthesis: All Fmoc amino acids reagents were purchased from Chem-Impex Inc (Wood Dale, IL). All peptides were either purchased from GenScript USA (Piscataway, NJ) or synthesized in the laboratory. In the latter case, peptides were synthesized using microwave-assisted synthesis on a CEM Liberty Blue solid-phase peptide synthesizer instrument. The peptides were capped (N-acetyl, C-amide), and PAL-PEG-PS resin from Applied Biosystems was used for synthesis. Standard Fmoc protection chemistry was used with coupling, based on standard procedures from HBTU/DIEA-mediated activation protocols and base-induced deprotection (20% piperidine in *N*,*N*-dimethylformamide with 0.1 M hydroxybenzotriazole). The peptides were purified thru reversed-phase high-pressure liquid chromatography (HPLC) on a C18 column with a gradient of water-acetonitrile (0.1% trifluoroacetic acid) gradient. The purified peptides were lyophilized and then stored at –20 C. The purity was evaluated by analytical MALDI-TOF mass spectrometry by means of an Applied Biosystems AB4700 Proteomics analyzer in reflectron positive ion mode using α-Cyano-4-hydroxycinnamic acid as the matrix. The results are below.

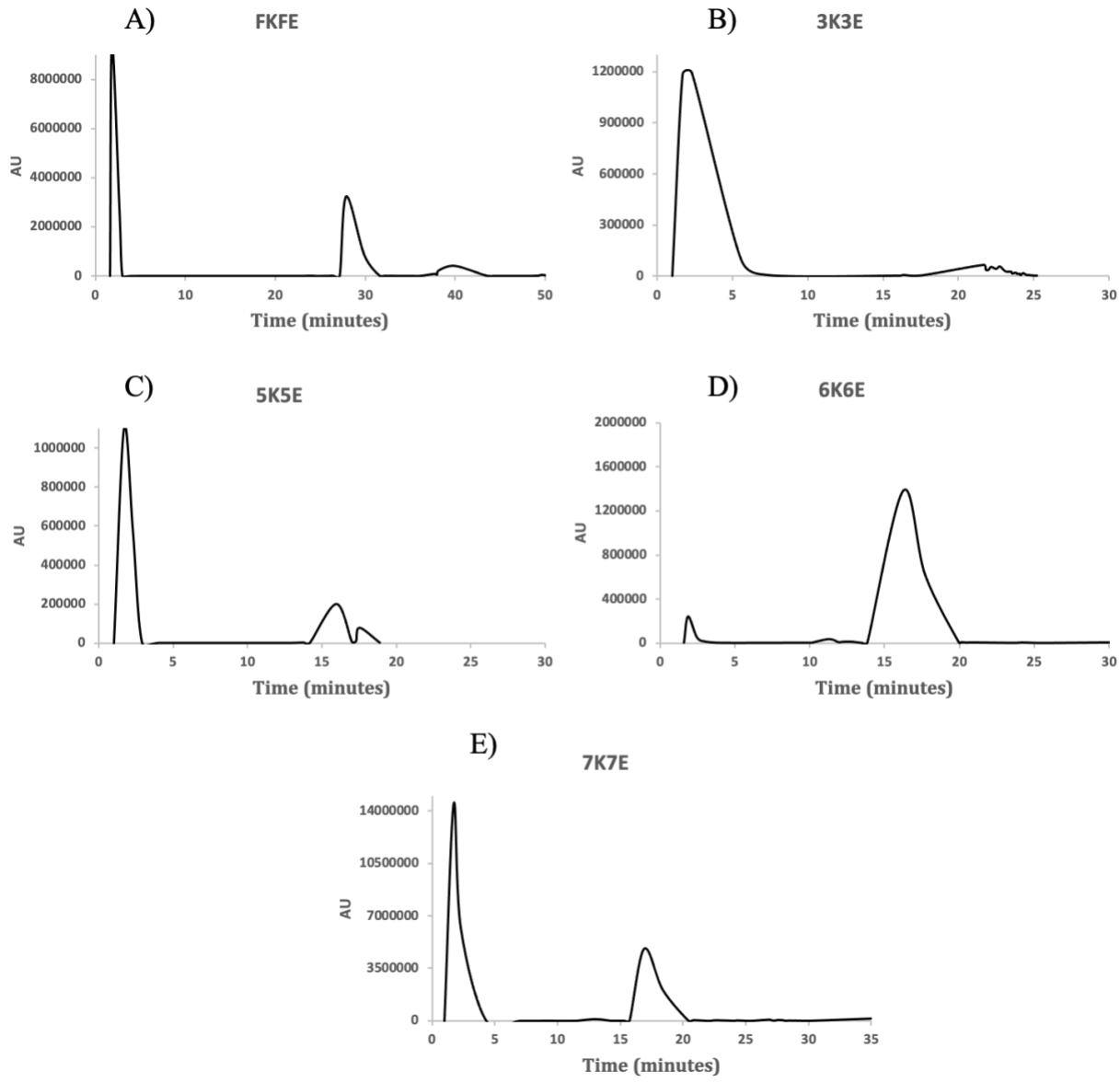

Figure 2.13 Preparative HPLC trace of A) FKFE; B)3K3E; C)5K5E; D)6K6E; E) 7K7E.

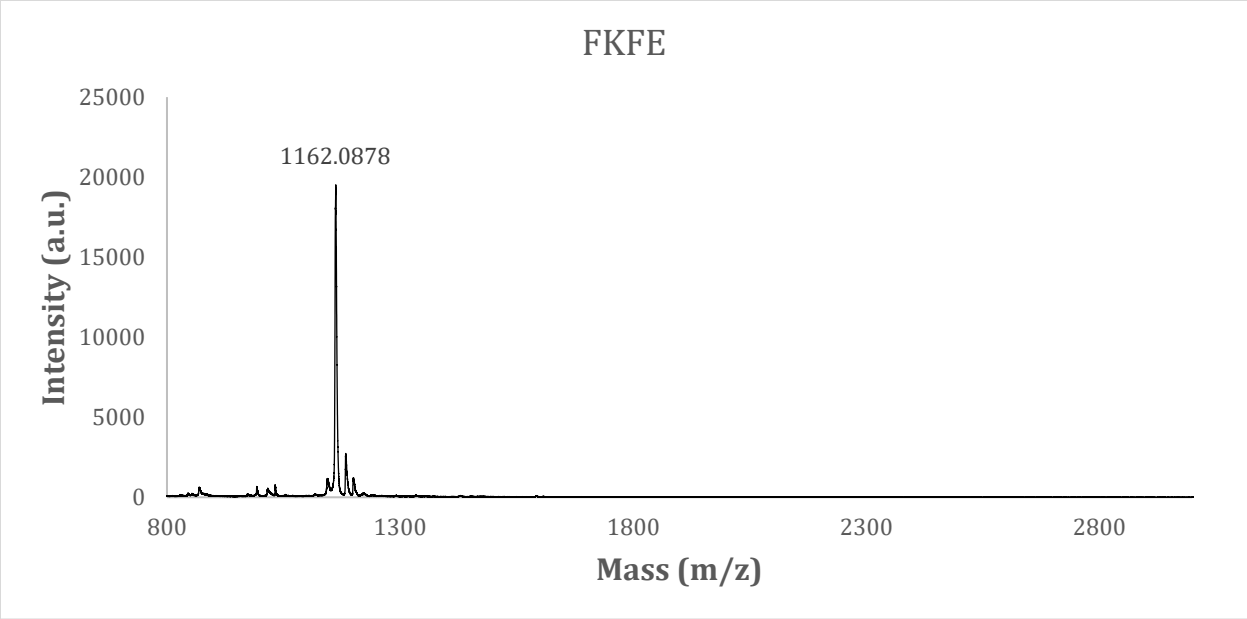

Figure 2.14 MALDI-TOF MS spectrum of the synthesized FKFE peptide. *Expected mass = 1162 g/mol.*

