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:

_____ _____

Charles C. Modlin Date

Engineering Self-Assembling Peptide Systems

By

Charles C. Modlin
Doctor of Philosophy

Chemistry

_____
Dr. Vincent P. Conticello
Advisor

_____
Dr. Stefan Lutz
Committee Member

_____
Dr. Emily Weinert
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

Engineering Self-Assembling Peptide Systems

By

Charles C. Modlin
B.S., University of Virginia, 2012

Advisor: Vincent P. Conticello, Ph.D.

An abstract of
A dissertation 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

Doctor of Philosophy
in Chemistry

2018

Abstract

Engineering Self-Assembling Peptide Systems

By Charles C. Modlin

Chemical self-assembly is an invaluable tool used in the bottom-up creation of *de novo* systems that otherwise would be impossible to construct in a top-down approach given the size regime in which many biomaterials exist. Currently, the ability to engineer self-assembling systems that rival the level of complexity observe in nature is highly limited. However, rational design enables researchers to push the current limits of complexity within *de novo* self-assembling systems, specifically within the scope of self-assembling peptide biomaterials. Often, this involves using what is observed in nature as a foundation for further design. To this end, numerous research projects will be presented, which hinge upon the rational design and alteration of existing systems to produce new and novel systems. These projects include the use of an arginine staple motif as a means of engineering nanotubular structures with variable diameters and structures, the use of computational design to create entirely new sequences which self-assemble into solenoidal nanotubes, and the addition of DNA to peptide systems as a means of increasing the complexity of self-assembling systems beyond what is possible for single-species assembly. To verify the hypotheses regarding these systems, a wide spectrum of biophysical analyses was used. The results of these projects help widen the structural ‘tool box’ for self-assembling peptide materials, and help provide better understanding of the key principles in peptide self-assembly.

Engineering Self-Assembling Peptide Systems

By

Charles C. Modlin
B.S., University of Virginia, 2012

Advisor: Vincent P. Conticello, Ph.D.

A dissertation 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

Doctor of Philosophy
in Chemistry

2018

Acknowledgements

I would like to sincerely thank my advisor Professor Vincent Conticello for his support, patience and guidance over the course of my PhD path. He has led by example, setting high standards for academic research as well as personal growth, and has served as a role model in and out of the laboratory. I am grateful to have had him as my mentor for six years and will carry forth the lessons he has taught me wherever I travel in life. I would also like to express gratitude for those who have served as my committee members, Professor Stefan Lutz and Professor Emily Weinert. They have elevated the quality of my research significantly with their wisdom, insight and direction throughout my PhD trajectory. I must also thank the members of the Conticello lab, who have helped me immensely: Rebecca Bartlett, Spencer Hughes, Shengyuan Wang, Dr. Andrea Merg, Avi Tuachi, and Gavin Touponse. These people have significantly improved both my research and my everyday life, providing a wonderful work environment as well as assistance with experimentation, insightful discussions, and encouragement. I am also grateful for former lab members as well, Dr. Chunfu Xu, Dr. Tao Jiang, Dr. Elizabeth Magnotti and Dr. I-Lin Wu. They too have provided great assistance and encouragement along the way. I would also like to acknowledge a graduate students not in the Conticello lab, but who have been critical to my success, Samantha Iamurri and Thomas Pickel. I thank them both for their insight and all of the wonderful discussions regarding research.

A special thanks goes to all the members of the Robert. P. Apkarian Integrated Electron Microscopy Core, Professor Elizabeth Wright, Ms. Hong Yi, Ms. Jeannette Taylor, and Mr. Art McCanna. Their mentorship helped me learn so much in the field of electron microscopy, and it

was an honor to serve as the Service Instructor for two years. Doing so substantially improved my research and knowledge of the field.

I would like to express my gratitude for all of our wonderful collaborators over the past six years. Professor Ed Egelman (University of Virginia) has been instrumental in helping us solve the structures of some of the most complex peptide systems. Dr. Xiaobing Zuo (Argonne National Lab) has provided constant support in the field of small-angle X-ray scattering, and without his assistance, much of our work would not be possible. Professor Gevorg Grigoryan (Dartmouth University) provided the inspiration and sequences for many of our computationally based projects. Professor Elizabeth Wright (University of Wisconsin) was key in many of our microscopy pursuits, particularly in the field of cryo-electron microscopy. Dr. Fred Strobel has been a constant source of assistance with mass spectrometry over the years. Without the help of all of these people, my research and PhD career would be profoundly lacking.

I must also thank those who inspired my scientific career, and decision to pursue a PhD. Professor Linda Columbus (University of Virginia) and Professor Gary Shiflet (University of Virginia) were monumental in my decision to pursue a career in science. I thank them for being such positive role models.

Lastly, I acknowledge my family, who has provided incredible and unwavering support throughout my life. My parents have worked tirelessly to help put me into positions and schools where I could succeed and grow. Without their love and support, I would not be where I am today. My sister has always been a source of support as well, and eternally brings a smile to my face during difficult times. I thank my grandparents as well, for their love, support and encouragement in all aspects of my life.

Table of Contents

List of Figures

List of Tables

No tables listed.

Chapter 1: Introduction

1.1 The role and importance of self-assembly in biological systems

Chemical and biological self-assembly, and more explicitly, the self-assembly of peptide- and protein-based systems has been noted as one of the most important fundamental quandaries facing today's scientific community.[4] "How do biological structures and systems self-assemble" has long been a question of paramount importance, predating even Watson and Crick's structural elucidation of deoxyribonucleic acid.[5] Biological self-assembly is ubiquitous in all forms of life, and nearly every biochemical process involves biological self-assembly in some fashion.[6] Many of these self-assembling systems consist of globular protein domains interacting with and docking into each other, while others are built from many copies of the same protein fold. One example of the latter motif: helical and superhelical self-assembling systems, which will be covered in more depth later.

Nature often uses self-assembly of helical molecules to produce complex nanomachines capable of enacting physical and chemical changes in astonishingly high yields, with little waste production, and at speeds sometimes approaching the diffusion limit. Moreover, rather than paying the metabolic tax of keeping these many subunit machines in working order at all times, living organisms are constantly recycling and rebuilding both the simple monomers and the supramolecular assemblies. Though these structures are present in every domain of life, those from archaea and bacteria are more deeply understood. This is largely due to the relative ease of purification and isolation of the individual components, which can then be studied *in vitro* rather than *in vivo.* For this reason, this introduction will focus on the helical assemblies whose structures have been solved at high resolution, via x-Ray crystallography[7] and cryo-electron microscopy (cryo-EM) based reconstruction.[8]

A better understanding of the natural rules and methodologies employed by biological systems (to create the fundamental building blocks of life) affords scientists the ability to create *de novo* structures and systems, such as peptide-based nanotubes,[9] nanosheets,[10-11] and more. A working comprehension of the rules that govern chemical self-assembly is indeed critical for *de novo* designs. Though the engineering world possesses and skillfully employs a wealth of top-down approaches for the construction of macro-scale systems (i.e. machines), there is a size limit which limits, or alters, the efficacy of the use of tools in the shaping of products and systems. For example, while an engineer can employ physical tools to construct a large device such as a battery, very few tools exist which can successfully build highly complex systems in a top-down fashion on the nanoscale size regime, such as a ribosomes,[12] actin,[13] or histones.[14]

Therefore, chemical-assembly becomes vital in the pursuit of *de novo* chemical systems. Briefly defined, chemical self-assembly is the study of the methods in which smaller molecules organize and pack to form larger, more-ordered architectures in a hierarchical manner. Generally speaking, chemical self-assembly, and in the context of this thesis – peptide self-assembly, is governed by thermodynamic principles, wherein energy is minimized in a defined system. For example, hydrophobic collapse occurs in many peptide-based systems as this mitigates the thermodynamic penalty of hydrophobic residues facing an aqueous barrier, effectively minimizing the energy within a closed system.[15] Self-assembly occurs across numerous size regimes. Individual molecules self-assemble into cellular nanostructures, nanostructures into cells, cells into tissues, tissues into organisms, and so on. However, for the purpose of this thesis, only the nanometer and micrometer scales will be considered, as these are the scales in which peptide-based structures primarily exist.

1.2 Self-Assembling Helical Biomolecules

The majority of this thesis will focus on self-assembling helical assemblies, therefore it is prudent to survey the current field of helical macromolecules. In the pursuit of rationally designing new helical biomaterials, it is important to take inspiration from systems that already exist in nature and those engineered *de novo* by other researchers within the field. Of note, helical biomolecule are relatively prevalent in nature, perhaps the most notorious being the DNA double helix. Even on the macroscale, helical structures are over represented, from plant tendrils an seed pods to sea shells and horns; the helix is a ubiquitous structural motif.[16]

1.2.1 Native Helical Biomolecules

One such self-assembling structure is Tobacco Mosaic Virus (TMV), the first virus ever crystallized [17]. TMV has been the subject of extensive scientific interest due to the wide range of of virulence against humans (which confers ease of handling), high isolated yield, and regularity of self-assembly. Fully assembled TMV is 300 nm long, 18 nm thick, and possesses a hollow lumen of 4 nm diameter. The TMV assembly consists of single-stranded RNA inserted between successive turns of a right-handed coat protein superhelix. It was revealed that TMV does not assemble by successively adding single coat protein copies onto the growing superhelix; the TMV assembly protomer is actually a "disk aggregate" comprising roughly 34 coat protein subunits (roughly 2 superhelical turns)[18]. This assembly mode better accounts for the observed rapid self-assembly of TMV than a single subunit addition model. This use of an assembly protomer is similar to and served as inspiration for a computationally designed 'protomer' approach described in Chapter 5.

The development of direct electron detectors for cryo-EM microscopes has made it possible to obtain near-atomic resolution structures for helical assemblies like TMV. Recently, Fromm et al. published a 3.35 Å resolution structure of TMV rods[19]. Though pushing towards better resolution is itself a noble goal, higher quality structures also allow one to explore docking small molecules into the lumen of a helical assembly, or even mutating the sequence to alter the structure in a specific fashion. Thus, high resolution structure solutions may pave the way for the development of antiviral treatments and the biomaterials to interface with native biological machinery or manmade, inorganic machinery. Within a more immediate scope, high resolution cryo-EM has served as a pinnacle technique for numerous research projects described herein.

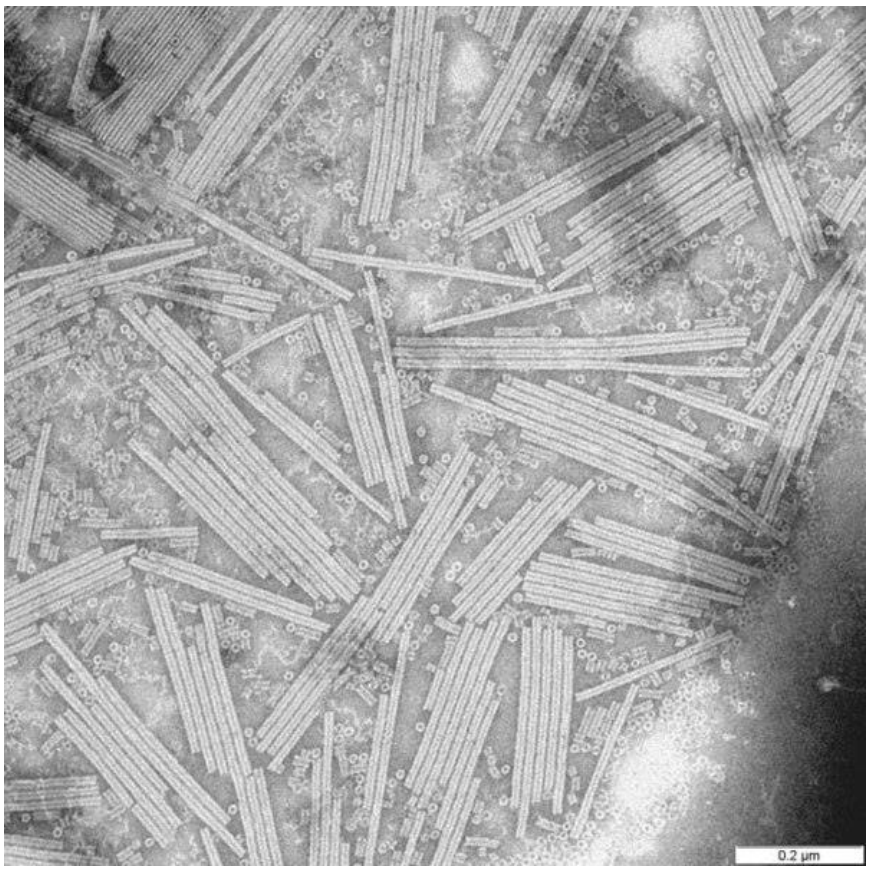


Figure 1.1 Negative-stained TEM of fully assembled TMV particles[1]. Scale bar is 100 nm. Image is in the public domain.

As it stands now, TMV is used as a mass-per-length standard in negative-stain scanning transmission electron microscopy (STEM)[20]. Briefly, protein assemblies of interest are incubated with TMV, which has a well-known mass-per-length and mass-per-area. Using the contrast between the TMV filaments and the stained background, it is possible to correlate sample brightness with mass measurements. Thus, comparing a protein assembly's apparent brightness to TMV in the same micrograph can allow for the determination of mass-per-length (which gives insight into protein stoichiometry for 1-dimensional materials such as filaments) or mass-per-area (which gives insight into protein stoichiometry for 2-dimensional protein materials).

Because of its repetitive and well-ordered structure, TMV has recently been investigated for use in antigen display. Virus-like particles (VLPs) have emerged as very effective antiviral therapies, owing to their ability to display many copies of a peptide antigen on a single particle[21-24]. Smith et al. modified TMV to have a reactive lysine residue exposed on the surface of the filament using a library based approach, allowing for biotinylation of the TMV particles[25]. Following this, they were able to demonstrate the fusion of GFP-streptavidin to the TMV-biotin construct, and finally they fused a peptide-antigen-streptavidin moiety to the TMV-biotin construct. This strategy was employed because the peptide antigen sequence could not be mutated into the TMV protein sequence without disrupting the structure. It was then shown that the TMV-antigen conjugate was significantly more immunogenic than the free antigen itself.

TMV has also been explored as a candidate structure for increasing surface area in microbatteries for microelectronic machines (MEMs)[26-33]. Simply put, the ability to scale down battery size is limited by available surface area. Recently, a strategy for conjugating modified TMV coat proteins to inorganic electrodes was developed for use in MEMs, taking advantage of the huge available surface area in a hollow cylinder such as TMV[34]. Briefly, the wild type TMV

sequence was point-mutated so that one exposed cysteine per coat protein was displayed on the outer surface of the nanotube. This TMV1cys protein was then allowed to self-assemble onto a gold plate, activated with a palladium catalyst, and finally coated with nickel. The nickel-coated TMV batteries were shown to have a six-fold higher capacity than that of analogous nickel batteries with no viral component. This study paves the way for future increases in microbattery capacity, and will allow for the generation of even smaller MEMs devices.

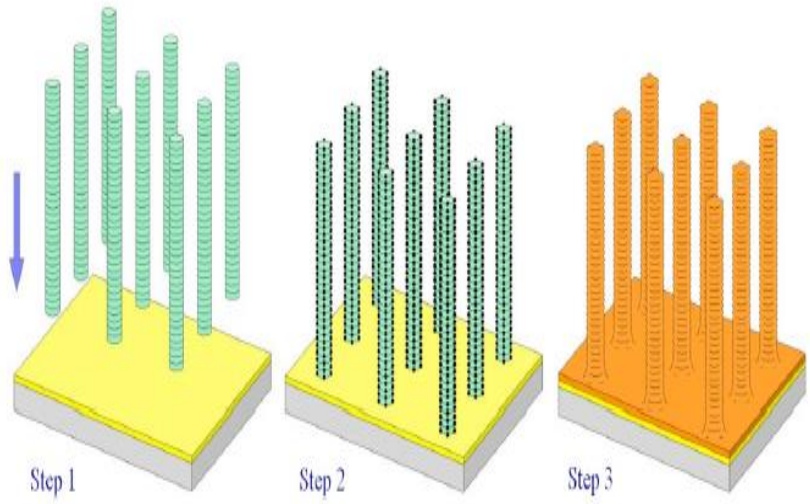


Figure 1.2. Schematic of TMV1cys incorporation into a MEMs device.[34] Step 1 shows the conjugation of TMV1cys to a gold surface. TMV1cys is activated with palladium in Step 2, and coated with nickel in Step 3. From Gerasopolous et al., 2008.

Central to the fields of structural biology and infectious disease research are the pili found in archaea, bacteria, and even eukaryotic cells. The structure of a chaperone-usher pilus was recently published and reveals more information on the general function and assembly mode of the most common pili structures of clinical interest[35]. Chaperone-usher (CU) pili are a class of pilus that are displayed on the surface of bacterial pathogens and are responsible for host-cell adhesion and recognition. These CU pili have also been implicated in biofilm formation.

Generally, these pili comprise two main components: a short tip fibrillum with three to four subunits, and a superhelical rod comprising ~1000 copies of a single pilin protein, which can extend 1-2 microns from the bacterial cell surface. One of the two types of pili that are commonly found in uropathogenic *E. coli*, the P type pilus' structure was solved to 3.8 Å resolution using cryo-EM and helical reconstruction. CU pili are so named for the outer-membrane embedded usher (PapC for the P pili) and periplasmic chaperone (PapD for the P pili) that assist in pilus assembly. Briefly, the PapA subunits that form the helically-wound pilus rod are stored in the inner membrane until they are ejected by a transporter protein to initiate pilus assembly. The chaperone captures the PapA subunits as they exit the transporter, and donates a strand to form a complete immunoglobulin fold with the PapA. This chaperone:subunit complex then docks to the outer-membrane usher, where PapA polymerization occurs, resulting in the secretion of the pilus. The usher serves to facilitate polymerization by promoting donor strand exchange, whereby the chaperone's donated strand is replaced by an N-terminal extension from a new PapA subunit. This leads to the dissociation of the chaperone, which allows it to recruit another PapA subunit and continue extension of the pilus out of the membrane. It is currently unknown what causes the pilus to emerge from the membrane, as there is no ATP in the periplasm, and no chemical gradient to drive this process. The resulting pilus rod has been described as "spring-like", as atomic force microscopy experiments have shown reversible uncoiling of the superhelical structure. This is hypothesized to allow the uropathogenic bacteria to continue adhering to the urinary tract under the high shear forces present during urination. Through solving the structure via cryo-EM, the molecular basis of this rod uncoiling was revealed. It was found that while the quaternary structure of the helical rod was largely held together by polar interactions, these interactions would progressively break down under high shear forces. In fact, this does occur, and while quaternary

structure of the pilus does indeed fall apart, the integrity of the pilus is unchanged due to the very strong, hydrophobic interactions that drive donor strand exchange. It has also been hypothesized that the formation of the quaternary structure during pilus extrusion drives the extrusion process, but further experimentation is required to say this with little doubt.

Another such structure, the bacterial sex F pilus, which is responsible for exchanging genetic material (notably plasmid DNA), has allowed the mapping of the entire *Escherichia coli*

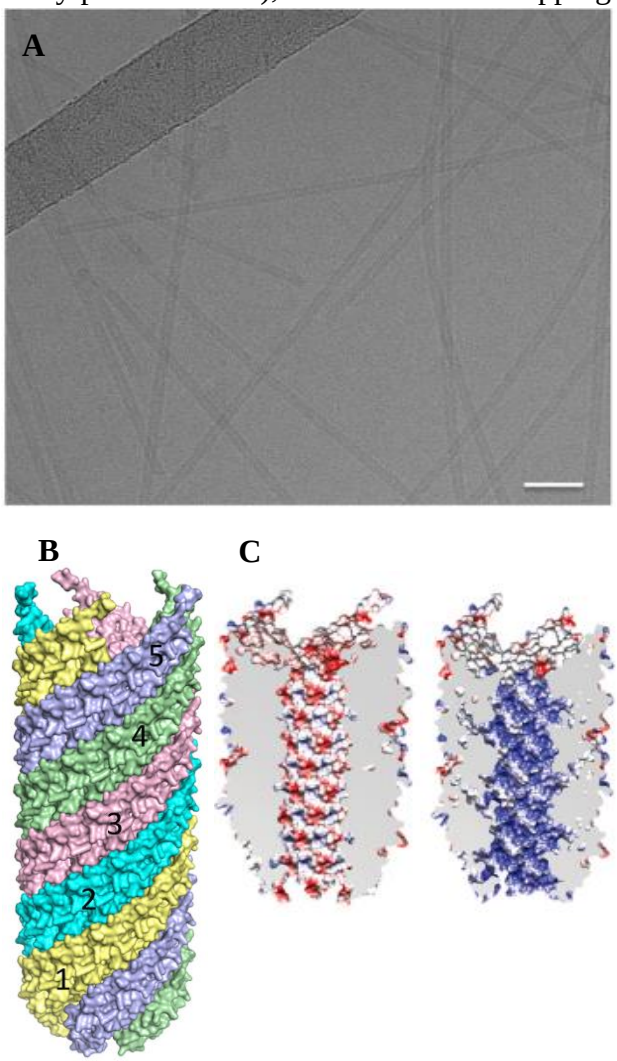


Figure 1.3: Structure of the sex pilus. (A) Cryo-electron micrograph of the sex pilus (scale bar is 40 nm). (B) Pilin assembly constitutes the outer surface of the pilus. (C) Difference in electrostatic potential between the lumen of the pilus with and without the phospholipid core[3] From Costa et al, 2017.

genome and greatly furthered our understanding of molecular biology and genetic engineering. Interestingly, the structure of this ever-important bacterial feature was unknown until 2016, when two related F pili structures were solved to 3.6 and 5.0 Å resolution using cryo-EM and iterative helical real space reconstruction (IHRSR)[3]. Emergent during the reconstruction process was the existence of a non-protein density, later determined to be phosphatidylglycerol, which follows the contour of the pilus helical array. The helical assembly is held together both by pilin subunit interactions, and phospholipid-pilin interactions, with 70.3% of the lipid surface area buried within the protein assembly, and the lipid head groups facing out towards the 28 Å lumen.

Interestingly**,** the phospholipid content within the pilus is distinct from the phospholipid composition of the membrane, meaning there is preferential binding of pilus protein to certain phospholipids. This specificity has not previously been observed for lipid-protein polymer complexes, and suggests an important role of the lipid. Three hypotheses were laid out in the study: 1) the lipid serves to facilitate pilus insertion into the host membrane, allowing the pilus to deliver genetic material into the host 2) the lipid aids in reinserting the pilus subunits into the bacterial membrane for pilus retraction/disassembly and 3) the lipid increases substrate specificity, with different species of conjugative pili exhibiting preferences for different phospholipids.

1.2.2 Non-Native Helical Assemblies of Synthesized Peptides

Non-native helical assemblies, particularly from synthetic peptide systems, is the primary focus of this thesis. As such, much can be gleaned from extant peptide systems engineered by others in the field. The Conticello group and Egelman group have engineered several alpha-helical peptide nanotube assemblies. The first of such systems, published in 2013, is the peptide 7HSAP1. Xu *et al.* utilized **GCN4-pAA** as a scaffold for sequence manipulation in the creation of 7HSAP1.

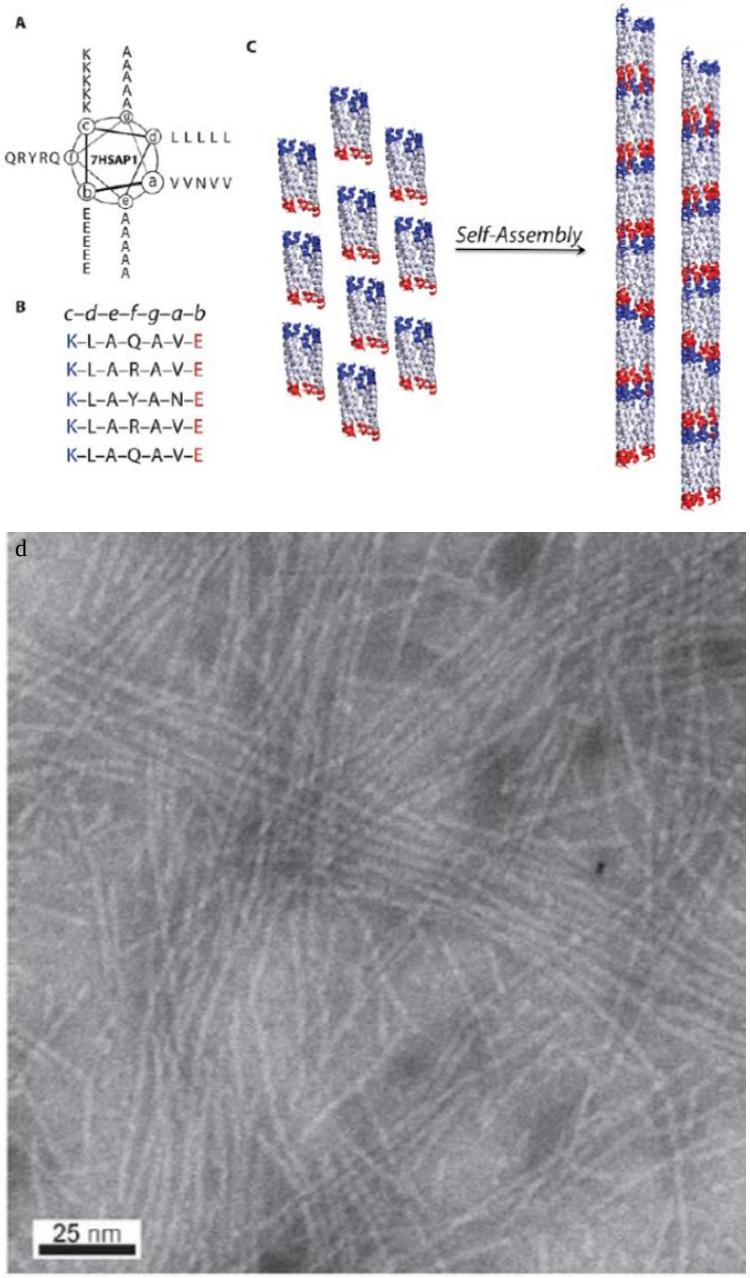


Figure 1.4 Design and assembly of 7HSAP1. (a) Helical wheel representation of the 7HSAP1 amino acid sequence derived from GCN4-pAA. (b) Linear representation of the amino acid sequence of 7HSAP1. (c) Graphical representation of the self-assembly process for 7HSAP1 wherein charge complementary termini stack in a fashion mimetic of lock washers stacking. The positively charged N-termini are indicated via red, while negatively charged C-termini are shown in blue. (d) STEM of 80 μM 7HSAP1 nanotubes in 10mM MES buffer at pH 6.0.

GCN4-pAA, a *de novo* engineered peptide derived from the leucine zipper portion of GCN40 [36] (a transcription factor in *S. cerevisiae)* forms a heptameric bundle of alpha-helices with an inner channel measuring 7 Å in diameter. This heptameric bundle assembly occurs identically in both solution and crystalline space, and the resulting supramolecular assembly is reminiscent of a lock washer [37]. Xu *et al.* recoded the amino acid sequence of **GC4-pAA** to generate positively charged N-terminal heptads and negatively charged C-terminal heptads, promoting self-assembly via linear propagation of helical bundles associating head-to-tail, similar to the stacking of lock washers. Each subunit stacks upon the next with a slight offset in register by one amino acid. This offset increases the surface area available for complementary interactions between lock washer subunits. The linear end-to-end association of lock washer subunits creates a helical nanotube with a discernible inner channel. Xu and colleagues posit that self-assembly of the 7HSAP nanotube is largely driven by charge complementarity of the opposing termini between subunits. The authors posit that this linear propagation is further assisted by burial of solvent-accessible surface area between helical bundle subunits [37]. 7HSAP1 was left uncapped in an effort to promote head-to-tail association of helical lock washer subunits. Interestingly, capping the sequence diminishes but does not fully abrogate fibrillar assembly of 7HSAP1.

Circular dichroism (CD) studies confirm the hypothesis that 7HSAP1 assemblies are alpha-helical and transmission electron microscopy (TEM) studies of 7HSAP1 reveal fibrillar assemblies in 10mM 2-(N-morpholino)ethanesulfonic acid (MES) buffer at pH 6.0. Fibrillar structures were observed for peptide concentrations ranging from 25 μM to 4.0 mM. Figure 4 shows a STEM of 7HSAP assembled in the aforementioned conditions, and clearly indicates successful fibrillar assembly. STEM analysis, corroborated by cryo-TEM, indicates a fibril diameter of ca. 3 nanometers. Mass per length analysis of the STEM data indicate that 7HSAP likely assembles as

a seven-helix bundle as hypothesized (given its parent structure **GCN4-pAA**). It should be noted though, that the experimental error inherent in this study was significant enough that a six-helix bundle could not be fully ruled out as a possible oligomeric identity for 7HSAP1 fibrils. A q^{-1} power law was observed for the small-angle region in the experimental small angle x-ray scattering (SAXS), indicative of a rod-like form in solution. Pair distance distribution function [38] approximates wall thickness at ca 10 Å, middle cylindrical shell diameter at 20 Å, and hollow lumen diameter at 10 Å. The radius of gyration of the fiber cross-section was calculated to be 12.4 Å. Solid-state NMR ^{13}C-^{15}N corroborated the hypothesized one residue registry shift. With **7HSAP1** Xu *et al.* provide a stellar example of employing selective recognition between structurally complementary interfaces in order to promote self-assembly of high-aspect-ratio nanotubular structures. Through a broader lens, it is also a good example of tailoring existing natural structural motifs to engineer *de novo* systems, and provides a good foundation for future research into helical systems.

A system that has been of considerable interest to the medical community is Aβ, the amyloid beta protein that, along with tau, is largely responsible for plaques found in the brains of deceased patients who suffered from neurodegenerative diseases, e.g. Alzheimer's. It is thought that these proteins misfold in a prion-like fashion, seeding the misfolding of other copies, leading to plaque build-up. To understand the folding and misfolding of Aβ, the structural biology community has studied the self-assembly of small domains of Aβ. Using residues 16-22 from the 42 residue protein, the Lynn group has been producing nanotubes of the wild-type fragment and conservatively mutated versions in an attempt to probe the self-assembly mode[39-40]. They were able to show that there was a small energetic difference that led to either the formation of fibers in Aβ (16-22) or nanotubes with a diameter of ~50 nanometers.

Recently, Chen *et al.* generated a double-walled β-sheet macrocycle from the 16-22 fragment and solved its structure to 2.1 Å via x-Ray crystallography[2]. Briefly, they synthesized the KLVFFAE fragment (the "K-E strand", with a *p*-iodophenylalanine at position 19 to allow for determination of x-ray crystallographic phases through single-wavelength anomalous diffraction

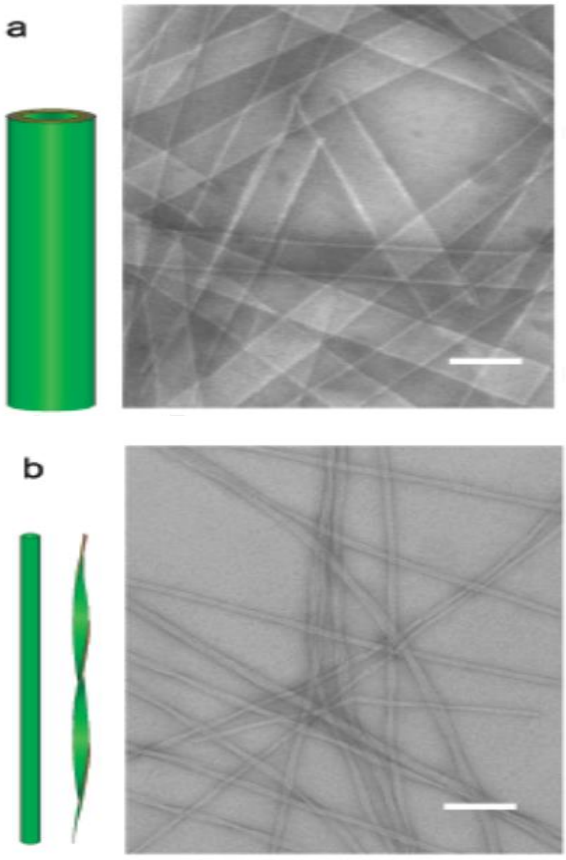


Figure 1.5 Assembly of Aβ (16-22) into tubes (A) under acidic conditions and into fibers (B) under neutral pH conditions.

of iodine. This fragment is connected to the retro-peptide sequence of KLVFFAE (the "E-K strand") through ornithine residues on each end. The "E-K" strand contains an N-methyl-phenylalanine residue to prevent uncontrolled hydrogen bonding, and to discourage amyloid fibril formation (see Figure 1.6).

The macrocyclic peptide chain folds into a β-hairpin structure, which forms hydrogen-bonded dimers and hydrogen-bonded tetramers. The dimers form a relatively flat, four strand anti-parallel beta-sheet, with 9 hydrogen-bonding pairs from the K-E strands dictating the arrangement. The dimers form the inner wall of the double-walled nanotube. The tetramers, on the other hand,

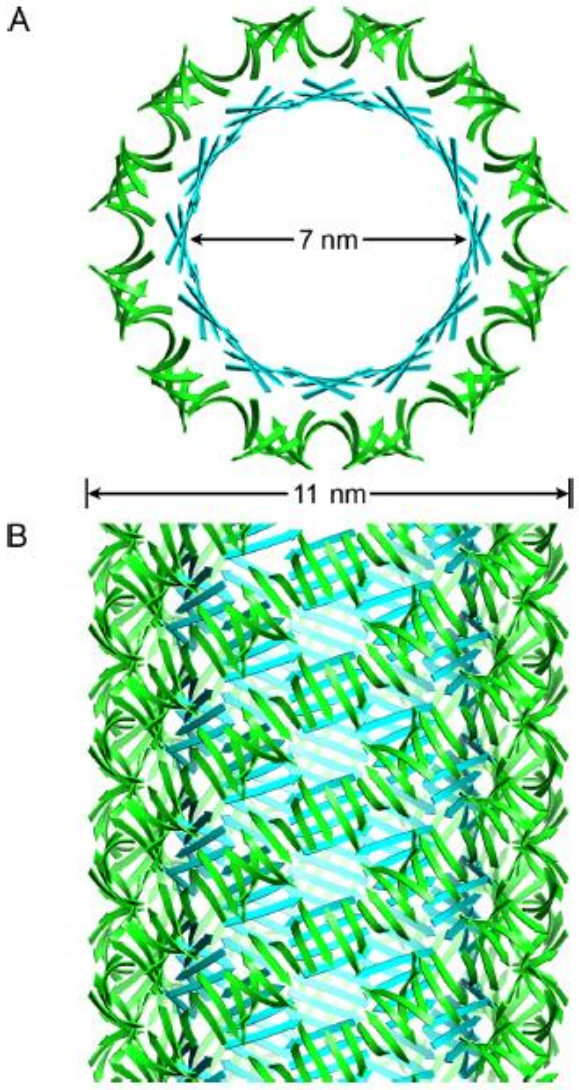


Figure 1.6: Model of the macrocycle nanotube. (A) Top view highlighting the large inner lumen diameter of the double-walled nanotube. (B) Side view of the nanotube, illustrating the packing of the dimeric and tetrameric structures against each other.[2] From Chen et al, 2017.

are twisted structures that rely on hydrophobic interactions as well as hydrogen-bonding for stabilization. The tetrameric protomers form the outer wall of the 11 nm diameter nanotube. This double-walled structure is almost as large as some of nature's largest homogeneous tubular assemblies (TMV has a diameter of 18 nm). The authors admit though that the structure was not possible to predict by looking at the sequence, in contrast to some of the α-helical peptide assemblies that we have highlighted.

1.3 Helical Assemblies Beyond Peptides

Though peptide-based approaches to supramolecular assembly predominate the field, recent advances by several groups demonstrate the efficacy and benefits of non-peptide "foldamers." In many cases, these foldamers distinguish themselves from peptide-based structures via chemically unique carbon backbones, though foldamers may still retain proteinaceous or amino-acid-like sidechains. Non-peptide foldamers offer several advantages when compared to their peptide-based counterparts, namely, foldamers: (a) are biomimetic yet often resistant to proteases [41], (b) follow unique folding rules distinct from protein folding [42], but (c) still retain the inherent link between sequence and structure, affording a level of control over secondary, tertiary and quaternary structure via sequence modulation. As noted, the rules governing the relationship between sequence and structure can be unique and dissimilar to protein folding rules. Within the foldamers field, the Guichard group reports *de novo* aliphatic, alpha-helicomimetic oligourea foldamers which self-assemble into either discrete helical bundles or superhelical nanotubes (sequences **H1** and **H2** respectively). The aforementioned supramolecular assemblies are selected for via oligourea primary sequence manipulation, wherein **H1** features two contiguous urea moieties Leu^{u} (urea with an uncharged, leucine-like sidechain) at the *a* and *d* positions of a

pentameric 2.5 helix, while **H2** features Leu^{u} at the *a* and *c* positions, with Ala^{u} and Pro^{u} at the *d* position. **H2** effectively extends the hydrophobic face of the 2.5 helix by one position, as compared to **H1** (see figure). These differences in hydrophobic distribution/density are largely what govern the structurally distinct supramolecular assemblies **H1** and **H2** described below.

H1 assembles non-cylindrically as a helical bundle, comprising a hexamer of oligourea 2.5 helices (see figure). It is thought that hydrophobic collapse contributes significantly to supramolecular formation, as the inner core of the hexamer bundle is populated primarily by Leu^{u} residues, similar to folding principles observed in many peptide-based macromolecular assemblies. Within the hexamer bundle, three distinct **H1** antiparallel dimers are observed, with the following local interactions govern their packing: (a) Leu^{u}-Leu^{u} hydrophobic packing, (b) reciprocal hydrogen bonding between the Lys^{u}8 side-chain amine and the C-terminal Leu^{u}11 uncapped carbonyl group, (c) canonical 'knobs-into-holes' packing of the inner Leu^{u}-rich core. No salt bridges are detected in the **H1** hexamer bundle despite the presence of four charged residues. Though comparatively small, an inner channel exists within the hydrophobic core of the assembly (495.0 Å^3). The stoichiometry of the six-helix bundle was evaluated via ESI-MS and matched the crystallographic data well. The stoichiometry was further substantiated with a negative control sequence **H3.** Oligourea **H3** features a Leu^{u} to Asn^{u} mutation at position 6 to test the effects of interrupting the wholly hydrophobic inner channel with a hydrophilic residue. As anticipated, ESI-MS and CD melting data indicate a reduction in stability of the **H3** assembly when compared to **H1**, and importantly, data did not indicate formation of a hexameric bundle. All aforementioned structural data were obtained via crystallography resolved to 1.25 Å.

H2 self-assembles in a decidedly distinct manner from **H1**, folding to produce a nanotube with an extended, water-filled inner lumen of 17Å diameter, much larger than that of **H1**. Additionally, different from **H1** is the distribution and location of hydrophobic residues, wherein

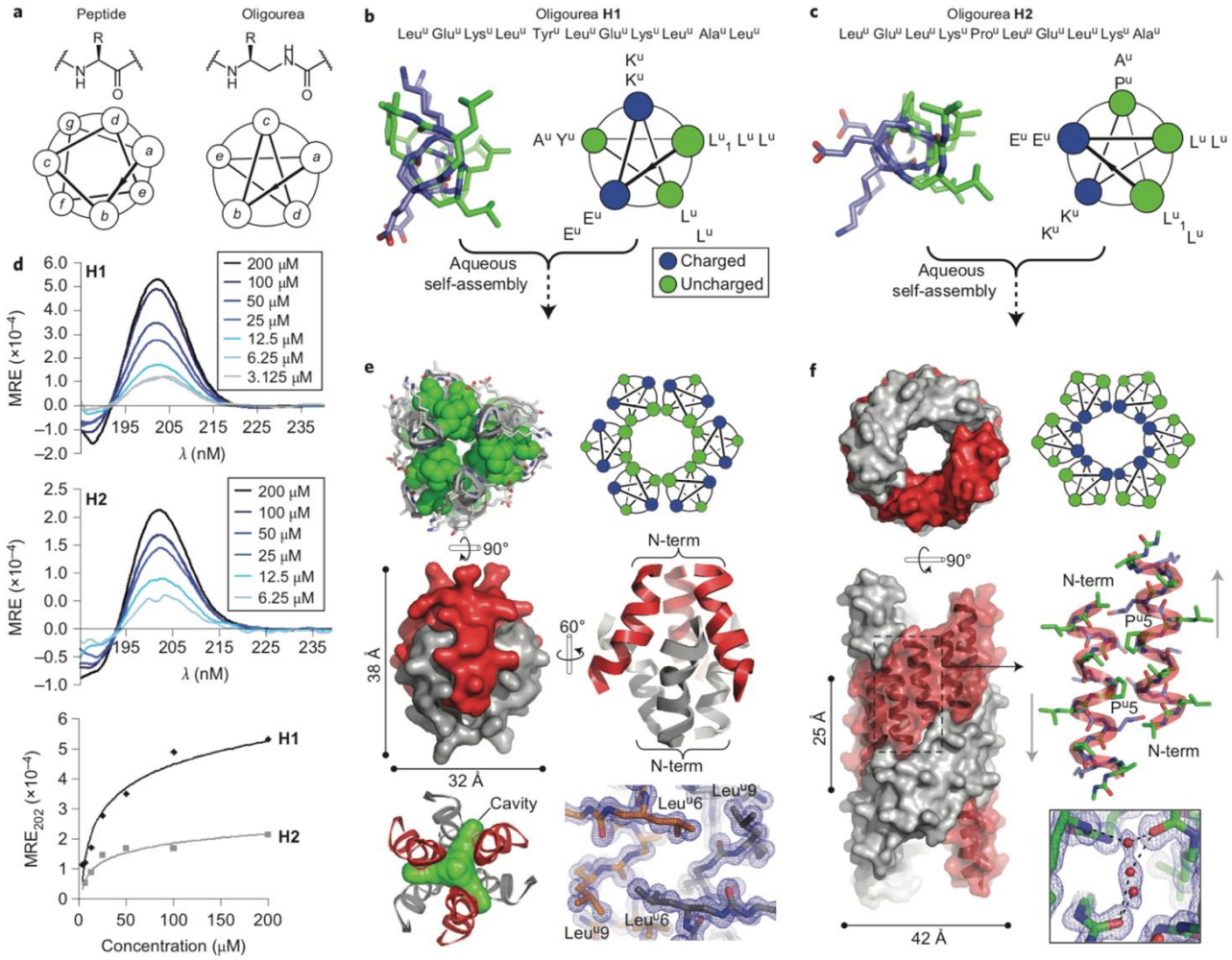


Figure 1.7: (a) Helical wheel diagrams delineating the geometric differences between alpha-peptides and aliphatic oligoureas. (b) and (c) primary amino acid sequence, top-down view of the crystal structure, and helical wheel indicating side-chain population at each position of the 2.5 helix for H1 (b) and H2 (c). (d) Variable concentration circular dichroism analysis of both peptides in water, shown in mean residue ellipticity. From Collie et al., 2015,

H2 demonstrates hydrophobic Leu^{u}, Ala^{u} and Pro^{u} residues populate the external face of the nanotube. The charged Lys^{u} and Glu^{u} residues fill the inner-channel, enabling penetration of water molecules into the lumen. Whereas **H1** assembles as a discrete hexameric bundle of oligourea 2.5 helices, the helices of **H2** self-assemble into entwined dimeric right-handed superhelices which pack adjacently in a staggered fashion, wherein 12 oligourea helices complete one superhelical turn and fully enclose the nanotube. Collie *et al* posit that hydrophobic interactions are also key in the assembly of **H2**; however, dissimilar to **H1**, **H2** demonstrates extensive salt bridges to stabilize the lumen. All aforementioned structural data was obtained via crystallography and resolved to 1.4Å. **H2** indicates that an extended, contiguous hydrophobic face is necessary for selective tubular formation rather than helical bundles. To corroborate this hypothesis, Guichard *et al* synthesized sequence **H5**. Like **H2,** it features three contiguous uncharged residues at the *a, b*, and *d* positions and contiguous charged residues at the *e* and *c* positions, analogous to **H2.** Additionally, the **H5** oligourea is 2 residues longer than **H2**, with both additional residues populating the hydrophobic face. Crystallographic data for **H5** confirm six superhelices form a tubular assembly with an inner channel measuring 26 Å in diameter, larger than that of **H2.** Though the stoichiometry of **H2** and **H5** are different, their modes of assembly are highly similar, with both featuring staggered, laterally associated helices held in place via salt-bridges and hydrophobic collapse.

Guichard *et al* report the first instance of helical, 'channel-type' assemblies from oligourea foldamers. This is an important step in expanding what is known about the relationship between structure and function. Moreover, due to their protease resistance and relatively large water-filled inner channel, the oligourea assemblies described by the Guichard group possess potential for

numerous downstream applications, including but not limited to small molecule (i.e. drug) encapsulation[43-45] and ion conductance across a membrane[46].

In addition to oligourea foldamers, peptoids represent one of the most promising non-protein moieties capable of forming extended helical assemblies[47]. . Whereas peptides have side-chains connected to the α-carbon, peptoids instead have side-chains attached to the amide

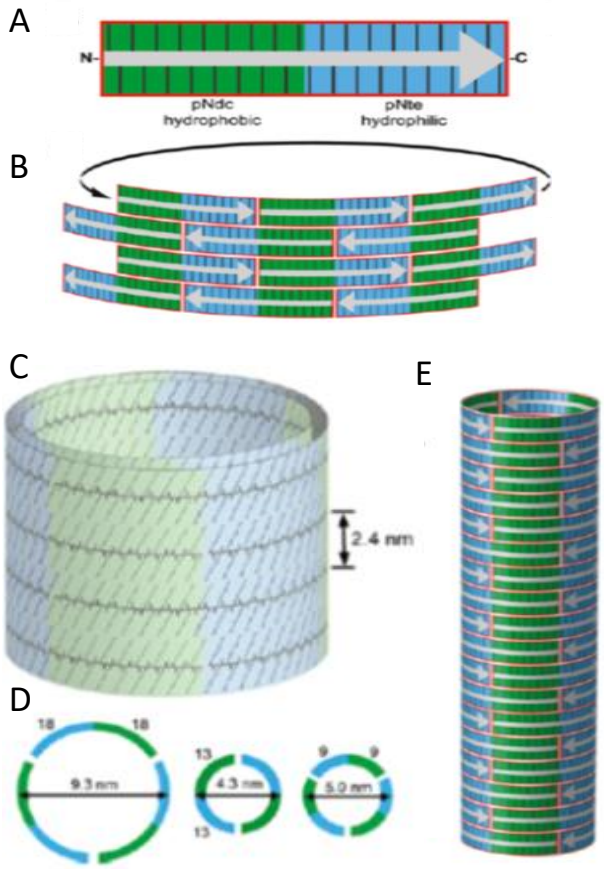


Figure 1.8 Structural schematic of the peptoid nanotubes. (A) organization of the peptoid diblock, with the hydrophobic pNdc block at the N-terminus and the pNte hydrophilic) block at the C-terminus. (B) The peptoid chains align anti-parallel, with the hydrophobic domains and hydrophilic domains stacking on top of themselves. (C) The nanotubes comprise many stacked rings, 2.4 nm in height, with the length of the peptoid chain determining the tube diameter (d). (e) Full side-on view of the peptoid nanotube, held together by longitudinal side-chain side-chain interactions. [52] **From Sun et al, 2016.**

nitrogen[48-49]. Among other features, peptoids are achiral, and do not rely on hydrogen bonding to hold together secondary structure, as is the case in peptides. Despite this, through appropriate side-chain selection, peptoids can form secondary structures[50-51], as well as super-secondary structures. The peptoid nanotubes developed by Sun et al. consist of a hydrophobic poly-N-decylglycine (pNdc) domain and a hydrophilic poly-N-2-(2-(2-methoxyethoxy)ethoxy)ethylglycine (pNte) domain which arrange into disks[52]. These nanotubes are unique from other diblock copolymer assemblies in that the hydrophobic core does not collapse to exclude water, rather, the hydrophobic domains from different peptoid chains stack on one another, aligning the strands antiparallel to one another, and forming nanotubes of discrete size (5-10 nm diameter). While many DNA and protein helical assemblies have been characterized, this is the first peptidomimetic assembly that forms a hollow nanotube. Amazingly, unlike their biological counterparts, the peptoid nanotubes do not rely on electrostatic interactions or hydrogen bonding networks to confer their rigidity or structural specificity. Rather, the structure is dictated by side-chain packing and van der Waals forces in between

1.4 Conclusion

Perhaps the most encouraging theme in the self-assembly of helical structures from macromolecules is the diversity in design. Through a few billions of years of evolution, nature has developed many ways to generate large aspect ratio structures from many copies of a single protein. Using these model systems, and with much tinkering, the scientific community has developed methodologies to mimic and even go beyond these natural structures. Finally, it has become possible to test the boundaries of biomacromolecular self-assembly by generating self-assembling biomimetic systems from non-biological building blocks. The development of x-ray

crystallography methods and ever-increasing high resolution cryo-EM technologies have allowed groups to solve the structure of these beautiful assemblies to atomic resolution. With precise atom placement comes the opportunity for specific modification for tailored applications. This introduction offers insight into peptide-specific helical biomolecules found in nature, surveys some current work in the field of helical self-assembling peptide systems, and delves into self-assembling systems that incorporate non-peptide subunits. The remainder of this thesis will follow the arc laid out in the introduction: first describing an established, published helical nanotubes project which lays a foundation for subsequent work into self-assembling helical nanotubes described herein. Following this, computationally designed self-assembling solenoidal nanotubes engineered via computational design will be described. Lastly, this thesis will describe work which pushes beyond peptide-specific work and incorporates DNA to create complex, multi-component nanowires.

1.5 References

1. Harris, J. R., Transmission electron microscopy in molecular structural biology: A historical survey. *Arch Biochem Biophys* **2015,** *581*, 3-18.

2. Chen, K. H.; Corro, K. A.; Le, S. P.; Nowick, J. S., X-ray Crystallographic Structure of a Giant Double-Walled Peptide Nanotube Formed by a Macrocyclic beta-Sheet Containing A beta(16-22). *J Am Chem Soc* **2017,** *139* (24), 8102-8105.

3. Costa, T. R. D.; Ilangovan, A.; Ukleja, M.; Redzej, A.; Santini, J. M.; Smith, T. K.; Egelman, E. H.; Waksman, G., Structure of the Bacterial Sex F Pilus Reveals an Assembly of a Stoichiometric Protein-Phospholipid Complex. *Cell* **2016,** *166* (6), 1436-+.

4. Anzini, P.; Xu, C. F.; Hughes, S.; Magnotti, E.; Jiang, T.; Hemmingsen, L.; Demeler, B.; Conticello, V. P., Controlling Self-Assembly of a Peptide-Based Material via Metal-Ion Induced Registry Shift. *J Am Chem Soc* **2013,** *135* (28), 10278-10281.

5. Watson, J. D.; Crick, F. H., Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **1953,** *171* (4356), 737-8.

6. Stupp, S. I.; Zha, R. H.; Palmer, L. C.; Cui, H.; Bitton, R., Self-Assembly of Biomolecular Soft Matter. *Faraday discussions* **2013,** *166*, 9-30.

7. Wierman, J. L.; Lan, T. Y.; Tate, M. W.; Philipp, H. T.; Elser, V.; Gruner, S. M., Protein crystal structure from non-oriented, single-axis sparse X-ray data. *IUCrJ* **2016,** *3* (Pt 1), 43-50.