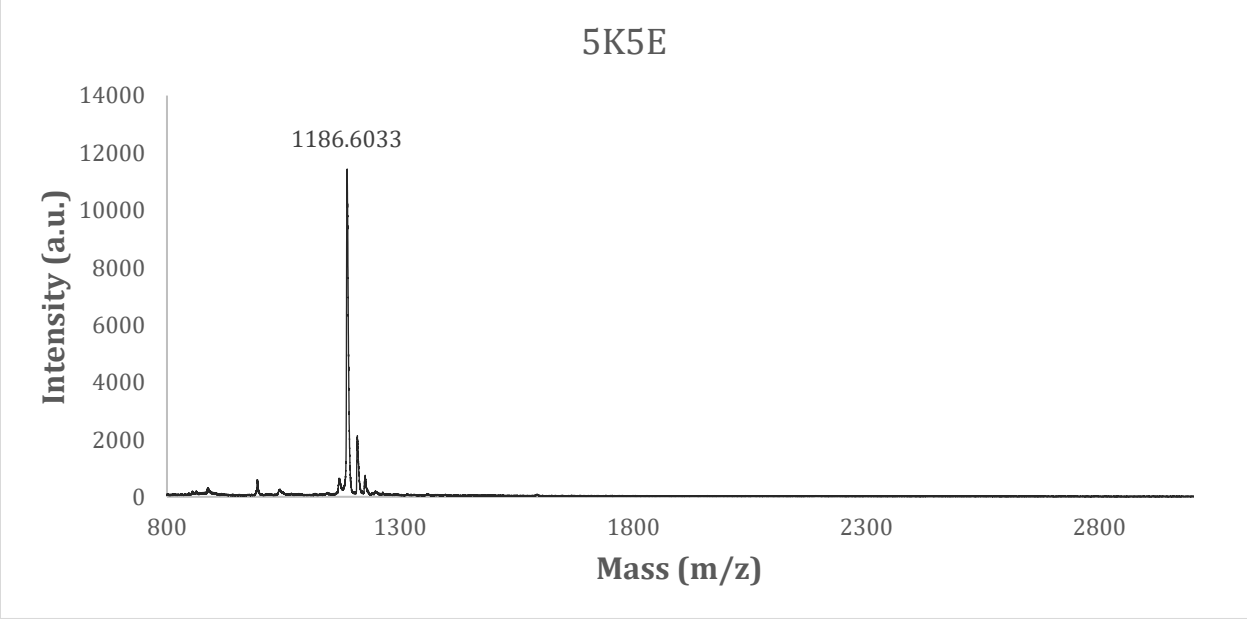

Figure 2.15 MALDI-TOF MS spectrum of the synthesized 5K5E peptide. *Expected mass = 1186 g/mol.*

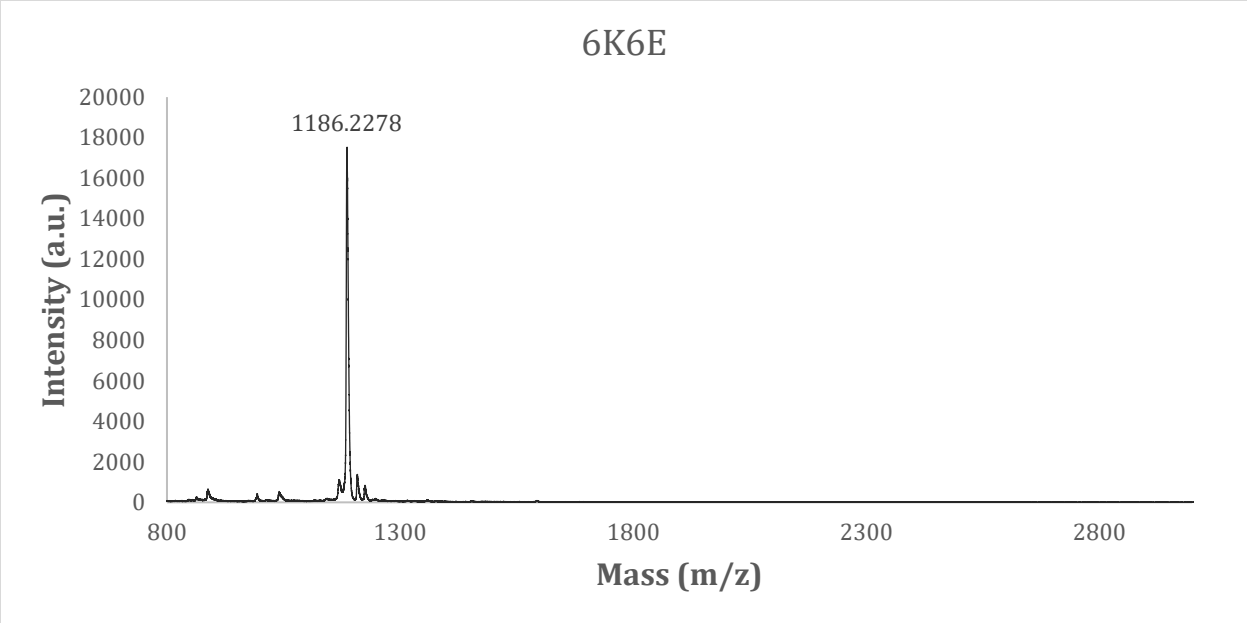

Figure 2.16 MALDI-TOF MS spectrum of the synthesized 6K6E peptide. *Expected mass = 1186 g/mol.*