8. Guo, F.; Jiang, W., Single particle cryo-electron microscopy and 3-D reconstruction of viruses. *Methods in molecular biology (Clifton, N.J.)* **2014,** *1117*, 401-43.

9. 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.

10. Jiang, T.; Xu, C. F.; Liu, Y.; Liu, Z.; Wall, J. S.; Zuo, X. B.; Lian, T. Q.; Salaita, K.; Ni, C. Y.; Pochan, D.; Conticello, V. P., Structurally Defined Nanoscale Sheets from Self-Assembly of Collagen-Mimetic Peptides. *J Am Chem Soc* **2014,** *136* (11), 4300-4308.

11. Jiang, T.; Vail, O. A.; Jiang, Z. G.; Zuo, X. B.; Conticello, V. P., Rational Design of Multilayer Collagen Nanosheets with Compositional and Structural Control. *J Am Chem Soc* **2015,** *137* (24), 7793-7802.

12. Wilson, D. N.; Doudna Cate, J. H., The Structure and Function of the Eukaryotic Ribosome. *Cold Spring Harbor Perspectives in Biology* **2012,** *4* (5).

13. Dominguez, R.; Holmes, K. C., Actin structure and function. *Annual review of biophysics* **2011,** *40*, 169-86.

14. Ramakrishnan, V., Histone structure and the organization of the nucleosome. *Annual review of biophysics and biomolecular structure* **1997,** *26*, 83-112.

15. Lapidus, L. J.; Yao, S.; McGarrity, K. S.; Hertzog, D. E.; Tubman, E.; Bakajin, O., Protein Hydrophobic Collapse and Early Folding Steps Observed in a Microfluidic Mixer. *Biophysical Journal* **2007,** *93* (1), 218-24.

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

17. Stanley, W. M., Chemical studies on the virus of tobacco mosaic III Rates of inactivation at different hydrogen-ion concentrations. *Phytopathology* **1935,** *25* (5), 475-492.

18. Butler, P. J. G., Self-assembly of tobacco mosaic virus: the role of an intermediate aggregate in generating both specificity and speed. *Philos T Roy Soc B* **1999,** *354* (1383), 537-550.

19. Fromm, S. A.; Bharat, T. A. M.; Jakobi, A. J.; Hagen, W. J. H.; Sachse, C., Seeing tobacco mosaic virus through direct electron detectors. *J Struct Biol* **2015,** *189* (2), 87-97.

20. Freeman, R.; Leonard, K. R., Comparative Mass Measurement of Biological Macromolecules by Scanning-Transmission Electron-Microscopy. *J Microsc-Oxford* **1981,** *122* (Jun), 275-286.

21. Harper, D. M.; Franco, E. L.; Wheeler, C.; Ferris, D. G.; Jenkins, D.; Schuind, A.; Zahaf, T.; Innis, B.; Naud, P.; De Carvalho, N. S.; Roteli-Martins, C. M.; Teixeira, J.; Blatter, M. M.; Korn, A. P.; Quint, W.; Dubin, G.; Study, G. H. V., Efficacy of a bivalent L1 virus-like particle vaccine in prevention of infection with human papillomavirus types 16 and 18 in young women: a randomised controlled trial. *Lancet* **2004,** *364* (9447), 1757-1765.

22. Villa, L. L.; Costa, R. L. R.; Petta, C. A.; Andrade, R. P.; Ault, K. A.; Giuliano, A. R.; Wheeler, C. M.; Koutsky, L. A.; Malm, C.; Lehtinen, M.; Skjeldestad, F. E.; Olsson, S. E.; Steinwall, M.; Brown, D. R.; Kurman, R.; Ronnett, B. M.; Stoler, M. H.; Ferenczy, A.; Harper, D. M.; Tamms, G. M.; Yu, J.; Lupinacci, L.; Railkar, R.; Taddeo, F. J.; Jansen, K. U.; Esser, M. T.; Sings, H. L.; Saah, A. J.; Lupinacci, L., Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial. *Lancet Oncol* **2005,** *6* (5), 271-278.

23. Bachmann, M. F.; Rohrer, U. H.; Kundig, T. M.; Burki, K.; Hengartner, H.; Zinkernagel, R. M., The Influence of Antigen Organization on B-Cell Responsiveness. *Science* **1993,** *262* (5138), 1448-1451.

24. Bachmann, M. F.; Zinkernagel, R. M.; Oxenius, A., Cutting edge commentary: Immune responses in the absence of costimulation: Viruses know the trick. *J Immunol* **1998,** *161* (11), 5791-5794.

25. Smith, M. L.; Lindbo, J. A.; Dillard-Telm, S.; Brosio, P. M.; Lasnik, A. B.; McCormick, A. A.; Nguyen, L. V.; Palmer, K. E., Modified Tobacco mosaic virus particles as scaffolds for display of protein antigens for vaccine applications. *Virology* **2006,** *348* (2), 475-488.

26. Pushparaj, V. L.; Shaijumon, M. M.; Kumar, A.; Murugesan, S.; Ci, L.; Vajtai, R.; Linhardt, R. J.; Nalamasu, O.; Ajayan, P. M., Flexible energy storage devices based on nanocomposite paper. *P Natl Acad Sci USA* **2007,** *104* (34), 13574-13577.

27. Zhang, Z. J.; Dewan, C.; Kothari, S.; Mitra, S.; Teeters, D., Carbon nanotube synthesis, characteristics, and microbattery applications. *Mat Sci Eng B-Solid* **2005,** *116* (3), 363-368.

28. Yuan, Y. F.; Tu, J. P.; Wu, H. M.; Li, Y.; Shi, D. Q., Size and morphology effects of ZnO anode nanomaterials for Zn/Ni secondary batteries. *Nanotechnology* **2005,** *16* (6), 803-808.

29. Nam, K. T.; Kim, D. W.; Yoo, P. J.; Chiang, C. Y.; Meethong, N.; Hammond, P. T.; Chiang, Y. M.; Belcher, A. M., Virus-enabled synthesis and assembly of nanowires for lithium ion battery electrodes. *Science* **2006,** *312* (5775), 885-888.

30. Lee, S. Y.; Royston, E.; Culver, J. N.; Harris, M. T., Improved metal cluster deposition on a genetically engineered tobacco mosaic virus template. *Nanotechnology* **2005,** *16* (7), S435-S441.

31. Royston, E.; Ghosh, A.; Kofinas, P.; Harris, M. T.; Culver, J. N., Self-assembly of virus-structured high surface area nanomaterials and their application as battery electrodes. *Langmuir* **2008,** *24* (3), 906-912.

32. Dujardin, E.; Peet, C.; Stubbs, G.; Culver, J. N.; Mann, S., Organization of metallic nanoparticles using tobacco mosaic virus templates. *Nano Lett* **2003,** *3* (3), 413-417.

33. Yi, H. M.; Nisar, S.; Lee, S. Y.; Powers, M. A.; Bentley, W. E.; Payne, G. F.; Ghodssi, R.; Rubloff, G. W.; Harris, M. T.; Culver, J. N., Patterned assembly of genetically modified viral nanotemplates via nucleic acid hybridization. *Nano Lett* **2005,** *5* (10), 1931-1936.

34. Gerasopoulos, K.; McCarthy, M.; Royston, E.; Culver, J. N.; Ghodssi, R., Nanostructured nickel electrodes using the Tobacco mosaic virus for microbattery applications. *J Micromech Microeng* **2008,** *18* (10).

35. Hospenthal, M. K.; Redzej, A.; Dodson, K.; Ukleja, M.; Frenz, B.; Rodrigues, C.; Hultgren, S. J.; DiMaio, F.; Egelman, E. H.; Waksman, G., Structure of a Chaperone-Usher Pilus Reveals the Molecular Basis of Rod Uncoiling. *Cell* **2016,** *164* (1-2), 269-278.

36. Oshea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T., X-Ray Structure of the Gcn4 Leucine Zipper, a 2-Stranded, Parallel Coiled Coil. *Science* **1991,** *254* (5031), 539-544.

37. Xu, C. F.; Liu, R.; Mehta, A. K.; Guerrero-Ferreira, R. C.; Wright, E. R.; Dunin-Horkawicz, S.; Morris, K.; Serpell, L. C.; Zuo, X. B.; Wall, J. S.; Conticello, V. P., Rational Design of Helical Nanotubes from Self-Assembly of Coiled-Coil Lock Washers. *J Am Chem Soc* **2013,** *135* (41), 15565-15578.

38. Svergun, D. I., Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J Appl Crystallogr* **1992,** *25*, 495-503.

39. Liang, Y.; Pingali, S. V.; Jogalekar, A. S.; Snyder, J. P.; Thiyagarajan, P.; Lynn, D. G., Cross-strand pairing and amyloid assembly. *Biochemistry-Us* **2008,** *47* (38), 10018-10026.

40. Lu, K.; Jacob, J.; Thiyagarajan, P.; Conticello, V. P.; Lynn, D. G., Exploiting amyloid fibril lamination for nanotube self-assembly. *J Am Chem Soc* **2003,** *125* (21), 6391-6393.

41. Johnson, L. M.; Gellman, S. H., alpha-Helix Mimicry with alpha/beta-Peptides. *Method Enzymol* **2013,** *523*, 407-429.

42. Collie, G. W.; Pulka-Ziach, K.; Lombardo, C. M.; Fremaux, J.; Rosu, F.; Decossas, M.; Mauran, L.; Lambert, O.; Gabelica, V.; Mackereth, C. D.; Guichard, G., Shaping quaternary

assemblies of water-soluble non-peptide helical foldamers by sequence manipulation. *Nat Chem* **2015,** *7* (11), 871-8.

43. Yadav, M. K.; Redman, J. E.; Leman, L. J.; Alvarez-Gutierrez, J. M.; Zhang, Y. M.; Stout, C. D.; Ghadiri, M. R., Structure-based engineering of internal cavities in coiled-coil peptides. *Biochemistry-Us* **2005,** *44* (28), 9723-9732.

44. Liu, R. Y.; Loll, P. J.; Eckenhoff, R. G., Structural basis for high-affinity volatile anesthetic binding in a natural 4-helix bundle protein. *Faseb J* **2005,** *19* (6), 567-576.

45. Ghirlanda, G.; Hilcove, S. A.; Pidikiti, R.; Johansson, J. S.; Lear, J. D.; DeGrado, W. F.; Eckenhoff, R. G., Volatile anesthetic modulation of oligomerization equilibria in a hexameric model peptide. *Febs Lett* **2004,** *578* (1-2), 140-144.

46. Joh, N. H.; Wang, T.; Bhate, M. P.; Acharya, R.; Wu, Y. B.; Grabe, M.; Hong, M.; Grigoryan, G.; DeGrado, W. F., De novo design of a transmembrane Zn2+-transporting four-helix bundle. *Science* **2014,** *346* (6216), 1520-1524.

47. Drexler, K. E., Peptoids at the 7th Summit: Toward Macromolecular Systems Engineering. *Biopolymers* **2011,** *96* (5), 537-544.

48. Zuckermann, R. N.; Kerr, J. M.; Kent, S. B. H.; Moos, W. H., Efficient Method for the Preparation of Peptoids [Oligo(N-Substituted Glycines)] by Submonomer Solid-Phase Synthesis. *J Am Chem Soc* **1992,** *114* (26), 10646-10647.

49. Culf, A. S.; Ouellette, R. J., Solid-Phase Synthesis of N-Substituted Glycine Oligomers (alpha-Peptoids) and Derivatives. *Molecules* **2010,** *15* (8), 5282-5335.

50. Kirshenbaum, K.; Barron, A. E.; Goldsmith, R. A.; Armand, P.; Bradley, E. K.; Truong, K. T. V.; Dill, K. A.; Cohen, F. E.; Zuckermann, R. N., Sequence-specific polypeptoids: A diverse

family of heteropolymers with stable secondary structure. *P Natl Acad Sci USA* **1998,** *95* (8), 4303-4308.

51. Stringer, J. R.; Crapster, J. A.; Guzei, I. A.; Blackwell, H. E., Extraordinarily Robust Polyproline Type I Peptoid Helices Generated via the Incorporation of alpha-Chiral Aromatic N-1-Naphthylethyl Side Chains. *J Am Chem Soc* **2011,** *133* (39), 15559-15567.

52. Sun, J.; Jiang, X.; Lund, R.; Downing, K. H.; Balsara, N. P.; Zuckermann, R. N., Self-assembly of crystalline nanotubes from monodisperse amphiphilic diblock copolypeptoid tiles. *P Natl Acad Sci USA* **2016,** *113* (15), 3954-3959.

Chapter II. Form I and Form II: A Foundation for Production of Helical Nanotubes

2.1 Introduction

Native helical assemblies (e.g., filamentous phage capsids,[1] flagellar filaments,[2] type IV pili,[3] and the type III secretion system needle[4]) envelope many diverse functions that would be desirable to mimic in synthetic protein- and peptide-based assemblies. This diversity of functions includes dynamic switching, locomotion, controlled release, directional transport, and selective catalysis. However, current limitations in the ability to reliably predict the relationship between sequence and resultant supramolecular structure present a substantial challenge to the *de novo* design of novel self-assembled protein architectures. At the time of publication of this project, the prediction of structure from sequence remained rather limited due to current capabilities of computational structural prediction. This phenomenon is due in part to the fact that, while the tertiary structure of proteins is robust with respect to many changes in amino acid sequence,[5] the quaternary structure of proteins can be very sensitive to slight variations in sequence and to slight changes in environment. Examples exist such as hemoglobin[6] or bacterial flagellar filaments[7] in which conserved protein folds can assemble into significantly different higher-order forms. Model systems, such as coiled-coil a-helical bundles of the Rop protein, have also shown an unanticipated degree of structural plasticity.[8] Since quaternary structures arise from interactions between protomers, these observations raise a significant question regarding the designability of the protein-protein interfaces within biological assemblies.[9] This issue of designability will remain a core issue throughout this thesis.

We designed two peptides, Form I and Form II, based on a coiled-coil structural motif, to self-assemble into extended tubular assemblies on the nanoscale. While coiled coils have been postulated as potential substrates for the creation of cylindrical assemblies,[10] the corresponding

formation of synthetic, large diameter nanotube assemblies has never successfully been accomplished. We demonstrate that these short synthetic peptides can self-assemble into high-aspect-ratio fibrillar structures that correspond to helical nano- tube assemblies. Using electron cryomicroscopy (cryo-EM) with a newly available direct electron detection capability[11-17] we can achieve near-atomic resolution of these thin structures when imaging the equivalent of a single tube that is ~17 mm long containing only 80,000 molecules. We show that the structure of these nanotubes can be changed dramatically with only one or two amino acid changes. This peptide system provides direct insight into the designability of extended assemblies based on biologically derived structural motifs, which has potential significance for the de novo design of synthetic peptide-based nanomaterials in the future.

2.2 Sequence Design

The sequences for forms I and II, as published in *Cell* [18], were designed using the structural model outlined by Walshaw and Woolfson [10]. This model postulates the helix-helix packing for bifaceted coiled-coils, and has been used to explain the packing of the constituent alpha-helices within protein tolC (a subunit of a pump system found in gram negative bacteria). It was hypothesized that the Forms I and II form alpha-helices, and these helices would subsequently pack in a fashion concurrent with the proposed structural model, wherein the helices would self-associate at the hydrophobic faces (see Figure 2.1) in an antitypic fashion. Hydrophobicity is provided by leucine/isoleucine residues populating the *c/d* positions of the canonical alpha-helical heptad repeat, as well as alanine residues at the *a/f* positions in order to enlarge the surface area of each hydrophobic stripe. The employment of alanine residues at these peripheral positions has been proven to encourage higher order coiled-coil oligomers [19-21]. Strategically placing the non-

polar residues at these heptad positions means the two hydrophobic faces of each helix are offset by 154 degrees, leading to a putative method of assembly in which alpha-helices aggregate parallel to the cylinder axis, driven by hydrophobic collapse. Charged residues are placed at the *b/e* positions in order to select against potential, unwanted beta-sheet formation. Form I differs from Form II by four residues. The positive charges in Form I are supplied by arginine residues, whereas the positive charges in Form II are in the form of lysine residues. The remainder of the two sequences is kept identical. The helical wheel representations and the linear amino acid sequences of both can be viewed in the figure. Both are chemically capped at their termini to discourage end-to-end association of helices (acetylated N-terminus and amidated N-terminus).

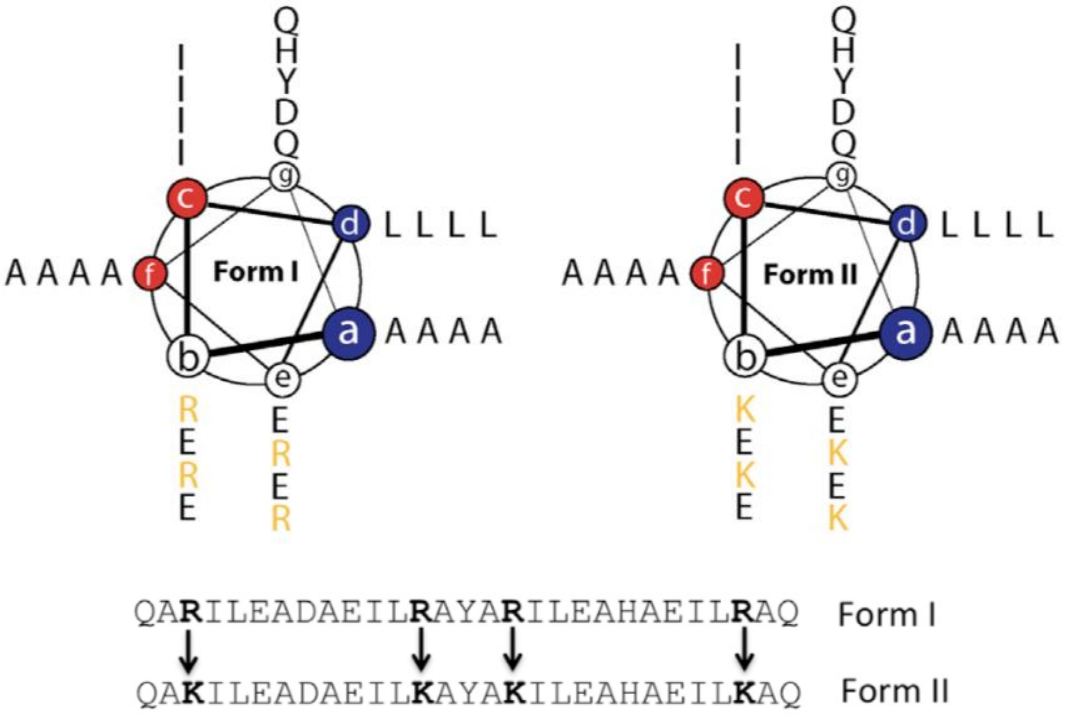


Figure 2.1: (top) helical wheel representations of Form I and Form II and (bottom) linear amino acid sequences of Form I and Form II.

2.3 Results and Discussion

2.3.1 Circular Dichroism

Confirmation of secondary structure is critical, as the self-assembly of peptide into nanotubes is dependent upon successful folding into alpha-helical character. When probed via circular dichroism, both Forms I and II indicate alpha-helical secondary character, as is seen via the minima in both traces at 208 nm and 222 nm. The CD traces are presented in units of mean residue ellipticity (MRE) such that the two spectra can be compared against one another. Peptides were originally solubilized at a concentration of 3 mg/mL, as this was shown to be an optimal concentration for the promotion of nanotube self-assembly, but were diluted to 100 µM for circular dichroism studies.

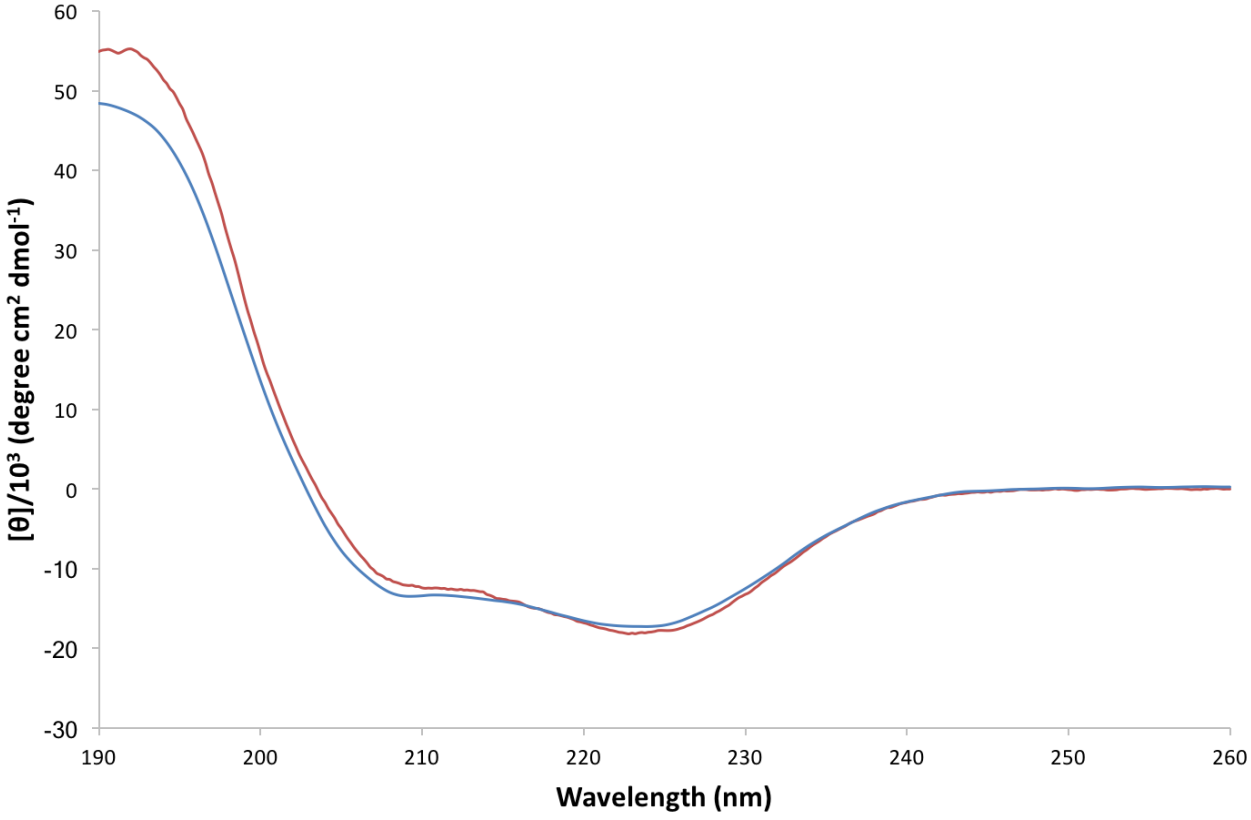


Figure 2.2: circular dichroism spectra for (red) form I and (blue) form II assembled at 100 µM peptide concentration in 10 mM acetate buffer at pH 4.0

2.3.2 Transmission Electron Microscopy

Transmission electron microscopy experiments confirm both forms self-assemble into high aspect-ratio helical nanotubes under the aforementioned assembly conditions (3 mg/mL peptide concentration in 10 mM acetate buffer at pH 4.0). Importantly, this data represents the first instance of self-assembling, large diameter (≥ 5 nm), *de novo* tubular assemblies based on the

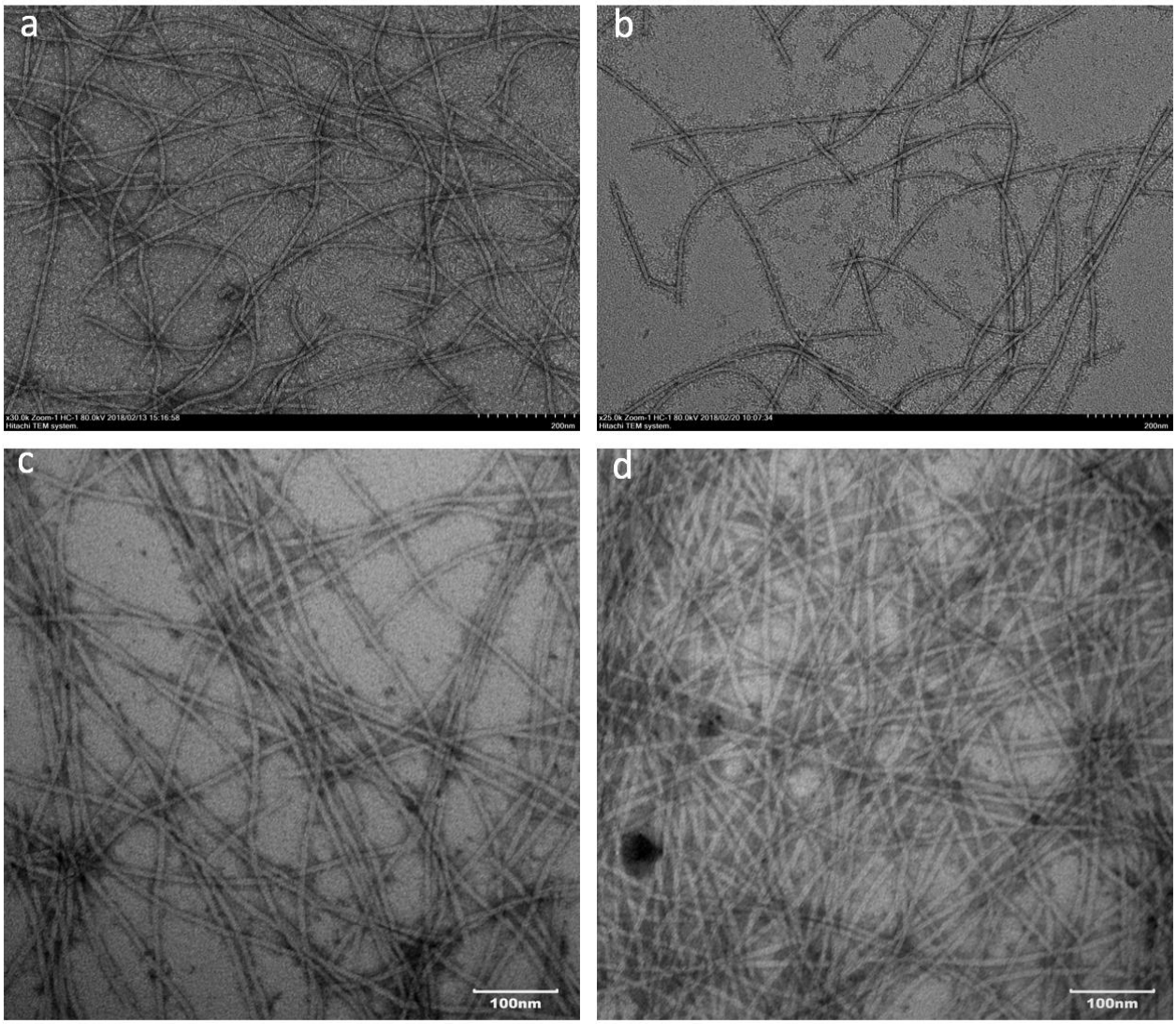


Figure 2.3: negative stain transmission electron microscopy of Forms I and II. (A, B) form I nanotubes, wherein the cross-sectional fiber diameter is approximated at 60 angstroms. (C, D) form II nanotubes, wherein cross-sectional fiber diameter is approximated from EM at 100 angstroms.

coiled-coil motif. All micrographs were produced using negative stain, in this case 1% uranyl acetate solution. Initial estimates of tube diameters were approximated at 6 nm for form I and 10 nm for form II, though this value would later be corrected to 12 nm with finer analysis via cryo-electron microscopy.

2.3.3 Small- and Wide-Angle X-Ray Scattering:

Small- and wide-angle x-ray scattering (SAXS) experiments provide further confirmation of structural details for both forms. Synchotron small-angle X-ray scattering measurements were performed on both assemblies. The scattering intensity profiles were measured as a function of the momentum transfer, *q*, within the range of .005 $Å^{-1}$ to 2.3 $Å^{-1}$. Both forms demonstrated a scattering intensity profile with a linear dependence on q^{-1} within the power law region. The mass fractal value for both of the Forms was measured to be ~-1, which is consistent with extended rod-like forms. The resultant values for the cross-sectional radii of gyration, or R_c values, were calculated to be 28.8 Å for Form I and 40.7 Å for Form II. These R_c values correlate to diameters of 8.12 nm and 11.5 nm for Form I and Form respectively. The resultant diameter from Guinier analysis for Form II matches reasonably well with the diameter observed from TEM analysis. However, Form I deviates from what is observed via TEM, with a Guinier calculated diameter value roughly 21 Å larger than the diameter from electron microscopy. It is hypothesized that a concentration-dependent conversion between a monolayer and bilayer tube structure is occurring for Form I. This is also later observed in the data for a mutant based off the sequences of both Form I and Form II. At lower concentrations, such as in TEM and STEM, there is insufficient peptide to form a bilayer tube, and Form I assembles as a four-start helical tube with a single alpha-helix as the subunit in the alpha-stack. However, as SAXS necessitates higher peptide

concentrations for reliable data, there could be sufficient peptide in solution for the formation of a bilayer tube, with a dimer of alpha-helices serving as the subunit in the alpha-helix stack. In contrast, the equivalent data for Form II do not demonstrate any change when comparing low concentration analyses to higher concentration analysis.

The scattering intensity profiles displayed a series of oscillations on *q* that are indicative of hollow cylindrical structures and the position of the first minima at these oscillations is measured. The first oscillation occurs at a *q* value of 0.106 $Å^{-1}$ for Form I and 0.065 $Å^{-1}$ for Form II**.** These values correspond to 59.3 Å and 96.6 Å respectively, both of which fall ~20 Å short of the Guinier-derived diameter values. This could indicate that the formation of a bilayer tube affects the position

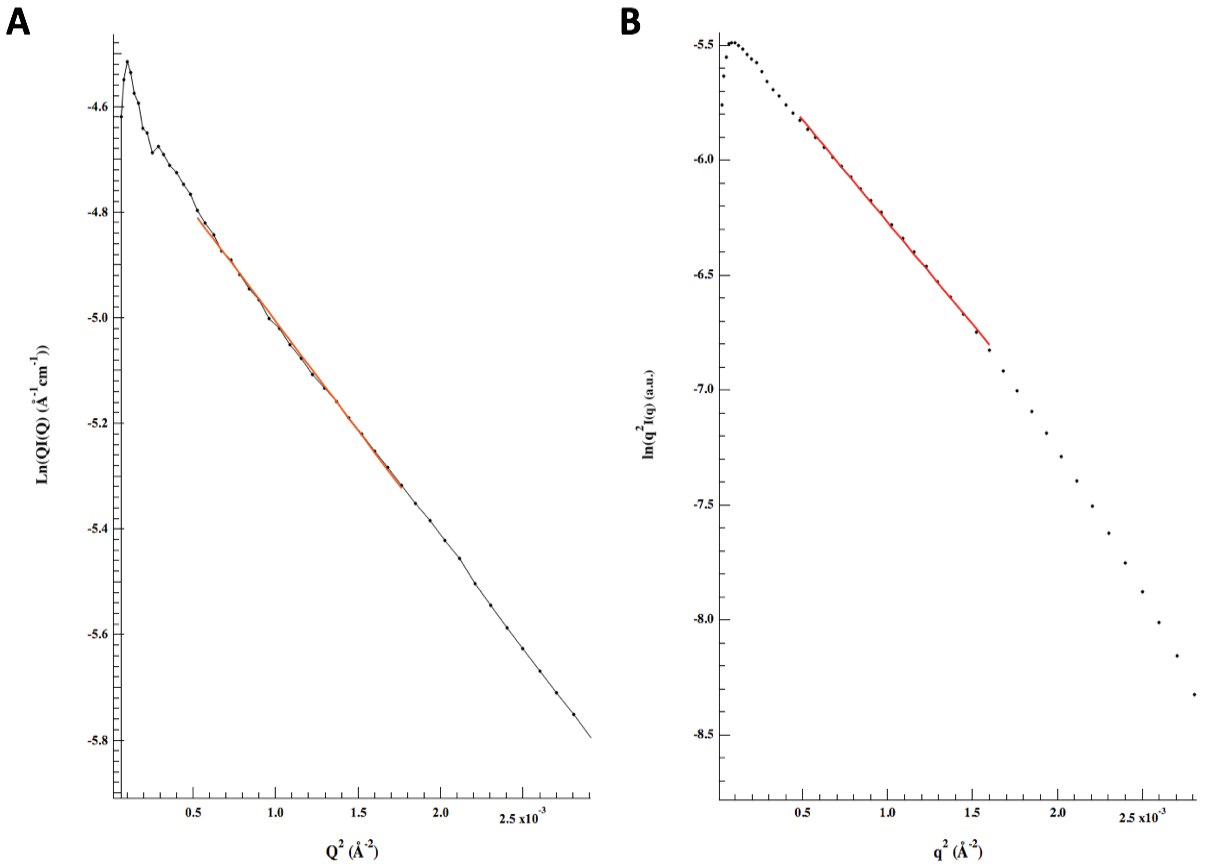


Figure 2.4: Modified Guinier plot, ln(q^2I(q)) vs q for Form I (A) and Form II (B) with the region used for Guinier derived tube diameters indicated in red.

of the first minima. If the Form I tubes were not forming bilayer tubes at high concentrations required for SAXS, then it is hypothesized that the values from the Guinier analysis and the first minima would be equivalent for Form I. The fact that we are seeing a similar difference in values consistent with what is observed between the Guinier analysis and first minima for Form II may intimate that Form I is forming a bilayer tube at sufficient peptide concentration. This is important

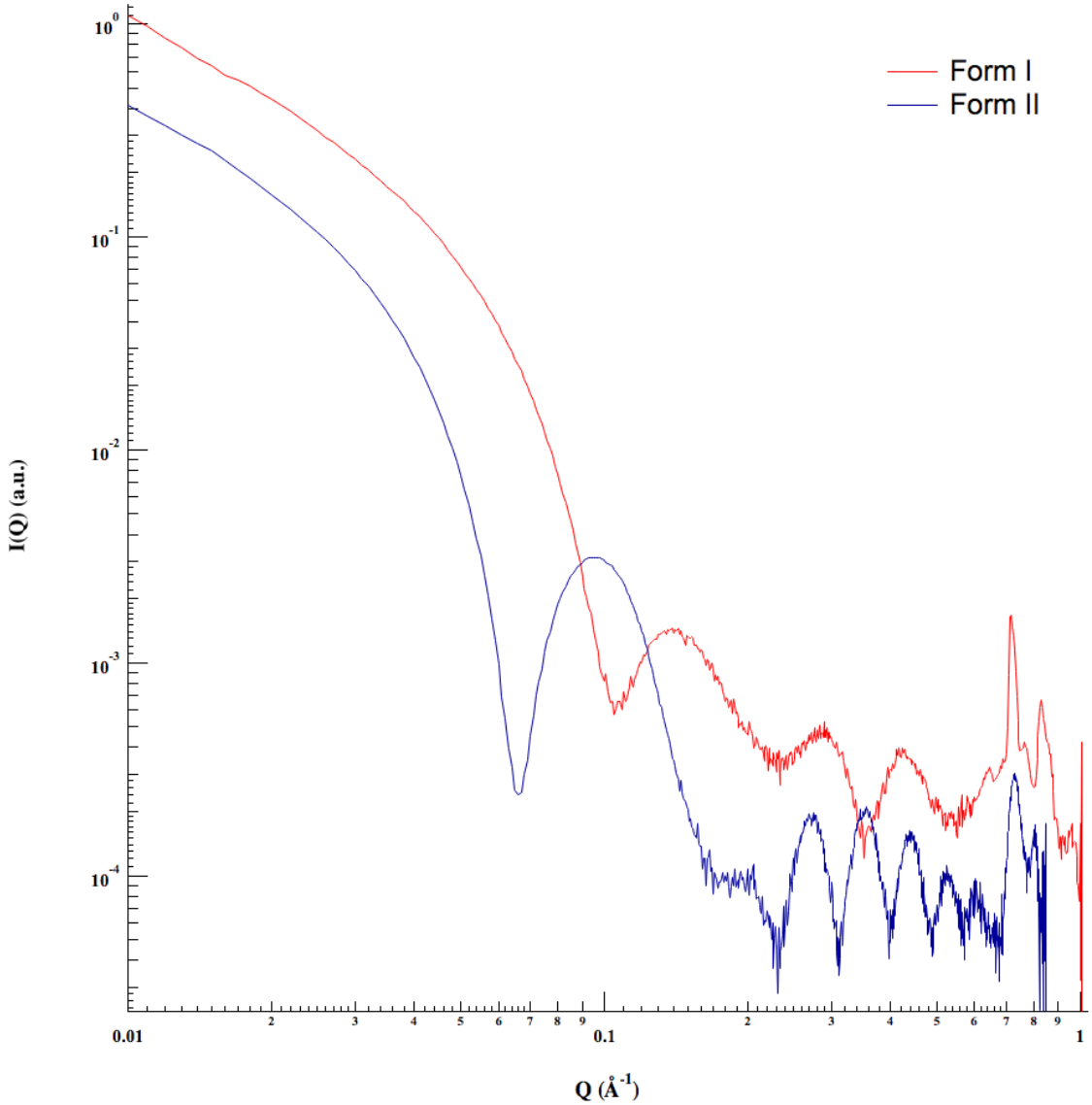


Figure 2.5 Small-angle X-ray scattering curve for solutions of Form I (red) and Form II (blue) at > 3 mg mL^{-1} concentrations in 10mM acetate, pH 4.0.

both for the analysis of Form I and Form II, but will also prove to be critical in the evaluation of subsequent Form mutants based off of the original two Form sequences and will be discussed later.

2.3.4 Cryo-Electron Microscopy and Helical Reconstruction

From cryo-electron microscopy, coupled with direct electron detection capability, atomic level structural models of both Forms I and II were produced. These models are critical for both the understanding of the heavy influence of sequence-space upon resultant tube morphology, and for rationalizing and engineering future mutations (which will be discussed in coming sections). An intimate understanding of the packing of helices into Form I and II nanotubes allows for a well-informed, highly strategic approach to designing new sequences which use the sequence identities of the original two forms as starting points. Therefore, the structural models elucidated by Egelman and colleagues (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*

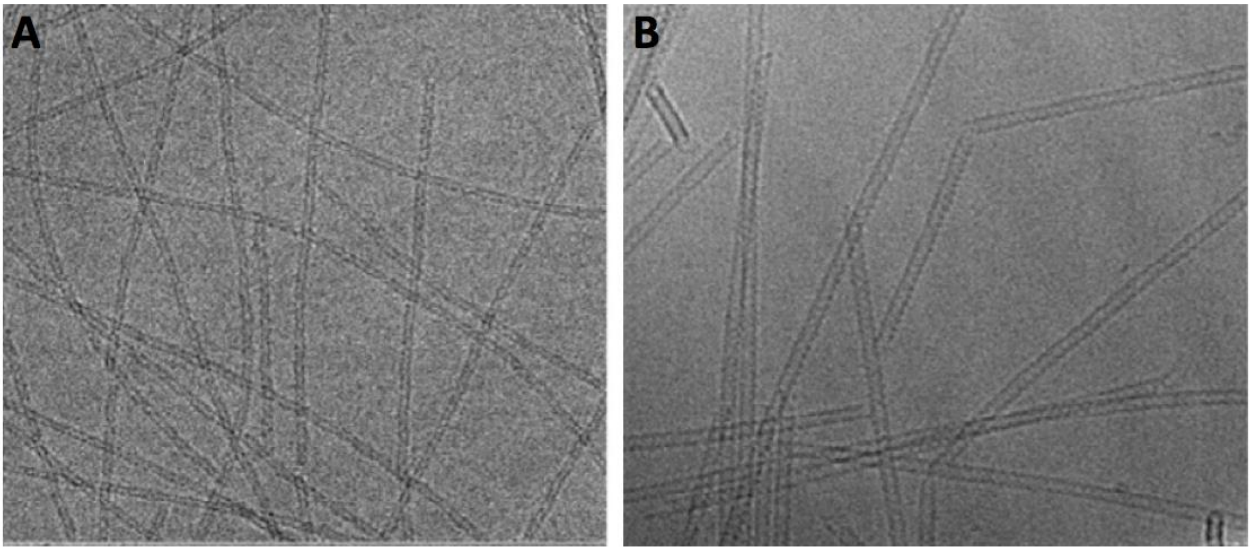


Figure 2.6: Cryo-electron micrographs for the ~6 nm Form I filament (A) and the ~12 nm Form II filament.

Assemblies. Structure, 2015. **23**(2): p. 280-289) will be described hereafter in detail in order to provide a contextual foundation for more current sequences in the Form family.

Cryo-electron microscopy is a linchpin technique in the structural determination of fibrillar biological materials. Importantly, cryogenic preparation of samples balances the capacity for high-resolution structural data with the ability to visualize samples in native or near-native conditions [11-12, 15-16, 22-24]. Increasingly, electron microscopy is being used as a means of structural resolution for helical fibers. This is due in part to the inability of traditional means of high-resolution structural determination, namely X-Ray scattering, are innefective techniques for probing filamentous structures, as fibers and tubes *typically* do not pack with high-order crystalline symmetry [25-26]. Current cyro-EM methodologies employ direction electron detectors to achieve near atomic resolution, allowing for highly detailed forays into the fine structure of ordered biological materials.

Egelman et al. have used cryo-electron microscopy data as a means to fully solving the helical structures at sub-nanometer resolution. Direct electron detection enables near-atomic resolution of the two forms when imaging a section of tube approximately 17 micrometers in length comprising 80,000 molecules [18]. Determinatino of helical symmetry in both forms was initiated via indexing the power spectra. Bessel orders are assigned to a layer line, which is a straightforward and robust process in the case of an array of atoms populating exclusively the same radius. However this method of indexing runs into substantial difficulties when applied to filamentous structures featuring a distribution of density across a radial range. As a result, this aforementioned indexing methodology produces ambiguities, wherein a multiplicity of helical symmetries will all yield an identical power spectrum. Given that both forms I and II are based upon the coiled-coil motif, a singular radial distance is unachievable. Therefore, it is assumed that

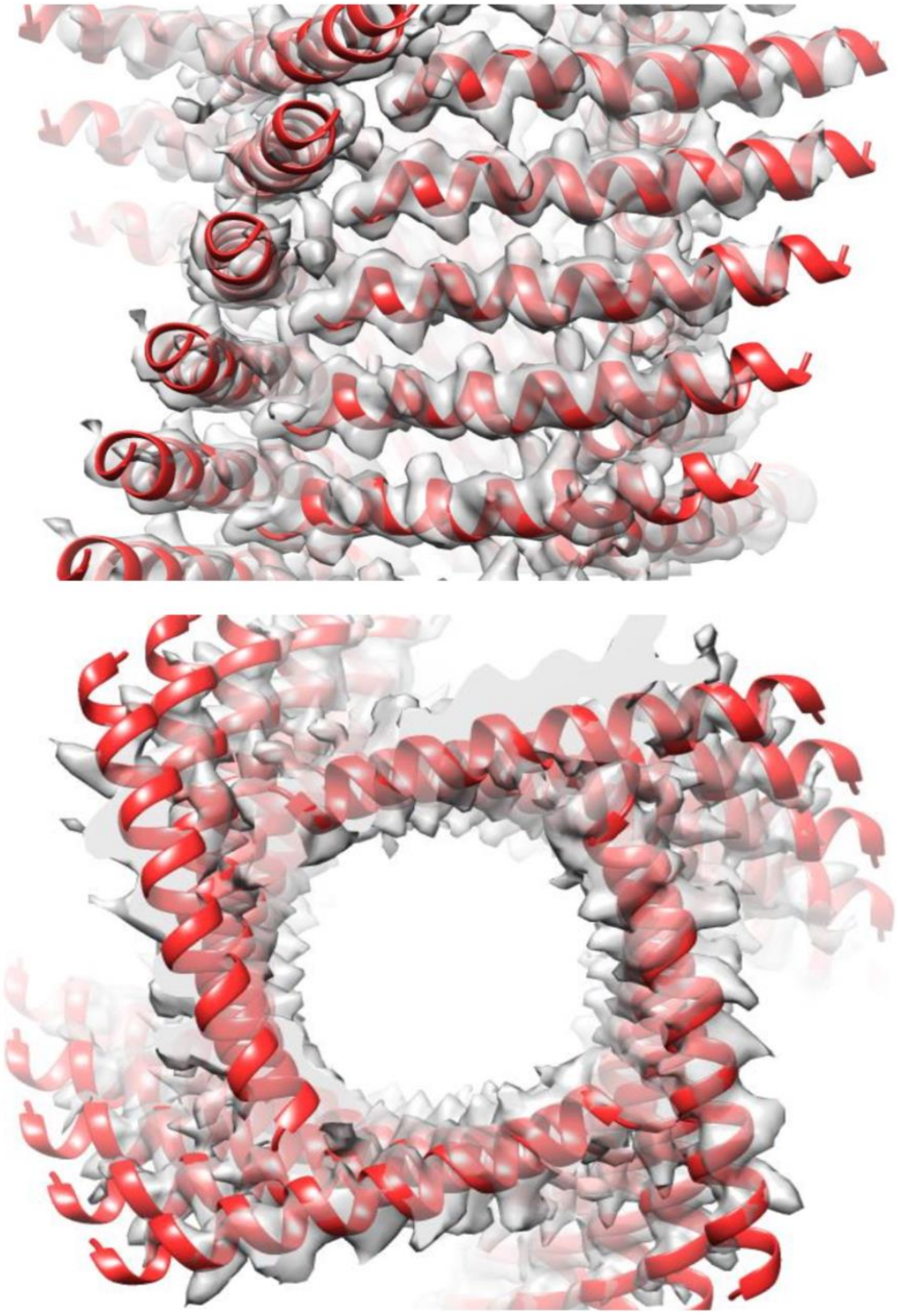


Figure 2.7: Side and top view of cryo-EM structure of the Form I nanotube at circa 3.6 Å resolution. The density map shows that the wall of the nanotube is composed of a single layer of helices, distinct from the structure of Barrel3CLys nanotubes. From Egelman et al. 2013.

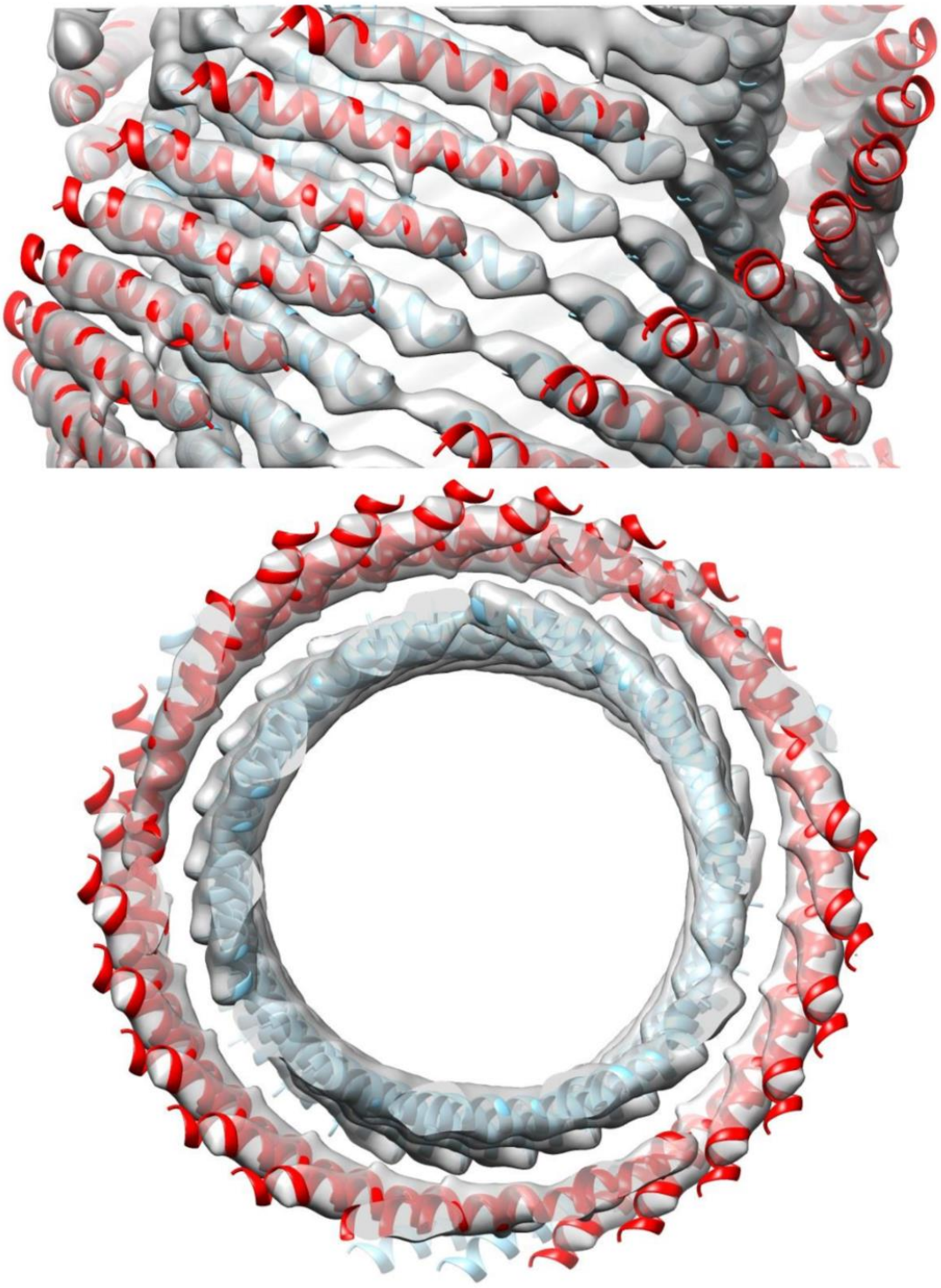


Figure 2.8: Side and top view of cryo-EM structure of the Form II nanotube at circa 5-6 Å resolution. The density map shows that the wall of the nanotube is composed of a bilayer of helices, distinct from the structure of Barrel3CLys nanotubes. From Egelman et al. 2013.

a multiplicity of potential structural identities exist for both forms, rendering structural determination via power spectra analysis unreliable. The iterative helical real space reconstruction method, established by Egelman and colleagues at the University of Virginia, provides a better determination of helical symmetries [26].