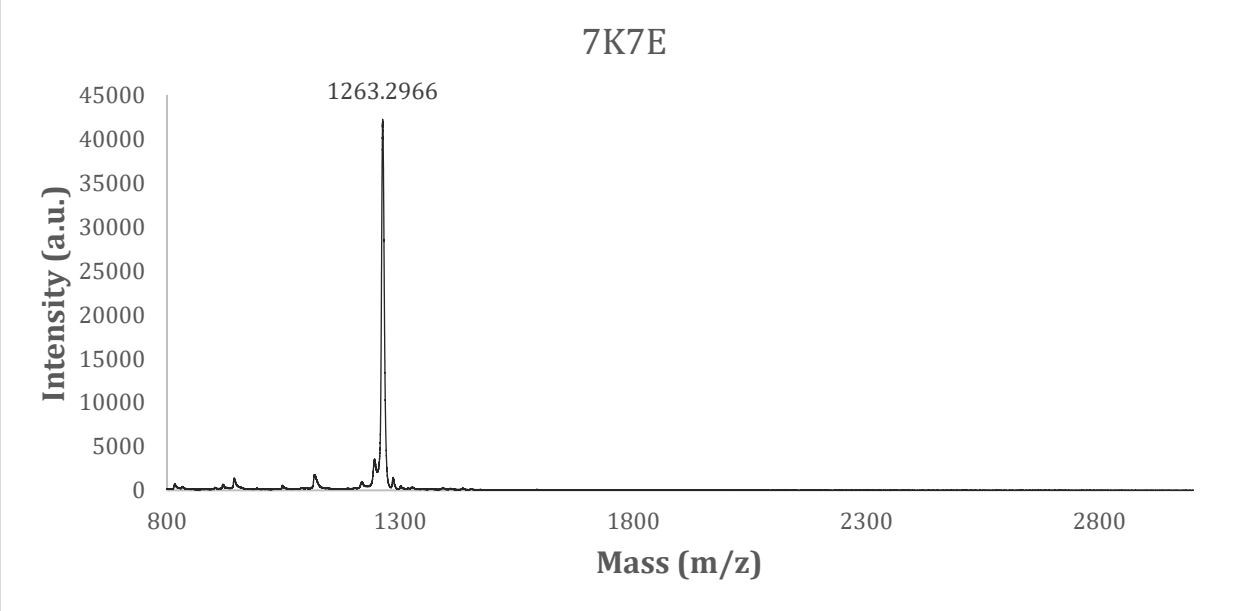

Figure 2.17 MALDI-TOF MS spectrum of the synthesized 7K7E peptide. *Expected mass = 1263 g/mol.*

Peptide assembly: Stock solutions of the FKFE series variants were prepared by dissolving the lyophilized peptide as 3mg/mL in each of the following: 10 mM TFA, pH 2.0; 10 mM Acetate, pH 4.0; 10 mM MOPS, pH 7.0; 10 mM CAPS, pH 10.0. Annealed samples were thermally annealed using the following thermal cycler protocol: initially heated up to 90 ℃ for 30 minutes; this was followed by cooling to 25℃ at a rate of 0.2℃ per minute. Unannealed samples were left at ambient temperature.

Circular Dichroism: The CD measurements were performed on a Jasco J-1500 CD spectropolarimeter in 0.1 mm quartz cells (Hellma Analytics), at various peptide self-assembly conditions. Spectra were documented at a scanning rate of 100 nm/min and a data pitch of 0.2 nm from 260 to 190 nm. For each spectrum, three scans were made and averaged.

Electron microscopy: Negatively stained TEM imaging was achieved on the Hitachi H-7700 microscope at an accelerating voltage of 80 kV. Samples were prepared by being deposited onto a on 200-mesh carbon coated copper grids from Electron Microscopy Sciences (Hatfield, PA). The sample was placed on the TEM grid for 30 seconds, then excess liquid was whisked away, and the samples were incubated with 1% uranyl acetate for 1 minute before being whisked away with qualitative filter paper.

2.6 References

1. Bowerman, C. J.; Ryan, D. M.; Nissan, D. A.; Nilsson, B. L., The effect of increasing hydrophobicity on the self-assembly of amphipathic beta-sheet peptides. *Mol Biosyst* **2009,** *5* (9), 1058-69.