Forms I and II were both analyzed via cryoelectron microscopy and subsequent data was subjected to reconstruction via IHRSR. Form I demonstrated a discerinble central lumen, leading credence to the central hypothesis that the previously described coiled-coil forming peptides can

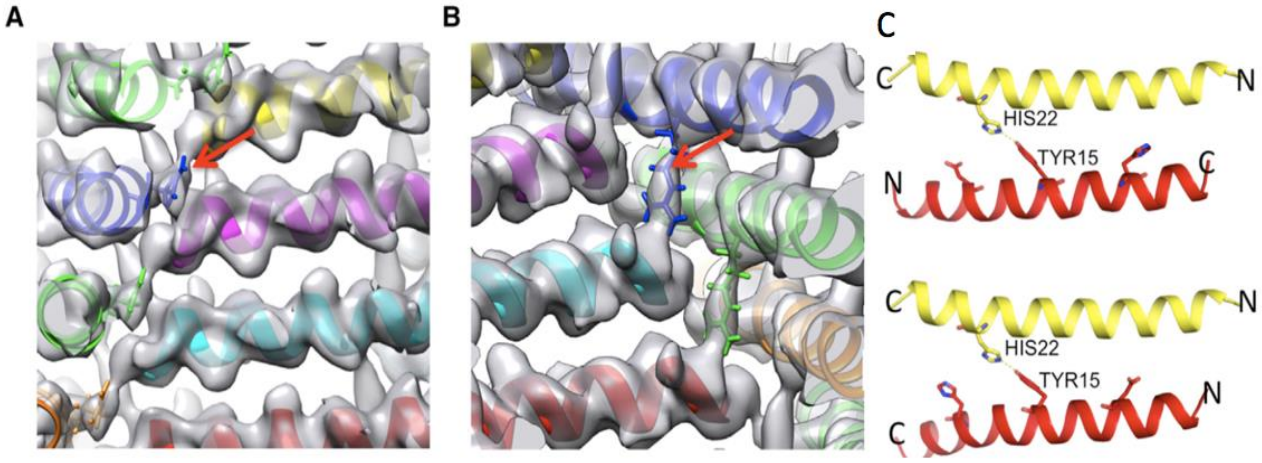


Figure 2.9: 3D models from the near-atomic resolution cryo-electron microscopy data indicating (A) Arg13 in the Form I filament demonstrates the most highly-ordered side chain, which is packed between two other subunits at each corner of the square, as seen in the exterior view of the filament. The side chain is indicated in blue and the two stabilizing subunits are shown in magenta and yellow. (B) Arg17 terminally caps the adjacent alpha-helix at the c-terminus, thereby stabilizing the interior corners of the square. (C) The lower resolution, coupled with the nearly palindromic character of the Form II sequence leads to uncertainty into whether the helices pack parallel or anti-parallel. In both models, His22 (yellow) interacts with Tyr15 (red).

self-assemble into nanotubes. While both forms exhibit random variability in twist, it was observed that form I tubes contain a single alpha-helix in the assymetric unit, while form II tubes contain two alpha-helices per symmetric unit. After sorting twist, a subset o fdata were used to generate three dimensional reconstructions for Form I (resolved at approximately 3.6 angstroms) and Form II (resolved at approximately 5-6 angstroms).

The Form I reconstruction derives from 62,122 overlapping segments correlating to an approximated 200,000 unique copies of the peptide. The *subunit*s of form I arrange as a one-start helix. One-start helices are helices which pass through every subunit [27]. Subseqently, two-start helices feature two strands that both pass through every other subunit, three-start helices pass through every third subunit and so on. Within form I, the *subunits* are arranged in a left-handed one-start helix with ~4.1 subunits per turn. Given this, the full form I structure can be described as four stacks of alpha-helical subunits, wherein each stack forms a four-start helix. This yields an almostly perfectly square cross-sectional geometry. This geometry displays an axial rise per subunit of 2.2 angstroms.

2.3.5 Power Spectra and Reference-Free Class Averages

From cryo-electron microscopy data, averaged power spectra and reference-free class averages were prodcued. The layer line bessel orders are n =4 for Form I and n=3 and n=6 for Form II. The distance to the first layer line for Form I is measured at ~1/72 $Å^{-1}$. This value is consistent with a four-start helix. Reference-free class averages were calculated from boxed segments of both form filaments from cryo-electron micrographs. These are shown and indicate the overall geometry and symmetry of the filaments. Both the power spectra and the reference-free class averages are useful for comparison against later peptide sequences based on the original Form

sequences. Form II demonstrates approximately 3.1 subunnits per turn with an axial rise per assymetric unit (two helices per unit) of 2.1 angstroms. Unique from form I, the form II cross-section is approximately circular, as the subunits pack at a larger radius and exhibit intramolecular curvature. Laslty, form II fibers comrpise two discrete helical layers, an inner and outer core. Each of these layers is formed from stacks of alpha-helical subunits forming three-start helices.

From the high resolution structural data of both supramolecular structures, several key interactions can be identified, and therefore manipulated to create new mutants based off of the original two Form sequences. These interactions are shown in Figure 2.9. Within the Form I filament, Arg13 is the most highly-ordered residue within the helix. This residue is sandwiched between two adjacent subunits at each corner junction of the square stack of alpha-helices. Further, Arg17 proves to be a critical residue. Arg17 provides a C-terminal cap for the adjacent alpha-helix, subsequently providing stability for the inside junction of the four-start stack of alpha-helices. His22 and Tyr15 demonstrate an interaction with one another in the Form II filament. Due to the semi-palindromic nature of the Form I sequence and the locations of the aforementioned residues within the alpha-helix, it cannot be determined whether the helices pack in parallel or anti-parallel.

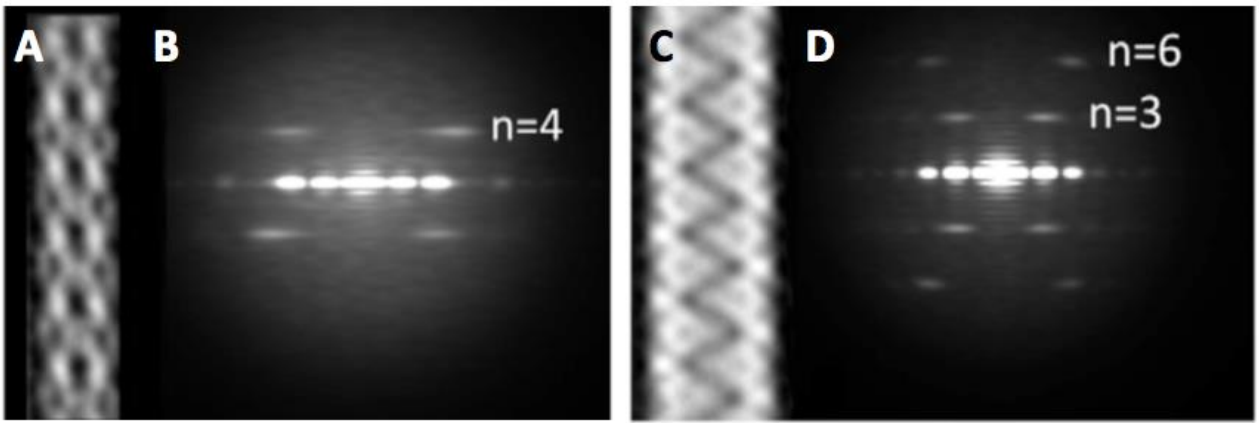


Figure 2.10: Reference-free class averages for (A) Form I and (C) Form II. Averaged power spectra, with layer line Besssel orders indicated by n, for (B) Form I and (D) Form II.

2.4 Conclusion

Detailed structural resolution and characterization of peptide-based supramolecular assemblies at atomic or near-atomic resolution has historically presented significant technical hurdles. This has hindered the development of knowledge underlying the principles that govern the interactions between subunits that stabilize self-assembling macromolecular structures. Though crystallography has traditionally been one of the primary tools in for the generation of atomic resolution detail of said assemblies, it is hindered by its requirement to have exactly two, three, four or six subunits per turn, otherwise it cannot be crystallized with every subunit in an equivalent environment. Recent advancements in imaging hardware, reconstruction algorithms, and computational processes for structural refinement enable near-atomic resolution structures of synthetic helical assemblies. Despite these improvements in the aforementioned fields, the structural analysis of Form I and Form II highlights a key challenge in the *de novo* design of structurally defined supramolecular peptide materials. Form I and II once again demonstrate that though tertiary peptide structure may be robust with respect to sequence, quaternary structure remains quite labile with respect to comparatively minute fluctuations within sequence-space, especially for residues located at the peptide-peptide or protein-protein interface. The manner of interactions between helices stacked upon each other deviated significantly from the hypothesized manner of helix packing. However, this data serves as an important foundation for later sequences which are based off of the sequence and resultant supramolecular structures of Form I and Form II. Moreover, as greater numbers of atomic-level structurally resolved supramolecular peptide assemblies become available, more information becomes available to inform and improve the reliability of designing novel peptide nanomaterials with defined structure and controllable function. Lastly, these nanotubes have great potential to serve as scaffolding for future applications

such as small molecule encapsulation within the lumen (for potential drug delivery) or functionalization at the outer surface area. To this end, Collier and colleagues have employed Form I as a platform for a vaccine system in which the Form I tubes are labelled with epitopes for CD4+ T-cell and CD8+ T-cell responses in mice. Multiple epitopes were used, including a cancer B-cell epitope from the epidermal growth factor receptor class III (EGFRvIII), a universal CD4+ T-cell epitope PADRE, and the model CD8+ T-cell epitope SIINFEKL. All variants were successfully incorporated in a modular manner into supramolecular multi-epitope nanofibers using the Form I tube as a scaffold.[28] Lastly, the identified RxxxR structural motif will serve as a lynchpin structural component in the design of novel helical nanotubes in the next chapter.

2.5 Methods

2.5.1 Peptide Synthesis

Peptides were either purchased from GenScript USA (Piscataway, NJ) or synthesized in-house. In the latter case, peptides were prepared as the capped (N-acetyl, C-amide) derivatives using microwave-assisted synthesis on a CEM Liberty solid-phase peptide synthesis instrument using PAL-PEG-PS resin from Applied Biosystems. Standard Fmoc protection chemistry was used with coupling cycles based on HBTU/DIEA-mediated activation proto- cols and base-induced deprotection (20% piperidine in *N*,*N*-dimethylformamide with 0.1 M hydroxybenzotriazole) of the Fmoc group. The peptides were purified via reversed-phase high-pressure liquid chromatography (HPLC) on a C18 column with a gradient of water-acetonitrile (0.1% trifluoroacetic acid). The purity was assessed to be above 95% by analytical HPLC (Figure S1). Peptide mass was confirmed using MALDI mass spectrometry (Figure S2). The peptides were lyophilized, sealed, and stored at –20 C.

For analytical measurements, peptides were dissolved in acetate buffer (10 mM, pH 4.0). Peptide concentration was determined spectrophotometrically from measurement of the absorbance at 280 nm (*A*280). For peptides containing Tyr, Trp, or Cys residues, the peptide concentration can be calculated from Equation 1:

$$\text{MW } 3\ A280 = 1280nY + 5690nW + 120nC\ c \qquad \text{(Equation 1)}$$

in which *c* is the concentration of peptide in mg/ml, and *n*Y, *n*W, and *n*C are the numbers of tyrosine, tryptophan, and cysteine residues, respectively, in the peptide sequence (Gill and von Hippel, 1989). Since the peptide derivatives contain only a single tyrosine residue per molecule, then *c* = MW 3 *A*280/ 1280. To eliminate error in determination of absorbance that could arise as a result of UV light scattering due to peptide self-assembly, aqueous solutions of peptide were mixed with 6 M guanidinium chloride in 1:9 v/v ratio and seated at room temperature for 2 hours to completely denature the sample prior to per- forming the absorbance measurements. Samples for EM were heated to 95 C in acetate buffer (10 mM, pH 4.0) and slowly cooled to 4 C to promote assembly of the nanotubes. Analytical HPLC and ESI-mass spectrometry were employed to confirm the purity and mass of the peptides, and can be found in the supplemental information of Egelman et al. 2015.

2.5.2 EM and Image Analysis

Samples (2.5 ml) were applied to lacey carbon grids and vitrified in a Vitrobot Mark IV (FEI). They were imaged in a Titan Krios (FEI) at 300 keV using a K2 Summit direct electron detector (Gatan), with a magnification yielding 1.02 A˚/pixel. Images were collected in the movie mode,

with a total exposure of 50 e /A˚2. The full integrated exposure was used for determination of the contrast transfer function (CTF) using CTFFIND3 (Mindell and Grigorieff, 2003) as well as for boxing filaments using the e2helixboxer routine in EMAN2 (Tang et al., 2007). The defocus range used was from 1.4 to 4.4 mm. Overlapping segments that were 400 pixels long were then cut from these filaments, using a shift of 7 pixels (98% overlap), which was possible due to the small axial rise of the subunits (~2 A˚). The SPIDER software package (Frank et al., 1996) was used for most image processing steps, including the implementation of the IHRSR algorithm (Egelman, 2000). For both the form I and the form II filaments, the correct helical symmetry was able to be determined only by a brute-force trial-and-error approach, as the information contained in the power spectra (Movies S1 and S2) led to a multiplicity of possible solutions, only one of which would be correct. Only one solution for each of the two forms generated recognizable alpha-helices, establishing that this was the correct solution.

For the form I filaments, a total of 113,581 segments were extracted from 260 micrographs. After sorting by the pitch of the four-start helices (Figure 3A), 24,111 segments were used for a 3D reconstruction. Initially, frames totaling a dose of 30 e /A˚2 were used in the IHRSR procedure. For the last cycle, however, the alignment parameters (shifts and Euler angle assignments) from the higher-dose frames were applied to initial frames with a dose of 10 e /A˚2, which yielded a significant improvement in the reconstruction due to minimizing both movement and radiation damage. At- tempts to directly apply motion correction to the frames (Li et al., 2013) failed due to the weak scattering of the filaments and the absence of carbon in many images. Due to the nonspecific binding of unpolymerized peptides to the hydrophobic outer surface of the filaments, a mask was imposed (Figure 4C) to remove some of this peripheral noise (seen in Figure 1D), which becomes evident only at lower thresholds. In contrast, the lumen of the tubes contained no

noise at any of the thresholds used. A small improvement in the map was obtained by using a broader range of twist (Figure 3A) containing 62,122 segments, showing that the improvements in signal to noise from using more segments were offset by the greater heterogeneity of the larger set.

A similar approach was used for the form II filaments, but because of the greater disorder, an additional search was made to optimize the box length. Using lengths of 192, 288, and 400 pixels for all procedures, starting with sorting by the pitch of the three-start helices (Figure 3B), it was found that the best resolution was obtained with the 288 pixel long boxes. A total of 26,896 segments (each 288 pixels) were used for the final reconstruction.

We suggest that the appearance of the map and comparison with the model is the best measure of resolution, as it provides a reality-based standard. In contrast, the standard Fourier Shell Correlation (FSC) is at best a measure of internal consistency and does not necessarily show the true resolution of a map. That is why one can impose the wrong symmetry on a helical structure and obtain an FSC curve that is completely meaningless (Wu et al., 2014; Xu et al., 2014). Using the FSC measure for helical filaments is further complicated by the fact that the helical filaments must be masked at the top and bottom to prevent large artifacts from arising due to the discontinuity in density at the ends of the filament. Nevertheless, Figure S3 shows a standard FSC curve for the form I filaments. If one uses the FSC = 0.143 criterion (Rosenthal and Henderson, 2003) after dividing the data set into two halves and using a common alignment, then an overly optimistic resolution of 3.3 A˚ is obtained. This FSC has actually been lowered by the peripheral density on the outside of the filament. Masking this density can produce even better FSC curves but these all depend upon the details of the mask that is used.

For the negatively stained samples, 2% uranyl acetate was used for staining. Grids were imaged on a Tecnai T12 microscope at an accelerating voltage of 120 keV and recorded on film. Micrographs were scanned with a Nikon Cool- scan 9000 densitometer with a raster of 4.2 A˚ /pixel.

2.5.3 Atomic Models

To construct models of the form I assembly, a single polyalanine helix was first docked into the density. Helical assembly parameters were determined from the experimental data, and a symmetric polyalanine assembly was constructed following these parameters. Since the placement of N and C termini of the models were ambiguous from the data, 14 different threadings of the designed sequence onto this polyalanine helix were considered: seven with the N termini inward and seven with the N termini outward. Following previous work, particles were split into two sets and independent reconstructions of each set were constructed. Each of these 14 models was refined against one of the reconstructions using Rosetta's cryo-EM refinement protocol (DiMaio et al., 2013); all refinement was done in the context of the helical assembly. Refined models were then evaluated against the reconstruction not used in the fitting process, assessing agreement against this independent set in high-resolution shells (10 4 A˚). Looking at the average model-map agreement of five models refined for each threading gives a clear signal for one particular threading.

A similar strategy was used for model determination of the form II assembly. As the asymmetric unit contains two helices and the lower resolution of the data made helix placement ambiguous, additional threadings were considered. For each helix, 22 different models were generated: 11 threadings in both N-terminal to C-terminal directions. Then, all 484 combinations

of these helices were refined (in the helical assembly) using the same proto- col as form I. Evaluating models of the inner and outer layers separately, we see a clear signal for the configuration of the inner layer (Figure S5). The he- lix-helix interactions of the inner layer are identical to that of the form I, but with additional interactions between GLN1 and GLN29 between adjacent he- lices stabilizing the assembly. The configuration of the outer layer (due to relatively lower local resolution and the nearly palindromic nature of the sequence) was ambiguous under our analysis (Figures 4D and S5b). Assess- ing agreement of model and map to independent reconstructions identifies one particular configuration of helices in the outer layer, but whether helices in this layer run parallel or antiparallel to helices in the inner layer is unclear (Figure S5c). In both models of the form II assembly, the helix-helix interac- tions within the outer layer are similar to those of the inner layer but shifted by two turns.

2.5.4 Synchrotron SAXS/WAXS Measurements

Synchrotron SAXS/wide-angle X-ray scattering (WAXS) measurements were performed at the 12-ID-B beamline of the Advanced Photon Source at Argonne National Laboratory. A SAXS/WAXS simultaneous setup was utilized, and the sample-to-detector distances were set such that the overall scattering momentum transfer *q* range was achieved from 0.003 to 2.4 A˚1, where *q* = 4psinq/l, 2q denoting the scattering angle and l the X-ray wavelength. The wavelength was set at 1.033 A˚during the measurements. Scattered X-ray intensities were measured using a Pilatus 2 M (DECTRIS) detector for SAXS and Pilatus 300K for WAXS. SAXS/WAXS measurements were performed on aqueous solutions of the form I and form II peptide assemblies at concentrations of 4 mg/ml (circa 1.2 mM) in acetate buffer (10 mM, pH 4.0) at 25 C. A flow cell equipped with a quartz capillary (1.5 mm diameter) was used to pre- vent radiation damage.

Twenty images were collected for each sample and buffer. The 2D scattering images were converted to 1D SAXS curves through azimuthally averaging after solid angle correction and then normalizing with the intensity of the transmitted X-ray beam using the software package at beamline 12-ID-B. The 1D curves of the samples were averaged and subtracted with the background measured from the corresponding buffers. The simulated SAXS curves were calculated using the program CRYSOL (Svergun et al., 1995).

2.5.5 ACCESSION NUMBERS

Models and volumes for form I have been deposited with IDs 3J89.PDB and EMDB-6123, respectively.

2.6 References

1. Wang, Y. A.; Yu, X.; Overman, S.; Tsuboi, M.; Thomas, G. J.; Egelman, E. H., The structure of a filamentous bacteriophage. *J Mol Biol* **2006,** *361* (2), 209-215.

2. Yonekura, K.; Maki-Yonekura, S.; Namba, K., Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* **2003,** *424* (6949), 643-650.

3. Craig, L.; Volkmann, N.; Arvai, A. S.; Pique, M. E.; Yeager, M.; Egelman, E. H.; Tainer, J. A., Type IV pilus structure by cryo-electron microscopy and crystallography: Implications for pilus assembly and functions. *Mol Cell* **2006,** *23* (5), 651-662.

4. Loquet, A.; Sgourakis, N. G.; Gupta, R.; Giller, K.; Riedel, D.; Goosmann, C.; Griesinger, C.; Kolbe, M.; Baker, D.; Becker, S.; Lange, A., Atomic model of the type III secretion system needle (vol 486, pg 276, 2012). *Nature* **2012,** *488* (7413).

5. Smith, J. M., Natural Selection and Concept of a Protein Space. *Nature* **1970,** *225* (5232), 563-+.

6. Coates, M. L., Hemoglobin Function in Vertebrates - Evolutionary Model. *J Mol Evol* **1975,** *6* (4), 285-307.

7. Galkin, V. E.; Yu, X.; Bielnicki, J.; Heuser, J.; Ewing, C. P.; Guerry, P.; Egelman, E. H., Divergence of quaternary structures among bacterial flagellar filaments. *Science* **2008,** *320* (5874), 382-385.

8. Amprazi, M.; Kotsifaki, D.; Providaki, M.; Kapetaniou, E. G.; Fellas, G.; Kyriazidis, I.; Perez, J.; Kokkinidis, M., Structural plasticity of 4-alpha-helical bundles exemplified by the puzzle-like molecular assembly of the Rop protein. *P Natl Acad Sci USA* **2014,** *111* (30), 11049-11054.

9. Zhang, J.; Zheng, F.; Grigoryan, G., Design and designability of protein-based assemblies. *Curr Opin Struc Biol* **2014,** *27*, 79-86.

10. Walshaw, J.; Woolfson, D. N., Open-and-shut cases in coiled-coil assembly: alpha-sheets and alpha-cylinders. *Protein Sci* **2001,** *10* (3), 668-673.

11. Bai, X. C.; Fernandez, I. S.; McMullan, G.; Scheres, S. H., Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *Elife* **2013,** *2*, e00461.

12. Bammes, B. E.; Rochat, R. H.; Jakana, J.; Chen, D. H.; Chiu, W., Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3/4 Nyquist frequency. *J Struct Biol* **2012,** *177* (3), 589-601.

13. Fernandez, I. S.; Bai, X. C.; Murshudov, G.; Scheres, S. H. W.; Ramakrishnan, V., Initiation of Translation by Cricket Paralysis Virus IRES Requires Its Translocation in the Ribosome. *Cell* **2014,** *157* (4), 823-831.

14. Lu, A.; Magupalli, V. G.; Ruan, J.; Yin, Q.; Atianand, M. K.; Vos, M. R.; Schroder, G. F.; Fitzgerald, K. A.; Wu, H.; Egelman, E. H., Unified Polymerization Mechanism for the Assembly of ASC-Dependent Inflammasomes. *Cell* **2014,** *156* (6), 1193-1206.

15. Li, X.; Mooney, P.; Zheng, S.; Booth, C. R.; Braunfeld, M. B.; Gubbens, S.; Agard, D. A.; Cheng, Y., Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* **2013,** *10* (6), 584-90.

16. Liao, M.; Cao, E.; Julius, D.; Cheng, Y., Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **2013,** *504* (7478), 107-12.

17. Wu, B.; Peisley, A.; Tetrault, D.; Li, Z. L.; Egelman, E. H.; Magor, K. E.; Walz, T.; Penczek, P. A.; Hur, S., Molecular Imprinting as a Signal-Activation Mechanism of the Viral RNA Sensor RIG-I. *Mol Cell* **2014,** *55* (4), 511-523.

18. 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), 280-289.

19. Liu, J.; Zheng, Q.; Deng, Y. Q.; Cheng, C. S.; Kallenbach, N. R.; Lu, M., A seven-helix coiled coil. *P Natl Acad Sci USA* **2006,** *103* (42), 15457-15462.

20. Thomson, A. R.; Wood, C. W.; Burton, A. J.; Bartlett, G. J.; Sessions, R. B.; Brady, R. L.; Woolfson, D. N., Computational design of water-soluble alpha-helical barrels. *Science* **2014,** *346* (6208), 485-488.

21. Zaccai, N. R.; Chi, B.; Thomson, A. R.; Boyle, A. L.; Bartlett, G. J.; Bruning, M.; Linden, N.; Sessions, R. B.; Booth, P. J.; Brady, R. L.; Woolfson, D. N., A de novo peptide hexamer with a mutable channel. *Nat Chem Biol* **2011,** *7* (12), 935-941.

22. Dubochet, J., Cryo-Electron Microscopy of Viruses in Native-State. *Micron Microsc Acta* **1987,** *18* (4), 333-334.

23. Fernandez, I. S.; Bai, X. C.; Murshudov, G.; Scheres, S. H.; Ramakrishnan, V., Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. *Cell* **2014,** *157* (4), 823-31.

24. Liu, J.; Zheng, Q.; Deng, Y.; Cheng, C. S.; Kallenbach, N. R.; Lu, M., A seven-helix coiled coil. *Proc Natl Acad Sci U S A* **2006,** *103* (42), 15457-62.

25. Egelman, E. H., Single-particle reconstruction from EM images of helical filaments. *Curr Opin Struc Biol* **2007,** *17* (5), 556-561.

26. Egelman, E. H., Reconstruction of Helical Filaments and Tubes. *Methods in Enzymology, Vol 482: Cryo-Em, Part B: 3-D Reconstruction* **2010,** *482*, 167-183.

27. Klug, A.; Crick, F. H. C.; Wyckoff, H. W., Diffraction by Helical Structures. *Acta Crystallogr* **1958,** *11* (3), 199-213.

28. Wu, Y. Y.; Norberg, P. K.; Reap, E. A.; Congdon, K. L.; Fries, C. N.; Kelly, S. H.; Sampson, J. H.; Conticello, V. P.; Collier, J. H., A Supramolecular Vaccine Platform Based on alpha-Helical Peptide Nanofibers. *Acs Biomater Sci Eng* **2017,** *3* (12), 3128-3132.

Chapter III: Expanding the library of helical nanotubes via the implementation of an arginine staple motif (Forms III, 0, and IV)

3.1 Introduction

Peptide-based assemblies have been employed with great success over the last two decades for fabrication of structurally ordered materials within the nano-scale size regime for applications in biomedicine and nanotechnology.[1-4] The molecular design of these materials has relied thus far on relatively simple sequence-structure relationships derived from analysis of native protein folds. The mode of assembly usually involves lateral and axial self-association of protomers (i.e., structural subunits) into long non-covalent polymers that display helical symmetry. However, few structures of these helical assemblies have been characterized at near-atomic resolution and, in the infrequent cases in which high resolution structural analysis has been performed, the observed structures can differ significantly from the models that were employed as the basis for the design.[4, 6-9] Near-atomic resolution structural information is essential for understanding the inter-subunit contacts that stabilize the lateral and axial interfaces of peptide assemblies, especially as these interactions are critical to functionally important properties such as mechanical response. Moreover, structural polymorphism is frequently observed for synthetic peptide assemblies as well as for native biological protein assemblies prepared *in vitro*. Many specimens display structural polymorphism that results in a diverse population of assemblies from the same protein sequence. Selection for monomorphic variants of defined and predictable structure represents a significant challenge to *de novo* design of peptide-based materials. In one recently reported analysis, which may not be unusual, at least ten distinguishable structural variants were observed using cryo-EM from *in vitro* self-assembly of an amyloidogenic peptide sequence.[10] Slightly different synthetic conditions can result in the formation of completely different structural variants in a manner that

has yet to be understood in terms of molecular level determinants.[11] Therefore, structural information that results from near-atomic resolution analyses of helical assemblies is essential for interrogating the limits of our knowledge of peptide design as well as for reverse engineering of structure to generate novel assemblies reliably and predictably using a materials informatics approach.

We recently described the structural analysis of helical assemblies derived from two related coiled-coil peptides, Form I and Form II (Figure 3.1).[9] The three-dimensional reconstructions were generated using Rosetta modeling of the peptides into the EM density map derived from cryo-electron microscopy with direct electron detection.[12] The sequences of the two peptides differed solely in the structurally conservative substitution of arginine residues in Form I for lysine residues in Form II. However, the observed structures differed significantly in terms of helical symmetry and the nature of the cohesive interactions between protomers. The Form I structure was based on an asymmetric unit corresponding to a single helix, while the Form II structure was derived from a pair of helices. The Form I structure comprised a left-handed one-start helix with 4.1 subunits per turn with an axial rise of 2.2 Å and a diameter of approximately 6 nm (as observed via cryo-EM). In contrast, the Form II structure could be defined in terms of a left-handed one-start helix with 3.1 asymmetric units per turn, an axial rise of 2.1 Å per asymmetric unit, and a diameter of 12 nm. Each assembly epitomizes an unusual packing motif recently designated as a cross-alpha structure[13] (alternatively, an alpha-sheet structure[14]), in which the chain axis of the helical protomers is oriented nearly perpendicular to the long axis of the helical assembly. The structures of the Form I and the Form II assemblies are composed of four and three cross-alpha helical stacks, respectively. Within a cross-helical stack, the peptides pack very similarly in the Form I and Form II structures. The mode of association between helical stacks is the main distinguishing feature

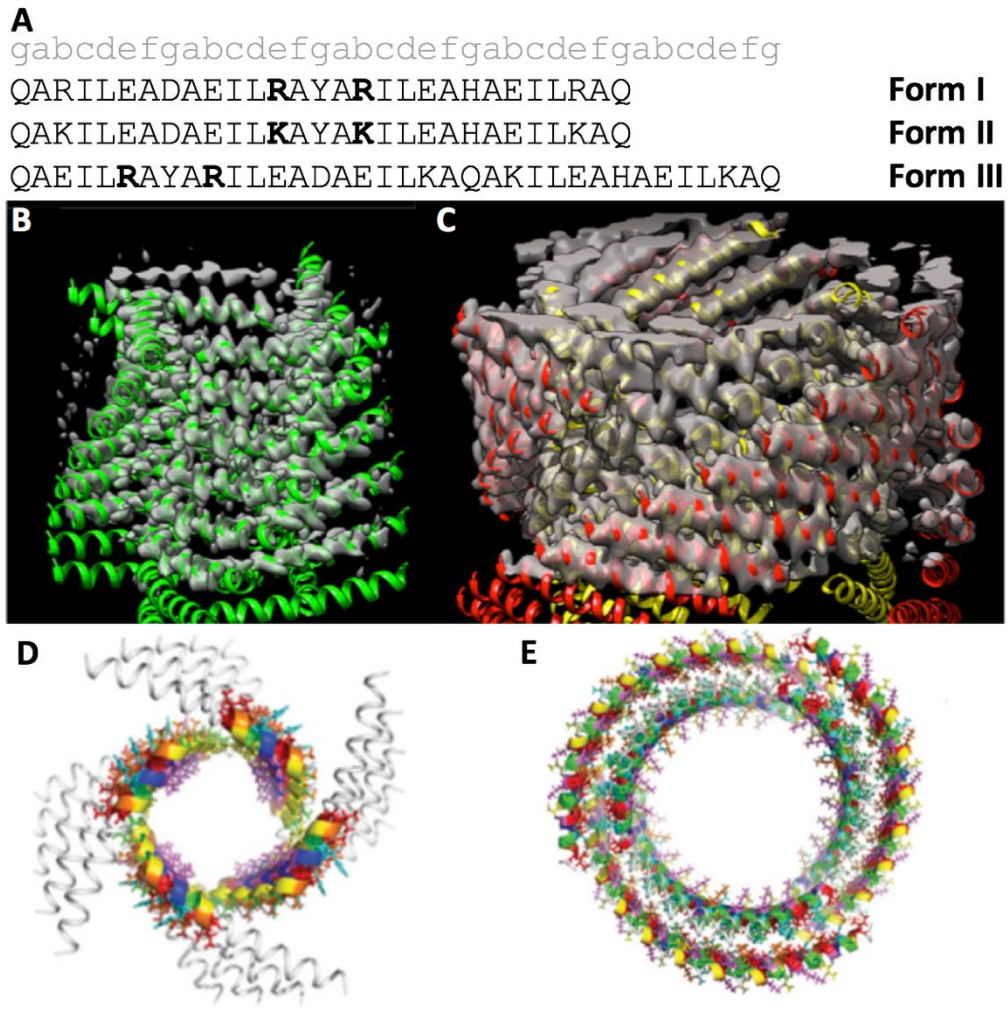


Figure 3.1: (A) Sequences of Forms I, II, and III with key sequence changes shown in bold. Above the sequences, the positions of the heptad repeat are indicated in grey. Near atomic resolution structural model of Form I (B, D) and Form II (C, E) from cryo-electron microscopy are shown in side-on and top down conformations. 3D models are from Egelman et al. 2013.

between the two assemblies. The Form I structure displays a unique interaction involving a pair of arginine residues, R13 and R17, that form a network of hydrogen bonds and electrostatic interactions with the *C*-terminal residues of helices in an adjacent cross-alpha stack. Mutagenesis of either of these residues to lysine results in conversion of the assembly to a Form II structure.

Conversely, a K13R, K17R double mutant of the Form II sequence resulted in conversion of the resultant assembly to a Form I structure, which suggested that the presence of both arginine residues at the appropriate position was a necessary pre-requisite for formation of the Form I structure. High resolution structural data led in this case to identification of critical structural interactions that underlie the differences between the Form I and Form II assemblies, as well as the ability to directly interrogate the importance of these interactions through site-directed mutagenesis.

These results suggested that the **RxxxR** motif, in which the two arginine residues, R13 and R17, are oriented i, i+4 within the Form I sequence, define a side-to-end helical interaction motif within the assembly. While interactions involving individual **RxxxR** motifs may be relatively

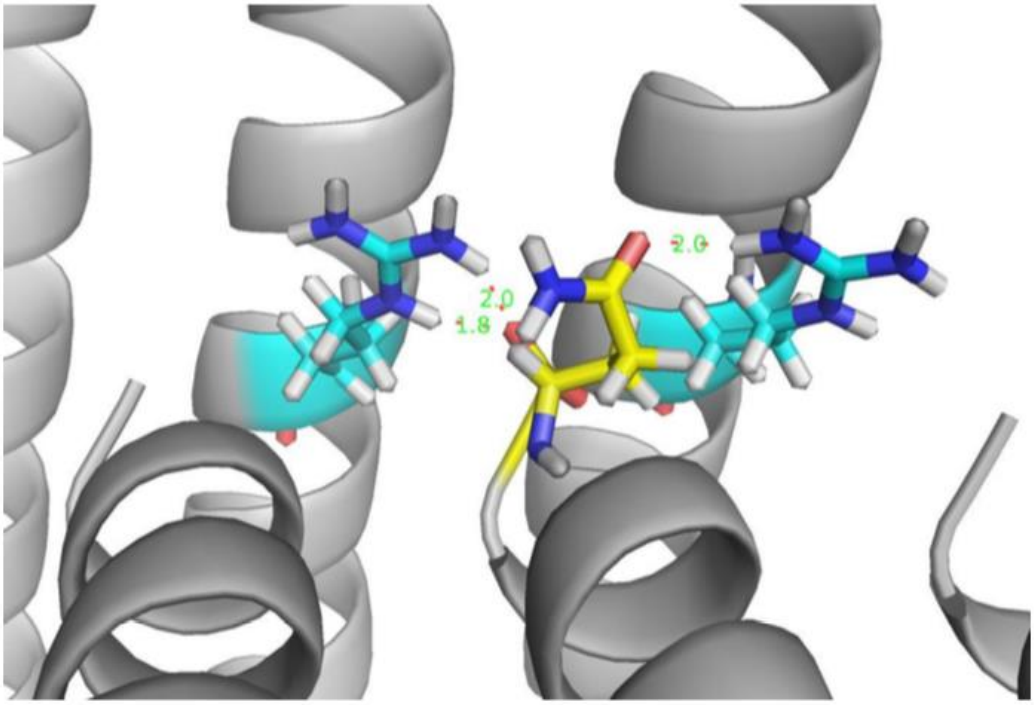


Figure 3.2: 3D close-up of the ribbon-ribbon interface of Form I tubes. The C-terminal glutamine is shown in yellow and forms hydrogen bonds with the two arginine 13 side chains of adjacent helices from neighboring ribbons.[5] Another example of over-representation of arginines in peptide interfaces.

weak, many such interactions occur with the structural context of the Form I assemblies such that their influence on the stability of assembly is magnified significantly. Numerous instances of the stabilizing effect of structural interactions between appropriately substituted i, i+4 residues have been described within the sequences of helical peptides.[15-22] In addition, specific interaction motifs such as the left-handed leucine zipper,[23-26] i.e., **LxxxLxx**, and the right-handed glycine zipper,[27-28] i.e., **GxxxG**, have been observed to promote and stabilize lateral (side-to-side) interactions in coiled-coils and transmembrane helical bundles, respectively. We posit that the **RxxxR** motif represents a potential mechanism to stabilize side-to-end interactions between helical peptides. The **RxxxR** motif defines a helix-helix crossing angle, Ω, of approximately 90° and, consequently, the interacting surface between helices occurs over much more limited contact area in comparison to leucine and glycine zipper motifs in which the crossing angle (Ω = -20° and +40°, respectively)[24, 27-30] that defines the side-to-side association is significantly more acute.

In recognition of this difference, we designate the **RxxxR** motif as the **arginine staple** (rather than a zipper), since it mediates a distinctive local interaction between helices. Evidence for the potential utility of the arginine staple in mediating interactions can be gleaned from previously reported structural analyses of proteins. Ghosh, et al.,[31] reported, from a bioinformatic analysis of helical structures in the Protein Data Bank, that arginines are over-represented among residues that interact with the *C*-termini of helices. Neves et al.[32] report that clusters of arginines have been observed in the form of rings and chains at the interacting interfaces between protein oligomers,[33] and arginine residues, in general, appear to be over-represented at protein-protein interfaces.[34] Further, Arginine accounts for 40% of salt-bridge formation in proteins between the arginine sidechain and the carboxylate moiety of neighboring residues.[35] Arginine's positive

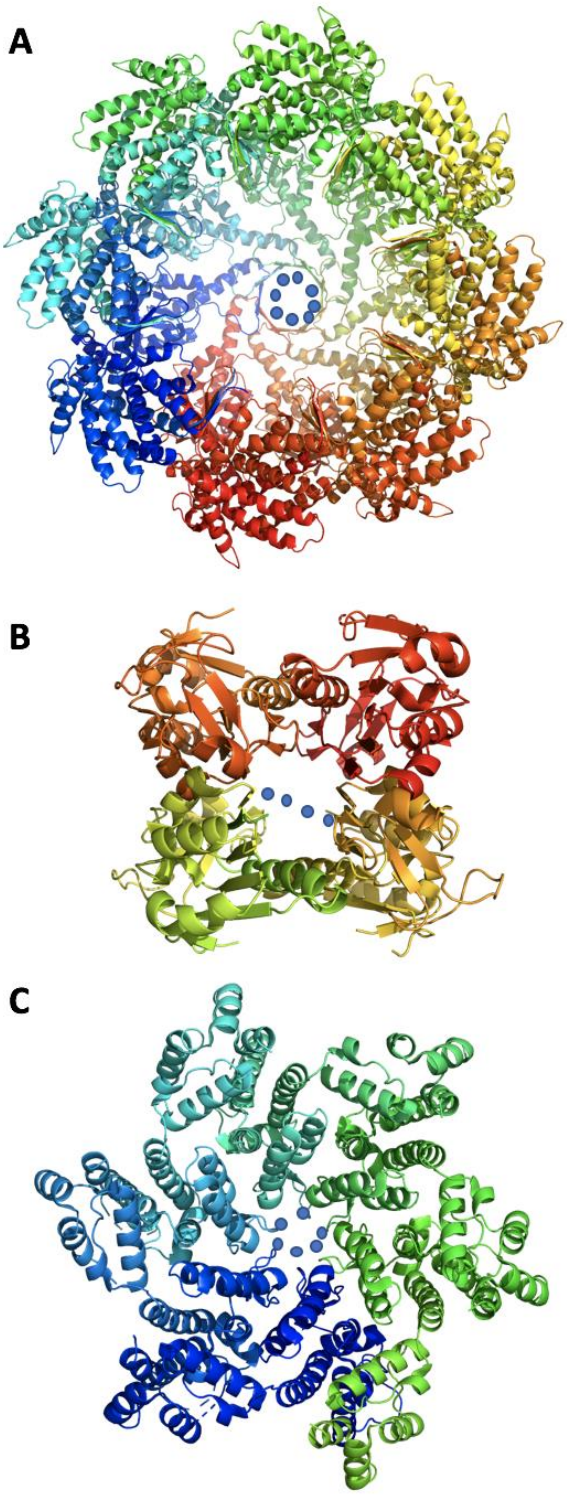


Figure 3.3: three examples of arginines over-represented at protein-protein interfaces, (A) Archael chaperonin, PDB ID 3KFB, (B) thymidine kinase 1, PDB ID 2WVJ, and (C) HIV CA protein, PDB ID 3H4E. The key arginine residues are indicated by blue spheres.

charge plays a critical role in both polar interactions with aromatic residues and planar stacking.[36-37] Given the structure and makeup of its sidechain, arginine is able to form multiple hydrogen bonds,[38] and they are known to make substantial energy contributions towards the stabilization protein-protein interactions, even being labelled as one of the most commonly identified amino acids at "hot spot" protein-protein interfaces.[39-40] All of these reasons lead to a much higher than random frequency of arginines populating the surface map of protein-protein interfaces, despite amino acids such as lysine, aspartic acid and glutamic acid being notably absent in equivalent spatial positions.[41] These contacts help govern the self-assembly of unique protein structures, such as rings, strings, and stacks, all of which feature interfacial interactions which are stabilized by a small number of dominant arginine interactions. Combining the aforementioned arginine-specific design principles with the near-atomic structural data of Forms I and II, new sequences can be produced in order to design *de novo* self-assembling nanotubes.

3.2 Sequence Design

In order to test the utility of the arginine staple in mediating interactions within a defined structural context, we designed a peptide sequence, Form III, that incorporated elements of the Form I and Form II sequences. The sequence of Form III comprised 36 amino acid residues and was consequently one heptad greater in length than Form I and Form II.[8] The arginine staple motif was located near the *N*-terminus of Form III, in a manner similar to Form I. Lysine residues occupied all of the other basic amino acid sites at *e*- and *g*-positions within the heptad sequence of Form III, such that, as in Form II, they contributed to stabilizing electrostatic interactions between peptides within a cross-alpha stack, but minimally to the cohesive interactions between stacks. We hypothesize that the interactions between helical stacks of the Form III assemblies should resemble

those of the Form I assemblies in that they will be driven primarily through arginine staple motif. If Form III self-associates through a similar mechanism to Form I, we would expect the formation of wider diameter nanotubes. The Form I assembly displays a nearly square cross-section in which the nanotube diameter is determined by the length of the interacting peptide segments in each symmetry-related helical stack (Figure 3.1). As the lengths of the interacting peptide segments are longer for Form III, we would expect a wider nanotube to result and, moreover, the width of the assembly should be predictable from position of the arginine staple motif and the structure of the Form I assembly.

3.3 Results and Discussion

3.3.1 Peptide Self-Assembly

The Form III peptide was synthesized using microwave-assisted solid-phase peptide synthesis and was purified to homogeneity using reverse-phase HPLC on a C-18 column. The molar mass was confirmed using ESI-mass spectrometry and was found to match the theoretical value of 4,043 g/mol. To promote the adoption of α-helical character, the peptide was dissolved in a 1:1 (V/V) mixture of aqueous buffer (10 mM acetate, pH 4.0) and 2,2,2-trifluroethanol to a final concentration of 3 mg mL^{-1}. TFE was then allowed to evaporate in a slow, controlled manner (using Parafilm ® pierced once with a small-bore needle to cover the assembly container) until the mixture volume was reduced by approximately half. Highly ordered nanotubes were observed via TEM as early as one day after the full evaporation of TFE from the assembly mixture. Filamentous assemblies were observed across a range of peptide concentrations, from ~ 0.3 mg mL^{-1} to 5 mg mL^{-1}, and demonstrated a high level of structural homogeneity across said concentration spectrum. Though remaining trace volumes of TFE cannot be fully ruled out, any remaining TFE did not prohibit or abrogate self-assembly of nanotubes. When assembled in the absence of TFE, Form III formed less-ordered filaments demonstrating a far greater degree of structural heterogeneity than Form III filaments assembled in the presence of TFE. Furthermore, TFE assisted in the solubility of the peptide, and the slow, controlled evaporation of TFE over multiple days meant that the concentration of peptide in the aqueous buffer (10 mM acetate, pH 4.0) increased gradually over time, potentially affording greater control over the assembly process. The pH of the solution was checked following full TFE evaporation and found to remain ~4.0, indicating no permanent pH shift due to the addition and subsequent evaporation of TFE from the assembly solution. The samples proved to be incredibly robust over time, showing no signs of structural degradation via

TEM over the span of several years, indicating that while the driving forces behind the self-assembly of the Form III nanotubes are largely non-covalent, they are quite stable.

3.3.2 Circular Dichroism and Flow Linear Dichroism

Circular dichroism spectropolarimetry (CD) was employed to study the conformational properties of Form III in aqueous buffer (10 mM acetate, pH 4.0, following full evaporation of TFE from the mixture). Form III displayed an alpha-helical signature that was consistent with the behavior of the Form I and Form II peptides (Figure 3.4), as was expected. The α-helical content

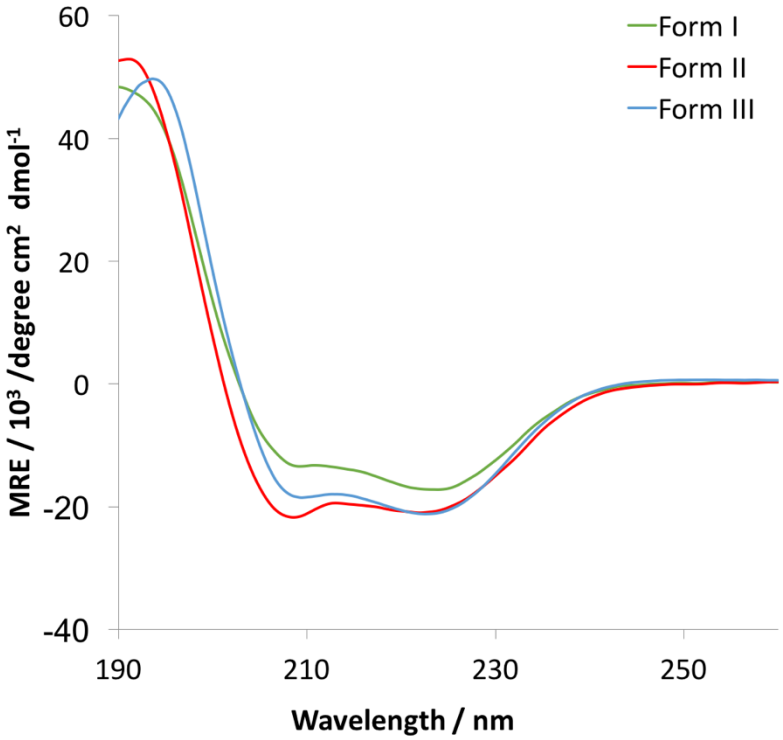


Figure 3.4: Circular dichroism spectra of 100 μM Form I (green), Form II (red), and Form III (blue) in 10 mM acetate buffer, pH 4.0.

was slightly greater than that observed for the Form I peptide, particularly when measured at the 208 nm minima. This is expected, as Form III features a longer sequence. In either case, the M.R.E. values fell below that which would be expected for a fully helical conformation. The *N*-terminus

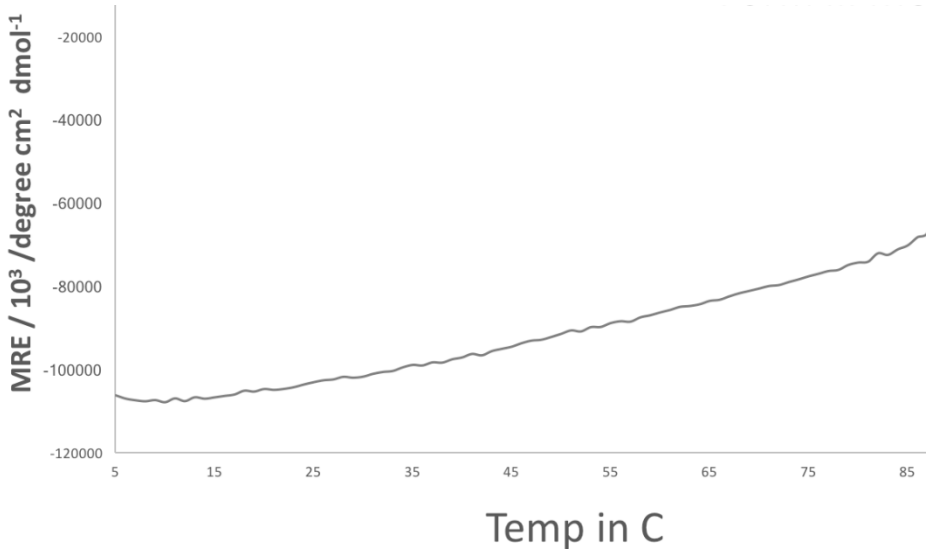


Figure 3.5: thermal denaturation curve for Form III in 10 mM acetate pH 4.0. No melting transition is observed.

of the Form I assemblies was found to project into solution and was not as well structured as the remaining peptide segment that was packed into the cross-helical stack. This structural ambiguity may contribute to the relatively weak CD intensities observed for the peptides. Alternatively, the aforementioned weak intensities may simply be attributed to the conditions under which the experiment was performed. The peptide was assembled at 750 μM, but examined via CD at 100 μM, which could lead to a partial melting of the helical structure due to the reduction in peptide concentration. The Form III peptide did not display a melting transition with the temperature range from 5 °C to 90 °C, although the signal for the α-helical conformation gradually diminished concurrent with increasing temperature. A sharp increase in the melting curve is observed at ~ 90°C, however, this is not attributed to thermal denaturation of the sample but rather loss of solution from the cuvette due to boiling.

After confirmation via TEM that Form III self-assembled into tubular structures, flow linear dichroism[21, 42-45] was employed under dilute conditions (25 μM) to survey the degree of

anisotropy that developed within the peptide samples under 5,000 RPM flow alignment in a Couette cell. Solutions of Form III demonstrated positive signals at 192 nm, and a weaker, broader positive signal around 220 nm. This was corroborated across three independently assembled solutions of Form III. This data would suggest that the electron transitions in the sample are aligned parallel with the direction of flow. Therefore, the helices are packed perpendicularly to the long axis of the Form III filaments. This matches what was observed for the Form I, and to a lesser degree, the Form II peptide systems. Form I was assembled at 50 µM and demonstrated a signal intensity at the 192 nm maximum of ~0.035 dOD, whereas Form III was assembled at 25 µM and demonstrated a maximal signal intensity of ~0.020 dOD at the equivalent wavelength. The broader peak around 222 nm is slightly blue shifted to ~217 nm, but is more pronounced than the equivalent peak for Form I.

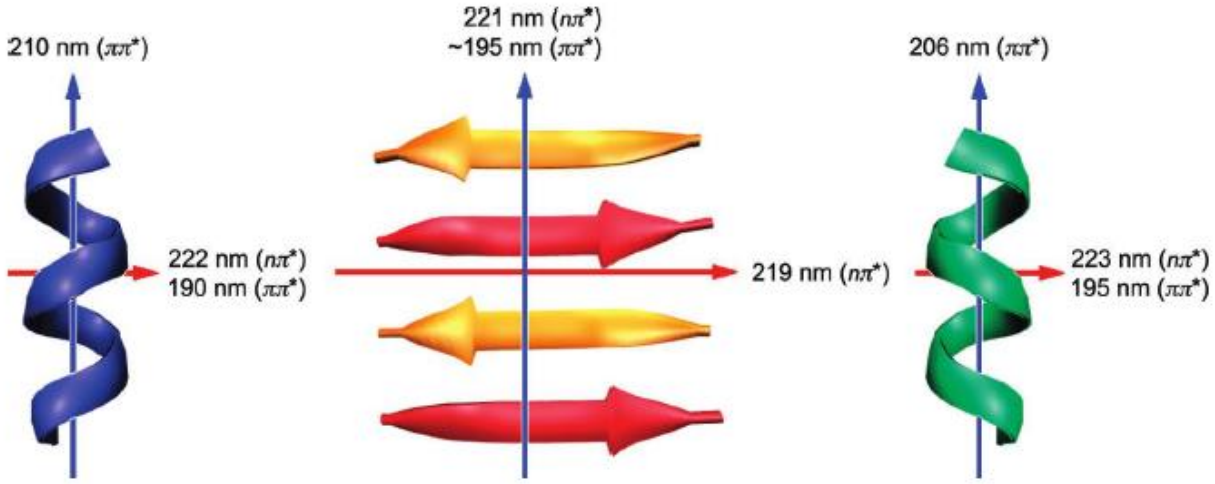


Figure 3.6: Polarizations of transitions in (purple) an alpha-helix, (red/yellow) a beta-sheet, and (green) a PII-helix.[43] Of note, the 222 nm nπ* and 190 nm ππ * transitions are polarized perpendicularly to the helical axis.

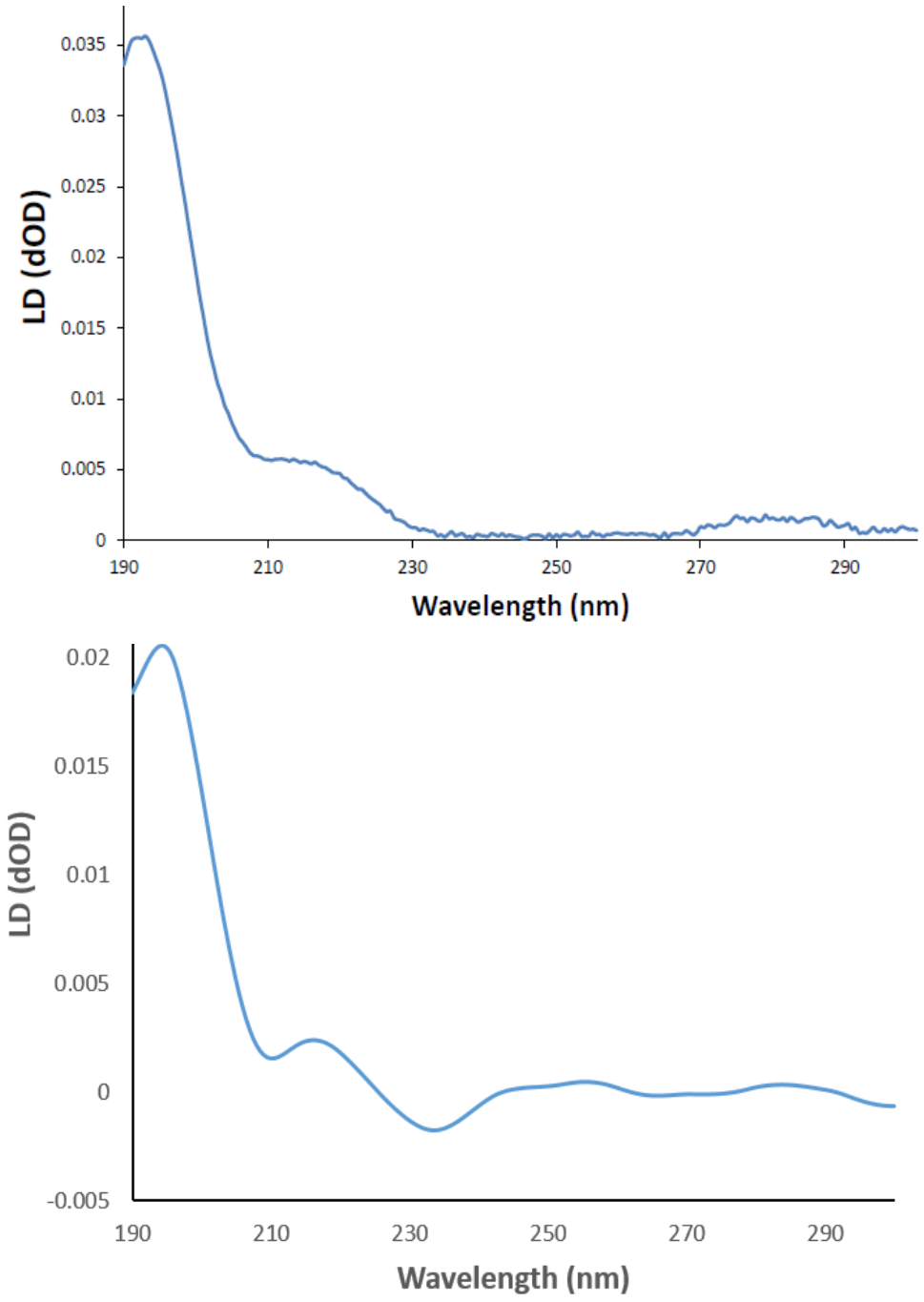


Figure 3.7: Flow LD spectra of (top) Form I and (bottom) Form III at 50 and 25 µM respectively in 10 mM acetate buffer at pH 4.0. Signal maxima are observed at 192 nm and 220 nm.

3.3.3 Transmission Electron Microscopy

Transmission electron microscopy (TEM) data from negatively stained samples of Form III indicated the presence of high aspect-ratio tubular assemblies with an apparent width of 9.8 ± 1.1 nm (Figure 3.8) from 100 measurements across three independently assembled samples. In

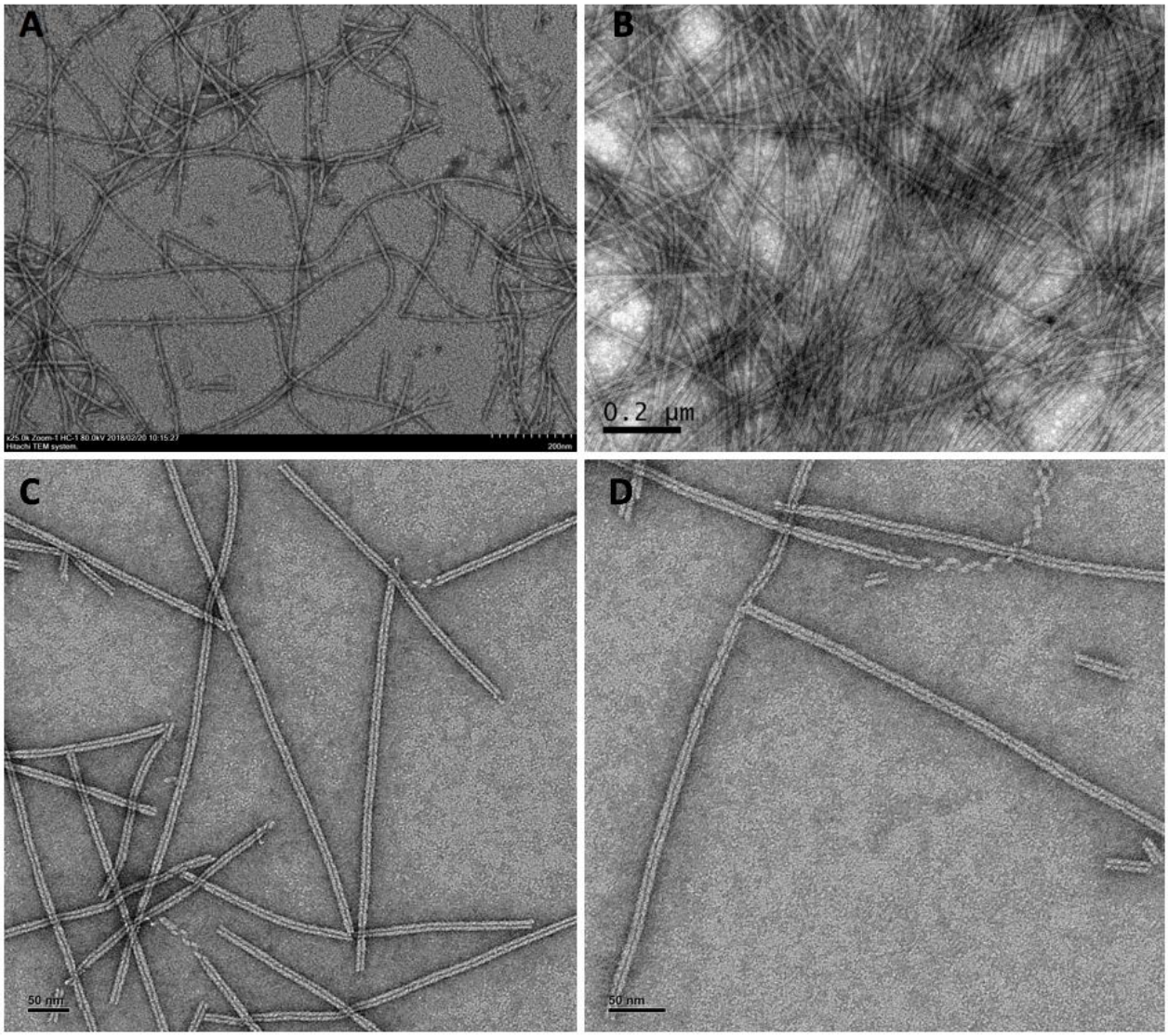


Figure 3.8: negatively stained transmission electron microscopy of the tubes from Form I (A) and Form II (B) for comparative reference. Micrographs for Form III (C and D). Micrograph D is included to indicate the helical unwinding of the filament that occurs in a minority of cases at the termini.

contrast to the Form I assemblies, Form III assemblies demonstrated an easily discernible inner lumen of sufficient cross-sectional size to allow for stain to permeate the inner channel (Figure 3.8). Form III tubes demonstrate moderate length polymorphism, though it is uncertain whether this occurs as a byproduct of self-assembly in solution, or is a result of filaments breaking during adsorption of sample to the grid. The averaged diameter for Form III assemblies was significantly greater than the 6 nm diameter observed for the single-walled Form I assemblies, but less than the 12 nm diameter observed for the double-walled Form II assemblies. The experimental diameter average for Form III corroborates the hypothesis that extending the **arginine staple** in sequence-

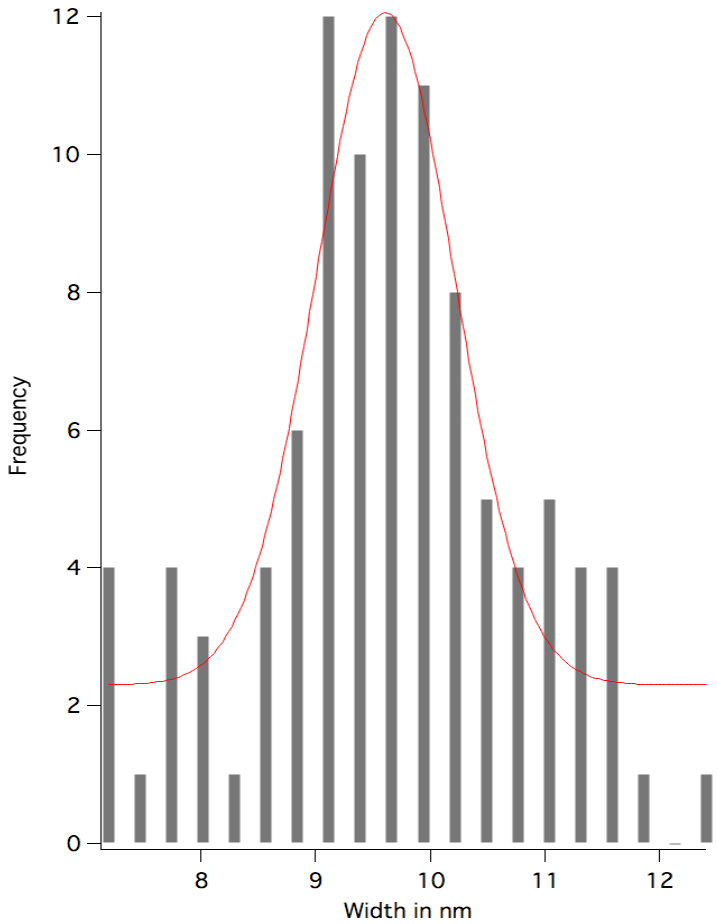


Figure 3.9: histogram of width measurements for Form III tubes measured from negatively stained electron micrographs via ImageJ image processing software.

space by one heptad repeat effectively lengthens the sidewall of the nanotube. The single wall ImageJ measurements for the nanotubular assemblies were observed to be 4.3 ± 0.77 nm. The average inner lumen diameter was measured to be 1.3 ± 0.26 nm. Instances of helical unwinding of the nanotubes at the termini can be observed in a minority of species (3.8D) and this localized unwinding of filaments is observed to occur exclusively at the termini. That these unwinding events demonstrate a helical, or corkscrew, morphology lends credence to the hypothesis that the tubes are assembling in a supramolecular helix with alpha stacks as the subunits.

3.3.4 Scanning Transmission Electron Microscopy

Scanning transmission electron microscopy (STEM) offers valuable structural data for one- and two-dimensional peptide assemblies in the form of mass per length (MPL) and mass per area (MPA).[44] These experimental values can be compared to what is expected based upon the near atomic resolution structural determination of Form I and Form II. STEM analysis of freeze-dried specimens of the Form III assemblies confirmed the presence of filaments of approximately 10 nm in diameter. Tobacco Mosaic Virus (TMV) was employed as an internal mass standard. Mass-per-length (MPL) measurements were performed on the Form III assemblies. A mean value for the experimental MPL of 3,431 ± 360 Da/Å was determined from a histogram plot of the STEM data. An MPL value of 1,838 Da/Å was estimated for a single-walled nanotube of Form III based on the peptide mass of 4,044 Da and the axial rise of 2.2 Å/peptide of the Form I structure. If the Form III helices are hypothesized to pack in a manner identical or mimetic to Form I, the differences in MPL values between the two supramolecular assemblies should originate solely from the added molar mass from the extra heptad repeat of Form III. The percentage difference of molecular weight (MW) between the two forms equates to 20.86% (4,043.72 g mol^{-1} / 3,345.85 g mol^{-1} Form

I). Subsequently, the MPL value for Form III should be ~1.2x larger than that of the parent peptide assuming a single-walled packing of constituent helices. From near atomic resolution cryo-electron reconstruction of the structure, Form I demonstrates a measured MPL value of 1,521 Da/Å. Applying the MW scaling factor of ~1.2 yields the anticipated MPL for a single walled tube (~1,838 Da/ Å) with Form III as the subunit. Surprisingly, the experimentally determined MPL value for the Form III assemblies was approximately double that calculated for a single-walled nanotube. The data argue that Form III nanotubes may actually be a double-walled nanotube. However, if this is true, it is not yet known exactly how the helices are packing. Future cryo-elecctron microscopy data and near-atomic structural resolution reconstruction will provide immense insight into the structural parameters of Form III, and guide the future use of **arginine staple** motifs.

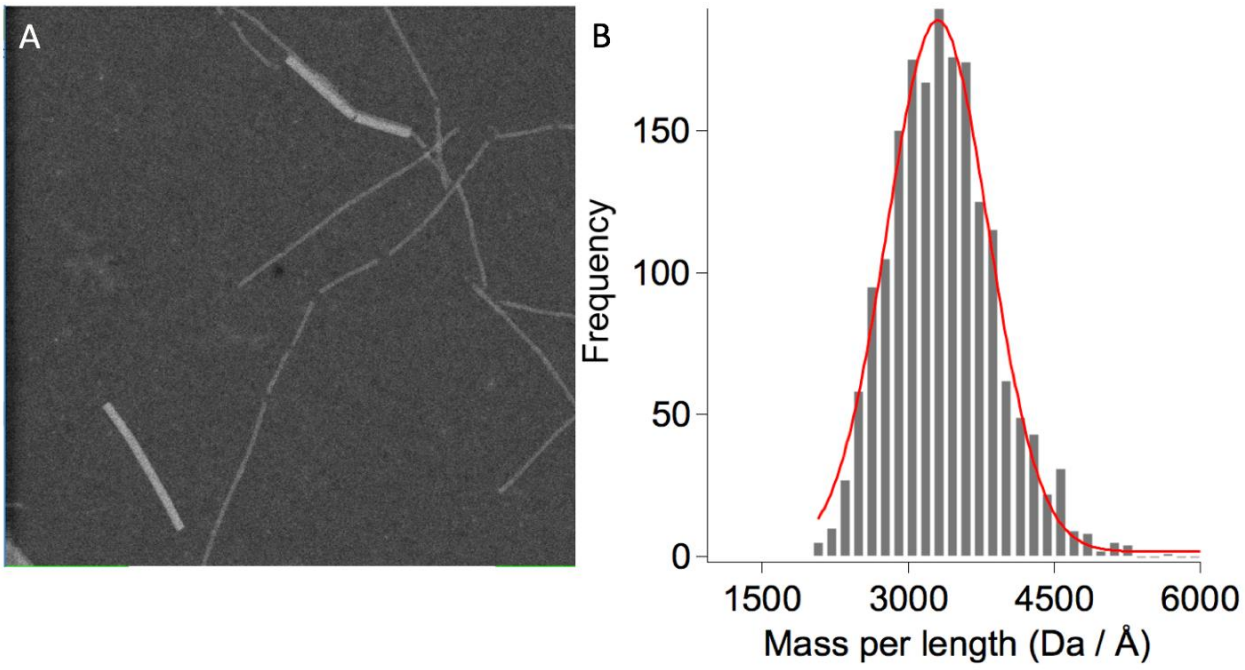


Figure 3.10: (A) Representative STEM image of freeze-dried Form III (smaller filaments) with TMV (larger filaments) as an internal standard. (B) STEM mass/length histogram for the freeze-dried nanotube specimens with Gaussian fit (red).

3.3.5 Power Spectra and Reference-Free Class-Averages

In order to confirm the bilayer hypothesis, averaged power spectra were calculated from uniform magnification negatively stained TEM images of the Form III assemblies. A layer line was observed at a space of 1/70 $Å^{-1}$ (Figure 3.11E) for Form III, which compares well to the position of the layer line observed for the four-start helix (Bessel order n=4) in the averaged power spectrum of the Form I assemblies at 1/62 $Å^{-1}$ (Figure 3.11A). A second set of layer-lines is observed in Form III at 1/35 $Å^{-1}$. Whereas the first layer lines indicate a Bessel order of n=4, the second layer lines indicate a Bessel order of n=8. In comparison, the averaged power spectrum for the Form II assemblies indicates the presence of a first order and second order layer line for the three-start helix at spacings of 1/52 $Å^{-1}$ and 1/26 $Å^{-1}$, respectively (Figure 3.11C). However, the layer-line for the presumed four-start helix of the Form III assemblies displays a pair of maxima, in contrast to the single banded maxima of Form I that is indicative of the presence of a double walled nanotube. Reference-free class-averages for each form are also provided for comparative purposes, allowing for a preliminary and rudimentary evaluation of symmetries within each tubular assembly. Of note, analysis of the reference-free class averages across all forms reveals Form III

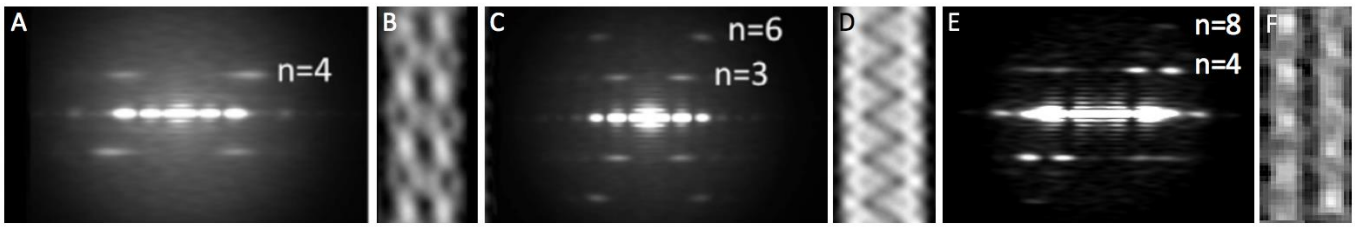


Figure 3.11: Averaged power spectra for negatively stained filaments of (A) Form I wherein a Bessel order of n=4 is observed in the layer line, (C) Form II wherein Bessel orders of n=3 and n=6 are observed in the layer lines, and (E) Form III wherein Bessel orders of n=4 and n=8 are observed in the layer lines. Reference free averages from negatively stained TEM samples for (B) Form I, (D) Form II, and (F) Form III.

sharing structural characteristics of both Form I and Form II (Figure 3.12). Similar to what is seen in the reference-free average for Form I, there are observable circular areas of low electron density running down each wall of the tube, each one offset from their adjacent point in the opposite wall. The reference-free average for Form III also demonstrates a similarity to that of Form II, wherein a zig-zag line of low electron density proceeds down the length of the tube. These data could point towards Form III self-assembling as a structural hybrid of Forms I and II, wherein the subunits of

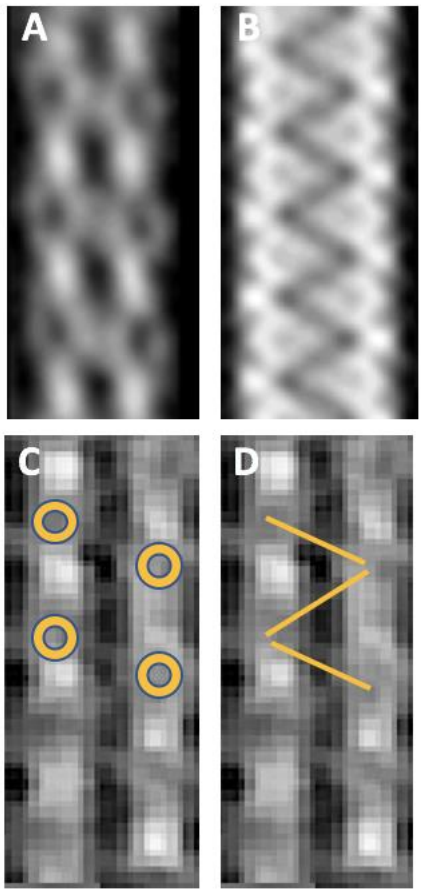


Figure 3.12: reference free power averages for (A) Form I and (B) Form II for geometrical comparison to that of Form III. The offset circles of low electron density, similar to what is seen for Form I are seen in Form III and indicated in orange in (C). The zig-zag pattern of low electron density, similar to what is seen in Form II, is seen in Form III and indicated in orange in (D).

Form III pack as a bilayer (as seen in Form II) in a four-start helix orientation (as seen in Form I). Potential electron density exposed at the solvent-facing regions of the Form I helix stacks may allow for dimerization of helices to a small extent, creating a minority species of bilayer tubes or tube segments at sufficiently high concentrations (≥ 3 mg mL^{-1}). Subsequently, a larger solvent-facing surface area derived from the greater contour length of the Form III helix may increase this effect, shifting the representative subunit from a helix monomer with highly limited instances of dimerization occurring only at significantly high peptide concentrations (Form I), to a helix dimer (Form III) which demonstrates bilayer formation even at low concentration (≤ 0.3 mg mL^{-1}).

3.3.6 Small- and Wide-Angle X-Ray Scattering

Small angle X-ray scattering (SAXS) offers critical insight into the shape and structure of colloidal assemblies. Synchrotron small-angle X-ray scattering measurements were performed on the Form III assemblies as a further point of comparison to the Form I and Form II assemblies. The scattered intensity profiles were measured as a function of the momentum transfer, q, within the range from .005 $Å^{-1}$ to 2.3 $Å^{-1}$. Similarly to the Form I and Form II, the scattered intensity displayed a linear dependence on q^{-1} within the power law region (Figure 3.13). The mass fractal value of -1.008 ± 0.0563 on q is consistent with extended rod-like assemblies. Notably, assemblies derived from all three peptides display a series of oscillations on q that are characteristic of hollow cylindrical assemblies. The position of the first minimum for the Form III assemblies is observed at a q value of 0.082 $Å^{-1}$ (corresponding to 76.6 Å), which lies between the corresponding q values of 0.106 $Å^{-1}$ (corresponding to 59.3 Å) and 0.065 $Å^{-1}$ (corresponding to 96.6 Å) for the Form I and Form II assemblies, respectively. Bragg diffraction peaks were observed in the high q region (inset of Figure 3.13) and correspond to (1) 53.2 Å which is commensurate with the length of the single

helix, (2) 20.4 Å, (3) 15.3 Å, (4) 8.58 Å, and (5) 7.85 Å. The Bragg-Diffraction peak value of 20.4 Å was also observed when performing linear power spectrum analysis of the Form III filaments from the STEM data. The scattering data in the mid-*q* region ≤ .002 were fit to the modified Guinier model for rod-like forms for all three of the Form peptide assemblies. Values for the cross-sectional radius of gyration, R_C, were determined for the Form I (28.8 Å), Form II (40.7 Å), and Form III (35.5 Å, Figure 3.11) assemblies, which corresponded to average cross-sectional rod diameters of 81.4 Å, 115 Å, and 100 Å, respectively. The calculated diameters from SAXS Guinier measurements as well as first minima calculations closely match the fiber widths calculated from EM data for the Form II and Form III data. Strangely, the rod diameter calculated from SAXS measurements of Form I is significantly greater than the widths determined from the EM images.

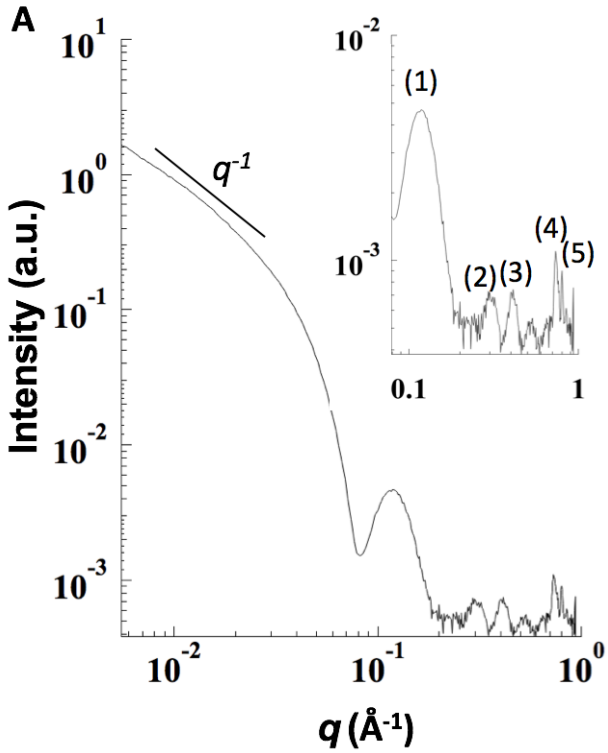


Figure 3.13: (A) Small-angle X-ray scattering curve for a solution of Form III (4 mg/mL) in acetate buffer (10 mM, pH 4.0). Inset: expansion of the high *q* region depicting the positions of the five diffraction peaks. Peaks (*d*-spacing): (1) 53.2 Å; (2) 20.4 Å; (3) 15.3 Å; (4) 8.58 Å; (5) 7.85 Å.

The rod diameter of 81.4 Å is also greater than the longest inter-atomic distance of circa 73 Å that can be measured from the 3.6 Å resolution three-dimensional reconstruction of the Form I filaments (PDB ID: 3J89). This phenomenon may be explained by peptide concentrations at which each of the aforementioned analytical technique is performed. SAXS typically requires high concentration sample solutions for high resolution data. In contrast, TEM (and STEM) as well as cryo-EM require more dilute sample conditions, typically at least 10x more dilute than what is requisite for SAXS. To this end, the higher concentrations may allow for excess, unordered, 'free' peptide in solution to bind to extant tubes, creating a bilayer tube for Form I. In the lower concentration techniques, sufficient peptide concentration may not exist for this bilayer formation to occur. Despite the diameter anomaly, the combined data from SAXS and EM measurements are

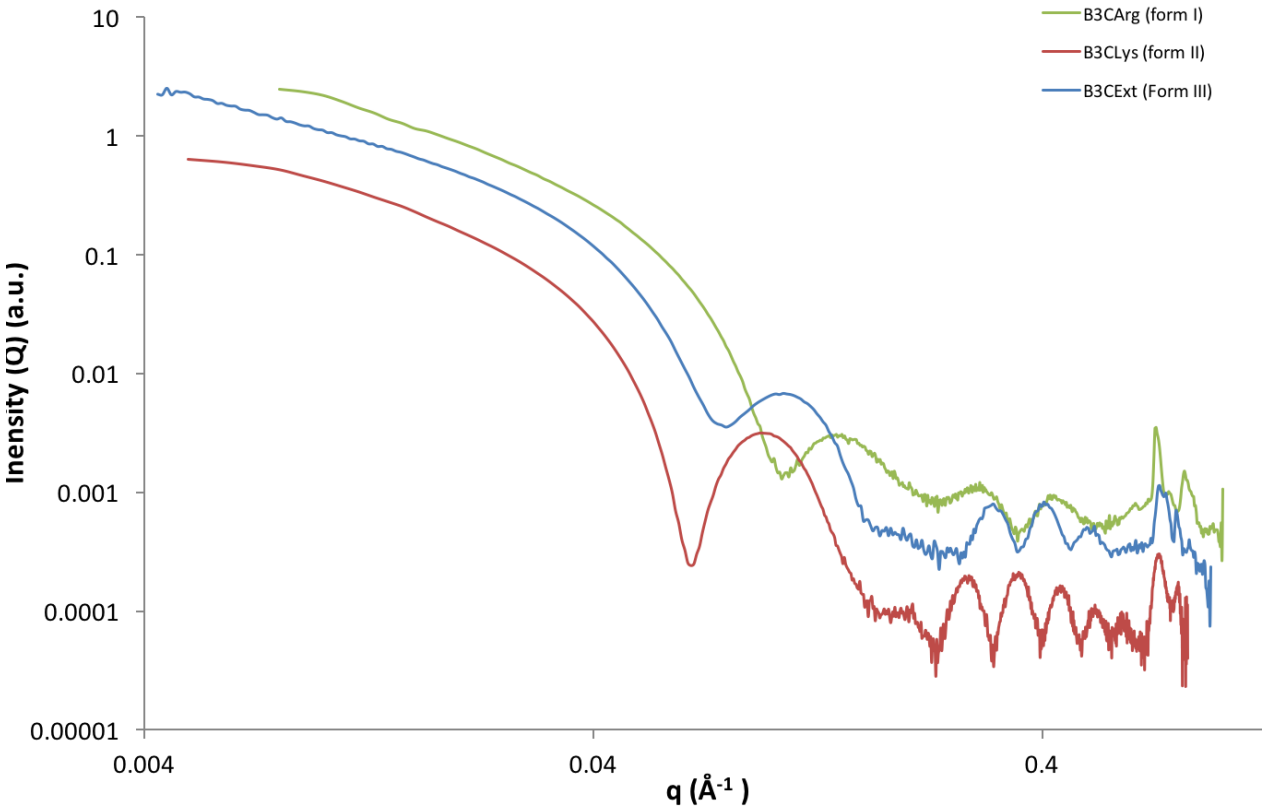


Figure 3.14: Small-angle X-ray scattering curves for Forms I, II, and III in green, red, and blue respectively.

consistent with the Form III assemblies being intermediate in size between the single-walled Form I filaments and the double-walled Form II filaments. If indeed the Form III assemblies are double-walled nanotubes, the smaller diameter of the Form III assemblies vis-à-vis the Form II assemblies is difficult to reconcile with the greater length of the Form III peptide. This difference must originate from differences in packing of the peptides in the corresponding helical assemblies of Form II and Form III. The Form II assemblies display an approximately circular cross-section due to curvature of the peptides within the three cross-helical stacks that comprise the filament. In contrast, the Form I assemblies display a nearly square cross-section that appears more compact. If the Form III assembly were to adopt a single-walled nanotube structure similar to the Form I assembly, we would anticipate that the diameter would be approximately 7.5 nm-based on the

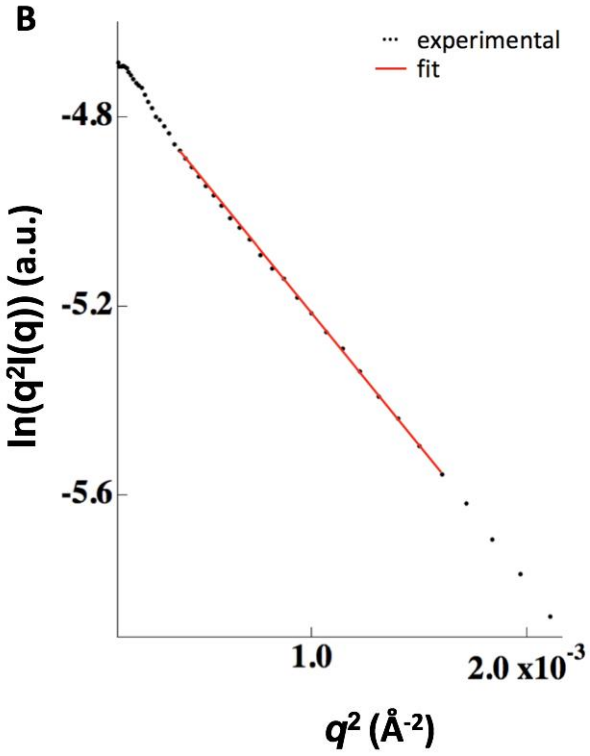


Figure 3.15: Modified Guinier plot of scattering data for Form III, 3 mg mL $^{-1}$ in aqueous solution with the Guinier region used for the diameter calculation indicated in red.

greater length of the Form III peptide sequence (36 residue/29 residue ~ 1.24 scaling factor). However, the apparent diameter derived from TEM, STEM, and SAXS measurements is approximately 10 nm. This value could be commensurate with a double-walled nanotube based on the Form I structure, as was similarly observed with the STEM data and power spectra results from TEM data.

3.4 Conclusion

Form III successfully implements an **arginine staple** as a critical structural motif for the self-assembly of alpha-stack tubular assemblies. This represents progress in expanding the library of structural tools in the pursuit of designing *de novo* one-dimensional assemblies with robust and predictable packing of subunits, which is still a significant challenge in peptide self-assembly. As was demonstrated with the original Forms (I and II), the plasticity of the quaternary tube structures remains a roadblock to highly accurate prediction of supramolecular structure from primary sequence.[4] Though based off both prior sequences (Form I and II), Form III was predicted to self-assemble in a fashion mimetic or similar to that of Form I, as it was thought the **arginine staple** would provide the majority of the driving force for association of helices in a side-to-end manner. However, the data for Form III deviate from what is hypothesized, but do indicate a combination of structural parameters of Form I and Form II. Data from TEM, STEM, and SAXS indicate that Form III tubes likely exist as a hybrid between the supramolecular structures of Forms I and II, rather than a tube which assembles mimetic to Form I with a larger cross-sectional diameter. This is not entirely unexpected, as Form III incorporates elements from the sequences of both Forms I and II. While the data corroborate implementation of an **arginine staple** within an alpha-helical peptide subunit provides a reliable structural motif for helical interactions, the resultant

supramolecular structure cannot be thoroughly predicted without full structural resolution via cryo-electron microscopy. Future data from cryo-electron microscopy and subsequent structural resolution will provide significant insight into the self-assembly of helical nanotubes based upon alpha-stacks, as well as insight into the use of an **arginine staple** as a structural motif in the pursuit of side-to-end association of alpha-helices within supramolecular structures. This **arginine staple** could be applied in the future to other helical designs, or, potentially to non-helical systems so long the two arginines are on the same face of the subunit and positioned reasonably close to one another.

3.5 Testing the Upper Size Limit of the Arginine Staple in Sequence Space: Form IV

3.5.1 Introduction and Sequence Design

Given the apparent success in implementing the previously described **arginine staple** as a key structural motif in the pursuit of self-assembly of peptide nanotubes, it is beneficial to know the limits of the staple as a driving structural component with regards terms of helix length. Longer helices inherently provide more surface area and potential interaction sites available for other subunits. Therefore, implementing longer constituent helices as the subunit for self-assembly of nanotubes tests the structural robustness of the **arginine staple** as it must contend with an increasing number competing interactions. To test this hypothesis, and whether or not the **arginine staple** remains a driving force (within a longer helix subunit) in self-assembly of peptide nanotubes similar to Form III nanotubes, a new sequence was engineered. Form IV was synthesized using microwave assisted solid-phase peptide synthesis and purified to homogeneity using reverse phase HPLC on a C-18 column. The sequence for the longer is shown in Figure 3.16

Form III: Acetyl-QAEIL**R**AYA**R**ILEADAEILKAQAKILEAHAEILKAQ-NH_2

Form IV: Acetyl-QAEIL**R**AYA**R**ILEADAEILKAQAEILKAQAKILEAHAEILKAQ-NH_2

Figure 3.16: the amino acid sequences of Form I and Form IV with the arginine staple highlighted in both.

The sequence for Form IV is largely conserved from Form III, with the key arginine staple motif remaining. The primary difference between the two variants is sequence length, with Form IV featuring an additional heptad occurring before the arginine staple, closest to the N-terminus.

Form IV features 43 total residues while Form III features 36. The isoelectric points of the peptides are nearly identical at 5.51 and 5.55, rendering potential supramolecular differences in structure unlikely to be caused by differences in peptide pI's. As was the case for Form III, Form IV is also chemically capped at both termini (amidated at the C-terminus and acetylated at the N-terminus).

3.5.2 Results and Discussion

3.5.2.1 Circular Dichroism

Initially, Form IV was self-assembled using the previously described assembly protocol for Form III, using TFE as a co-solvent with 10 mM acetate pH 4.0. Variants without TFE and annealed variants were also self-assembled. It is vital to test the secondary structure of Form IV to ensure that it is similarly alpha-helical in character to Form I, as was the case for Form III. Resultant supramolecular assemblies for Form III and IV cannot be reasonably compared in order to test the sequence-length hypothesis if Form IV does not adopt an alpha-helical secondary structure like the preceding Form variants. Circular dichroism spectropolarimetry of an aqueous solution of Form III in identical buffer conditions to Form III (10 mM acetate, pH 4.0, TFE evaporated) confirm that the secondary structure of Form IV is alpha-helical. As it is a longer helix, it is hypothesized that Form IV should demonstrate an alpha-helical signature higher in intensity than that of Form III. This is confirmed and is shown in Figure 3.17. Three variants are tested, all in acetate buffer (10mM, pH 4.0): a variant assembled with TFE and not annealed, a variant assembled without TFE and not annealed, and a sample assembled without TFE and annealed from 90 °C to 22 °C at a rate of -1 °C / 5 minutes. Each variant demonstrated alpha-helical secondary character, with the TFE variant demonstrating the highest intensity signature. This is not unexpected, as TFE is a known helical promoter.[45] Importantly, the intensity of all the

signals is equal to or greater than what is observed for Form III for all the Form IV variants, and greater than the intensity for all the Form I and II variants, which matches with what is expected given the greater helix length.

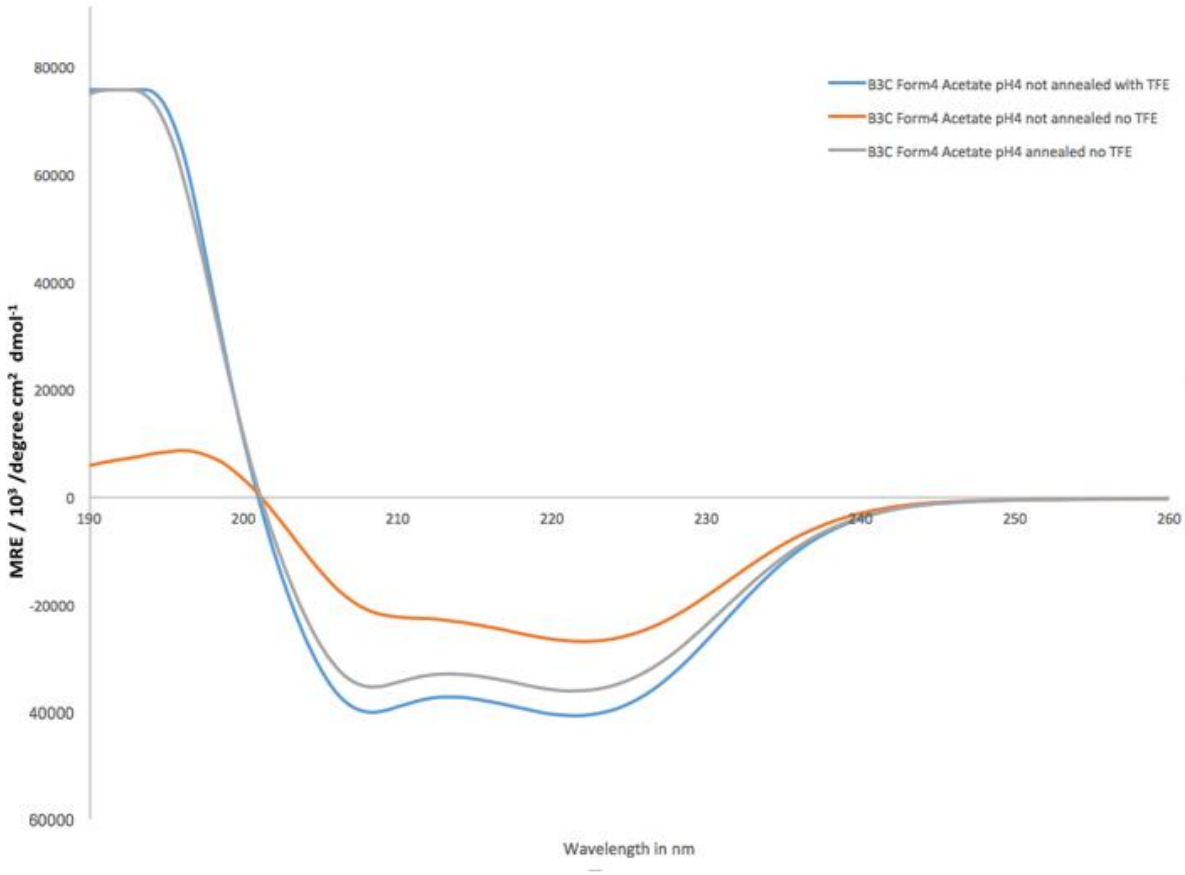


Figure 3.17: circular dichroism results for Form IV in 10 mM acetate (pH 4.0) with TFE (blue), without TFE, not annealed (blue), and without TFE, annealed (orange).

3.5.2.2 Transmission Electron Microscopy

Transmission electron microscopy results indicate that the **arginine staple** sequence Form IV does not self-assemble into nanotubular assemblies. Negatively stained transmission electron micrographs are shown in Figure 3.18 for 3 mg mL^{-1} solutions of Form IV in 10 mM acetate buffer

(pH 4.0) both with and without TFE and both annealed and not annealed. None of the aforementioned assembly conditions produced self-assemblies of nanotubular, or even filamentous, morphology. Peptide aggregates are observed for these conditions, wherein no degree of higher order is observed. Of note, a pH range from 4.0 to 8.0 (increasing in increments of 0.5) was tested to determine if Form IV demonstrated ordered, tubular self-assembly at other pH values. The only assembly condition with discernible order via TEM was found to be 3 mg mL^{-1} in

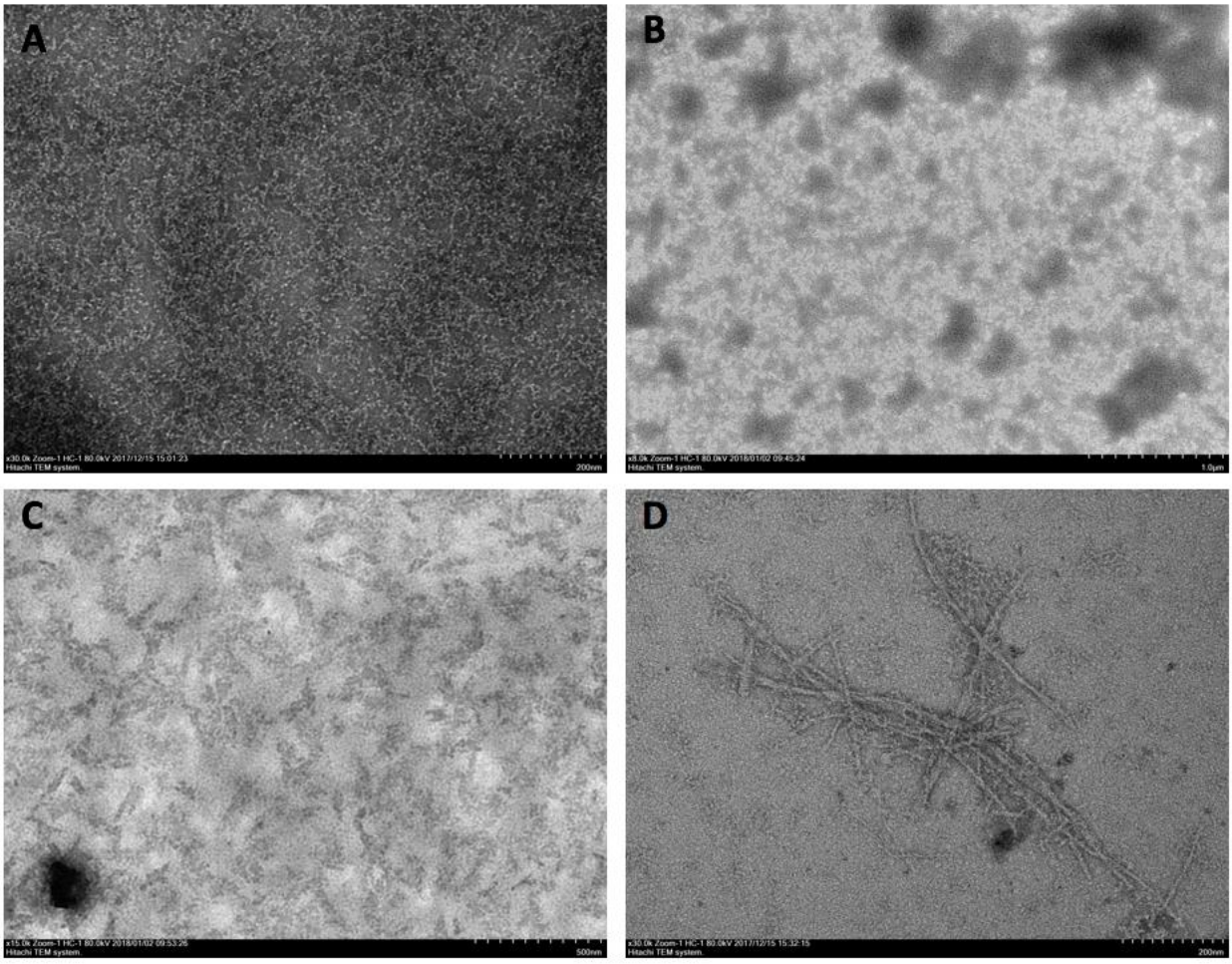


Figure 3.18: transmission electron micrographs for Form IV in 10 mM acetate (pH 4.0) with variants (A) not annealed and in the absence of TFE, (B) annealed and in the absence of TFE, and (C) not annealed with TFE and (D) in 10 mM TAPS (pH 7.5) annealed.

tris(hydroxymethyl)methylamino]propanesulfonic acid (TAPS) buffer (10 mM, pH 7.5) thermally annealed from 90 *°C to 22 °C* at a rate of -1 °C/5 minutes. This sample demonstrated a heterogeneous mixture of globular aggregate and short fibrillar assemblies with relatively low persistence lengths and little overall order. All assemblies were tested at time intervals up to one month, with no tubular assemblies discerned over this time span for all conditions tested. The TEM data indicate that a size limit may very well exist for the **arginine staple** as a driving force for self-assembly, at least within the confines of an alpha-helical structural subunit.

3.5.3 Conclusion

The data for **Form IV** intimate that there does exist a size limit for alpha-helices for which the **arginine staple** cannot out-compete other potential subunit interactions to drive the formation of nanotubular assemblies. The lack of supramolecular self-assembly (at least as it was hypothesized) could be attributed to myriad reasons. One potential reason for the observed lack of tubular self-assembly for Form IV may rely on the relatively small size of the **arginine staple** in comparison to the total sequence length. Further, though the CD data indicate a high degree of helicity for Form IV, if a segment of the sequence (particularly a segment occurring between the C-terminus and the **arginine staple** exists in a hon-helical state (i.e. random coil), this could position the **arginine staple** such that it cannot interact with the C-terminus of neighboring helices. Lastly, increasing the helix length may also increase the inherent flexibility of the resultant alpha-stack (hypothesized to comprise four alpha-helices, as in Form I nanotubes). This increased flexibility could also position the helices such that they cannot stack in a thermodynamically stable fashion along the fiber axis.

3.6 Form 0: Testing the Arginine Staple in Shorter Helices

3.6.1 Introduction and Sequence Design

The structural parameters and critical sequence components for Form I have been established, with the **arginine staple** motif being identified as a key structural linchpin in the self-assembly of the four-start superhelical Form I nanotube. Recent work by Joel Collier's group into the capacity for epitope labelling of Form I nanotubes (vide supra) has also inspired further exploration into what is required in sequence-space for Form I to self-assemble. Specifically, the Collier group examined the effects of the elimination of portions of the solvent-facing portion of the largely palindromic alpha-helix in the Form I nanotube structure.[46] The work of the Collier group implemented a 23 residue alpha-helical sequence based on Form I (Coil23). Similarly, their work using Form I as an epitope scaffold was termed Coil29. Using this work as inspiration, we aimed to test how important the solvent-facing portion of the alpha-helix is in the self-assembly of Form I mimetic nanotubes. In this pursuit, seven residues are deleted from the solvent-facing portion of the helix, including three hydrophilic residues, which could prove critical to the stability of the Form I structure. Form 0 is engineered, having 22 total residues (one full heptad less than Form I) to test the how critical the solvent-facing portion of Form I helices are to the self-assembly and subsequent stability of nanotubular assemblies. This deletion may subsequently test the structural robustness of the **arginine staple** as a driving force in self-assembly.

Form I: Acetyl-QARILEADAEIL**R**AYA**R**ILEAHAEILRAQ-NH_2

Form 0: Acetyl-QAEIL**R**AYA**R**ILEAHAEILRAQ-NH_2

Figure 3.19: amino acid sequence of Form 0 and its parent peptide sequence Form I, with the arginine staple indicated in bold and the deleted residues indicated in orange.

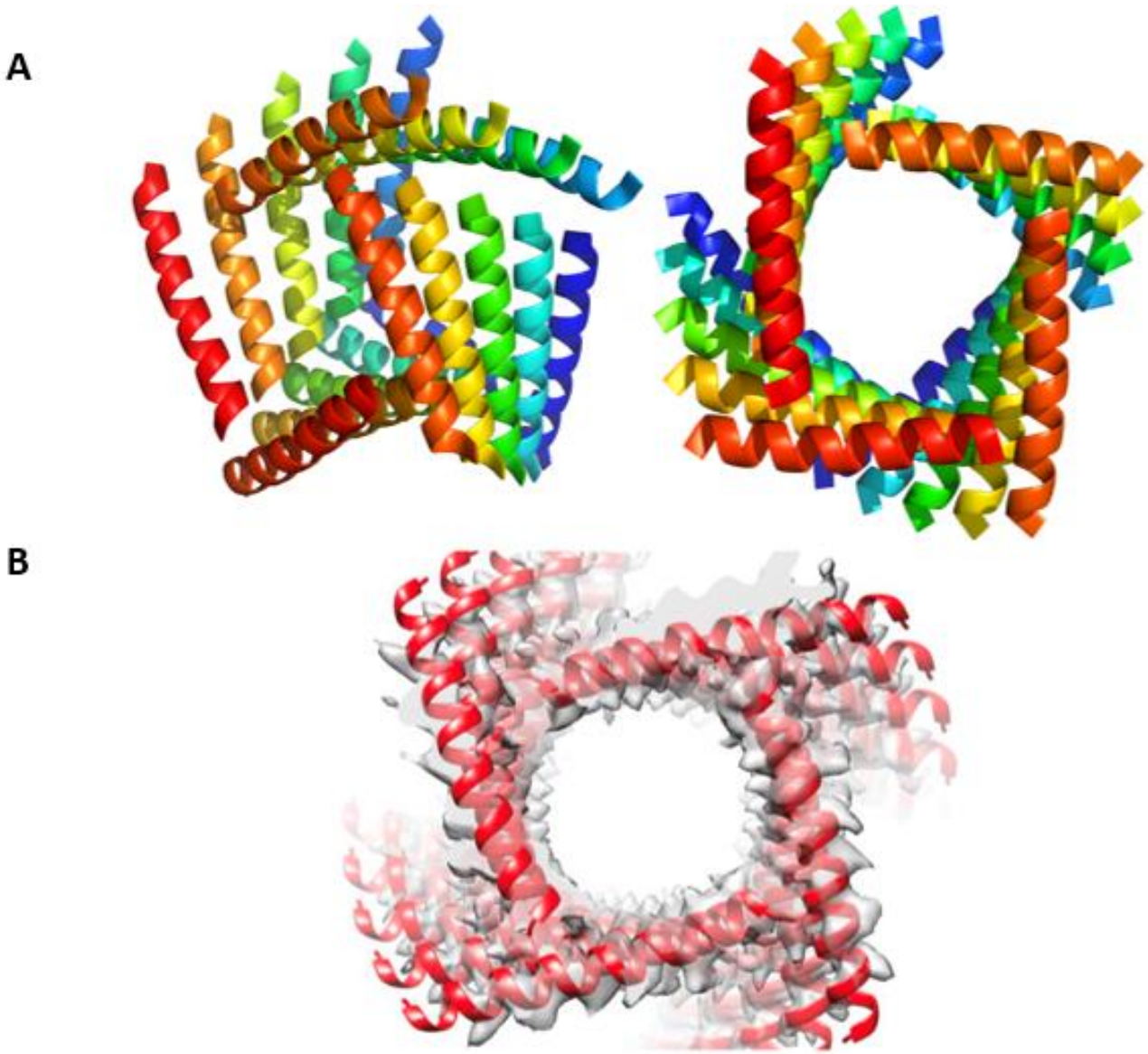


Figure 3.20: (A) PyMol rendering of the hypothesized nanotubular structure from the self-assembly of Form 0 helices, with a much shorter helix length at the solvent exposed region of the tube than that of Form I (B).