2. Nilsson, B. L.; Doran, T. M., *Peptide Self-Assembly: Methods and Protocols*. Springer: 2018.

3. Lee, S.; Trinh, T. H. T.; Yoo, M.; Shin, J.; Lee, H.; Kim, J.; Hwang, E.; Lim, Y.-B.; Ryou, C., Self-Assembling Peptides and Their Application in the Treatment of Diseases. *International journal of molecular sciences* **2019,** *20* (23), 5850.

4. Inaba, H.; Matsuura, K., Peptide Nanomaterials Designed from Natural Supramolecular Systems. *Chem Rec* **2019,** *19* (5), 843-858.

5. Wang, Y. A.; Yu, X.; Overman, S.; Tsuboi, M.; Thomas, G. J., Jr.; Egelman, E. H., The structure of a filamentous bacteriophage. *J Mol Biol* **2006,** *361* (2), 209-15.

6. Chen, L.; Liang, J. F., Peptide fibrils with altered stability, activity, and cell selectivity. *Biomacromolecules* **2013,** *14* (7), 2326-31.

7. Bera, S.; Gazit, E., Self-assembly of Functional Nanostructures by Short Helical Peptide Building Blocks. *Protein Pept Lett* **2019,** *26* (2), 88-97.

8. Jiang, Y.; Zhang, W.; Yang, F.; Wan, C.; Cai, X.; Liu, J.; Zhang, Q.; Li, Z.; Han, W., Molecular design of stapled pentapeptides as building blocks of self-assembled coiled coil–like fibers. *Science Advances* **2021,** *7* (4), eabd0492.

9. Dai, B.; Li, D.; Xi, W.; Luo, F.; Zhang, X.; Zou, M.; Cao, M.; Hu, J.; Wang, W.; Wei, G.; Zhang, Y.; Liu, C., Tunable assembly of amyloid-forming peptides into nanosheets as a retrovirus carrier. *Proceedings of the National Academy of Sciences* **2015,** *112* (10), 2996-3001.

10. Ghosh, P.; De, P., Modulation of Amyloid Protein Fibrillation by Synthetic Polymers: Recent Advances in the Context of Neurodegenerative Diseases. *ACS Applied Bio Materials* **2020,** *3* (10), 6598-6625.

11. Tayeb-Fligelman, E.; Salinas, N.; Tabachnikov, O.; Landau, M., Staphylococcus aureus PSMalpha3 Cross-alpha Fibril Polymorphism and Determinants of Cytotoxicity. *Structure* **2020,** *28* (3), 301-313 e6.

12. Tayeb-Fligelman, E.; Tabachnikov, O.; Moshe, A.; Goldshmidt-Tran, O.; Sawaya, M. R.; Coquelle, N.; Colletier, J. P.; Landau, M., The cytotoxic Staphylococcus aureus PSMalpha3 reveals a cross-alpha amyloid-like fibril. *Science* **2017,** *355* (6327), 831-833.

13. Salinas, N.; Tayeb-Fligelman, E.; Sammito, M. D.; Bloch, D.; Jelinek, R.; Noy, D.; Uson, I.; Landau, M., The amphibian antimicrobial peptide uperin 3.5 is a cross-alpha/cross-beta chameleon functional amyloid. *Proc Natl Acad Sci U S A* **2021,** *118* (3).

14. Honcharenko, D.; Juneja, A.; Roshan, F.; Maity, J.; Galan-Acosta, L.; Biverstal, H.; Hjorth, E.; Johansson, J.; Fisahn, A.; Nilsson, L.; Stromberg, R., Amyloid-beta Peptide Targeting Peptidomimetics for Prevention of Neurotoxicity. *ACS Chem Neurosci* **2019,** *10* (3), 1462-1477.

15. Petty, A. J.; Keate, R. L.; Jiang, B.; Ameer, G. A.; Rivnay, J., Conducting Polymers for Tissue Regeneration in Vivo. *Chemistry of Materials* **2020,** *32* (10), 4095-4115.

16. Yao, L.; He, M.; Dongfang, L.; Liu, H.; Wu, J.; Xiao, J., Self-assembling bolaamphiphile-like collagen mimic peptides. *New Journal of Chemistry* **2018,** *42*.

17. Loquet, A.; Habenstein, B.; Lange, A., Structural investigations of molecular machines by solid-state NMR. *Acc Chem Res* **2013,** *46* (9), 2070-9.