3.6.2 Results and Discussion

3.6.2.1 Circular Dichroism

Circular dichroism spectropolarimetry was employed to probe the effects of the sequence mutation on resultant secondary structure. CD data indicate that Form 0 (assembled at 3 mg mL^{-1} in 10 mM acetate, pH 4.0, diluted to ~100 µM for experimentation) maintains some of the alpha-helical character of Form I, with significant changes between the two forms occurring at the minima at 208 nm. It is anticipated that, given the shorter sequence of Form 0 when compared to Form I, a lower alpha-helical signal intensity will occur. Form 0 demonstrates a much weaker intensity at the 208 nm minimum and the 195 nm maximum, indicating that the Form 0 peptide

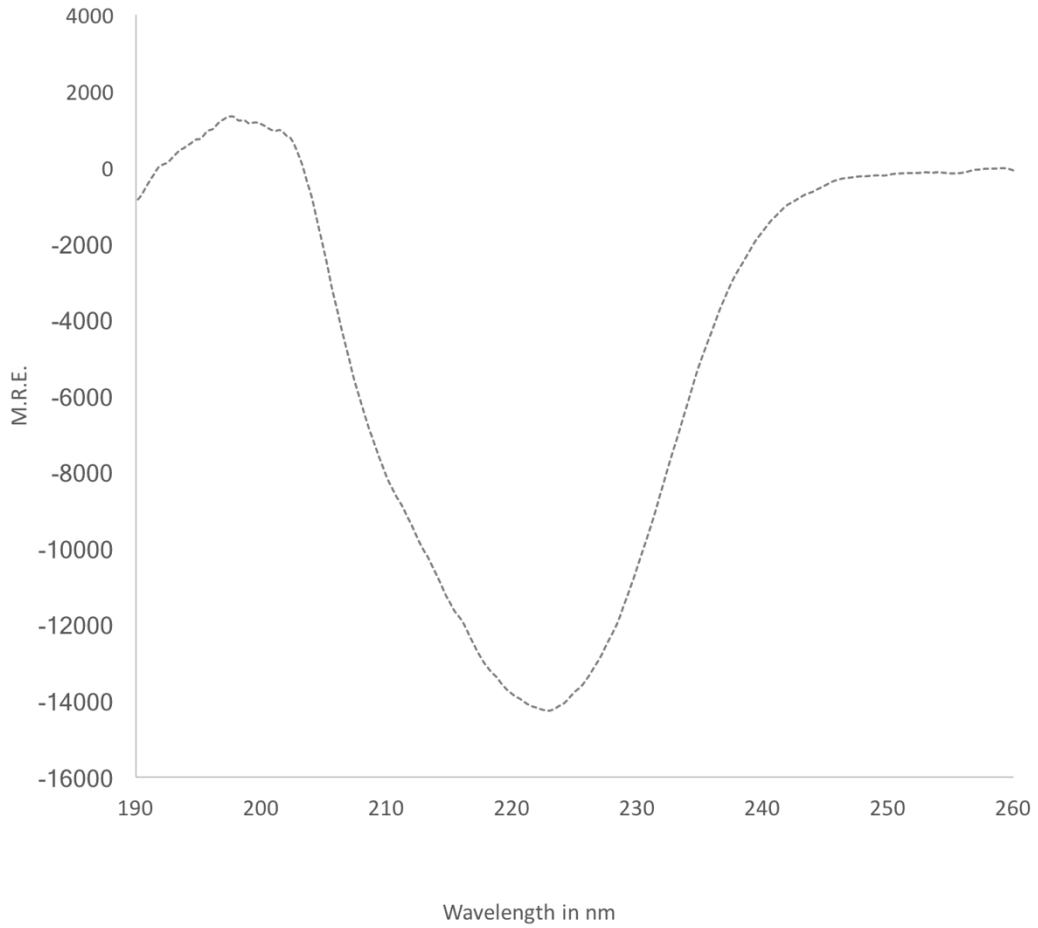


Figure 3.21: circular dichroism results for Form 0 self-assembled in 10 mM acetate, pH 4.0.

may be a hybrid of multiple secondary structures. This may indicate that deleting a portion of the solvent-facing segment of the helix may affect the resultant stability and identity of the remainder of the sequence. This transition in secondary structure as witnessed via CD was also observed in the Coil23 sequence made and characterized by the Collier Group.

3.6.2.2 Transmission Electron Microscopy

TEM data indicated the presence of high aspect-ratio fibrillar assemblies from the self-assembly of Form 0 peptide at 3 mg mL^{-1} in 10 mM acetate buffer (pH 4.0). The resultant fibrillar structures did not change when TFE was used in the assembly process or when the assembly solution was thermally annealed from 90 °C to 22 °C at a rate of -1 °C / 5 minutes. No stain permeation into the fiber was observed, indicating that the filaments are either non-tubular in morphology (thus lacking a lumen), or that the filaments are tubular but the lumen diameter is sufficiently small enough to prohibit the entrance of stain into the fiber. This is not uncommon for

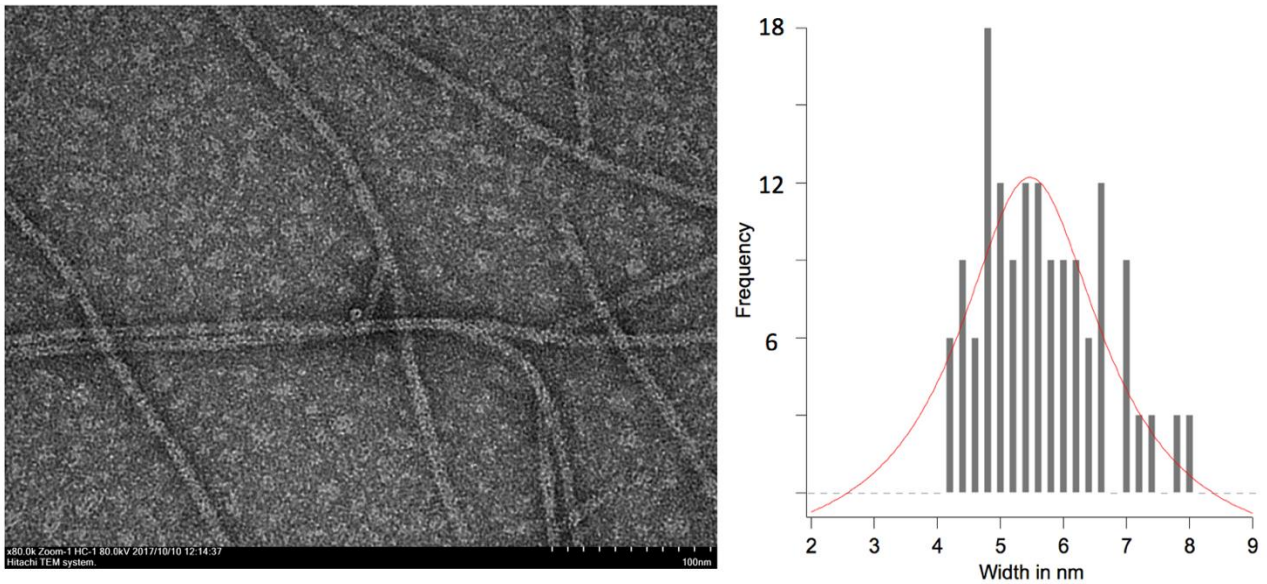


Figure 3.22: negatively stained transmission electron micrograph of Form 0 filaments and a histogram of measured fibril widths via ImageJ.

peptide-based tubular structures with lower diameters. Fibril widths were measured in ImageJ and the apparent width average was calculated to be 5.76 ± 0.95 nm. Fifty data points across three independently assembled samples were used to derive this value. This value is in line with what is hypothesized. Given that the single-walled Form I fiber diameter is governed predominantly by the length of the constituent helix (due to the nearly perpendicular packing of the alpha-helix stacks to the fiber axis), shortening the helix length should shorten the resultant tube diameter assuming Form 0 assembles mimetically to its parent peptide Form I. The sequence length of Form 0 is 79% of the length of Form I. The resultant fiber diameters are ~5.76 nm and ~6.00 nm, making the diameter for Form 0 tubes 96% of the diameter for Form I. Though this does not match identically to the constituent helix length reduction, the transmission electron microscopy data do corroborate that the Form 0 peptide will still self-assemble into fibers with resultant tube diameter less than that of the parent peptide. The discrepancy in the predicted vs. measured diameter values could be attributed to a non-circular, non-uniform cross section of the tubes when analyzed via TEM. In other words, the tubes could be flattening, and thus widening in cross section with respect to the viewing angle of TEM, as a result of deposition of sample onto the TEM grid.

3.6.2.3 Small-Angle X-Ray Scattering

Synchrotron small-angle X-ray scattering measurements were performed on the Form 0 assemblies and the scattered intensity profiles were measured as a function of the momentum transfer, q, within the range from .007 $Å^{-1}$ to 2.3 $Å^{-1}$. Similarly to the Form I, Form II and Form III, the scattered intensity displayed a linear dependence on q^{-1} within the power law region (Figure 3.24). The mass fractal value of -1.402 on q is consistent with extended rod-like assemblies, though not as explicitly close to a mass fractal value of 1 as the previous Forms. Notably, the SAXS data

for Form 0 is lacking Bragg diffraction peaks, and does not show the typical oscillation found in the mid q region that Forms I-III do. The fitting of low q data using the modified Guinier equation for rod-like forms produced an R_c value of 26.3 ± 4.44 Å, yielding a filamentous diameter of ~74 Å. This value is larger than what is observed for TEM, indicating the adsorption of Form 0 filaments to the grid may affect the cross-sectional geometry or that the TEM measured results do not include side-chain volume. This larger SAXS derived diameter may also mirror what was seen in prior Form peptides which employ the **arginine** staple, a potential concentration-dependent switch from single-walled to double-walled tubes. Overall, the SAXS data indicate rod-like

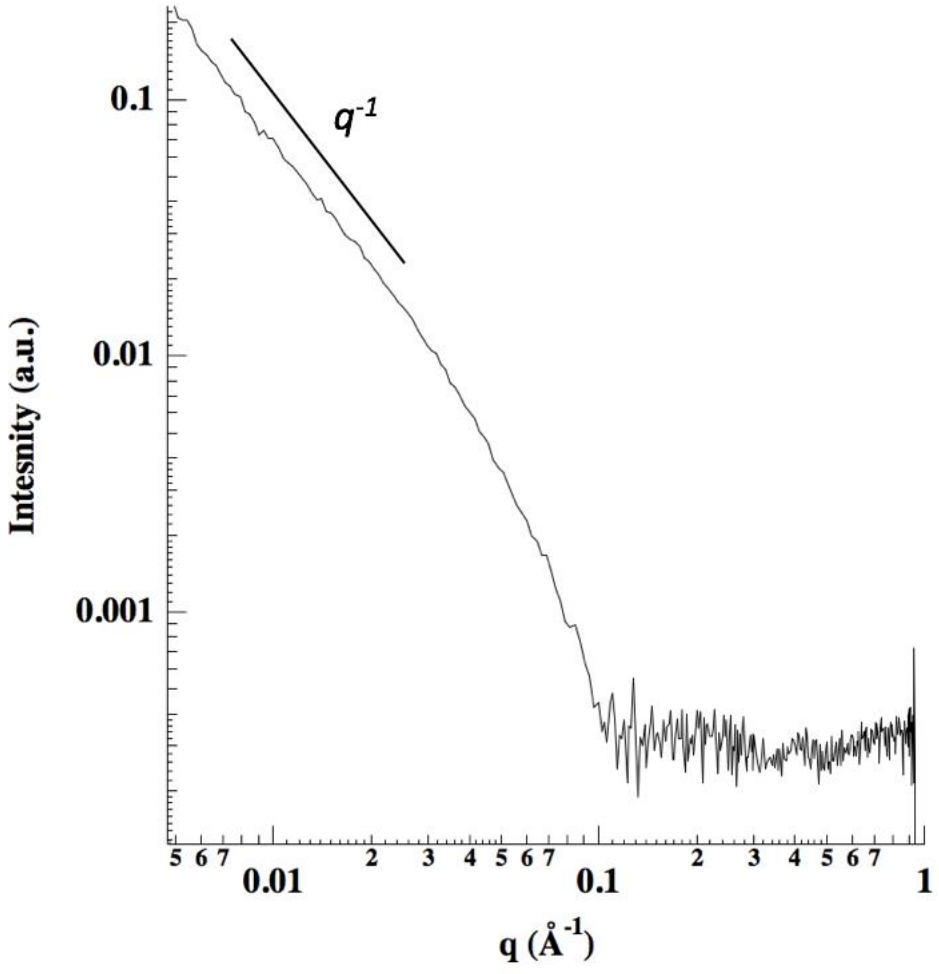


Figure 3.23: Small-angle X-ray scattering curve for Form 0 in 10 mM acetate, pH 4.0.

assembly of Form 0 helices, with a calculated diameter approximately equivalent to what is observed via TEM. The lack of diffraction in the high *q* region, however, prohibits the understanding of finer structural detail in the Form 0 nanotubes, subsequently also prohibiting comparison to nanotubes assembled from Forms I-III.

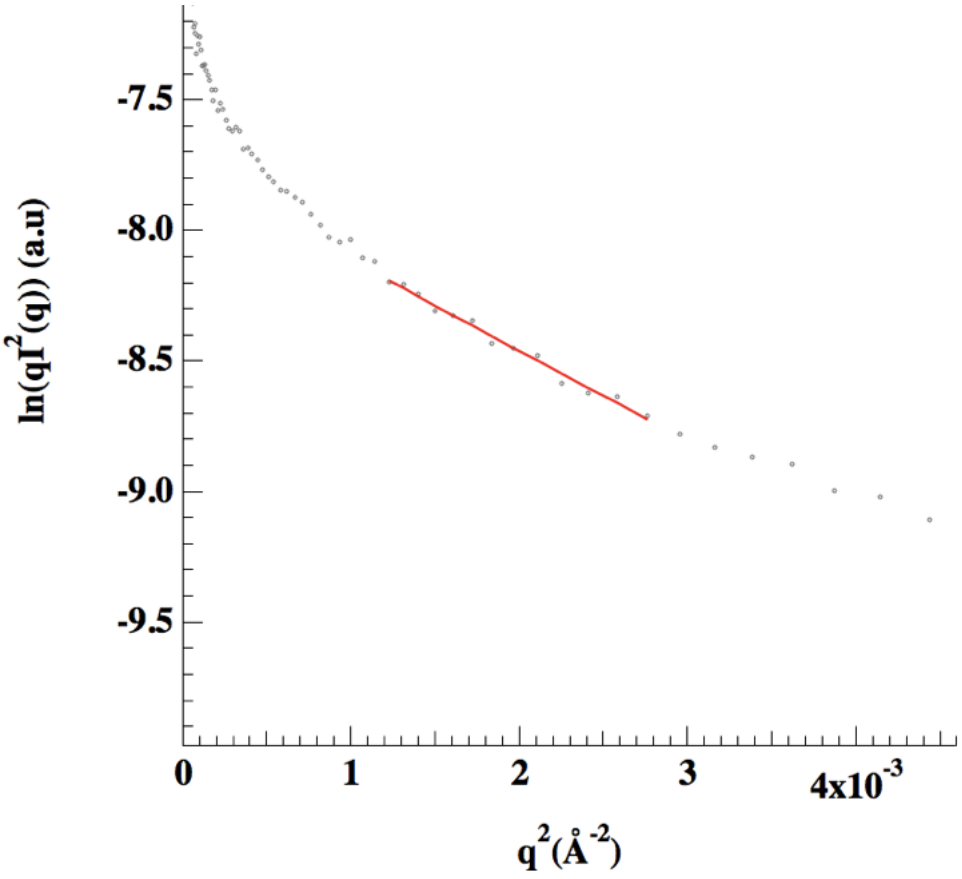


Figure 3.24: Modified Guinier plot of scattering data for Form 0 with the region used to calculate the R_c value from the modified Guinier equation for rod-like forms indicated in red.

3.6.4 Conclusion

The data for Form 0 confirm that, while the **arginine staple** did not maintain its capacity for driving self-assembly of nanotubes in the context of longer constituent helices, it does work in smaller helical sequences, at least when the distance from the c-terminus to the **arginine staple** is maintained. The Form 0 data also corroborate that the solvent-facing portion of the helix is not paramount and required for the self-assembly of filaments, further indicating that the **arginine staple** is likely the predominant driving force in self-assembly. This work is already being applied as a scaffold for epitope-labeling for viral response studies in murine samples.[46] Moreover, this work sets a foundation for numerous downstream applications wherein the exterior, solvent-facing area of the nanotubular structure is replaced with functional moieties, given that said moieties do are sufficiently hydrophilic and do not alter the peptide's capacity to self-assemble into nanotubular assemblies.

3.7 Methods

3.7.1 Materials. Unless otherwise stated, all chemical reagents were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO) or Anaspec, Inc. (Fremont, CA). All peptides were synthesized by and purchased from Genscript, USA (Fremont, CA) or from Synpeptide, China.

3.7.2 Peptide Self-Assembly. As previously stated, peptides were purchased from GenScript, USA (Fremont, CA) or Synpeptide (China). Peptides were purified via reverse phase high performance liquid chromatography (RP-HPLC) using a C-18 column and a gradient of water (0.065% trifluoroacetic acid)-acetonitrile (0.05% trifluoroacetic acid). Peptides were modified at the N- and C-terminus with acetyl (CH_3CO) and methyl-amide (NHMe) capping groups

respectively. 10 mM acetate buffer pH 4.0 was added to the lyophilized GenScript peptides at approximately 3 mg mL^{-1} peptide concentrations for all experimentation except for small angle x-ray scattering (SAXS), for which peptides were assembled at ≥ 4 mg mL^{-1}. Peptide concentrations were determined spectrophotometrically from measurements of the absorbance at 280 nm (A_{280}). For peptides containing Tyr, Trp, or Cys residues, the peptide concentration can be calculated from Equation 1:

$$MW \times (A_{280}/c) = 1280n_Y + 5690n_W + 120n_C \quad \text{(Equation 1)}$$

in which c is the concentration of peptide in mg mL^{-1}, and n_Y, n_W, and n_C are the numbers of tyrosine, tryptophan and cysteine residues, respectively, in the peptide sequence (Gill and von Hippel, 1989). 6M guanidinium chloride is added in a 1:9 v/v ratio to each aqueous specimen solution before CD analysis to eliminate error in determination of absorbance due to the scattering of UV light due to peptide self-assembly. The **Form IV** peptide was found to assemble optimally

in 10 mM acetate at pH 4.0 with a 1:1

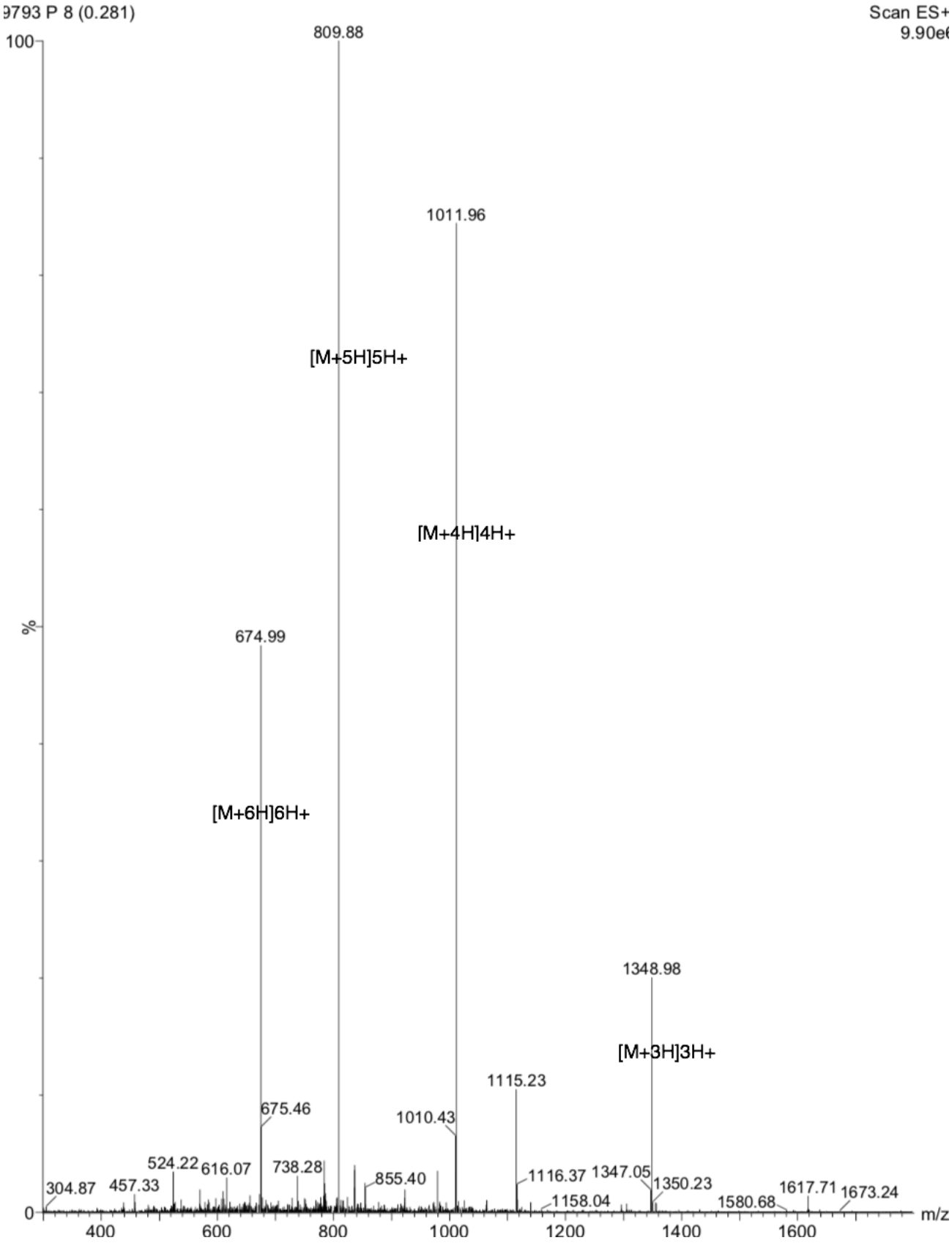


Figure 3.25: ESI-Mass Spectrometry confirmation of the mass for the Form III peptide.

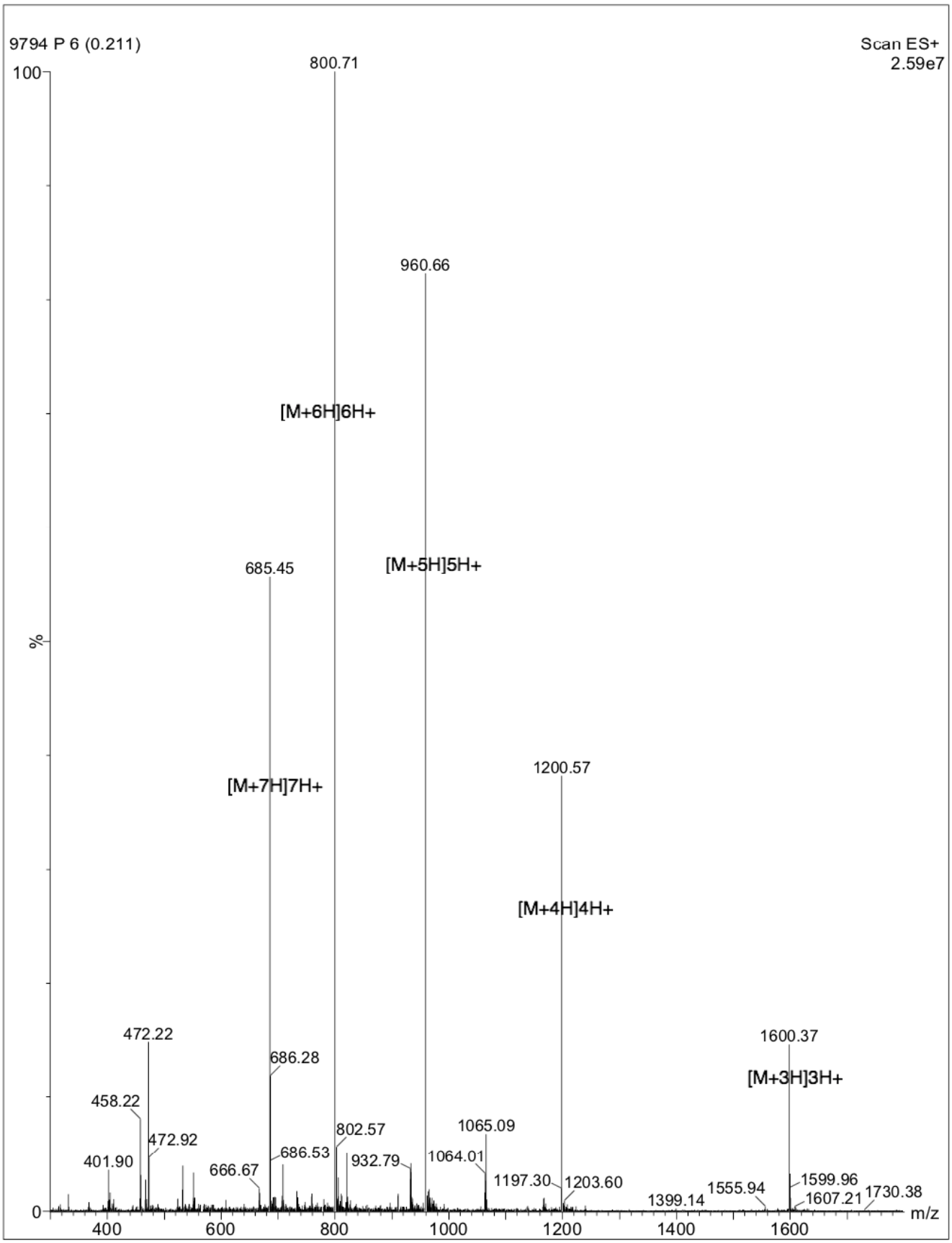


Figure 3.26: ESI-mass spectrometry confirmation of the mass for Form IV.

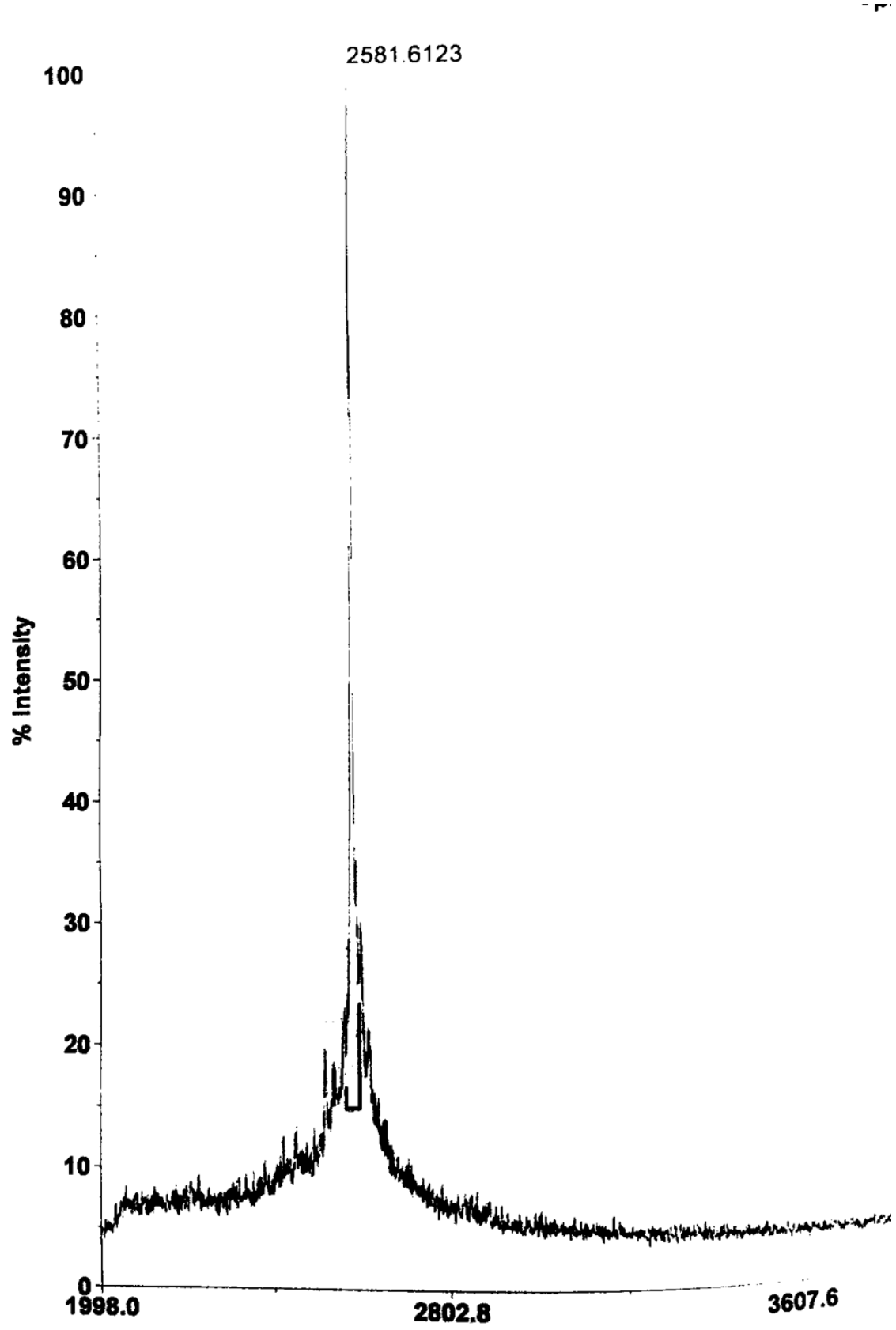


Figure 3.27: MALDI-Mass Spectrometry confirmation of the mass for the Form 0 peptide.

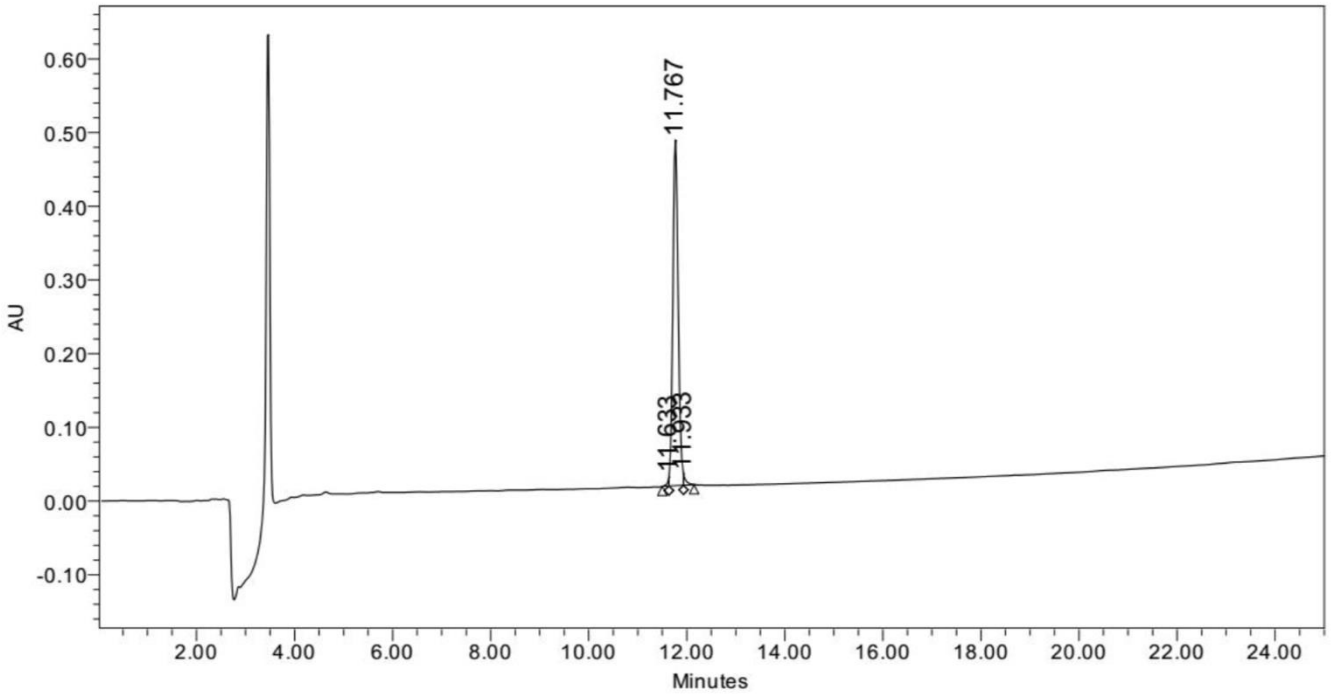


Figure 3.28: Analytical HPLC trace for Form III indicating purity.

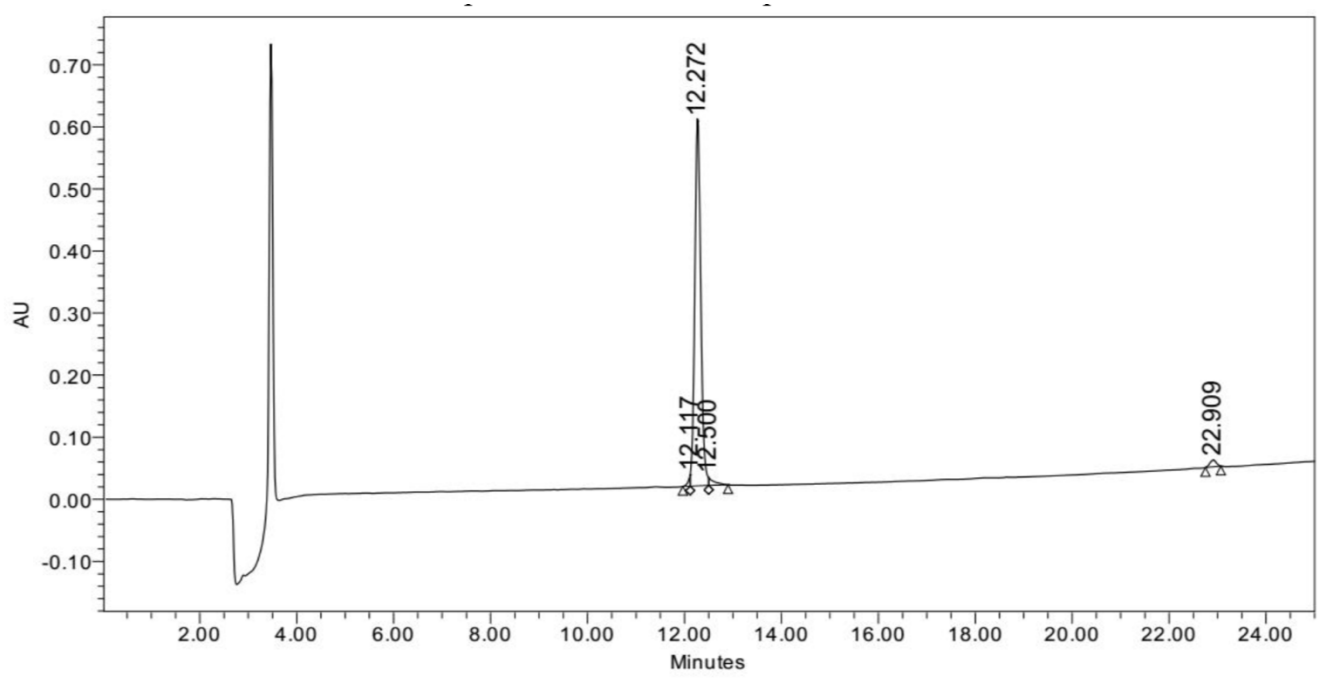


Figure 3.29: Analytical HPLC trace for Form IV indicating purity.

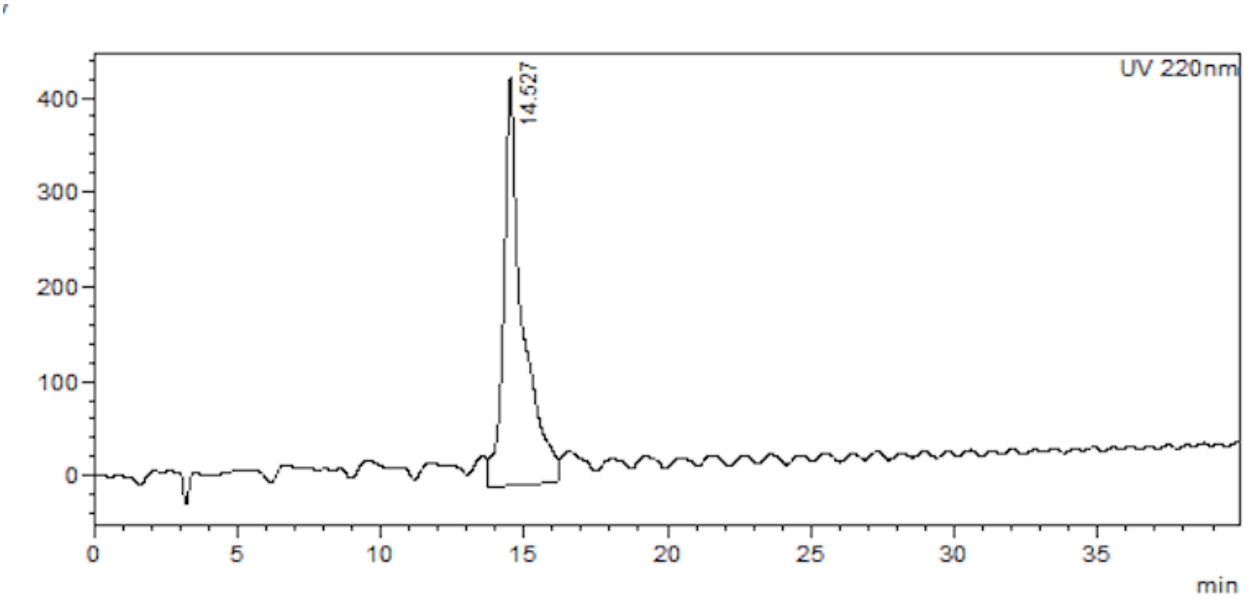


Figure 3.30: Analytical HPLC trace for Form 0 indicating purity.

addition of 2,2,2,-trifluoroethanol (TFE) to promote helicity and solubility of **Form III**. (TFE was not required in the assemblies of the other Forms.) TFE was subsequently allowed to evaporate sufficiently such that the volume of the combined peptide assembly buffer and TFE was reduced by half, typically taking between 1 and 5 days depending on total volume. It is assumed that the aforementioned halving of the original assembly solution volume indicates near full evaporation of TFE, and that the TFE-acetate buffer mixture did not evaporate azeotropically to a significant degree. Eppendorf microcentrifuge tubes containing the assembly solutions + TFE were covered with perforated Parafilm ® to decrease chances of environmental contamination. The Parafilm ® was pierced once using a sterile needle, allowing for controlled evaporation of TFE. Following TFE addition, the assembly solution was incubated at room temperature for several weeks, with fibers visible via TEM as early as 1 day. No evidence of sample degradation was witnessed via TEM over a time scale of several years. **Form IV** was assembled similarly to **Form III,** however,

additional assembly conditions were tested for **Form IV**, including annealing the assembly solution from 90 °C to 22 °C at a rate of -1 °C / 5 minutes, and testing across a pH range from 4.0 to 8.0 at pH increments of 0.5. Form 0 can be assembled via either of the methodologies described above. It also demonstrated self-assembling structures which did not require the presence of TFE. Form 0 exhibits identical tube morphology regardless of TFE addition. Further, form 0 is tolerant of annealing (90 C to room temperature at a rate of 0.2 C/minute).

3.7.3 Circular Dichroism. Circular dichroism analyses of Form assemblies were performed using a JASCO J-810 circular dichroism spectropolarimeter. Peptide solutions were analyzed at a peptide concentration of ~100uM for both variable-wavelength scans and variable-temperature scans. For variable wavelength scans, Spectra were collected at 50nm/min from 190-260 nm with a data pitch of 0.2 nm. For variable temperature scans, spectra were collected at 222nm, with a 60°C/hour temperature progression, and a data pitch of 0.5°C. Variable temperature spectra were collected from 25-85°C. Quartz plates (0.10 mm thickness) from Hellma Analytics (Müllheim, Germany) were employed for all CD experiments.

3.7.4 Flow Linear Dichroism. Flow linear dichroism spectra were recoreded using a JASCO J-810 circular dichroism spectropolarimeter using an LD channel and a microvolume cuvette with a path length of 50 microns and a rotation speed of 5,000 rpm to establish Couette flow. The background scattering for each sample was obtained from the LD spectra of samples at 0 rpm. The LD spectra were taken after 15 minutes of rotation and were recorded using a wavelength span from 300 nm to 190 nm.

3.7.5 Transmission Electron Microscopy. Grids were prepared by depositing 4 μL of peptide solution (0.3 mg mL^{-1} peptide concentration in 10 mM acetate buffer pH 4.0) to 200 mesh carbon-coated copper grids purchased from Electron Microscopy Services (Hatfield, PA). The solution was allowed to incubate on the grid for 45 seconds before being wicked away using 55mm qualitative filter paper. The peptide solution was wicked from the grid surface in such a way that a thin film of liquid remained, then 4 μL of freshly prepared, filtered (0.2 micron Whatman filter from G.E. Healthcare Services) 1% aqueous uranyl acetate solution was applied to the grid to mix with and negatively stain the sample. After 45 seconds of incubation, the stain was fully wicked from the grid using the same methodology (*vide supra*). Grids were stored in a tabletop desiccator and sufficiently dried. All transmission electron microscopy was carried out on a JEOL JEM-1400 transmission electron microscope at an accelerating voltage of 120kV. Power spectra and reference-free averages were produced using EMAN2 software bundle. For Form III, 25 negatively stained micrographs were used to box out >150 segments of filaments to be used for the power spectrum and reference-free average. The electron microscopy data described here were gathered on a JEOL JEM-1400 (supported by National Science Foundation Major Research Instrumentation Grant 0923395). This study was in part supported by the Robert P. Apkarian Integrated Electron Microscopy Core (RPAIEMC) which is subsidized by the Emory College of Arts and Sciences and the Emory School of Medicine and is one of the Emory Integrated Core facilities. Additional support was provided by the National Center for Advancing Translational Sciences of the National Institutes of Health under award number UL1TR000454. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institutes of Health.

3.7.6 Scanning Transmission Electron Microscopy. STEM data were acquired at Brookhaven National Laboratory (BNL). Assembly solutions for each variant (0.3 mg mL^{-1} in 10 mM acetate buffer, pH 4.0) were deposited on thin carbon (ca. 2 nm thick) supported on a thicker holey carbon film mounted on a titanium grid using the wet-film, hanging drop method. Tobacco Mosaic Virus (TMV) is added to the grid as an internal control, then injection buffer is added, followed by specimen solution. These solutions are allowed to sit for one minute, and excess solution is wicked from the edge with filter paper. Following The grid is fast frozen by plunging into liquid nitrogen slush and stored under liquid nitrogen. All grids are freeze-dried overnight in an ion-pumped chamber with an efficient cold trap. Grids are then transferred under vacuum to the STEM cold stage, which is operated at -160 °C. Measurements were taken at 40keV with a scanning probe of <0.3 nm directed from a cold field emission source. Data was collected using bright field, small-angle dark field, and large-angle dark field scintillator photomultiplier detectors which operate at 0-15, 15-40, and 40-200 mRadian respectively. The large-angle signal is proportional to the mass of the atoms in the path of the beam. Specimen quality was tested and mass calibration was standardized using TMV. Mass per length (M/L) values were calculated for **Form III** (1,246 data points) using aforementioned TMV raft standards set at a theoretical m/L of 13.1 kDa/Å by using the software 'PCMass32' [47-48]. STEM data for **Form IV** were not obtained, as the beam caused significant sample degradation during experimentation, as decreasing signal intensity with each subsequent pulse was observed.

3.7.8 Small- and Wide-Angle X-ray Scattering Measurements. Synchrotron SAXS/WAXS analyses were performed at the Advanced Photon Source at Argonne National Laboratory (Beamline 12-ID-B). Assembly solutions of 4 mg mL^{-1} in 10mM acetate pH 4.0 were dialyzed

against pure 10mM acetate pH 4.0 such that the dialyzed sample is identical to the buffer standard in all manners except for peptide concentration. To achieve this, Slide-A-Lyzer dialysis cassettes (2,000 MWCO) from Thermo-Fisher were used to dialyze 0.5 mL of peptide assembly solution against buffer containing no peptide. SAXS/WAXS analysis was performed at 25 °C in a 1.5 mm quartz capillary flow cell (applying constant vertical agitation of the peptide assembly solution) so as to prevent radiation damage. For each Form variant, a minimum of thirty 2-dimensional shots were gathered and azimuthally averaged into 1-dimensional scattering spectra. Identically gathered data was obtained for the dialysis buffer (10mM acetate pH 4.0, no peptide present) and data were background subtracted to produce signal correlating exclusively to peptide-based assemblies in solution. Normalization against beam intensity and solid angle correction were performed by the in-house software suite (Igor Pro) present at Beamline 12-ID-B. Radii were calculated using Equation 2:

$$R_c = R_{exp}/\sqrt{2} \text{ (or) } R_{exp} = R_c * \sqrt{2}. \qquad \text{(Equation 2)}$$

3.8 REFERENCES

1. Acar, H.; Srivastava, S.; Chung, E. J.; Schnorenberg, M. R.; Barrett, J. C.; LaBelle, J. L.; Tirrell, M., Self-assembling peptide-based building blocks in medical applications. *Advanced drug delivery reviews* **2017,** *110-111*, 65-79.

2. Ekiz, M. S.; Cinar, G.; Khalily, M. A.; Guler, M. O., Self-assembled peptide nanostructures for functional materials. *Nanotechnology* **2016,** *27* (40), 402002.

3. Ramakers, B. E.; van Hest, J. C.; Lowik, D. W., Molecular tools for the construction of peptide-based materials. *Chemical Society reviews* **2014,** *43* (8), 2743-56.

4. 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.

5. Xu, C., Programmed Self-Assembly of Coiled-Coil Peptides. *Dissertation* **2013,** (Emory University).

6. Cormier, A. R.; Pang, X.; Zimmerman, M. I.; Zhou, H. X.; Paravastu, A. K., Molecular structure of RADA16-I designer self-assembling peptide nanofibers. *ACS nano* **2013,** *7* (9), 7562-72.

7. Nagy-Smith, K.; Moore, E.; Schneider, J.; Tycko, R., Molecular structure of monomorphic peptide fibrils within a kinetically trapped hydrogel network. *Proc Natl Acad Sci U S A* **2015,** *112* (32), 9816-21.

8. Lee, M.; Wang, T.; Makhlynets, O. V.; Wu, Y.; Polizzi, N. F.; Wu, H.; Gosavi, P. M.; Stohr, J.; Korendovych, I. V.; DeGrado, W. F.; Hong, M., Zinc-binding structure of a catalytic amyloid from solid-state NMR. *Proc Natl Acad Sci U S A* **2017,** *114* (24), 6191-6196.

9. Chen, K. H.; Corro, K. A.; Le, S. P.; Nowick, J. S., X-ray Crystallographic Structure of a Giant Double-Walled Peptide Nanotube Formed by a Macrocyclic beta-Sheet Containing Abeta16-22. *J Am Chem Soc* **2017,** *139* (24), 8102-8105.

10. Adamcik, J.; Mezzenga, R., Amyloid Polymorphism in the Protein Folding and Aggregation Energy Landscape. *Angew Chem Int Ed Engl* **2018,** *57* (28), 8370-8382.

11. Guenther, E. L.; Ge, P.; Trinh, H.; Sawaya, M. R.; Cascio, D.; Boyer, D. R.; Gonen, T.; Zhou, Z. H.; Eisenberg, D. S., Atomic-level evidence for packing and positional amyloid polymorphism by segment from TDP-43 RRM2. *Nature structural & molecular biology* **2018,** *25* (4), 311-319.

12. DiMaio, F.; Song, Y.; Li, X.; Brunner, M. J.; Xu, C.; Conticello, V.; Egelman, E.; Marlovits, T.; Cheng, Y.; Baker, D., Atomic-accuracy models from 4.5-A cryo-electron microscopy data with density-guided iterative local refinement. *Nat Methods* **2015,** *12* (4), 361-365.

13. 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.

14. Walshaw, J.; Woolfson, D. N., Open-and-shut cases in coiled-coil assembly: α-sheets and α-cylinders. *Protein Sci* **2001,** *10* (3), 668-73.

15. Lyu, P. C.; Gans, P. J.; Kallenbach, N. R., Energetic contribution of solvent-exposed ion pairs to alpha-helix structure. *J Mol Biol* **1992,** *223* (1), 343-50.

16. Huyghues-Despointes, B. M.; Scholtz, J. M.; Baldwin, R. L., Helical peptides with three pairs of Asp-Arg and Glu-Arg residues in different orientations and spacings. *Protein Sci* **1993,** *2* (1), 80-5.

17. Scholtz, J. M.; Qian, H.; Robbins, V. H.; Baldwin, R. L., The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide. *Biochemistry-Us* **1993,** *32* (37), 9668-76.

18. Cid, H., Gazituia F., Bunster, M., Interactions that regulate the helical fold in proteins. *Biol. Res.* **1996,** (29), 213-225.

19. Andrew, C. D.; Penel, S.; Jones, G. R.; Doig, A. J., Stabilizing nonpolar/polar side-chain interactions in the alpha-helix. *Proteins* **2001,** *45* (4), 449-55.