18. Lee, Y. S., *Self-assembly and nanotechnology systems : design, characterization, and applications*. Wiley: Hoboken, N.J., 2012; p xx, 459 p.

19. Barbosa, M. r. A.; Martins, M. C. L., *Peptides and proteins as biomaterials for tissue regeneration and repair*. Woodhead Publishing is an imprint of Elsevier: Duxford, United Kingdom, 2018; p xiv, 376 pages.

20. Egelman, E. H.; Xu, C.; DiMaio, F.; Magnotti, E.; Modlin, C.; Yu, X.; Wright, E.; Baker, D.; Conticello, V. P., Structural plasticity of helical nanotubes based on coiled-coil assemblies. *Structure* **2015,** *23* (2), 280-9.

21. Wang, F.; Gnewou, O.; Modlin, C.; Beltran, L. C.; Xu, C.; Su, Z.; Juneja, P.; Grigoryan, G.; Egelman, E. H.; Conticello, V. P., Structural analysis of cross alpha-helical nanotubes provides insight into the designability of filamentous peptide nanomaterials. *Nat Commun* **2021,** *12* (1), 407.

22. Haimov, B.; Srebnik, S., A closer look into the alpha-helix basin. *Sci Rep* **2016,** *6*, 38341.

23. Sun, X.; Lai, L., Protein Fibrils Formed by Rationally Designed alpha-Helical Peptides. *Langmuir* **2020,** *36* (22), 6126-6131.

24. Koutsopoulos, S., *Peptide applications in biomedicine, biotechnology and bioengineering*. Elsevier/Woodhead Publishing: Duxford, United Kingdom ; Cambridge, MA, United States, 2018; p xiv, 639 pages.

25. Worsfold, P.; Townshend, A.; Poole, C. F., *Encyclopedia of analytical science*. 2nd ed.; Elsevier Academic Press: Amsterdam ; Boston, 2005.

26. Forterre, Y.; Dumais, J., Materials science. Generating helices in nature. *Science* **2011,** *333* (6050), 1715-6.

27. Perczel, A.; Gaspari, Z.; Csizmadia, I. G., Structure and stability of beta-pleated sheets. *J Comput Chem* **2005,** *26* (11), 1155-68.

28. Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R., A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc Natl Acad Sci U S A* **2002,** *99* (26), 16742-7.

29. Caplan, M. R.; Moore, P. N.; Zhang, S.; Kamm, R. D.; Lauffenburger, D. A., Self-assembly of a beta-sheet protein governed by relief of electrostatic repulsion relative to van der Waals attraction. *Biomacromolecules* **2000,** *1* (4), 627-31.

30. Hwang, W.; Marini, D.; Kamm, R.; Zhang, S., Supramolecular structure of helical ribbons self-assembled from a β-sheet peptide. *The Journal of chemical physics* **2003,** *118*, 389.

31. Marini, D.; Hwang, W.; Lauffenburger, D.; Zhang, A. S.; Kamm, R., Left-Handed Helical Ribbon Intermediates in the Self-Assembly of beta-Sheet Peptide. *Nano Letters* **2002,** *2*, 295-299.

32. Rudra, J. S.; Sun, T.; Bird, K. C.; Daniels, M. D.; Gasiorowski, J. Z.; Chong, A. S.; Collier, J. H., Modulating adaptive immune responses to peptide self-assemblies. *ACS Nano* **2012,** *6* (2), 1557-64.

33. Lee, N. R.; Bowerman, C. J.; Nilsson, B. L., Effects of varied sequence pattern on the self-assembly of amphipathic peptides. *Biomacromolecules* **2013,** *14* (9), 3267-77.

34. Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J., Biosynthesis of a Periodic Protein Containing 3-Thienylalanine: A Step Toward Genetically Engineered Conducting Polymers. *Journal of the American Chemical Society* **1995,** *117* (1), 536-537.

35. Merg, A. D.; Touponse, G.; Genderen, E. V.; Blum, T. B.; Zuo, X.; Bazrafshan, A.; Siaw, H. M. H.; McCanna, A.; Brian Dyer, R.; Salaita, K.; Abrahams, J. P.; Conticello, V. P., Shape-Shifting Peptide Nanomaterials: Surface Asymmetry Enables pH-Dependent Formation and Interconversion of Collagen Tubes and Sheets. *J Am Chem Soc* **2020,** *142* (47), 19956-19968.

36. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **2006,** *1* (6), 2876-2890.