20. Shi, Z.; Olson, C. A.; Kallenbach, N. R., Cation-pi interaction in model alpha-helical peptides. *J Am Chem Soc* **2002,** *124* (13), 3284-91.

21. Andrew, C. D.; Bhattacharjee, S.; Kokkoni, N.; Hirst, J. D.; Jones, G. R.; Doig, A. J., Stabilizing interactions between aromatic and basic side chains in alpha-helical peptides and proteins. Tyrosine effects on helix circular dichroism. *J Am Chem Soc* **2002,** *124* (43), 12706-14.

22. Butterfield, S. M.; Patel, P. R.; Waters, M. L., Contribution of aromatic interactions to alpha-helix stability. *J Am Chem Soc* **2002,** *124* (33), 9751-5.

23. Crick, F., The packing of [alpha]-helices: simple coiled-coils. *Acta Crystallogr* **1953,** *6* (8-9), 689-697.

24. O'Shea, E. K.; Rutkowski, R.; Kim, P. S., Evidence that the leucine zipper is a coiled coil. *Science* **1989,** *243* (4890), 538-42.

25. O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T., X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **1991,** *254* (5031), 539-44.

26. Alber, T., Structure of the leucine zipper. *Current opinion in genetics & development* **1992,** *2* (2), 205-10.

27. Walters, R. F.; DeGrado, W. F., Helix-packing motifs in membrane proteins. *Proc Natl Acad Sci U S A* **2006,** *103* (37), 13658-63.

28. Russ, W. P.; Engelman, D. M., The GxxxG motif: a framework for transmembrane helix-helix association. *J Mol Biol* **2000,** *296* (3), 911-9.

29. Dieckmann, G. R.; DeGrado, W. F., Modeling transmembrane helical oligomers. *Curr Opin Struct Biol* **1997,** *7* (4), 486-94.

30. Walther, D.; Eisenhaber, F.; Argos, P., Principles of helix-helix packing in proteins: the helical lattice superposition model. *J Mol Biol* **1996,** *255* (3), 536-53.

31. Ghosh, T. S.; Chaitanya, S. K.; Sankararamakrishnan, R., End-to-end and end-to-middle interhelical interactions: new classes of interacting helix pairs in protein structures. *Acta crystallographica. Section D, Biological crystallography* **2009,** *65* (Pt 10), 1032-41.

32. Neves, M. A. C.; Yeager, M.; Abagyan, R., Unusual Arginine Formations in Protein Function and Assembly: Rings, Strings, and Stacks. *J Phys Chem B* **2012,** *116* (23), 7006-7013.

33. Neves, M. A.; Yeager, M.; Abagyan, R., Unusual arginine formations in protein function and assembly: rings, strings, and stacks. *J Phys Chem B* **2012,** *116* (23), 7006-13.

34. Kufareva, I.; Budagyan, L.; Raush, E.; Totrov, M.; Abagyan, R., PIER: protein interface recognition for structural proteomics. *Proteins* **2007,** *67* (2), 400-17.

35. Barlow, D. J.; Thornton, J. M., Ion-Pairs in Proteins. *J Mol Biol* **1983,** *168* (4), 867-885.

36. Burley, S. K.; Petsko, G. A., Dimerization Energetics of Benzene and Aromatic Amino-Acid Side-Chains. *J Am Chem Soc* **1986,** *108* (25), 7995-8001.

37. Flocco, M. M.; Mowbray, S. L., Planar Stacking Interactions of Arginine and Aromatic Side-Chains in Proteins. *J Mol Biol* **1994,** *235* (2), 709-717.

38. Guillen Schlippe, Y. V.; Hedstrom, L., Guanidine derivatives rescue the Arg418Ala mutation of Tritrichomonas foetus IMP dehydrogenase. *Biochemistry-Us* **2005,** *44* (50), 16695-700.

39. Bogan, A. A.; Thorn, K. S., Anatomy of hot spots in protein interfaces. *J Mol Biol* **1998,** *280* (1), 1-9.

40. Ben-Shimon, A.; Eisenstein, M., Computational mapping of anchoring spots on protein surfaces. *J Mol Biol* **2010,** *402* (1), 259-77.

41. Bordner, A. J.; Abagyan, R., Statistical analysis and prediction of protein-protein interfaces. *Proteins* **2005,** *60* (3), 353-66.

42. Marrington, R.; Dafforn, T. R.; Halsall, D. J.; MacDonald, J. I.; Hicks, M.; Rodger, A., Validation of new microvolume Couette flow linear dichroism cells. *Analyst* **2005,** *130* (12), 1608-1616.

43. Bulheller, B. M.; Rodger, A.; Hicks, M. R.; Dafforn, T. R.; Serpell, L. C.; Marshall, K. E.; Bromley, E. H. C.; King, P. J. S.; Channon, K. J.; Woolfson, D. N.; Hirst, J. D., Flow Linear Dichroism of Some Prototypical Proteins. *J Am Chem Soc* **2009,** *131* (37), 13305-13314.

44. Rodger, A.; Marrington, R.; Geeves, M. A.; Hicks, M.; de Alwis, L.; Halsall, D. J.; Dafforn, T. R., Looking at long molecules in solution: what happens when they are subjected to Couette flow? *Phys Chem Chem Phys* **2006,** *8* (27), 3161-3171.

45. Marrington, R.; Seymour, M.; Rodger, A., A new method for fibrous protein analysis illustrated by application to tubulin microtubule polymerisation and depolymerisation. *Chirality* **2006,** *18* (9), 680-90.

46. Magnotti, E. L.; Hughes, S. A.; Dillard, R. S.; Wang, S.; Hough, L.; Karumbamkandathil, A.; Lian, T.; Wall, J. S.; Zuo, X.; Wright, E. R.; Conticello, V. P., Self-Assembly of an alpha-Helical Peptide into a Crystalline Two-Dimensional Nanoporous Framework. *J Am Chem Soc* **2016,** *138* (50), 16274-16282.

47. Vymetal, J.; Bednarova, L.; Vondrasek, J., Effect of TFE on the Helical Content of AK17 and HAL-1 Peptides: Theoretical Insights into the Mechanism of Helix Stabilization. *J Phys Chem B* **2016,** *120* (6), 1048-59.

48. Wu, Y. Y.; Norberg, P. K.; Reap, E. A.; Congdon, K. L.; Fries, C. N.; Kelly, S. H.; Sampson, J. H.; Conticello, V. P.; Collier, J. H., A Supramolecular Vaccine Platform Based on alpha-Helical Peptide Nanofibers. *Acs Biomater Sci Eng* **2017,** *3* (12), 3128-3132.

49. S. Wall, J.; N. Simon, M., *Scanning Transmission Electron Microscopy of DNA-Protein Complexes*. 2001; Vol. 148, p 589-601.

50. ftp.stem.bnl.gov, P. i. a. a.

Chapter IV. A Concentration Dependent Switch Between Tube and Crystal: Form IA

4.1. Introduction

Given the depth of understanding of the packing of helices in both forms, as well as the intricacies of self-assembly which hinge upon multiple key residues, well-informed mutations to the original Form sequences can be made to engineer new tubes, and to test hypotheses regarding tube diameter, manner of packing, and overall morphology. Prior mutations have centered upon the conservation of the **arginine staple**, but important information can be garnered from the permutation or abrogation of this structural motif. It has been established that interconversion between Forms I and II requires minute sequence mutations. For conversion from Form I to Form II, only the R13K mutation is required. To switch back to the Form I filamentous morphology from Form II, the K13R, K17R double mutation is required. Further, it has been established that R13 demonstrates the single most ordered side chain, as it sandwiched between two other subunits at each corner of the four-start stack of alpha-helices, and changing this single residue is sufficient for the structural switch from Form I to Form II.[1] This makes R13 a critical residue for the maintenance of the four-start helix morphology of the Form I filaments. R17 is similarly important in the Form I structure, as it provides a C-terminal cap for the adjacent alpha-helix in the n+1 alpha stack, thereby further stabilizing the corners of the square helix stacks. Though it is known that the single R13K mutation promotes interconversion between Forms I and Form II, the effect of an R17X mutation on the resultant four-start helix structure of Form I has yet to be explored (X representing an ambiguous residue). To test the structural importance of the C-terminal capping provided by R17, Form 1A is designed. It maintains most of the sequence identity of Form I, with key mutations being R3E and R17E. The R17E mutation encompasses the primary hypothesis: that changing the charge and electrostatic identity of the residue, as well as the sidechain size, will

alter or nullify its capacity for C-terminal capping of the neighboring helices. The glutamic acid is small in size relative to the native arginine, not reaching as far into the adjacent alpha stack. This in turn may affect supramolecular stability and the resultant supramolecular assembly of the alpha-helices. Further, this mutation changes the electrostatic identity at the end-to-face helix juncture from a positively charged moiety to a negatively charged moiety. Lastly, mutating R17 removes the **arginine staple** motif as the key structural linchpin in the formation of nanotubes, potentially increasing the importance of R13 and its interaction with neighboring helices.

Form I: Acetyl-QA**R**ILEADAEIL**R**AYA**R**ILEAHAEILRAQ-NH_2

Form IA: Acetyl-QA**E**ILEADARIL**R**AYA**E**ILKAHAEILKAQ-NH_2

Figure 4.1: the amino acid sequences for Form I and Form IA with residue mutations between the two indicated in bold.

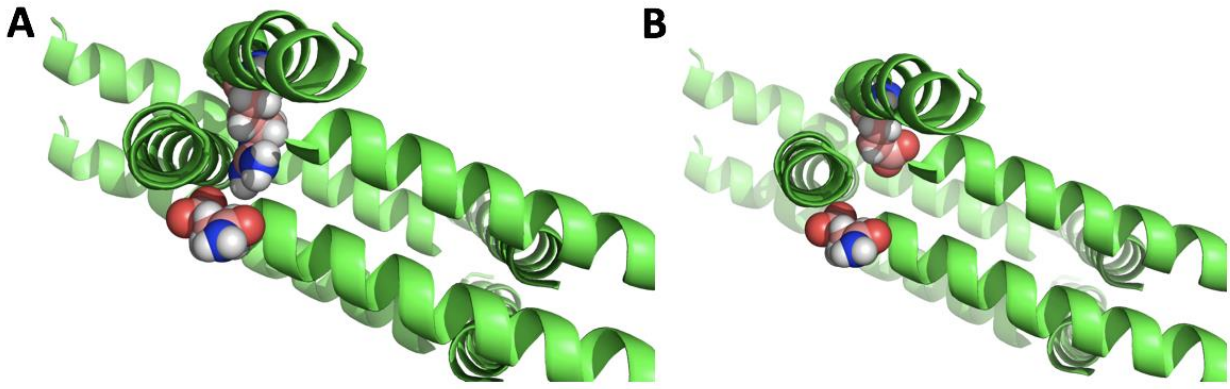


Figure 4.2: graphical representations of R17 in Form I (A) and E17 in Form IA (B) wherein differences in sterics and electrostatics are shown.

4.2 Results and Discussion

4.2.1 Peptide Self-Assembly

Form IA was produced via microwave solid-phase peptide synthesis, and purified to homogeneity via reverse-phase HPLC. Its mass was confirmed via ESI-mass spectrometry. Form IA, similar to the other variants in the Form series, was observed to self-assemble efficaciously in 10 mM acetate buffer, pH 4.0. As with other peptides in the Sol series, 2,2,2-trifluoroethanol (TFE) was found to assist in the self-assembly process. Form IA did not demonstrate ordered assemblies at other pH values, from 5-8. When TFE was employed in the self-assembly process, supramolecular structures were observed as early as one day after full evaporation of TFE from the assembly solution. Form IA self-assembled into significantly different structures depending on concentration, with higher concentrations ($\geq$ 6 mg mL^{-1}) leading to a highly ordered crystalline network and lower concentrations ($\leq$ 2 mg mL^{-1}) leading to nanotubes. Interestingly, in both the low concentration and high concentration states, assemblies were still observed in the absence of TFE, however only tubular structures were found in these cases, indicating that a slow, controlled process (i.e. TFE evaporation over time) is an energetic requirement for the formation of the crystalline morphology.

4.2.2 Circular Dichroism and Flow Linear Dichroism, Low Concentration

Circular dichroism provides critical insight into the secondary structure of peptides.[2] Circular dichroism data for this variant, assembled at 2 mg mL^{-1} in 10 mM acetate (pH 4.0), indicate, as hypothesized, an alpha-helical secondary structure. The peptide was analyzed after one week assembly time and diluted to ~100 µM directly before performing CD. The mutations that differentiate Form 1A from Form 1 are relatively minute (with only two residue mutations), so it

is reasonably hypothesized that Form 1A will adopt an alpha-helical morphology akin to its parent peptide Form 1. Since this peptide demonstrates remarkably different supramolecular assemblies at low and high concentrations as observed via TEM, circular dichroism was performed for both concentrations to probe whether a shift in secondary structure between the two variants is responsible for the resultant macromolecular change in structures. The CD results are shown in Figure 4.3. As anticipated, data demonstrate the canonical peaks which indicate alpha-helical secondary structure (minima at 208 nm and 222 nm and a maximum around 195 nm). For this

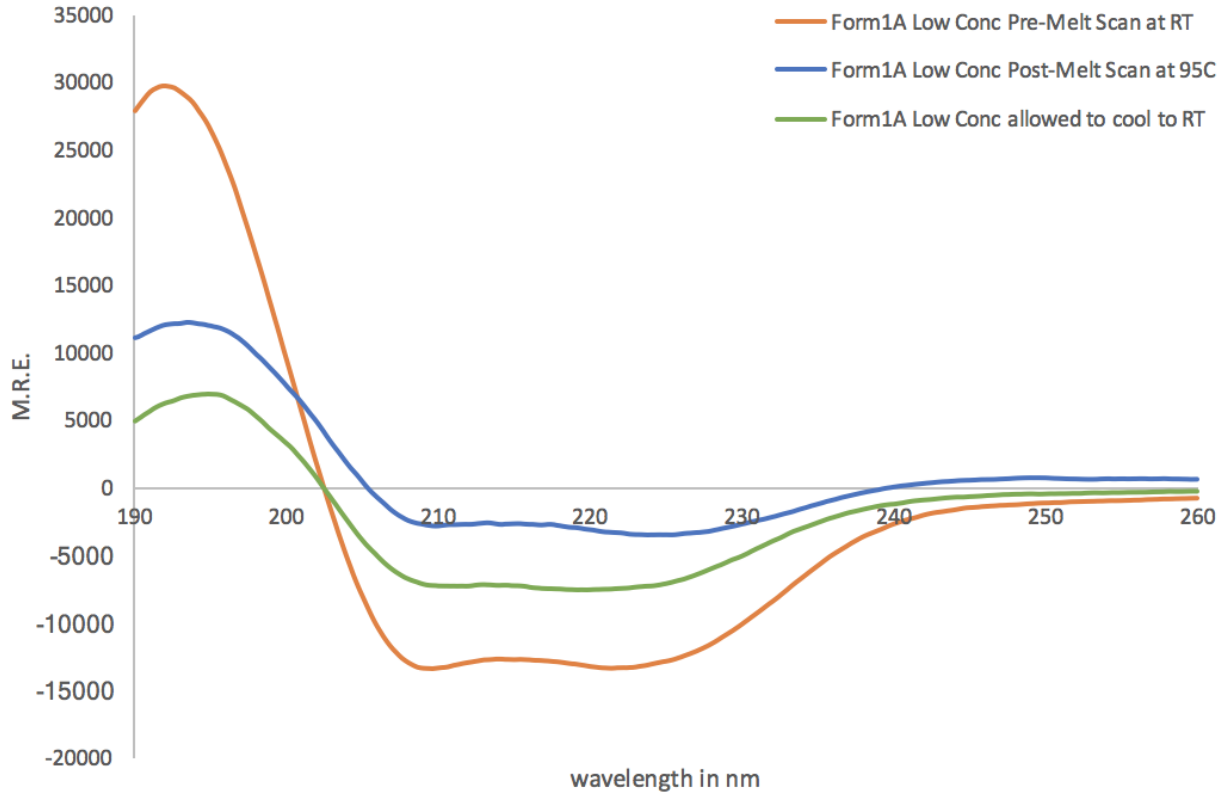


Figure 4.3: circular dichroism results for Form 1A in tubular form at low concentration. Sample was analyzed at room temperature (orange) and after a melting scan at 95 °C (blue). After cooling to room temperature post-melt, the sample was analyzed a final time, to determine if Form 1A can re-adopt its secondary structure after thermal annealing (green).

variant, the assembly was diluted to 100 µM for analysis after being allowed to assemble for one week. The assembly was analyzed via circular dichroism spectropolarimetry at room temperature, and then analyzed after a melting curve at 95 °C to determine the thermostability and percentage of alpha-helical character that remained. At 95 °C post-melt, form IA retains very little of its alpha-helical character. However, after cooling to 22 °C from 95 °C, Form IA re-adopted ~40% of its helical content within five minutes.

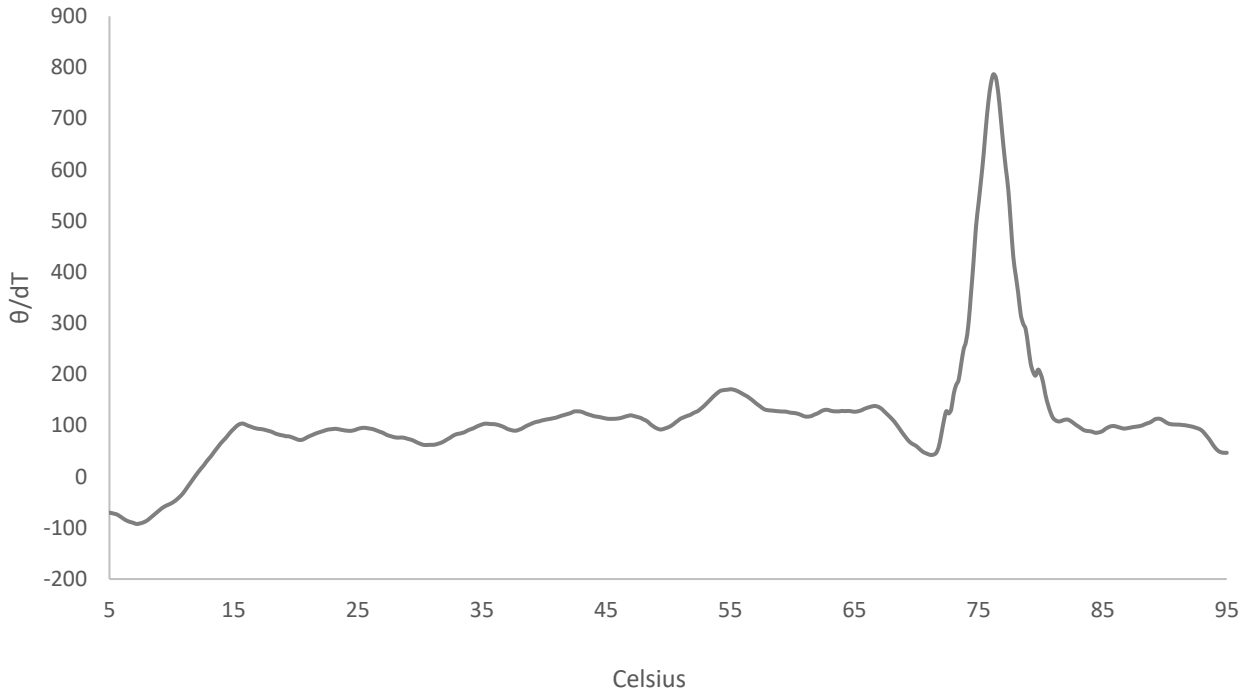


Figure 4.4: First derivative analysis of the melting curves for aqueous solution of Form1A assembled at low concentration (filamentous morphology). First derivative curves were calculated by following the maximal signal relative to temperature. From this analysis, the melting transition occurs at 76 °C.

Form IA at low concentration demonstrated a melting transition (Tm) at ~76 °C, as seen from the first derivative trace of Form IA followed at 222 nm from 5 °C to 95 °C (Figure 4.4).

As with prior tubular variants within this peptide series, flow linear dichroism was performed on the low concentration tubular variant of Form IA to determine the packing alignment of the helices within the filament. Flow linear dichroism[3-5] was employed under dilute conditions (25 µM) to survey the degree of anisotropy that developed within the peptide samples under 5,000 RPM flow alignment in a Couette cell. Solutions of Form IA demonstrated positive signals at 192 nm, and a weaker, broader positive signal around 220 nm. Unlike with Form III, limited sample volume and limited pure peptide yield limited the number flow linear dichroism scans performed for Form IA. In the case of Form IA filaments, one solution of uniform tubular assemblies was

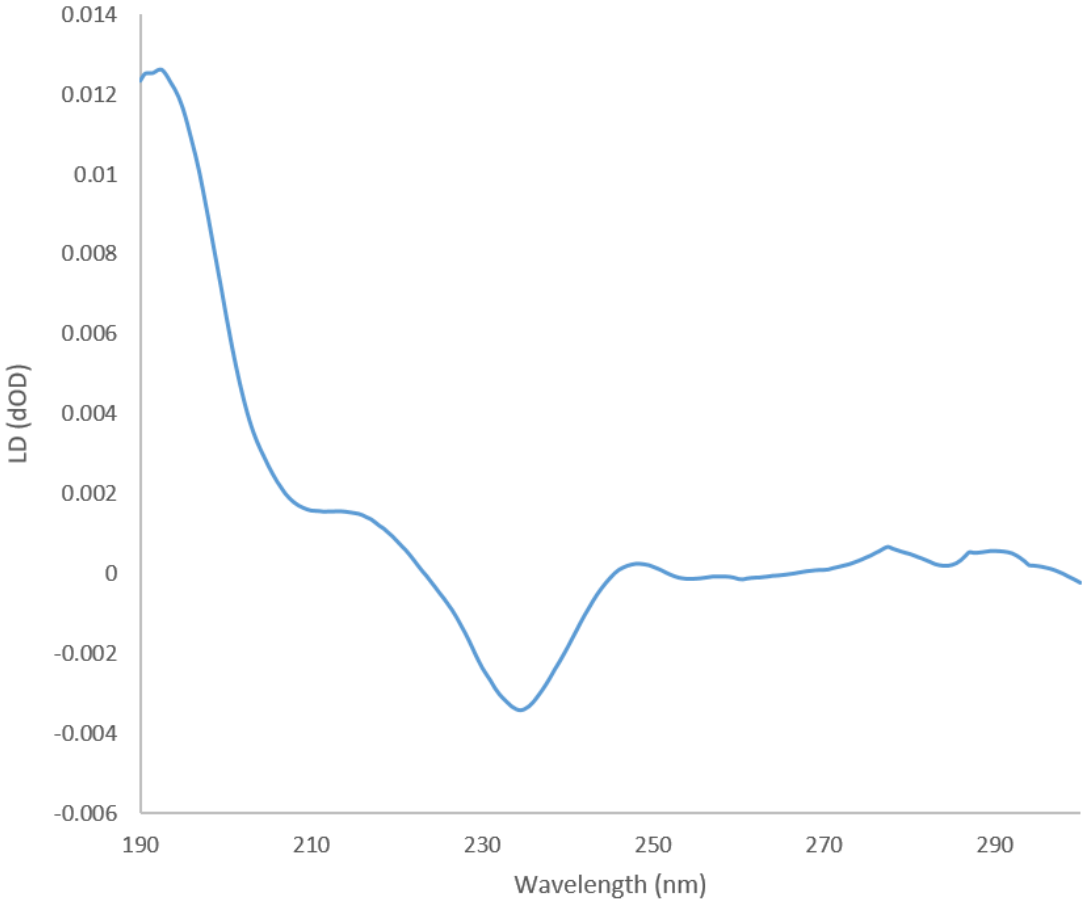


Figure 4.5: flow linear dichroism of low concentration Form IA tubular filaments demonstrating helices packing perpendicularly to the filament axis.

analyzed in triplicate. This data would suggest that the electron transitions in the sample are aligned parallel with the direction of flow. Therefore, the helices are packed perpendicularly to the long axis of the Form IA filaments, similar to what was observed for Forms I, II and III. Form I was analyzed at 50 μM and demonstrated a signal intensity at the 192 nm maximum of ~0.035 dOD, whereas Form IA was assembled at 25 μM and demonstrated a maximual signal intensity of ~0.013 dOD at the equivalent wavelength. The broader peak around 222 nm is slightly blue shifted to ~217 nm, but is more pronounced than the equivalent peak for Form I.

The confirmation of helical packing perpendicular to the fiber axis presents some interesting structural possibilities with regards to the packing gemeotry of the helices. Given a helix length of ~4.4 nm, it is difficult to predict a packing geometry in which the helices align perpnedicularly to thefilament axis that could produce such large diameter nanotubes. One possible helical arrangement may feature oligomerization of helices in a lateral fashion, potentially offset from one another. This could afford a coiled-coil subunit with a larger axial length than a single helix.

4.2.3 Circular Dichroism, High Concentration

Similar to the low concentration assembly, Form IA assembled at high concentration (6 mg mL^{-1}) also demonstrates an alpha-helical character (Figure 4.6). The minima, typically occurring at 208 nm and 222 nm for alpha-helices, are slightly red-shifted in the case of Form IA. The 222 nm minima (occurring at 225 nm for Form IA) demonstrates greater signal intensity than the 208 nm minima (occurring at 209 nm for Form IA). The higher concentration variant did not

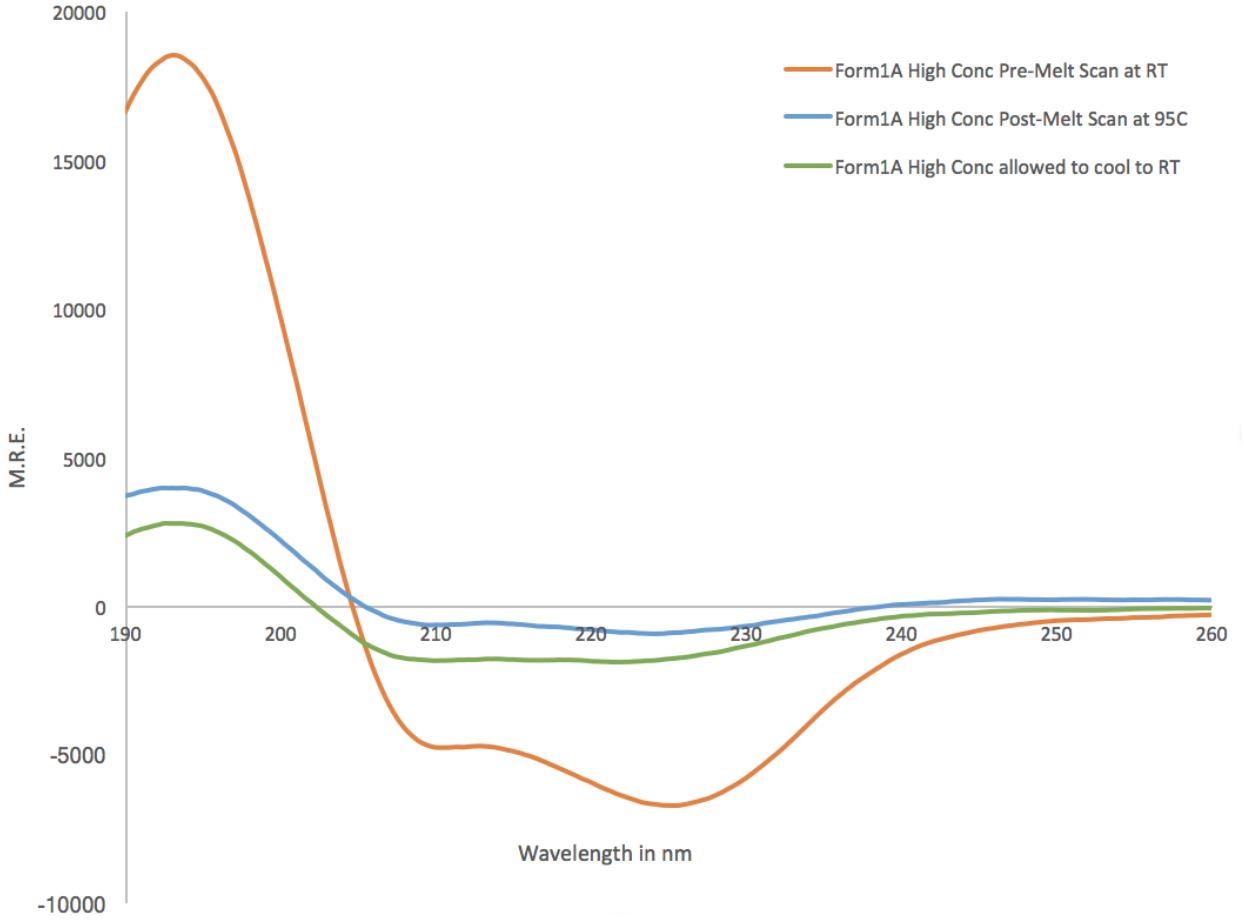


Figure 4.6: Circular dichroism results for Form 1A in crystalline form. Sample was analyzed at room temperature (orange) and after a melting scan at 95 °C (blue). After cooling to room temperature post-melt, the sample was analyzed a final time, to determine if Form 1A can re-adopt its secondary structure after thermal annealing (green).

refold as quickly as the lower concentration variant. This is not unexpected, given the greater degree of complexity observed for the higher concentration variant and its highly ordered, crystalline network structure. While the lower concentration variant refolded to ~40% of its original helical content within ten minutes of cooling from 95 °C, the higher concentration variant only refolded to ~20% of its original helical content (Figure 4.7).

Form IA at higher concentrations (crystalline morphology when observed via TEM) demonstrated a higher melting transition temperature (Tm) than what was observed for the lower concentration variant tubular morphology when observed for TEM). This is in line with what was hypothesized. The higher concentration should putatively feature a greater number of local interactions per subunit, if it is indeed packing into a crystalline or semi-crystalline state.

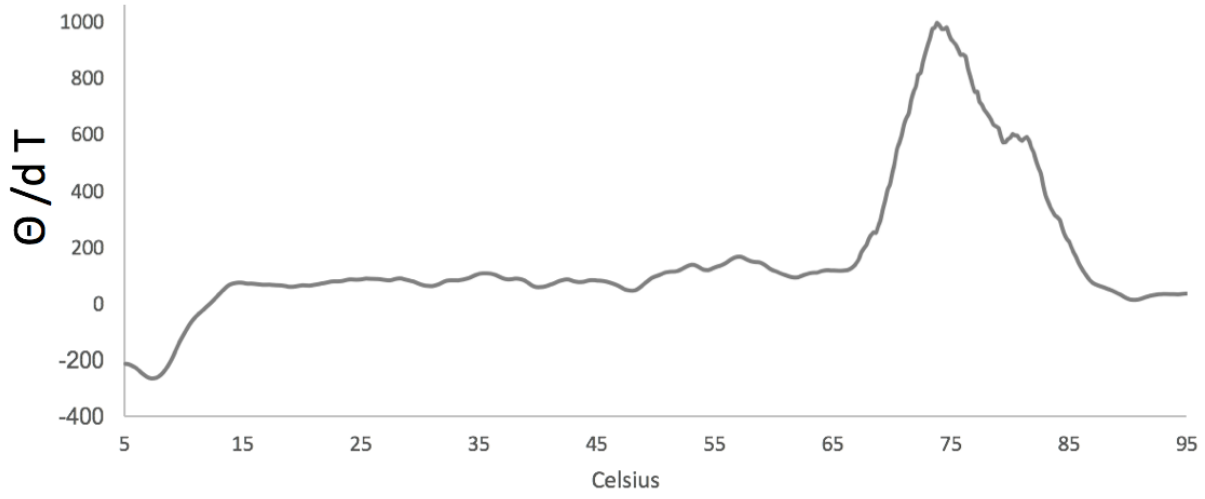


Figure 4.7: First derivative analysis of the melting curves for aqueous solution of Form1A. First derivative curves were calculated by following the maximal signal relative to temperature. From this analysis, the melting transition occurs at 74 °C. Of note, when compared to the melting profile of the lower concentration assembly, the transition occurs over a broader temperature range, and features a shoulder around 83 °C.

4.2.4 Transmission Electron Microscopy, Low Concentration Filaments

Transmission electron microscopy data confirm two structurally unique supramolecular assemblies. When assembled via the TFE assembly protocol (see materials and methods) at low concentrations (i.e. 2 mg/mL), the peptide self-assembles into nanotubes. The lumen of these tubes is substantive in size to allow for the observance of negative stain penetrating and filling the

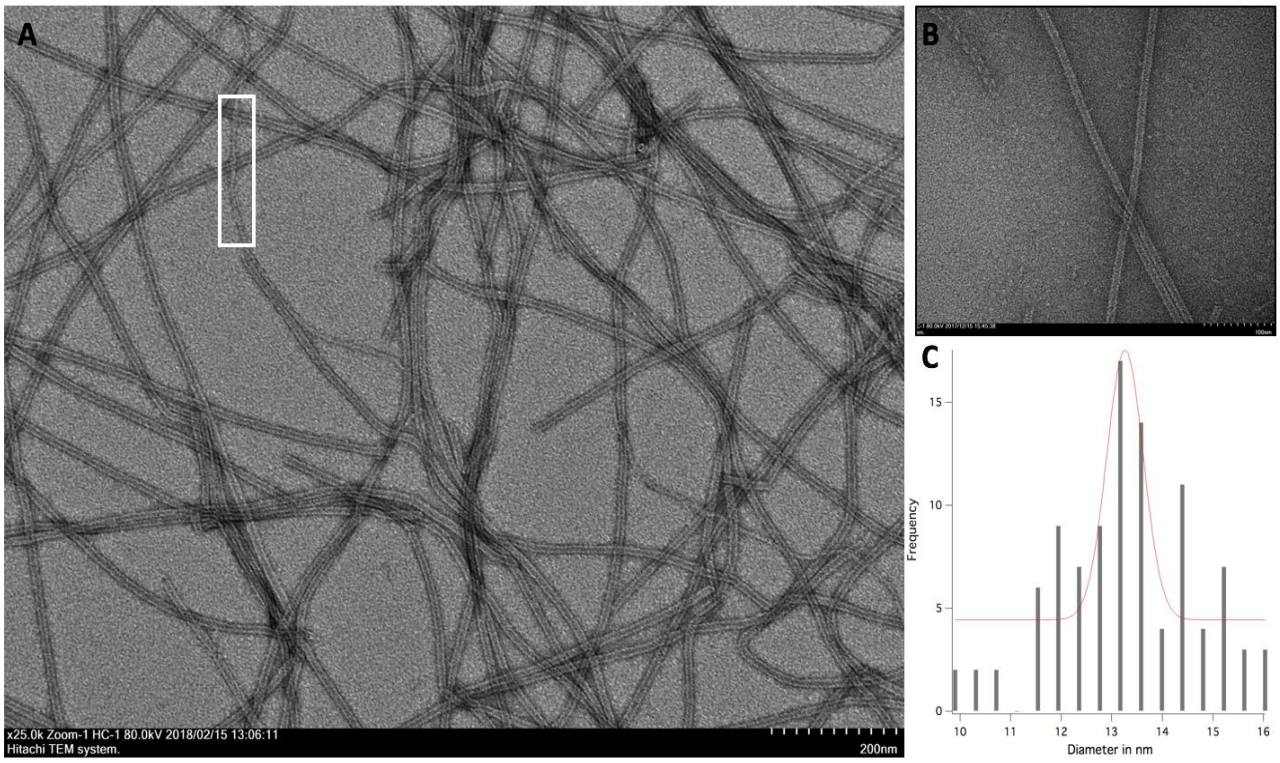


Figure 4.8: (A) transmission electron micrographs of Form 1A assembled at low concentrations (i.e. 2 mg/mL) which self-assembles into majority nanotubes with discernible inner lumens. The white box indicates a transition of the assembly from a double-walled tube to a singular filament. (B) transmission electron micrograph demonstrating a phenomenon seen with several peptides in the Form family: an apparent unwinding of the nanotube at the terminus which appears to occur in helical fashion. (C) Histogram of measured Form IA tube diameters via ImageJ.

interior of the tube. The nanotubes demonstrate a measured diameter (via ImageJ analysis program) of 13.5 ± 1.40 nm. This value is later corroborated by SAXS/WAXS data, a more accurate analytical technique which demonstrates a Guinier-derived diameter value of 14.0 nm using the Guinier equation for rod-like assemblies. A diameter of ~14 nm is significantly larger than the diameter of 6.0 nm (as is observed for the parent Form I peptide). Moreover, the diameter more closely resembles that of form II which demonstrates a measured diameter of 12 nm. However, this neither confirms nor denies that the constituent helices are packing in a manner mimetic to either Form I or Form II. In a minority of cases, helical unwinding of the tube at the termini is observed. This is a common phenomenon witnessed in multiple peptides in the Form family, which is not unexpected as the Form IA peptide is hypothesized to from a superhelical structure. Unique to Form IA is the occurrence of a single nanotube transitioning from a double-walled tube to a singular filament (shown in white in Figure 4.8). The measured diameter from ImageJ of the singular filaments is approximately 6 nm, which is equivalent to the calculated diameter for Form I nanotubes.

4.2.5 Transmission Electron Microscopy, High Concentration Crystals

When assembled at high concentrations (i.e. 6 mg/mL) via the TFE assembly protocol, the supramolecular identity changes drastically from singular, well resolved, high-aspect ratio nanotubular assemblies to highly ordered, crystalline lattice networks (Figure 4.10 and 4.11). The size of these crystals varies from hundreds of nanometers to a few microns at most. TEM data indicate that the helices may be packing in a tetragonal manner. When measured from center-to-center (see figure 4.11, red line) the crystalline network demonstrates a diagonal distance of 10 nm, and a distance between neighboring points of 6.7 nm. When measured from exterior-to-

exterior (see figure 4.11, green line) the diagonal distance is 13.45 nm, which is equivalent to what was measured for the lower concentration tubular assemblies.

The concentration-dependent difference in supramolecular assemblies may be due to one of two scenarios: (1) the helices are packing in completely distinctive fashions in each variant and the 'subunits' of each variant are completely unique from one another. (Subunit here is used to describe the minimal number of alpha-helices required to complete one period of the supramolecular assembly). (2) The subunits are assembling in the same manner, but said subunits exhibit differing proclivities for axial interactions vs. lateral interactions (See Figure 4.9). Two pieces of data indicate the validity of the latter case: firstly, the measured diameters for the low concentration tubular assemblies and the diagonal distance of one unit of the high concentration crystalline network are equivalent (~14 nm vs ~13.5 nm). Secondly, the scattering data, for the high concentration crystals, in the mid-*q* region ≤ .004 were fit to the modified Guinier model for rod-like forms for Form IA assemblies and a subsequent cross-sectional rod diameter of 14.0 nm (within a sample demonstrating only crystalline morphology) was calculated, matching what is observed for the lower concentration tubes in TEM and STEM data.

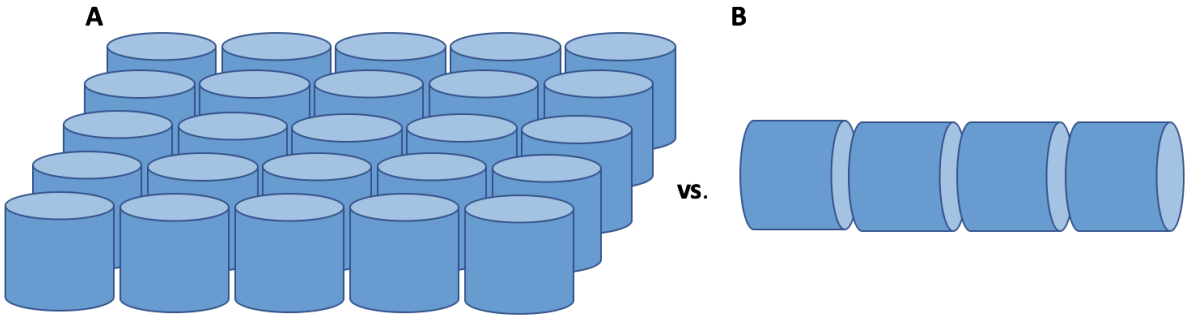


Figure 4.9: Graphical representation of the competing interactions between subunits (A) lateral association, as seen in the crystalline structures and (B) axial association, as seen in the tubular structures.

If the subunits are equivalent across both the low and high concentration assemblies, the concentration-dependent switch in supramolecular identity for Form IA suggests that there is competition between the stability of the lateral and axial packing of subunits, with lateral interactions of subunits favored at high concentrations and axial interactions of subunits favored

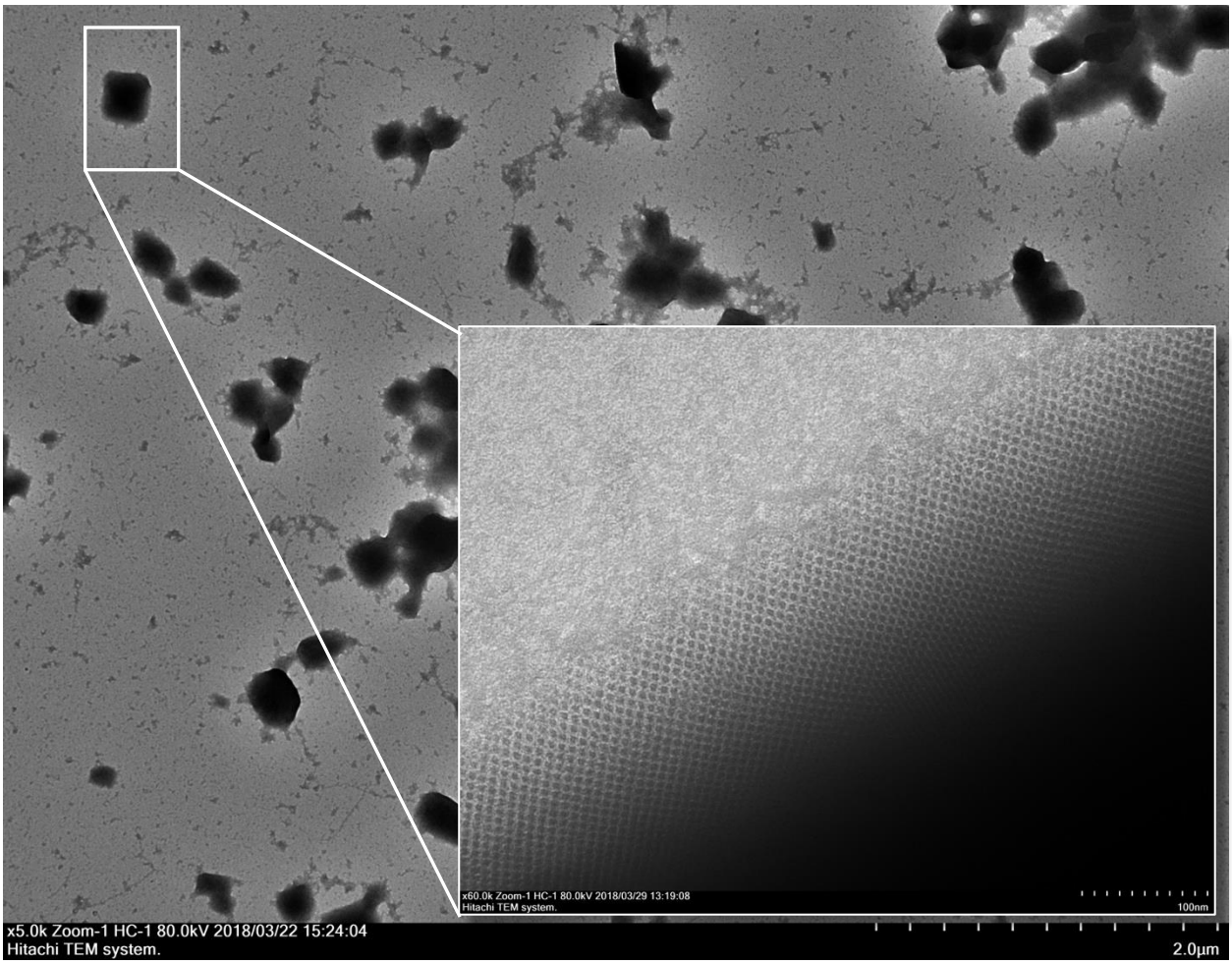


Figure 4.10: transmission electron micrographs of Form 1A assembled at high concentrations (i.e. 6 mg/mL) which self-assembles into a highly ordered crystalline network when assembled in the presence of equal parts TFE to aqueous buffer. The negative stained transmission electron micrographs above show a low magnification image of Form1A crystals. A higher magnification micrograph indicates the lattice packing style of the Form 1A peptide.

at low concentrations. It is hypothesized that the lower concentration tubes may represent a kinetically trapped state. Both variants are assembled in the presence of TFE, which effectively mimics numerous types of crystallization protocols which rely on sequential extraction of solvent (i.e. hanging drop vapor diffusion) to reach supersaturation. The nanotubes were observed to self-assemble more quickly than the crystals. The TFE-evaporation protocol used in the self-assembly

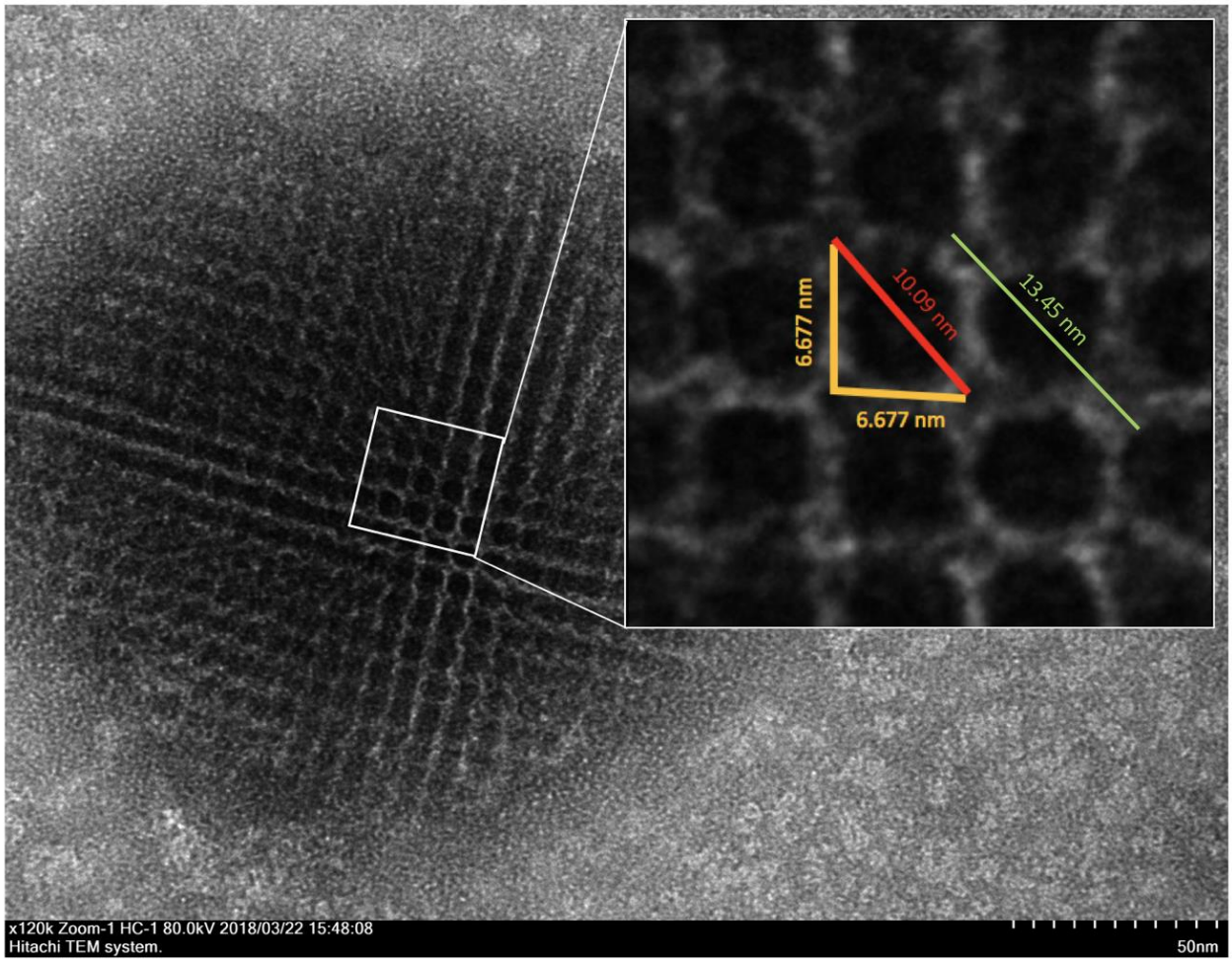


Figure 4.11: TEM analysis of Form IA crystalline lattice packing with approximated dimensions: (red) diagonal measurement from center point to center point between neighboring points is approximated at 10 nm. When measured from edge to edge (green), the diagonal value more closely resembles what is observed for the diameter of the low concentration nanotube assemblies. The edge-to-edge value was measured to be 13.45 nm

of many of the Form peptides replicates the slow, controlled increase in peptide concentration over time that is a hallmark of numerous crystallization techniques. It is observed that the tubes are not an intermediate step on the path to crystallization. However, when the TFE is removed via dialysis before all of the free peptide in solution has assembled into crystals, the remaining peptide readily self-assembles into nanotubes identical to those observed in the low concentration assemblies. When the TFE is removed via dialysis from the low concentration sample, the nanotubes still form, though at a slower rate.

Time points were taken incrementally for both variants, wherein at each time point, the TFE was dialyzed out of the assembly solution. For the high concentration crystals, multiple weeks were necessary for the full self-assembly crystals. Removing TFE before all of the available

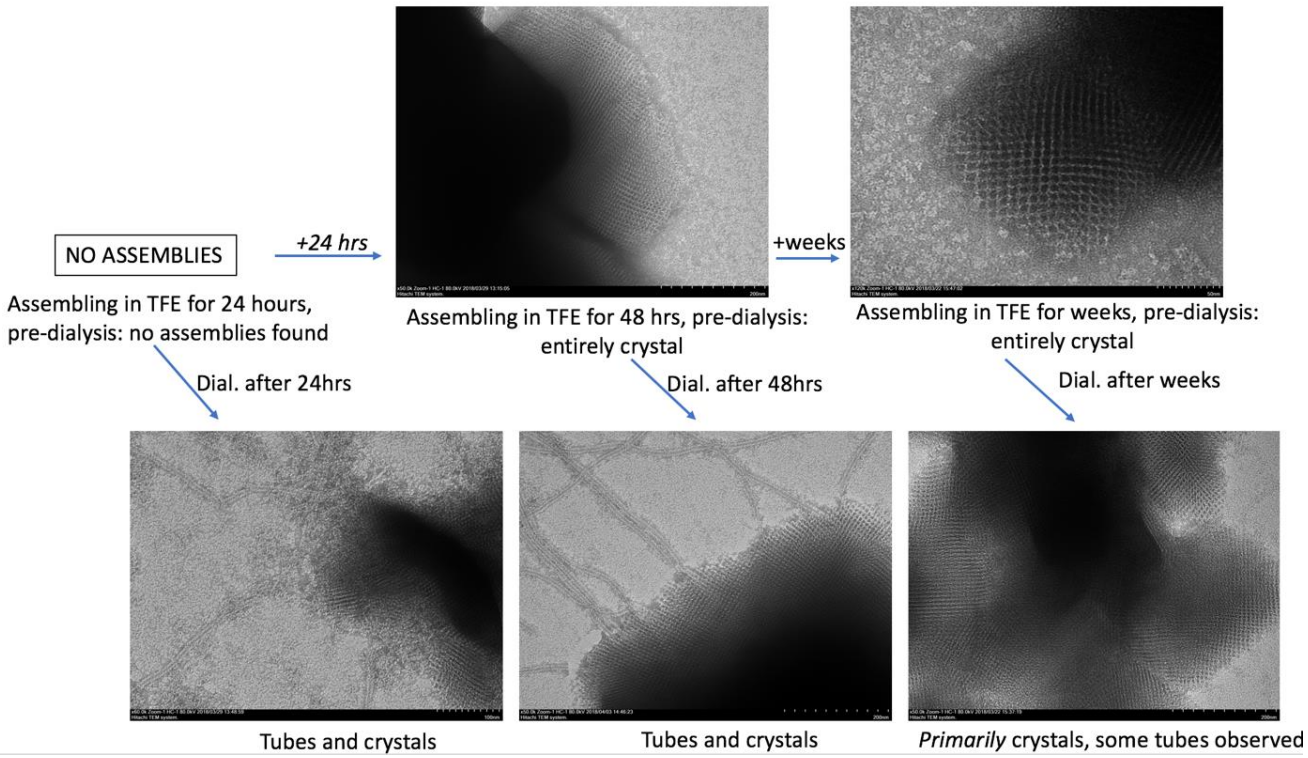


Figure 4.12: TEM time point analysis of TFE removal from solution for the high concentration crystal structures.

peptide had assembled into crystals spurred the self-assembly of the remaining peptide into tubular forms. For the low concentration variant, removing the TFE from the assembly solution had no effect at any of the time points tested, tubular morphology was all that was witnessed. Therefore, it is apparent that slow, controlled, crystallization-mimetic assembly conditions are required for the crystals to form. Lastly, the low concentration assembly solution (once fully formed tubes were observed via TEM) was allowed to evaporate until the resultant peptide concentration increased from 2 mg mL^{-1} to 6 mg mL^{-1} to determine whether or not crystals would arise as a result of the concentration increase. No change in morphology occurred, and no crystals were observed, indicating that the tubular state is not an intermediate on the folding pathway to crystal state. Ongoing work is being performed upon this system, particularly with regards to the production of larger crystals for the purpose of X-ray diffraction in order to glean critical structural data. Crystallographic data could provide insight into how the crystals are forming, and how the helices are packing within the crystalline network. The current size threshold for crystallographic analysis is ~30 microns. The current maximal crystal size, measured diagonally across the structure, is ~3 microns. Importantly, crystallographic data may indicate whether or not the subunits of the crystalline form and the tubular form are identical, or if the helices in each form are packing in fundamentally different geometries.

4.2.6 Scanning Transmission Electron Microscopy

Scanning transmission electron microscopy (STEM) data provide valuable insight into the structural parameters of tubular and sheet-like systems.[1, 6] STEM was performed on the lower concentration only, wherein the peptide forms filamentous assemblies with an inner lumen. The apparent diameter measured from freeze-dried specimens was found to be 17.4 nm, which is larger than the diameter values obtained from both TEM and SAXS (13.5 nm and 14.0 nm respectively). The mass per length (MPL) average for **Form IA,** from at least 3 discrete data sets, was calculated to be 4,726.322 Da/Å with a standard deviation 215.866 Da/Å (4.567% error). As the molar masses are nearly identical for Form I and Form IA, the MPL measurements should be approximately equivalent, *if* Form IA self-assembles in an identical or mimetic fashion to Form I. The MPL for Form IA is ~3.1 times greater than that of Form I. This could indicate the presence of a triple walled tube, self-assembling as a four-start helix similar to Form I, or Form IA could be self-assembling entirely differently than its parent peptide.

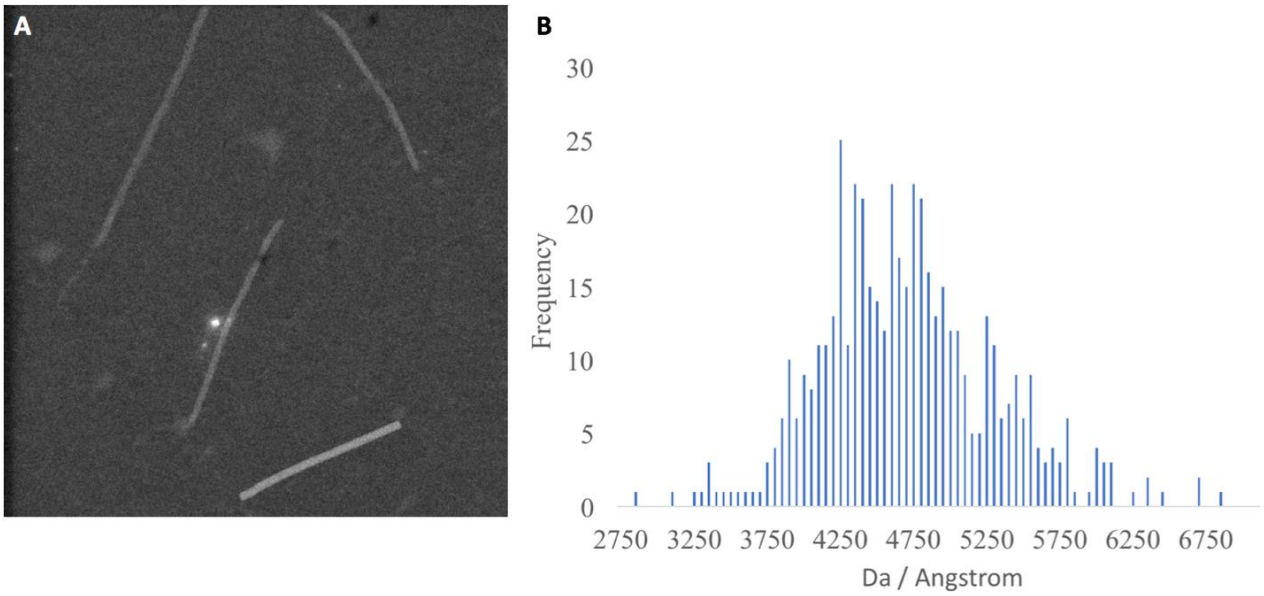


Figure 4.13: (A) STEM image from freeze-dried specimens of Form IA and (B) histogram of MPL measurements of Form IA.

4.2.7 Small-angle X-ray Scattering

In the case of the low concentration assembly solutions for Form IA, SAXS[7] data could not be obtained due to significant sample degradation in the x-ray beam, even when tested at lower pulse frequencies and in the presence of glycerol (added specifically to limit or abrogate degradation of sample by the beam). Obtaining SAXS data in the future for the low concentration tubes will be immensely beneficial in the determination of subunit identities compared across both variants. In turn, this data can be used to help determine whether or not the alpha-helices are packing into identical subunits with different interfacial proclivities (lateral vs. axial), or if each variant demonstrates entirely unique helical packing. SAXS data was obtained for the high concentration crystalline variant. Scattering data within the low *q* region (~<0.05 $Å^{-1}$) were fit to the modified Guinier equation for rod like forms. The cross-sectional radius of gyration R_c was

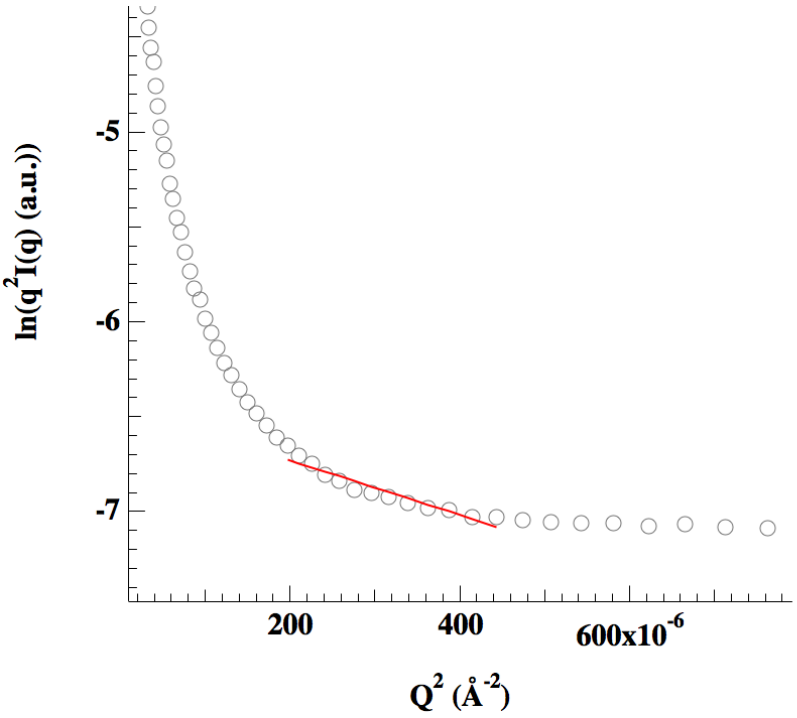


Figure 4.14: Modified Guinier plot of scattering data for Form IA using an equation for rod-like forms. An analysis using the equation for sheet-like forms was also used.

determined to be 49.55 ± 4.63 Å, corresponding to a calculated fibril diameter of 14.0 nm. When analyzed using the Guinier equation for sheet-like forms, the resultant thickness value was 50.15 ± 7.20 nm. Representative scattering profiles as a function of the momentum transfer *q* are shown. The scattering in the small-angle region (i.e. *q* analyzed at < 0.2 $Å^{-1}$) indicate that the global shape of the assemblies is likely a large three-dimensional sphere, with a power law exponent slightly greater than q^{-3}. Multiple Bragg diffraction peaks are observed, correlating to d-spacings of (1) 7.19 nm, (2) 6.72 nm, (3) 5.06 nm, (4) 4.08 nm, and (5) 3.24 nm. Typically a unimodal oscillation is observed in the mid *q* region, however the oscillation for the Form IA crystal is split. Further,

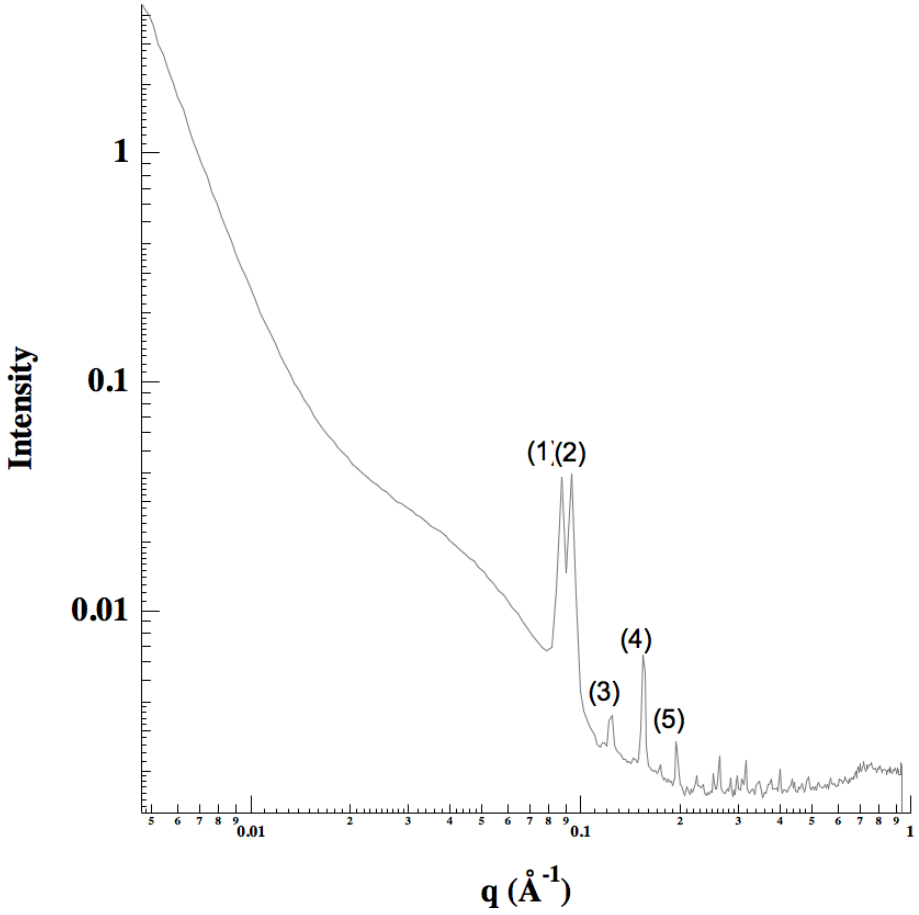


Figure 4.15: Synchrotron SAXS/WAXS Scattering profile for Form IA crystals with the calculated *d*-spacings for each diffraction peak indicated.

two of the Bragg diffraction peaks (7.19 nm and 5.06) are related by a factor of $\sqrt{2}$, which is canonically indicative of tetragonal packing of subunits due to the Pythagorean relation. Future work towards obtaining SAXS data for the low concentration tubular variant will be critical in helping determine subunit identities across both variants.

4.3 Conclusion

Form IA represents a novel peptide system wherein differences in concentration produce profound morphological effects. Previously, research into this peptide family has shown that small mutations to the amino acid sequence could produce similarly significant changes in supramolecular structure. This research portrays a system that is highly labile with respect to not only changes in sequence-space, but also to changes in environment, such as peptide concentration. Though the structures of both variants are yet unknown, this peptide provides a foundation for further research into concentration-dependent morphological switches in self-assembling peptide systems. Current (and future) work involves the optimization of crystallization parameters for this peptide using a library of conditions including variable polyethylene glycol (PEG) chain lengths (40, 400, 4,000) and PEG concentrations, salt concentrations, and peptide concentrations.[8] This system could be used in downstream applications such as small molecule encapsulation for drug delivery or,[9] in the case of the crystals, as a robust scaffold for functionalization.

4. 4 Methods

4.4.1 Peptide Synthesis.

Peptides were either purchased from GenScript USA (Piscataway, NJ) or synthesized in-house. In the latter case, peptides were prepared as the capped (N-acetyl, C-amide) derivatives using microwave-assisted synthesis on a CEM Liberty solid-phase peptide synthesis instrument using PAL-PEG-PS resin from Applied Biosystems. Standard Fmoc protection chemistry was used with coupling cycles based on HBTU/DIEA-mediated activation protocols and base-induced deprotection (20% piperidine in *N*,*N*-dimethylformamide with 0.1 M hydroxybenzotriazole) of the Fmoc group. The peptides were purified via reversed-phase high-pressure liquid chromatography (HPLC) on a C18 column with a gradient of water-acetonitrile (0.1% trifluoroacetic acid). The purity was assessed to be above 95% by analytical HPLC. Peptide mass was confirmed using MALDI mass spectrometry or ESI-Mass Spectrometry. The peptides were lyophilized, sealed, and stored at –20 C until ready for assembly. Dialyzing trace amounts of trifluoroacetic aid remaining from HPLC out of the sample did not have any discernible effect on the assemblies. Analytical HPLC was performed to confirm the purity of the peptides and the data can be seen below. Lastly, MALDI mass spectrometry or ESI-Mass Spectrometry were performed to confirm that the peptides were synthesized correctly and had the correct masses. The mass spectrometry data can also be seen below.

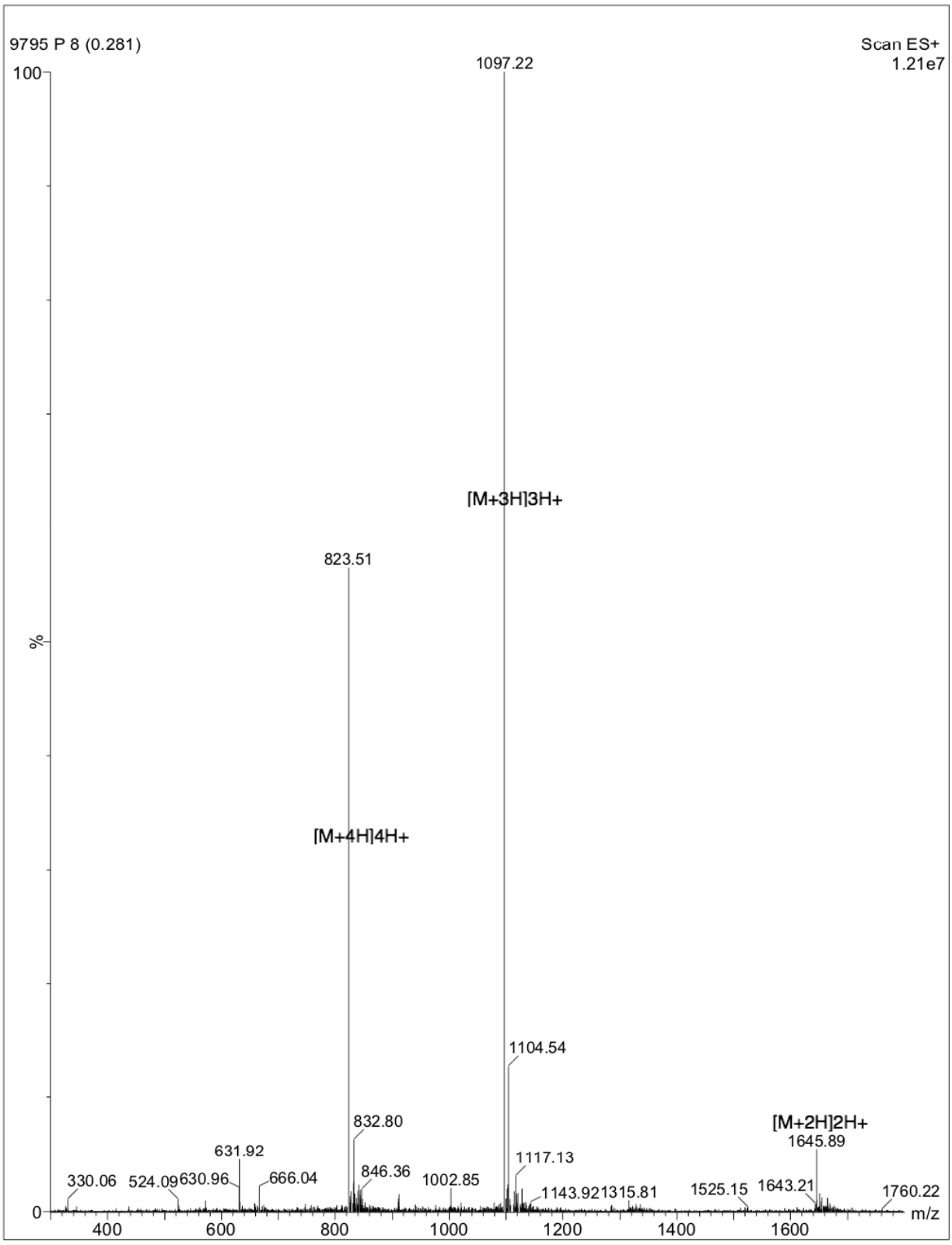


Figure 4.16: ESI Mass Spectrometry confirmation of molar mass for Form IA, from Synpeptide Company.

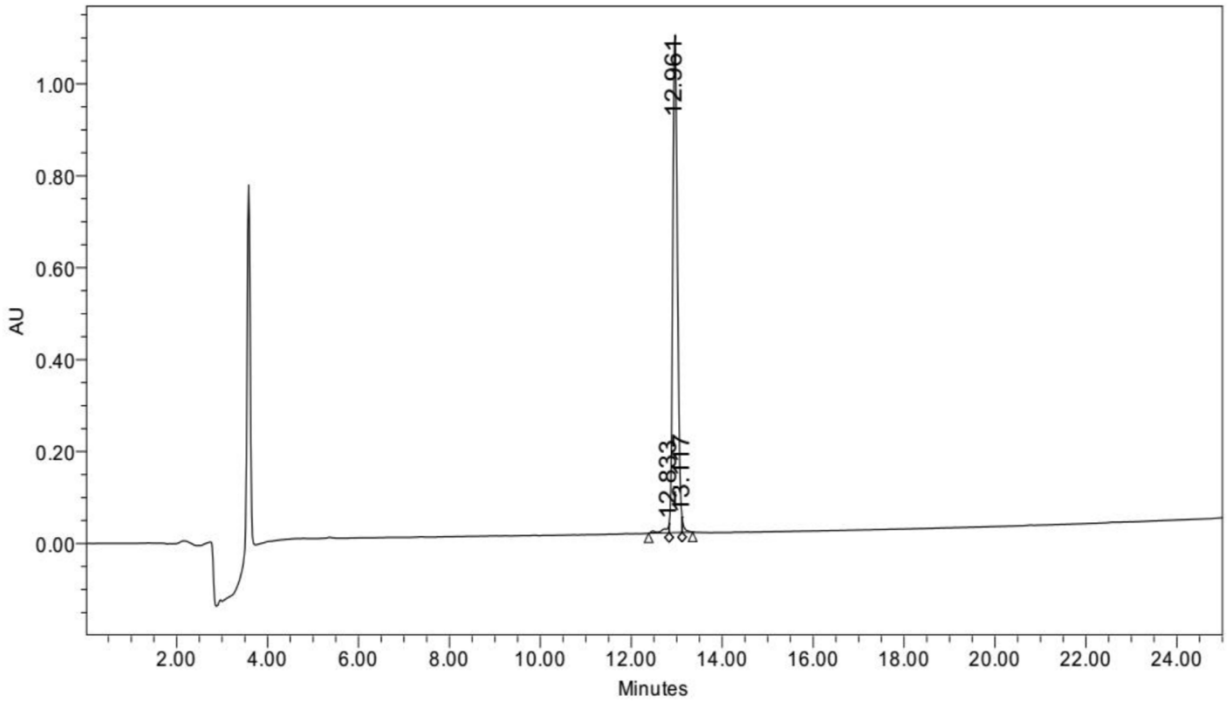


Figure 4.17: Form IA analytical HPLC spectrum, from Synpeptide Company.

4.4.2 Peptide Self-Assembly

Form IA was dissolved in acetate buffer (10 mM, pH 4.0) using the TFE assembly protocol (vide supra). For peptides containing aromatic Trp, Tyr, and/or Cys residues, peptide concentration was determined spectrophotometrically from measurement of the absorbance at 280 nm (*A*280). For peptides containing Tyr, Trp, or Cys residues, the peptide concentration can be calculated from Equation 1, as seen below:

286 Structure *23*, 280–289, February 3, 2015 a2015 Elsevier Ltd All rights reserved

$$\text{MW } 3\,A280 = 1280nY + 5690nW + 120nC\ c$$

(Equation 1)

in which *c* is the concentration of peptide in mg/ml, and *n*Y, *n*W, and *n*C are the numbers of tyrosine, tryptophan, and cysteine residues, respectively, in the peptide sequence (Gill and von Hippel, 1989). Since the peptide derivatives contain only a single tyrosine residue per molecule, then *c* = MW 3 *A*280/ 1280. To eliminate error in determination of absorbance that could arise as a result of UV light scattering due to peptide self-assembly, aqueous solutions of peptide were mixed with 6 M guanidinium chloride in 1:9 v/v ratio and seated at room temperature for 2 hours to completely denature the sample prior to per- forming the absorbance measurements. Samples for EM were heated to 95 C in acetate buffer (10 mM, pH 4.0) and slowly cooled to 4 C to promote assembly of the nanotubes.

Form IA was assembled by solubilizing (using a 1.5 mL Eppendorf tube) in 10 mM acetate buffer at (pH 4.0). Subsequently, the pH is measured, and corrected back to 4.0 if necessary using 0.1M sodium hydroxide solution. 2,2,2-trifluorouethanol (TFE) is added in a 1:1 ratio, such that the volume of buffer is equivalent to the volume of TFE. The Eppendorf tube is sealed in Parafilm ™, and a needle is used to perforate the film once to allow for controlled evaporation of TFE from the mixture over time. Of note, when this methodology is used, Form IA demonstrates divergent supramolecular morphologies dependent upon concentration of peptide in solution. Lower concentrations (2 mg/mL after full TFE evaporation) of Form IA assembled in the aforementioned fashion present nanotubes with discernible inner lumens when analyzed via transmission electron microscopy. Alternatively, nanotubes can also be produced by assembling at high concentrations (6 mg/mL) via the aforementioned protocol, but in the absence of TFE. Higher concentrations (6 mg/mL) are observed to pack into highly ordered crystalline networks. Lastly, dialyzing off the TFE at varying time points during the assembly process has significant effects on the resultant

supramolecular morphology. Assembly timescales and results are described in the results and discussion section.

4.4.3 Circular Dichroism.

Circular dichroism analyses of Form IA assemblies were performed using a JASCO J-810 circular dichroism spectropolarimeter. Peptide solutions were analyzed at a peptide concentration of ~100uM for both variable-wavelength scans and variable-temperature scans. For variable wavelength scans, Spectra were collected at 50nm/min from 190-260 nm with a data pitch of 0.2 nm. For variable temperature scans, spectra were collected at 222nm, with a 60°C/hour temperature progression, and a data pitch of 0.5°C. Variable temperature spectra were collected from 25-85°C. Quartz plates (0.10 mm thickness) from Hellma Analytics (Müllheim, Germany) were employed for all CD experiments.

4.4.4 Transmission Electron Microscopy.

Grids were prepared by depositing 4 μL of peptide solution (0.3 mg mL^{-1} peptide concentration in 10 mM acetate buffer pH 4.0) to 200 mesh carbon-coated copper grids purchased from Electron Microscopy Services (Hatfield, PA). The solution was allowed to incubate on the grid for 45 seconds before being wicked away using 55mm qualitative filter paper. The peptide solution was wicked from the grid surface in such a way that a thin film of liquid remained, then 4 μL of freshly prepared, filtered (0.2 micron Whatman filter from G.E. Healthcare Services) 1% aqueous uranyl acetate solution was applied to the grid to mix with and negatively stain the sample. After 45 seconds of incubation, the stain was fully wicked from the grid using the same methodology (*vide supra*). Grids were stored in a tabletop desiccator and sufficiently dried. All

transmission electron microscopy was carried out on a JEOL JEM-1400 transmission electron microscope at an accelerating voltage of 120kV. The electron microscopy data described here were gathered on a JEOL JEM-1400 (supported by National Science Foundation Major Research Instrumentation Grant 0923395). This study was in part supported by the Robert P. Apkarian Integrated Electron Microscopy Core (RPAIEMC) which is subsidized by the Emory College of Arts and Sciences and the Emory School of Medicine and is one of the Emory Integrated Core facilities. Additional support was provided by the National Center for Advancing Translational Sciences of the National Institutes of Health under award number UL1TR000454. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institutes of Health.

4.4.5 Scanning Transmission Electron Microscopy.

STEM data were acquired at Brookhaven National Laboratory (BNL). Assembly solutions for each variant (0.3 mg mL^{-1} in 10 mM acetate buffer, pH 4.0) were deposited on thin carbon (ca. 2 nm thick) supported on a thicker holey carbon film mounted on a titanium grid using the wet-film, hanging drop method. Tobacco Mosaic Virus (TMV) is added to the grid as an internal control, then injection buffer is added, followed by specimen solution. These solutions are allowed to sit for one minute, and excess solution is wicked from the edge with filter paper. Following, the grid is fast frozen by plunging into liquid nitrogen slush and stored under liquid nitrogen. All grids are freeze-dried overnight in an ion-pumped chamber with an efficient cold trap. Grids are then transferred under vacuum to the STEM cold stage, which is operated at -160 °C. Measurements were taken at 40keV with a scanning probe of <0.3 nm directed from a cold field emission source. Data was collected using bright field, small-angle dark field, and large-angle dark field scintillator

photomultiplier detectors which operate at 0-15, 15-40, and 40-200 mRadian respectively. The large-angle signal is proportional to the mass of the atoms in the path of the beam. Specimen quality was tested and mass calibration was standardized using TMV. Mass per length (M/L) values were calculated for Form IA using aforementioned TMV raft standards set at a theoretical m/L of 13.1 kDa/Å by using the software 'PCMass32' [10-11]. STEM data for Form IV were not obtained, as the beam caused significant sample degradation during experimentation, as decreasing signal intensity with each subsequent pulse was observed.

4.4.6 Small- and Wide-Angle X-ray Scattering Measurements.

Synchrotron SAXS/WAXS analyses were performed at the Advanced Photon Source at Argonne National Laboratory (Beamline 12-ID-B). Assembly solutions of 4 mg mL^{-1} in 10mM acetate pH 4.0 were dialyzed against pure 10mM acetate pH 4.0 such that the dialyzed sample is identical to the buffer standard in all manners except for peptide concentration. To achieve this, Slide-A-Lyzer dialysis cassettes (2,000 MWCO) from Thermo-Fisher were used to dialyze 0.5 mL of peptide assembly solution against buffer containing no peptide. SAXS/WAXS analysis was performed at 25 °C in a 1.5 mm quartz capillary flow cell (applying constant vertical agitation of the peptide assembly solution) so as to prevent radiation damage. For each Form variant, a minimum of thirty 2-dimensional shots were gathered and azimuthally averaged into 1-dimensional scattering spectra. Identically gathered data was obtained for the dialysis buffer (10mM acetate pH 4.0, no peptide present) and data were background subtracted to produce signal correlating exclusively to peptide-based assemblies in solution. Normalization against beam intensity and solid angle correction were performed by the in-house software suite (Igor Pro) present at Beamline 12-ID-B. Radii were calculated using Equation 2:

$R_c = R_{exp}/\sqrt{2}$ (or) $R_{exp} = R_c * \sqrt{2}$. (Equation 2)

4.5 References

1. 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.

2. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols* **2006,** *1* (6), 2876-90.

3. Rodger, A.; Marrington, R.; Geeves, M. A.; Hicks, M.; de Alwis, L.; Halsall, D. J.; Dafforn, T. R., Looking at long molecules in solution: what happens when they are subjected to Couette flow? *Phys Chem Chem Phys* **2006,** *8* (27), 3161-3171.

4. Marrington, R.; Dafforn, T. R.; Halsall, D. J.; MacDonald, J. I.; Hicks, M.; Rodger, A., Validation of new microvolume Couette flow linear dichroism cells. *Analyst* **2005,** *130* (12), 1608-1616.

5. Bulheller, B. M.; Rodger, A.; Hicks, M. R.; Dafforn, T. R.; Serpell, L. C.; Marshall, K. E.; Bromley, E. H. C.; King, P. J. S.; Channon, K. J.; Woolfson, D. N.; Hirst, J. D., Flow Linear Dichroism of Some Prototypical Proteins. *J Am Chem Soc* **2009,** *131* (37), 13305-13314.

6. Magnotti, E. L.; Hughes, S. A.; Dillard, R. S.; Wang, S. Y.; Hough, L.; Karumbamkandathil, A.; Lian, T. Q.; Wall, J. S.; Zuo, X. B.; Wright, E. R.; Conticello, V. P., Self-Assembly of an alpha-Helical Peptide into a Crystalline Two-Dimensional Nanoporous Framework. *J Am Chem Soc* **2016,** *138* (50), 16274-16282.

7. Cullity, B. D. S., S. R., *Elements of X-Ray Diffraction*. Prentice Hall: Upper Saddle River, NJ, 2001; Vol. 3rd Edition.

8. Spencer, R. K.; Nowick, J. S., A Newcomer's Guide to Peptide Crystallography. *Israel journal of chemistry* **2015,** *55* (6-7), 698-710.

9. Habibi, N.; Kamaly, N.; Memic, A.; Shafiee, H., Self-assembled peptide-based nanostructures: Smart nanomaterials toward targeted drug delivery. *Nano Today* **2016,** *11* (1), 41-60.

10. S. Wall, J.; N. Simon, M., *Scanning Transmission Electron Microscopy of DNA-Protein Complexes*. 2001; Vol. 148, p 589-601.

11. ftp.stem.bnl.gov, P. i. a. a.

Chapter V. Improving Designability and Predictability of Self-Assembling Peptide Systems: Computationally Designed Solenoidal Nanotubes Based on an Alpha-Loop-Beta Structural Motif

5.1 Introduction

Sequence-specific biomolecules, i.e., proteins and nucleic acids, have advantages as design elements for construction of nano-scale materials in that sequence-structure correlations can be established that promote the formation of structurally defined supramolecular assemblies. This information can be encoded at the molecular level within the sequences of the respective protomers to direct the formation of well-defined supramolecular assemblies through the progression of structural hierarchy. For example, the development of computational algorithms to predictably direct the self-assembly of oligonucleotides into defined 2D and 3D supramolecular structures (DNA origami[1] and nano-bricks[2-3]) has resulted in a revolution in nanotechnology with profound impact for applications in materials science. DNA has the advantage that the Watson-Crick base pairing rules provide a simple scoring function that can be employed to distinguish among alternative structural alignments in order to drive formation of the desired supramolecular structure. DNA interactions are "digital", in the sense that the interface between oligonucleotides can be predicted from knowledge of sequence alone. In contrast, the rules that govern protein-protein interactions are more complex and difficult to predict from sequence information alone [4]. However, the sequence specificity and structural diversity of protein-based assemblies should afford the opportunity to introduce functional complexity across length-scales, *if quaternary structure could be accurately predicted and effectively controlled from sequence information*.

Biological protein assemblies display exquisitely detailed tertiary and quaternary structures that underlie a complex range of native functions. The functional macromolecular machines of living organisms primarily arise from homomeric or heteromeric association of protein subunits (protomers) that involves selective recognition at structural interfaces. One of the most common modes of protein assembly is based on the application of a helical symmetry operation. In these helical assemblies, successive protomers are related by a characteristic angular twist (□) and axial translation [5-6]. The resultant assemblies comprise a supramolecular polymer in which the protomers associate primarily through non-covalent interactions. Helical protein assemblies encompass a range of functional roles in biological systems that would be desirable to emulate in simpler synthetic systems, including locomotion, controlled release, directional transport, dynamic switching, and shape-selective catalysis. Filamentous materials having helical symmetry are observed among viruses (e.g., filamentous phage capsids,[7-9] and tobamoviruses), cytoskeletal elements (e.g., actin,[10-11] and tubulin[12]), extracellular organelles (e.g., flagellar filaments,[13-14] type IV pili,[15-17], and the type III secretion system needle complex[18-20]), functional amyloids (e.g., curli fibrils[21-22] and other functional amyloids,[23-26]) and signaling complexes (e.g., pyrin domains in the inflammosome) [27].

The design of synthetic helical assemblies is complicated by the apparent infinite continuum of possible inter-protomer interfacial geometries, which presents a "chicken or the egg" problem for designing assemblies computationally. On one hand, to correctly anticipate the inter-protomer interface geometry, one needs to know the full protomer sequence and structure. On the other hand, the interfacial sequence cannot be designed until after the interface geometry is chosen. To help break this cycle, we utilize the observation that the universe of protein structure, rather than being a uniform continuum, is highly degenerate, with many frequently repeating structural motifs.

Secondary structure is certainly degenerate, with alpha-helices and beta-strands accounting for the majority of structured residues.[28] Relative geometries of secondary structural elements are also highly biased,[29-34] and even domains or inter-domain interfaces tend to recur in nature. Although several reasons may exist for this drastically varying motif utilization in nature[35-37] (e.g., incomplete knowledge of the structural universe or stochasticity in early evolution), an important reason is likely *designability*. Designability refers to the ease with which a structure or a structural motif can be realized with the 20 available native amino acids.[38-39] Because a protein of length *L* has a finite number of possible sequences, but an infinite continuum of physically plausible conformations, it must be that some conformations are not designable (i.e., are not ground states for any sequence), whereas others are. Structural motifs highly reused in nature are necessarily designable, whereas those absent or underused (but expected to have been well sampled based on their size) may well have low designability.

Herein we describe the use of protein designability,[30, 38, 40-43] estimated on the basis of native structural representation, as a mechanism to promote and control association of a small, folded protein domain based on an alpha-loop-beta structural motif. The alpha-loop-beta fold was selected on the basis of high designability, structural robustness, and ease of chemical synthesis. We chose to focus initially on the alpha-loop-beta fold since these sequences have not been investigated extensively for the creation of synthetic helical assemblies and, thus, represent relatively unexplored structural space. This structural motif occurs as a concatenated sequence feature within solenoid repeat proteins such as the alpha/beta leucine-rich repeat (LRR) proteins.[44-46] The high level of representation of these structures in the protein databank and the extensive degree of concatemerization within solenoid proteins[47] suggests that the interfaces between the repeat motifs may be designable and structurally stable. Moreover, the crystal structures of alpha/beta solenoid

repeat proteins display a natural helical curvature that suggests that the formation of extended helical assemblies can result from self-association of these structural motifs.[45] However, despite the curvature that results from extensive concatenation of the alpha/beta motifs, the corresponding solenoid proteins have not been observed to oligomerize into extended supramolecular assemblies under native conditions, which may be due to the presence of specific capping sequences that prevent further oligomerization. Removal of the capping sequences from native LRR sequences would likely result in uncontrolled polymerization due to strong interactions between the concatenated beta-strands of the LRR repeats. Moreover, the structural features of the native LRR sequences are not necessarily optimized to form helical nanotubes of the type that we envision. i.e., in which cohesive interactions are established at both the lateral and axial interfaces between protomers. Circular assemblies of alpha-loop-beta motifs have been computationally designed using the sequences of LRR repeats fit to a defined, structurally permissible closed geometry, which solely involves optimization of lateral interfaces [48]. While these assemblies were not observed to stack along the axial interface, the results suggest that the alpha-loop-beta motif is amenable to computational design.

5.2 Results and Discussion

5.2.1 Sequence and Assembly Design. We have previously demonstrated protein structure search algorithms that can be used to rapidly estimate the natural designability of any given motif based on its abundance in the Protein Data Bank (PDB).[39, 49] The central idea behind this approach is that any structure or interface can be decomposed into constituent local tertiary motifs, which we call TERMs (tertiary motifs), such that the designability of the overall structure can be judged based on the designability of its component TERMs. Further, sequence statistics from structural matches

for each TERM serve to identify sequence features important to the formation of the motif. Our design procedure starts with the choice of a protomer structure, which can be either an existing protein structure, a unit designed *de novo*, or even just an anchor motif around which to further organize the protomer. Whereas in the former case designability is ensured automatically, in the latter two cases high designability must be an explicit constraint. In the case of our prototype alpha-loop-beta assembly above, we initiated the promoter with a small alpha/beta motif, ensuring its designability by identifying many natively occurring motifs of similar structure (see Figure 5.1).

Once the protomer (or the protomer anchor) is chosen, the next task is to select a helical geometry as a target for the assembly design. A helical assembly can be fully described with six parameters– helical radius (*R*), pitch (*P*), and frequency ($\tilde{\omega}$), as well as three Euler angles defining the protomer's absolute spatial orientation in the helical frame of reference (see Figure 5.1). Furthermore, these parameters are not entirely independent. Given a certain protomer, its approximate size would dictate that only certain pitch values are admissible, as we ideally want adjacent turns of the helical assembly to form stable close contacts through axial interactions. Finally, given certain values of pitch and radius, and given the rough size of a protomer, only certain narrow range of helical frequencies would be admissible to allow for reasonable interaction between adjacent protomers (i.e., lateral interfaces). Thus, the total *effective* number of degrees of freedom is not quite six, but likely closer to four. Still, the space of possibilities is large. However, each combination of parameters leads to specific lateral and axial interfaces. So, applying our principle of designability, we can rapidly find the set of parameters that produce the most designable interfaces, by breaking each into its corresponding TERMs and describing their abundance individually (e.g., see Figure 1C). This concept has previously been demonstrated in the design of peptide-based assemblies for wrapping carbon nanotubes.[50] As a result, before the

final protomer sequence is known (or, in cases of choosing just the anchor motif, even before the full protomer structure is complete), we can already build the most promising assembly topology to target in design. We applied this procedure for designing our initial alpha-loop-beta assembly prototype, and it led to the geometry shown in Figure 5.1. If the final protomer structure remains incomplete (i.e., discontinuous in sequence space), as was the case with our simple alpha-loop-beta anchor motif, additional secondary-structure elements or stabilizing loops are added at this point, guided by the constraints of making additional designable contacts with the existing assembly template (red regions in Figure 1C). Finally, once the assembly geometry is complete, computational sequence design proceeds using standard tools (e.g., the Rosetta Design software suite was used for our initial alpha-loop-beta assembly). Importantly, the positional amino-acid preferences that emerge from the analysis of TERM designability serve to constrain amino-acid choices available to the computational design approach, focusing sequence exploration to the most promising region in sequence space. Importantly, this procedure is highly general, enabling it to be easily applied to a variety of different protomer structures and assembly parameter requirements (e.g., different radii; with or without a central channel). The ability to rapidly design large libraries of different assembly structures and sequences also enables post-selection based on properties more difficult to include during design, such as predicted mechanical properties of the assembly (e.g., persistence length) or opportunities for functionalization (e.g., the ability to transplant enzymatic or small-molecule/metal binding sites).

Three candidate sequences were computationally identified for the alpha-loop-beta assembly, **Sol1-3**, and subjected to energy minimization and molecular dynamics simulation to provide further assurances of their stability (Figure 5.2A). Computational models of the assembled state for these peptides (shown in Figure 5.1 for **Sol1**) serve as valuable benchmarks for

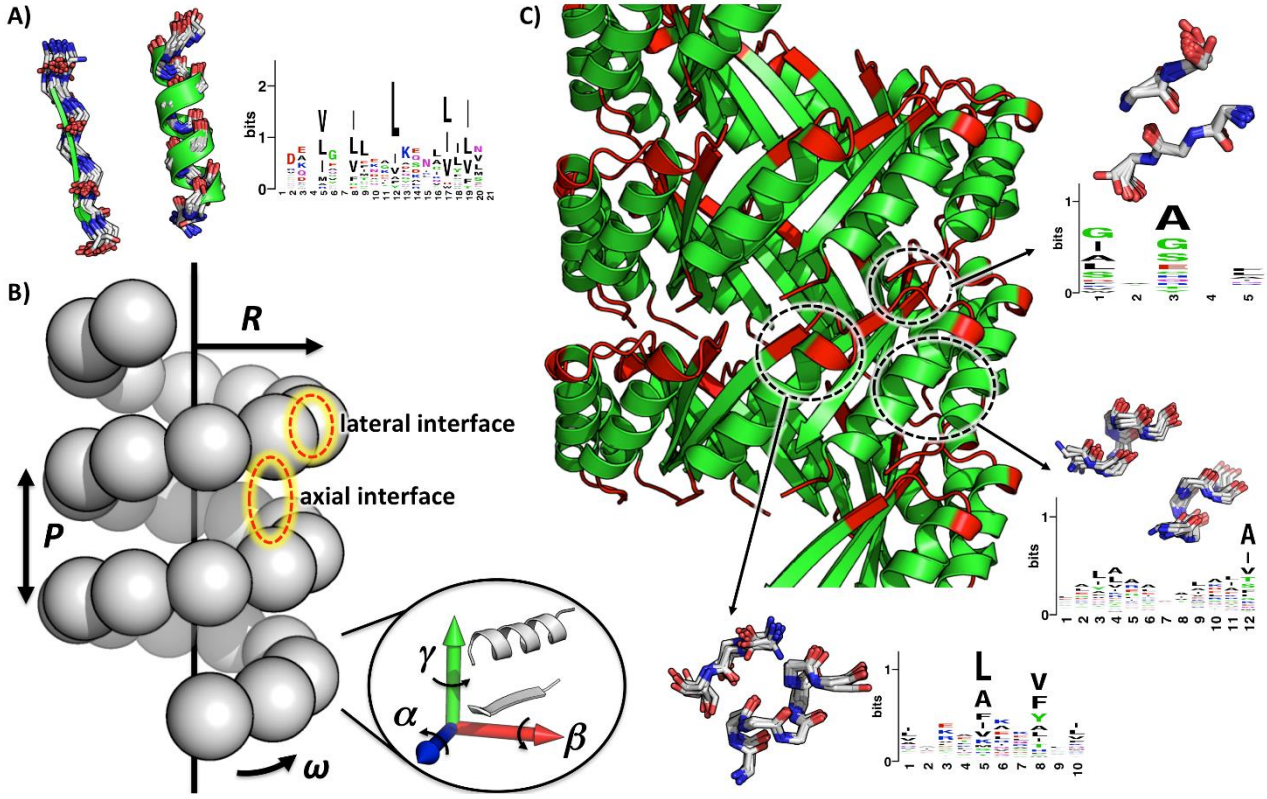


Figure 5.1: Computational design approach. (A) The anchor motif for the alpha-loop-beta assembly (green cartoon) along with closely matching native representatives (gray sticks) and the sequence logo capturing native amino-acid preferences (right). (B) Geometric parameters of a helical assembly. (C) Design model of the alpha-loop-beta assembly for peptide sol1 (red regions indicate additions to the alpha-loop-beta anchor motif made during the design process). Shown are three TERMs formed by various interfaces within the assembly (outlined with dashed circles) along with their native representatives (gray sticks) and resulting sequence logos.

comparison to experimentally derived metrical parameters. Each peptide comprises 29 amino acids in which an *N*-terminal alpha-helix is connected to a *C*-terminal beta-strand through a short, 3-residue loop sequence. The length of the peptide is comparable to the length of LRR repeat sequences of the alpha-loop-beta type that are observed for the porcine ribonuclease inhibitor (RI) and the NOD-like receptor protein (NLRP) sub-families. In each case, the peptides are capped at the *N*-terminus with an acetyl group and the *C*-terminus with an *N*-methyl amide group. These groups were selected based on the computational analysis to provide sufficient hydrophobic contacts to stabilize axial interactions between successive turns of the helical assembly. **Sol1** and **Sol2** differ in that alanine and glycine residues, respectively, occupy position 20 in the sequence, which resides within the turn sequence at the axial interface. Ala20 results in slight overpacking of this interface as it makes close contact with the *N*-acetyl group. These two peptides allow us to probe the effect of hydrophobic packing at the axial interface of the stability of the helical assembly. **Sol3** results from an I24T substitution within the sequence of **Sol1**, which was employed to lower the overall hydrophobicity and to improve the solubility of the peptide without compromising self-assembly behavior. The requirement for designable lateral and axial interfaces is apparent in the model, as it shows the beta-strand of the *n-1* peptide packing against the alpha-helix of the *n* peptide within the assembly (see Figure 5.2).

The computational model of the **Sol** series assemblies comprises a closed helical assembly, i.e., helical nanotube, with approximately 10 peptides per turn (ϕ = 36°), a pitch of circa 21 Å, and a radius of ~6 nm. The alpha-helices are located on the exterior (i.e., convex) surface and the beta-strands are located on the interior (i.e., concave) surface. The requirement for designable lateral and axial interfaces is apparent in the model in Figure 1C. For each protomer, the beta-strand of the *n-1* peptide packing against the alpha-helix of the *n* peptide within the assembly. This

mode of interaction represents a *de facto* domain swapping mechanism for self-assembly and suggests that folding of the protomer may be conflated with intermolecular association. Similar domain swapping has been observed to stabilize lateral association between protomers in biologically derived helical assemblies. Axial interactions occur between the turn sequence of the *n* protomer and the *N*- and *C*-terminal residues of the *n+11* protomer. The axial interface buries less surface area than the lateral interface (need to calculate both of these values and confirm this observation). Hydrogen bonds are observed between structurally adjacent parallel beta-strands along the concave surface of the lumen. A right-handed, upward shear corresponding to 2 residues occurs between beta-strands of the *n* and *n+1* protomers such that the *i-th* residue of the former

Sol1: Ac-YIPSEALRKLAEALKLAGVAFRIILQIEA-N(H)-CH_3
Sol2: Ac-YIPSEALRKLAEGLKLAGVGFRIILQIEA-N(H)-CH_3
Sol3: Ac-YIPSEALRKLAEALKLAGVAFRITLQIEA-N(H)-CH_3

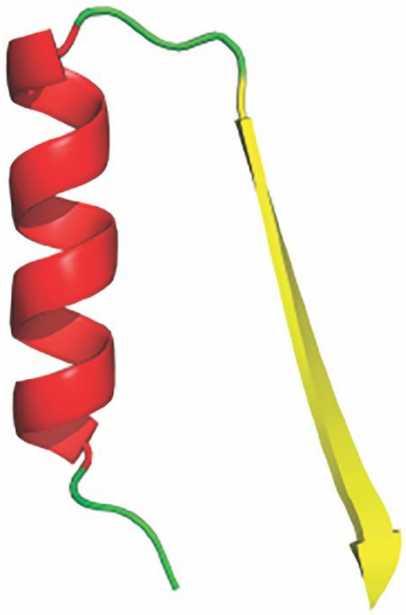


Figure 5.2. Sequences and corresponding protomer structure of the Sol series peptides. Sequence elements corresponding to the secondary structures are color coded as in the motif rendering. Residue differences between the peptides are underlined.

strand is directly adjacent to the *(i+2)-th* residue of the latter. The formation of this hydrogen-bonded network should reinforce the helical displacement between adjacent strands and strengthen the lateral interactions between protomers.

The peptides have been purified to homogeneity using preparative HPLC and the respective molar masses were confirmed using ESI-mass spectrometry. All three peptides exhibited limited solubility in conventional aqueous buffers over a pH range from 4 to 8, which could potentially be attributed to the strong aggregation potential of the beta-strands as well as a high content of hydrophobic residues (circa 65-70%). In order to weaken the lateral interactions and improve solubility, peptides were initially dissolved in 1:1 (V/V) mixtures of aqueous buffer, (pH titrated to 4.0) and of 2,2,2-trifluorethanol (TFE) to a final concentration of 0.3 mg mL^{-1}, which enabled efficient solubilization of peptide. TFE was subsequently allowed to slowly evaporate over time until the volume of the mixture was reduced by approximately 50%, typically taking between 1 and 5 days depending on total volume. This process coincided with the onset of fibril formation as detected using TEM image analysis. The most optimal buffer system consisted of a 1:1 mixture of TFE with aqueous acetate buffer (10 mM, pH 4.0). While the presence of trace amounts of TFE cannot be ruled out, it did not prohibit the subsequent self-assembly of fibrils.

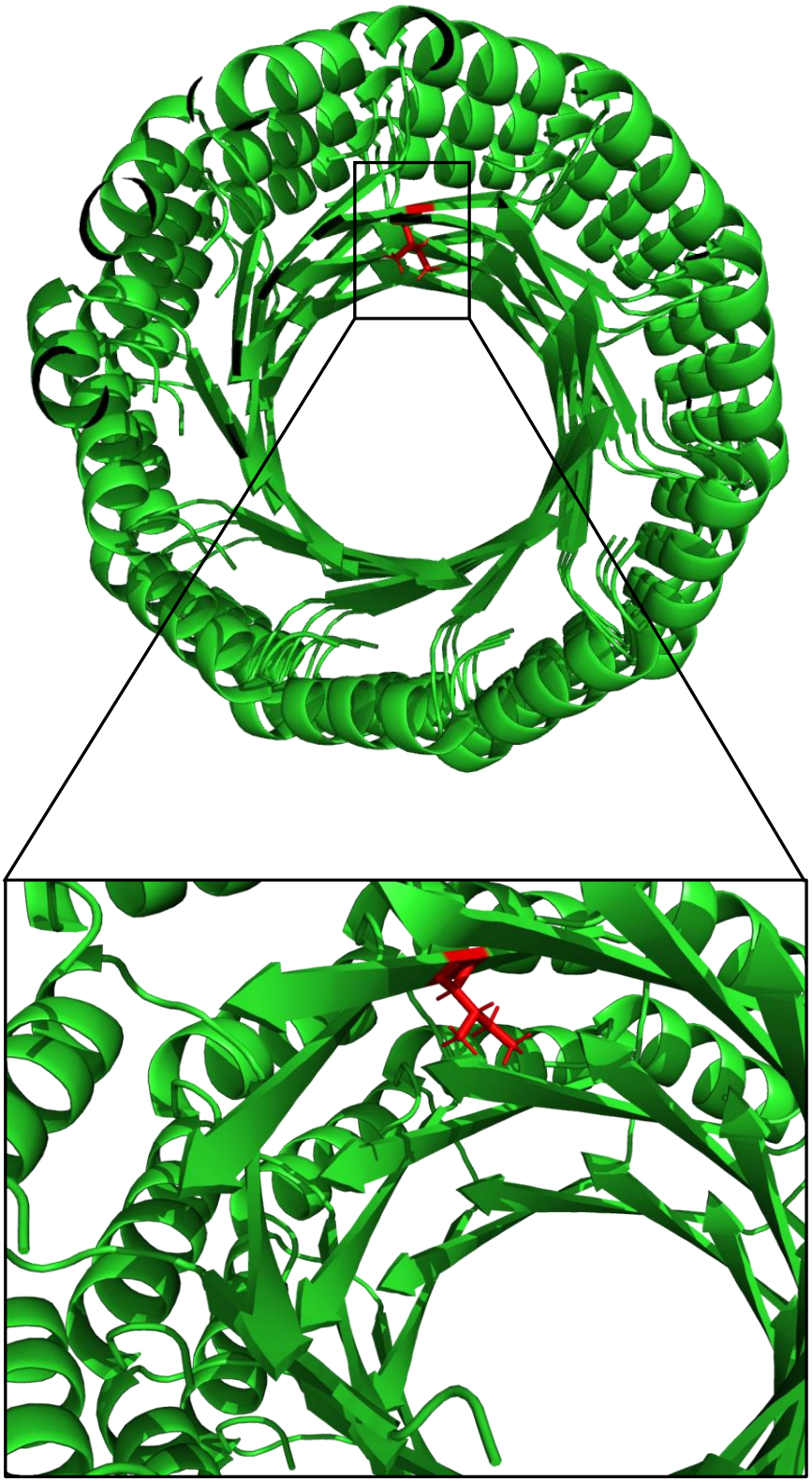


Figure 5.3: PyMol renderings of the Sol1 I→T (Sol3) mutation, which occurs at the hypothesized inner beta-sheet layer.

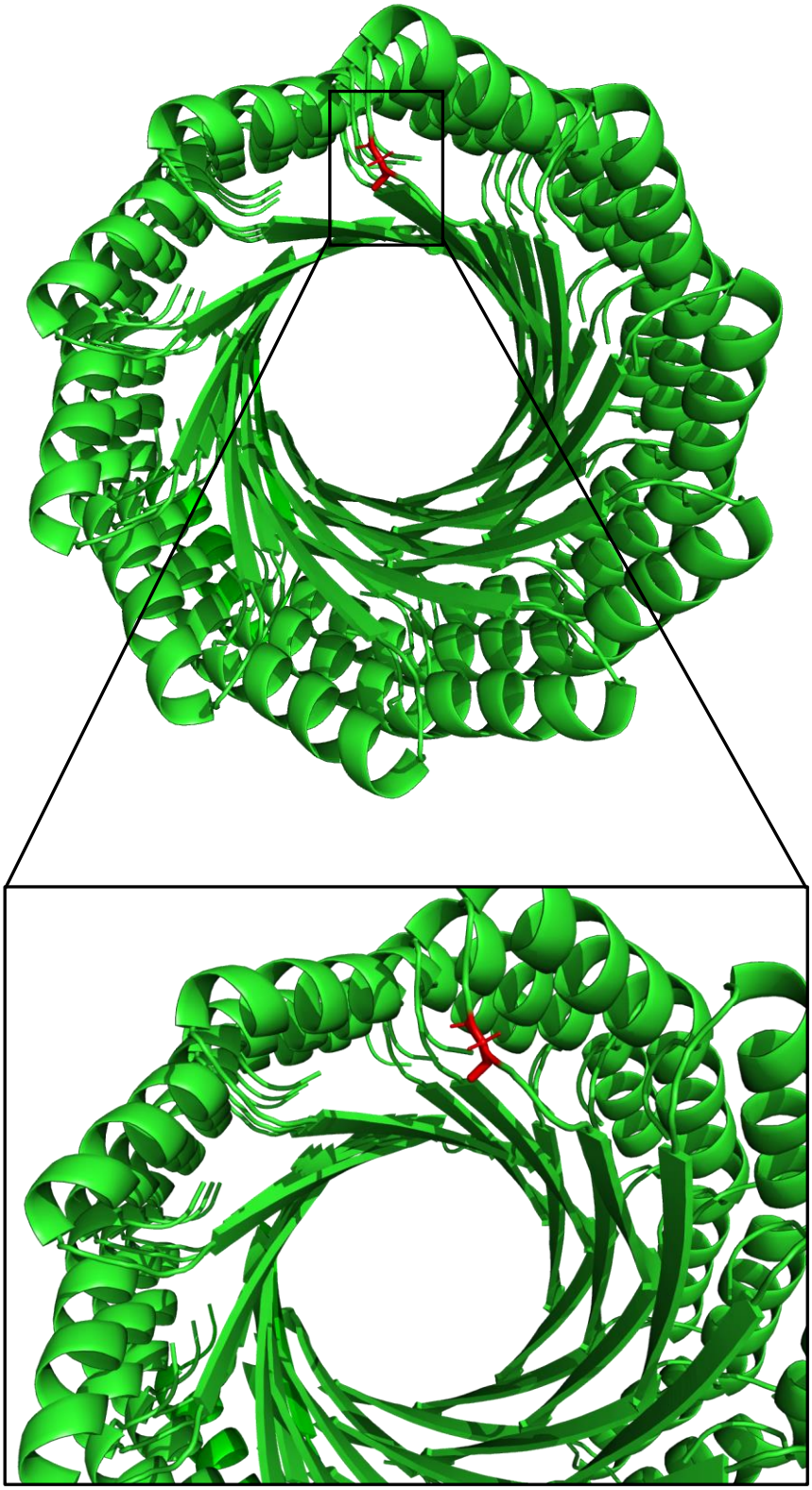


Figure 5.4: PyMol renderings of the Sol1 A→G (Sol2) mutation, which occurs at the hypothesized interstitial loop moiety.

5.2.2 Circular Dichroism

Circular dichroism spectropolarimetry of aqueous solutions of the three **Sol** variants, prepared as described above, displayed spectroscopic behavior indicative of a superposition of alpha-helical and beta-sheet signatures. The inherently low resolution of CD analysis limits insight into the conformational behavior of the peptides, however it is consistent with the structural model of Figure 1C wherein each subunit comprises an alpha-loop-beta structural motif. The computational model of **Sol1** was employed to calculate a CD trace using the program Dichrocalc (Figure 5.5),

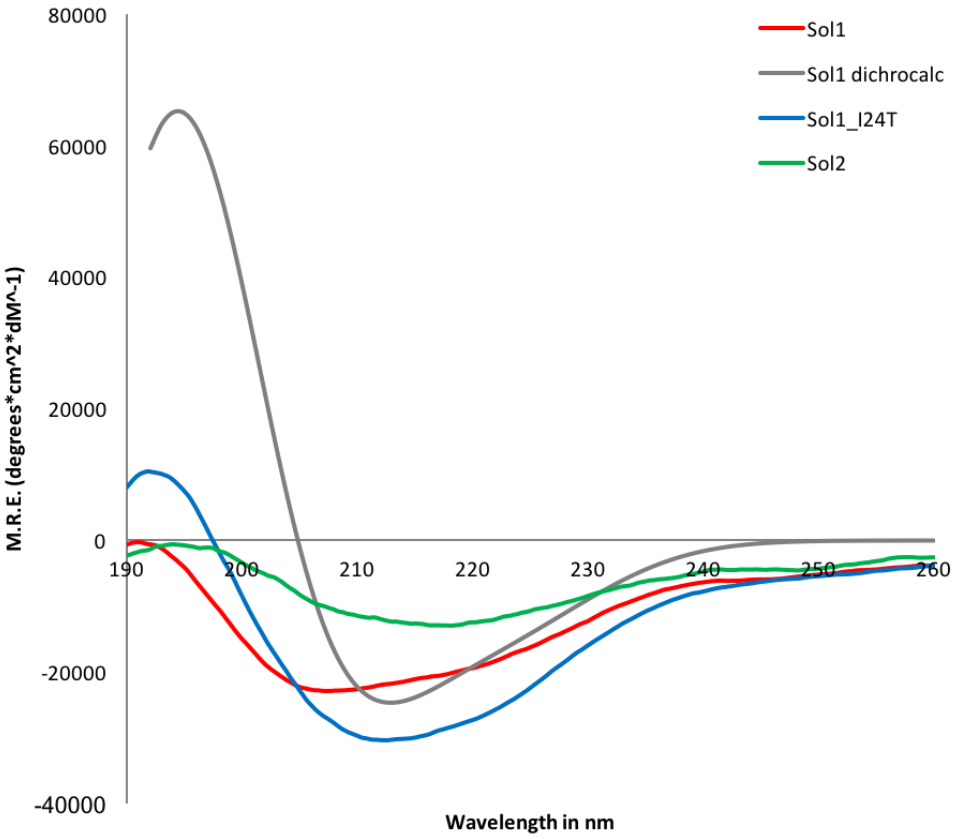


Figure 5.5: Circular dichroism data demonstrating experimentally derived spectra for each variant compared to a predicted trace based on the sequence and computationally modeled structure of the subunit. The predictive trace (shown in grey) is calculated via the Dichrocalc algorithm. http://comp.chem.nottingham.ac.uk/dichrocalc/

which shows a minimum at approximately 212 nm. However, a qualitative analysis of the experimental CD data shows signals blue-shifted with respect to the Dichrocalc prediction, though still between the canonical alpha-helical minima (222 and 208 nm) and beta-sheet minima (~218 nm). From analysis of the computational model, it is approximated that the subunit comprises 31% beta-sheet, 21% loop, and 48% alpha-helix. While qualitatively similar, significant differences are observed in the CD spectra among the Sol series peptides. **Sol2** exhibits the weakest and broadest CD spectrum, while **Sol1** and **Sol3** are more closely related to each other and to the expected curve of a hybrid alpha/beta peptide (Figure 5.5). The only difference between **Sol1** and **Sol3** resides in a conservative mutation of isoleucine to threonine at a solvent-exposed position

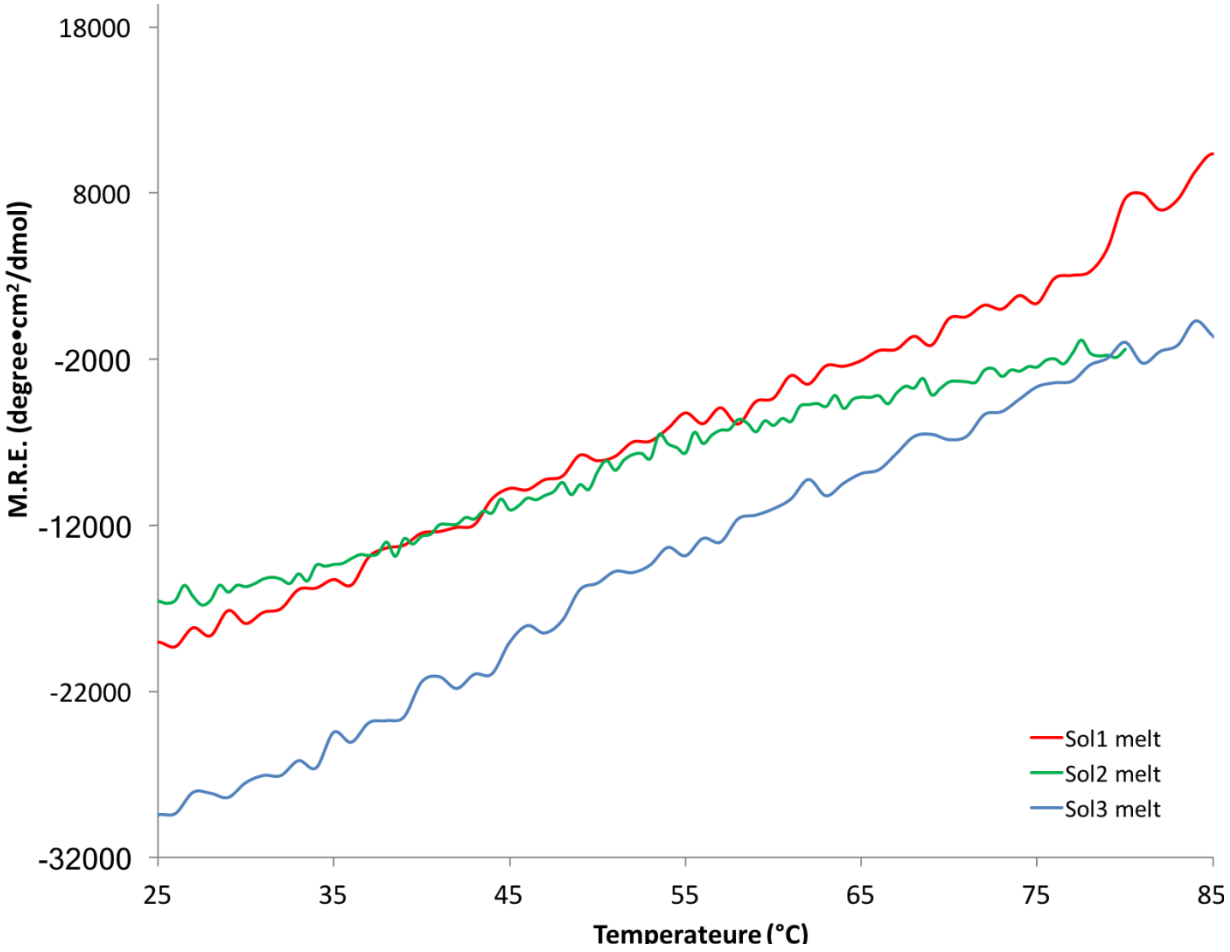


Figure 5.6: thermal denaturation curves for each Sol variant from 25 degrees Celsius to 85 degrees Celsius.

within the beta-strand. However, **Sol2** differs with both peptides in that a smaller glycine residue substitutes for alanine in the loop region that participates in the axial interaction between protomers. The relatively weak CD spectrum of **Sol2** may be consistent with the influence of the greater flexibility and lower hydrophobic surface area of the glycine residue, which could destabilize the target fold. In addition, thermal denaturation studies were performed using CD analysis for all three peptides. In each case, the melting curves displayed consistent loss of secondary structure with increasing temperature, but no evidence for a cooperative melting transition. This observation contrasts with the cooperative thermal denaturation of a computationally designed beta-alpha-beta assembly that forms an amyloid-like filament [51]. The observed differences between these two systems may be attributed to a combination of factors including extensive non-local interactions for the **Sol** series, solubility differences, and structural polymorphism.

5.2.3 Transmission Electron Microscopy. Transmission electron microscopy (TEM) data indicated the presence of high aspect-ratio fibrillar assemblies for all three of the **Sol** peptides. Fibrils of the **Sol** peptides display extensive length polymorphism, as expected for unrestricted polymerization of uniaxial filaments, with average lengths on the order of multiple microns. In contrast, the fibril diameters of **Sol** peptide assemblies were relatively uniform, as estimated from the apparent widths of negatively stained TEM images. Fibril widths were measured in ImageJ and the estimated averages were observed at values of 6.25 ± 0.81 nm, 6.64 ± 1.19 nm and 6.27 ± 1.04 nm for **Sol1**, **Sol2**, and **Sol3**, respectively (Figures 5.6-5.8). Measurements were taken in triplicate across three independently assembled samples for each **Sol** variant. The measured fibril

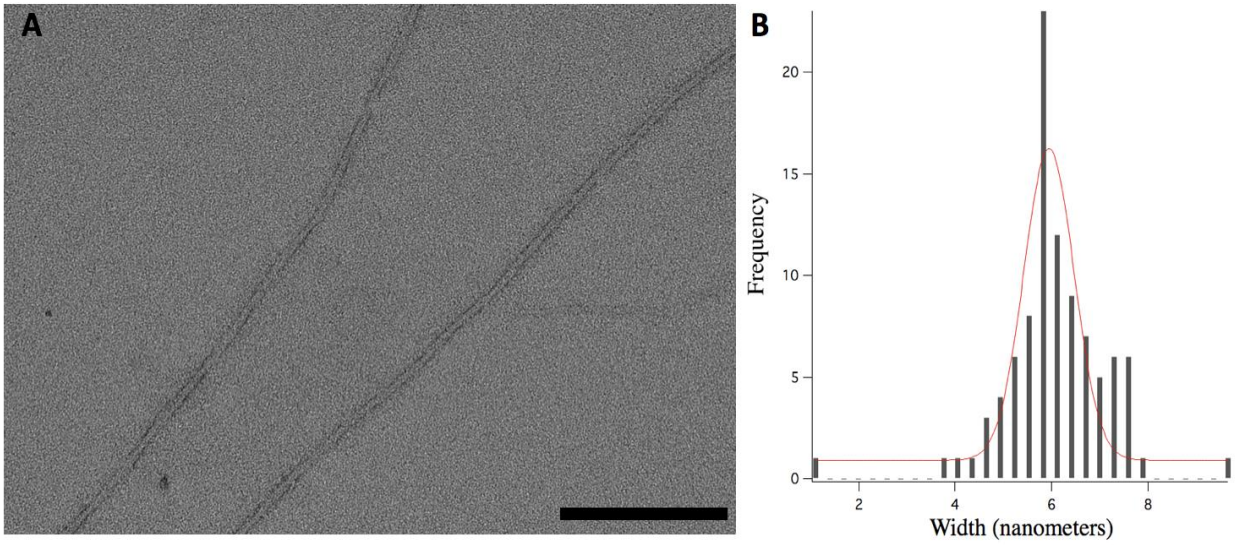


Figure 5.7: TEM image of Sol1 filaments (A) and histogram of the measured filament widths via ImageJ (B).

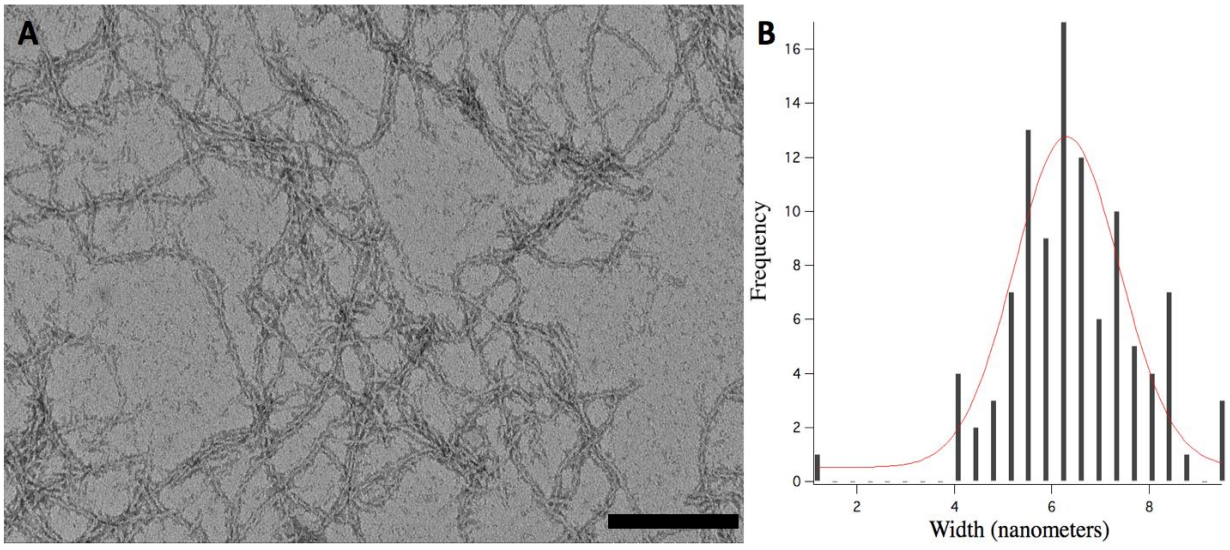


Figure 5.8: TEM image of Sol2 filaments (A) and histogram of the measured filament widths via ImageJ (B).

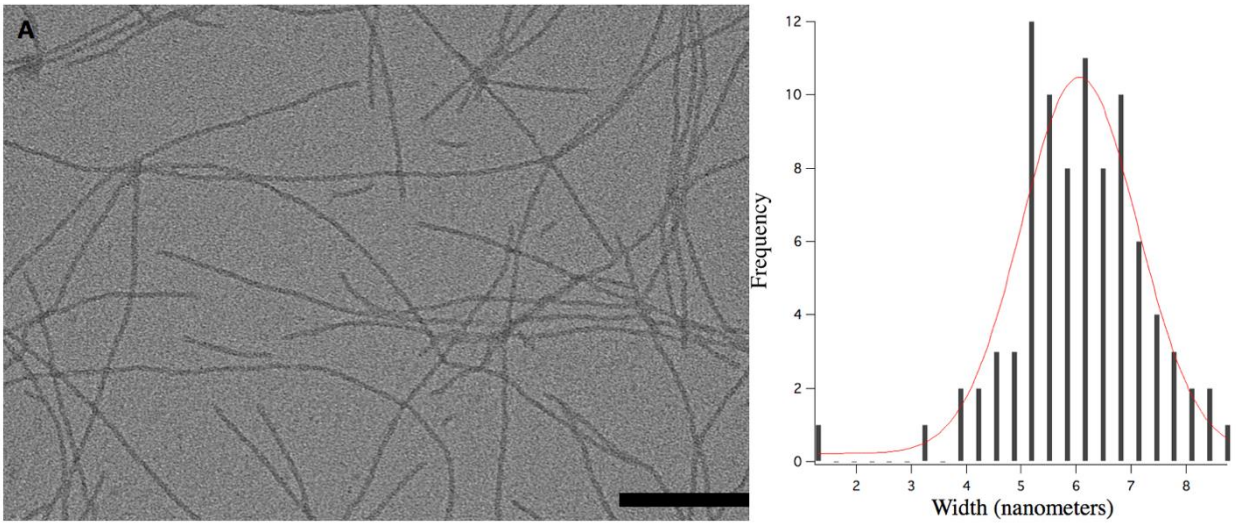


Figure 5.9: TEM image of Sol3 filaments (A) and histogram of the measured filament widths via ImageJ (B).

widths for the **Sol1**, **Sol2**, and **Sol3** assemblies were indistinguishable within experimental error. These values compare well to the diameter values calculated from the computational model of 6.26 nm for each of the models. These TEM data suggest that the **Sol** peptides form helical nanotubes of approximately similar diameter to that of the computational model. Despite the apparent similarity between the different **Sol** peptide variants, the presence of a significant number of narrow ribbon-like filaments could also be detected in samples of the **Sol2** assemblies (Figure 5.9). This observation can be explained in terms of a self-assembly mechanism in which lateral association is more facile than the axial interaction between successive turns of the helical fibril. The lateral interaction derives from the formation of a hydrogen-bonded network of the parallel beta-strands that is supplemented with hydrophobic contacts between alpha-helices and beta-strands on structurally adjacent protomers (Figure 5.1). In comparison, the axial interaction is primarily mediated through burial of hydrophobic surface area at the interface of protomers at successive turns of the helical assembly. Coiling of helical

assemblies into ribbons and subsequently into nanotubes is a common occurrence for peptides that adopt a cross beta-sheet conformation, in which hydrogen bond formation drives self-association and subsequent coiling occurs due to sheet stacking mediated primarily through hydrophobic interactions [52]. The formation of coiled ribbons that close into nanotubes has been observed in the self-assembly of amyloidogenic oligopeptides [53]. We expect that in the mechanism of coalescence

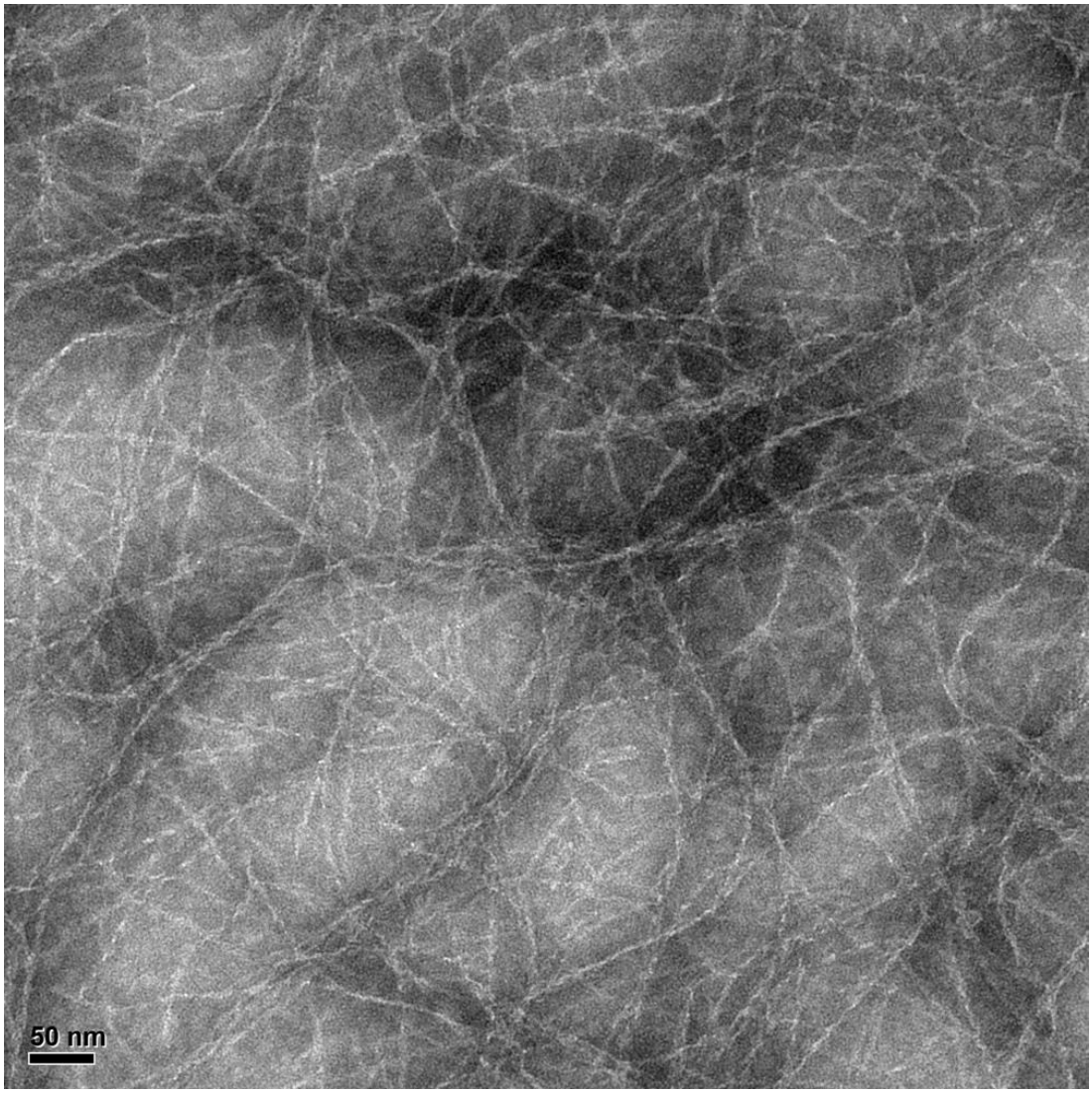


Figure 5.10: Negatively stained transmission electron micrograph of Sol2 demonstrating the prevalence of the 'unwinding' effect of the fibrils, wherein the fibers do not fully close down into a tightly packed solenoidal filament.

of the ribbons into tubes should differ between the **Sol** series of peptides and amyloidogenic sequences since one face of the beta-sheet is blocked by the alpha-helix for the **Sol** peptides and, therefore, is incapable of extended sheet stacking. The formation of twisted ribbons for **Sol2** may reflect the fact that insufficient surface area is buried at the axial interface due to the small size of the glycine residue.

5.2.4 Scanning Transmission Electron Microscopy

Scanning transmission electron microscopy (STEM) has proven to be a valuable tool for determining mass per length (MPL) values for one-dimensional and two-dimensional peptide assemblies [54-56]. These experimental values can be compared to values calculated for structural models as a method for validation of the corresponding model. MPL measurements were taken for each variant. The MPL values of 1,502 ± 154.8 Da/Å, 1,280 ± 113.2 Da/Å, and 1,421 ± 158.1 Da/Å were determined from STEM measurements for fibrils of **Sol 1**, **Sol 2**, and **Sol 3**, respectively (Figures 5.10-5.12). These data were compared to the values calculated from the corresponding computational models for the Sol assemblies (~1,500 Da/Å for each). The experimental MPL values for **Sol1** and **Sol3** fibrils were found to be comparable within experimental error to the MPL values calculated from the computational model. In contrast. The experimentally determined MPL value for the **Sol2** fibrils was significantly lower than that derived from the computational model. This observation is consistent with the presence of unwound (i.e., ribbon-like) segments of the helical assembly interspersed with closed (tube-like) segments. Notably, the MPL distribution for **Sol1** indicated the presence of a major peak (1526.7 Da/ Å, blue trace in the histogram for Figure 5.10) and a minor peak at a lower MPL value (1,257 Da/Å, green trace in the histogram for Figure 5.10), which may potentially be due to the presence of a minor component of locally unwound

filaments. Careful analysis of TEM images of the **Sol1** fibrils could detect unwound segments within an otherwise homogeneous sample preparation. These defect sites might result from

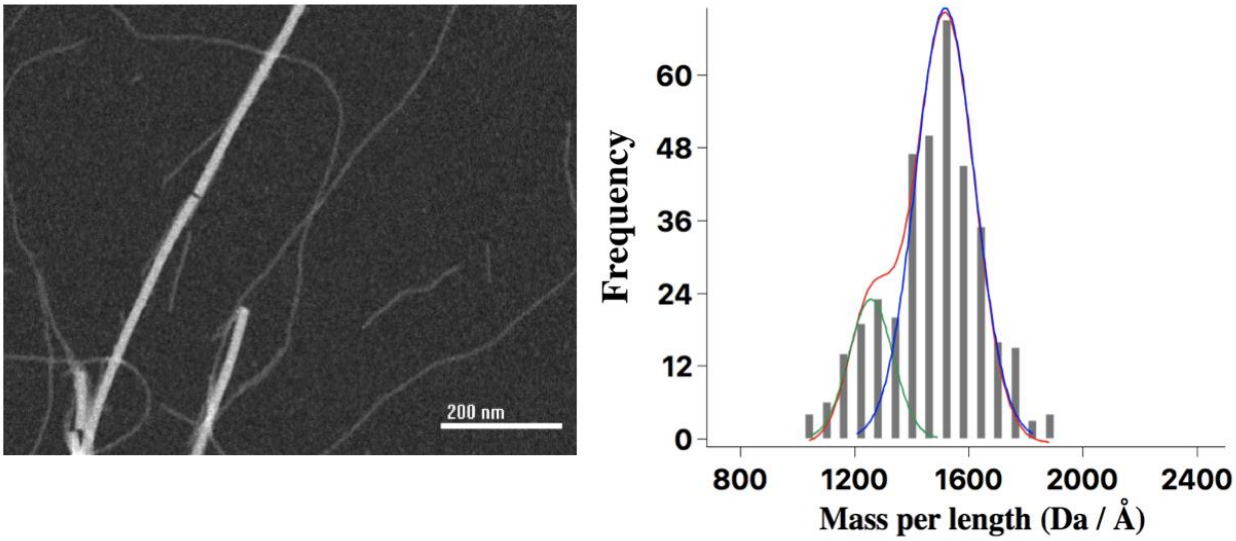


Figure 5.11: Scanning transmission electron micrograph of unstained, freeze-dried Sol1 and corresponding histogram of measured MPL values of the filaments.

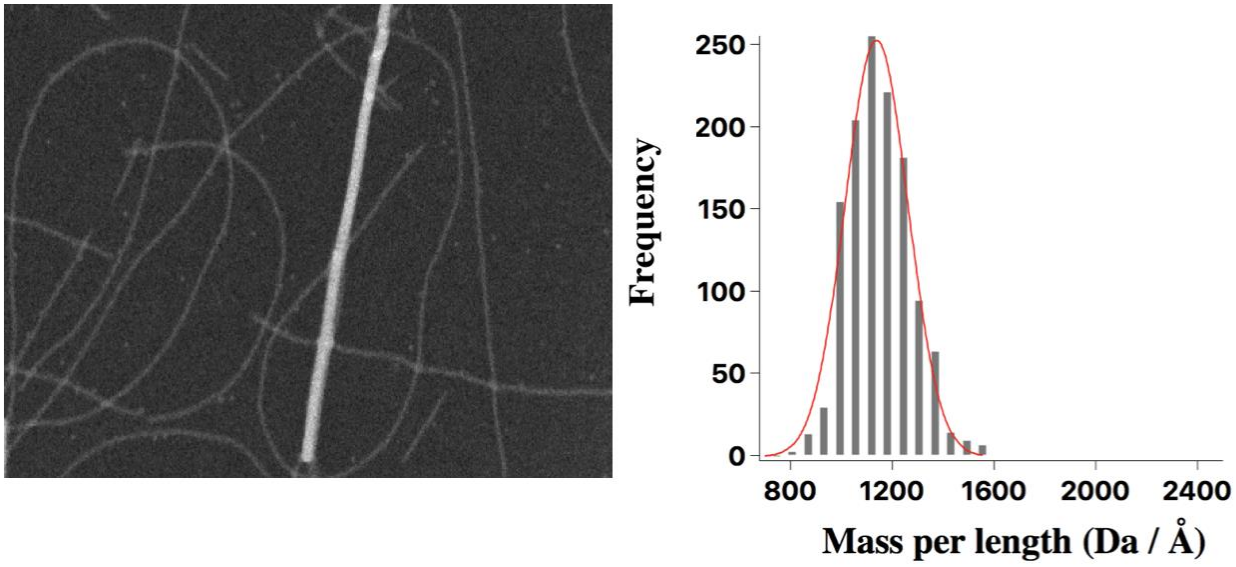


Figure 5.12: Scanning transmission electron micrograph of unstained, freeze-dried Sol2 and corresponding histogram of measured MPL values of the filaments.

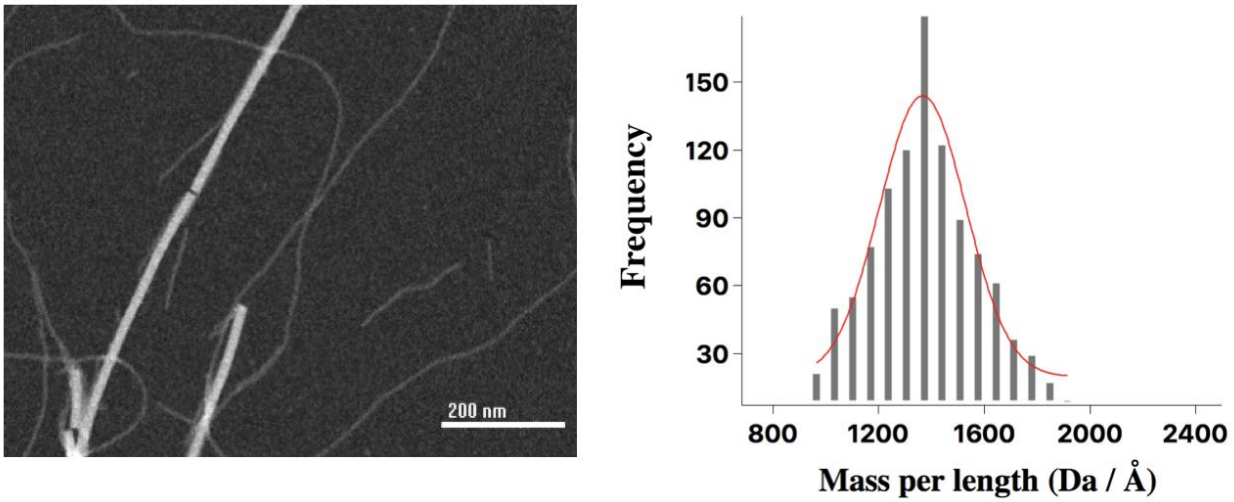


Figure 5.13: Scanning transmission electron micrograph of unstained, freeze-dried Sol3 and corresponding histogram of measured MPL values of the filaments.

interaction of the filaments with the surface of the grid or may represent kinetically trapped states upon sample adsorption. The **Sol1** peptide is notably less soluble than either **Sol2** or **Sol3**, and can undergo phase separation during sample preparation. These results demonstrate that STEM MPL measurements are a powerful method to gain insight into structural differences between closely related peptide assemblies, especially when corresponding structural models are available.

5.2.5 Small-angle X-ray Scattering

While the electron microscopy data provide evidence for the formation of fibrillar structures that matched the computational predictions for **Sol1** and **Sol3**, it did not confirm the presence of a tubular structure. Short of determination of an atomic-resolution structure, small-angle X-ray scattering (SAXS) represents one of the few methods that are available to interrogate the shape and structure of colloidal assemblies. Synchrotron SAXS measurements were performed on solutions of the **Sol** peptide assemblies. Representative scattering intensity profiles as a function of the momentum transfer, *q*, are depicted in Figure X.13A and X.13B for the **Sol1** and **Sol3**

assemblies, respectively. For **Sol1** and **Sol3**, the scattering in the small-angle region (i.e., '*q'* analyzed at < 0.2 $Å^{-1}$) indicates that the global shape of the assemblies is rod-like, as it approximately follows the q^{-1} law with values equaling ~0.983 ± 0.101 and ~1.166 ± 0.015 for **Sol1** and **Sol3**, respectively. In comparison, the scattering in the small-angle region for solutions of the **Sol2** assemblies did not indicate the presence of a rod-like assembly, instead demonstrating an I(q) greater than 2 at 2.540 ± 0.019 (X.14), which is incongruous with the q^{-1} law for rod-like structures. This observation may be attributed to the tendency of **Sol2** assemblies to uncoil from a suppositious helical nanotube to a twisted ribbon, as presumed from TEM (Figure 5.9) and STEM MPL measurements. Moreover, the strong deviation of the scattering data from rod-like behavior indicates that unwinding and resultant structural polymorphism of the **Sol2** assemblies is significantly greater in the solution state than in dried sample preparations for EM measurements.

Scattering data within the low q (~<0.05 $Å^{-1}$) region were fit to the modified Guinier equation for rod-like forms for each of the **Sol** variants. In each case, the cross-sectional radius of gyration, R_c, for the corresponding rod-like assemblies were determined and used to calculate the cross-sectional radius, R (= $\sqrt{2}R_c$). The Guinier fits of the scattering data for **Sol1** and **Sol3** yielded R_c values of 22.16 ± 0.93 Å (R = 31.34 ± 1.32 Å) and 23.37 ± 0.867 Å (R = 35.77 ± 1.23 Å), respectively (Figures 5.15A and 5.15B). The calculated values for the cross-sectional radii for **Sol1** and **Sol3** closely correspond to each other. In contrast, the calculated R_c value of 59.16 ± 2.11 Å (R = 83.35 ± 2.99 Å) for **Sol2** (Figure 5.16) differs significantly from the corresponding values for **Sol1** and

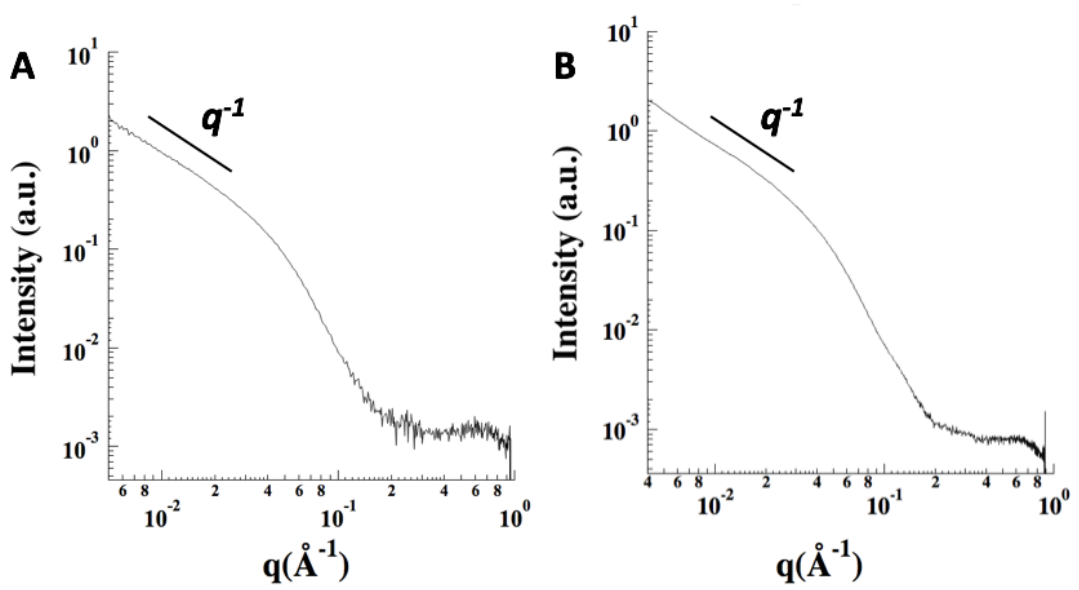


Figure 5.14: Small-angle X-ray scattering (SAXS) curve for solutions of the two variants which assemble as hypothesized into fully enclosed tubes- (A) Sol1 at 3.6 mg mL^{-1} and (B) Sol3 at 4.0 mg mL^{-1} in acetate buffer (10 mM, pH 4.0, TFE dialyzed from solution).

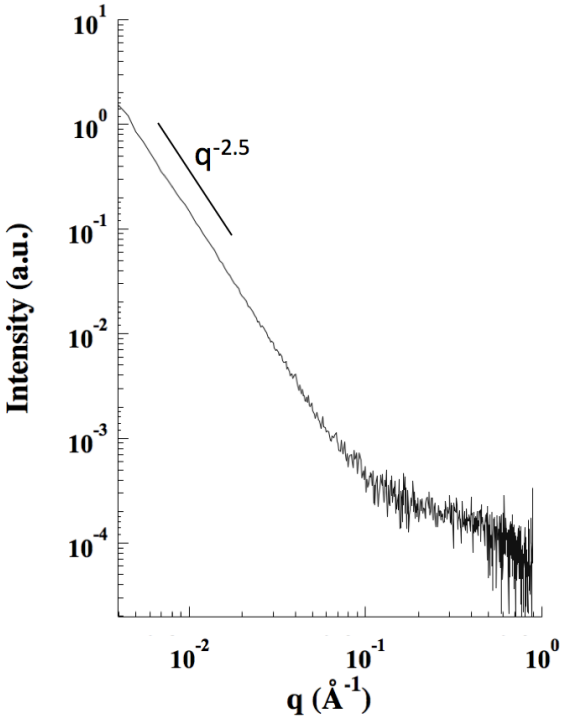


Figure 5.15: Small-angle X-ray scattering (SAXS) curve for the solution of the variant which did not assemble as hypothesized into fully enclosed tubes- Sol2 at 3.0 mg mL^{-1} in acetate buffer (10 mM, pH 4.0, TFE dialyzed out of solution.

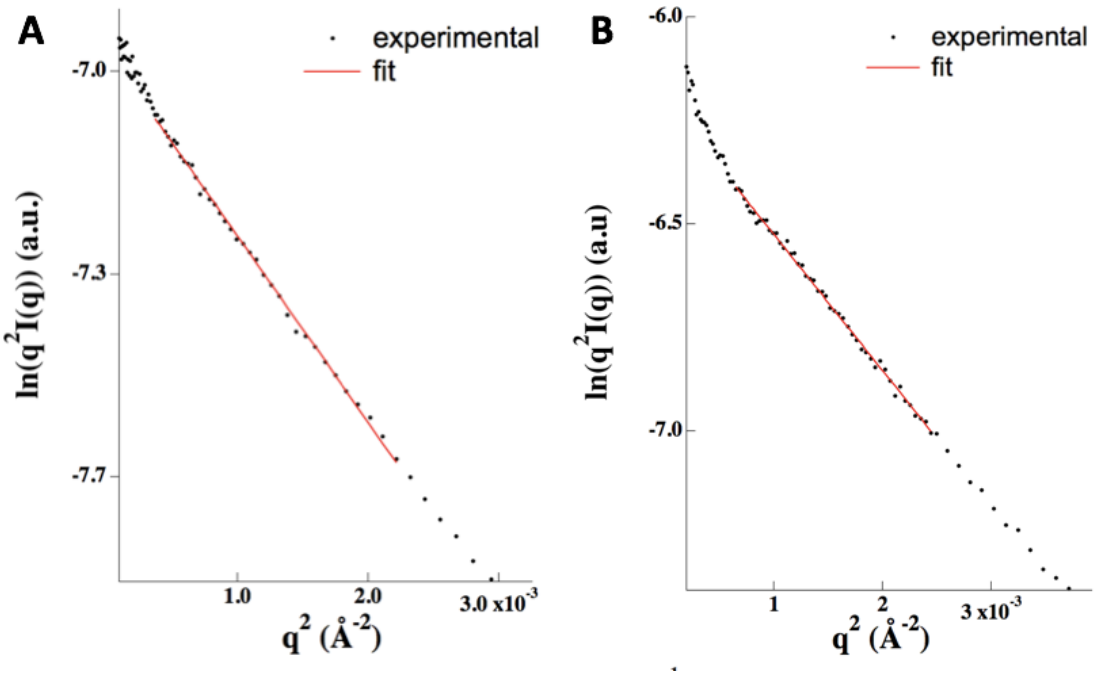


Figure 5.16: Modified Guinier plots, ln(q^2I(q)) versus q^2 of scattering data for the two variants which assemble as hypothesized into fully enclosed tubes- Sol1 (A) and Sol3 (B) from which Guinier-derived cross-sectional diameter values are calculated.

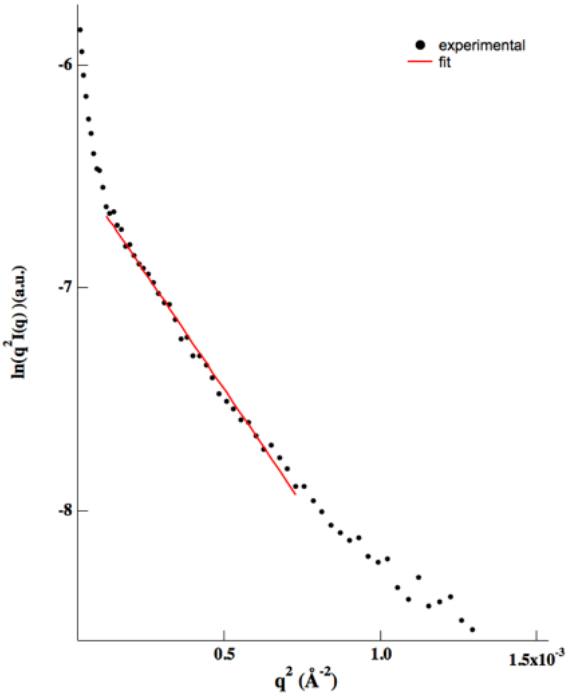


Figure 5.17 Modified Guinier plot, ln(q^2I(q)) versus q^2 of scattering data for the variant which does not assemble as hypothesized into a fully enclosed tube- Sol2.

Sol3, as well as the computational model. This result is not surprising in that the fractal dimension of the form factor scattering of the **Sol2** assemblies deviates significantly from expectations for a one-dimensional assembly. The calculated cross-sectional radii for the **Sol1** and **Sol3** are greater than those determined from TEM measurements, but still within experimental error, especially in consideration of the experimental differences between the solution and solid states.

Finally, a predicted scattering curve was produced using a 300 subunit PDB model of the computationally generated Sol1 Structure. This predicted trace can be seen in figure 5.17 (blue) and can be compared to the experimentally derived scattering curve (orange) which is largely featureless. The x-axis and y-axis are arbitrary, and the predictive curve is provided as a qualitative reference point and does not provide discrete Bragg diffraction peak values or corresponding d-

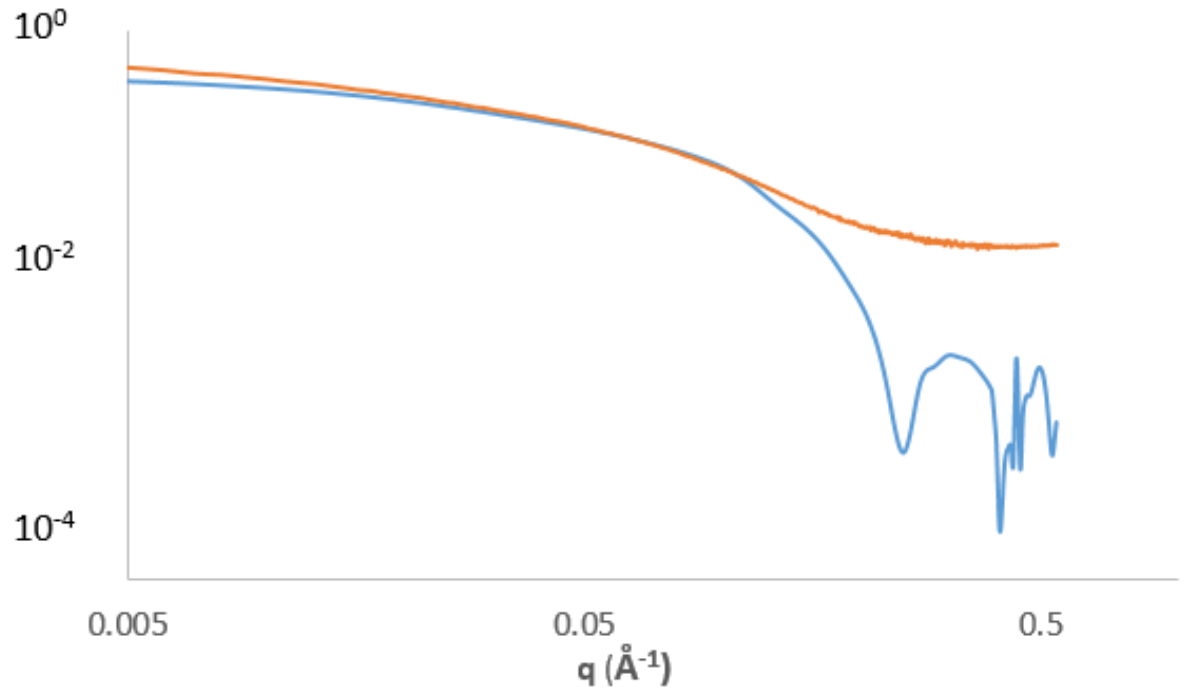


Figure 5.18: Crysol generated scattering curve (blue) for Sol1 generated from a 300 subunit PDB file compared against the experimentally derived scattering curve (orange).

spacings. The predictive curve was produced using Crysol program from Svergun and colleagues at EMBL-Hamburg.[57]

5.2.6 Cross-Sectional Pair Distance Distribution Functions

Cross-sectional Pair Distance Distribution Functions (PDDFs) were calculated from the scattering data for **Sol1** and **Sol3** (Figure 5.19). Significantly, the PDDFs determined from the experimental scattering data do not match the simulated data calculated from the corresponding computational model. A computational model of 300 subunits was employed to produce the aforementioned PDDF simulation data, as well as a simulation of X-ray scattering profile. This larger assembly should more closely approximate the aspect ratio of a helical filament, while yet remaining a computationally tractable model. The shape of the PDDF obtained from simulated data (blue) differs from that derived from the experimental data (red). The experimental PDDF has

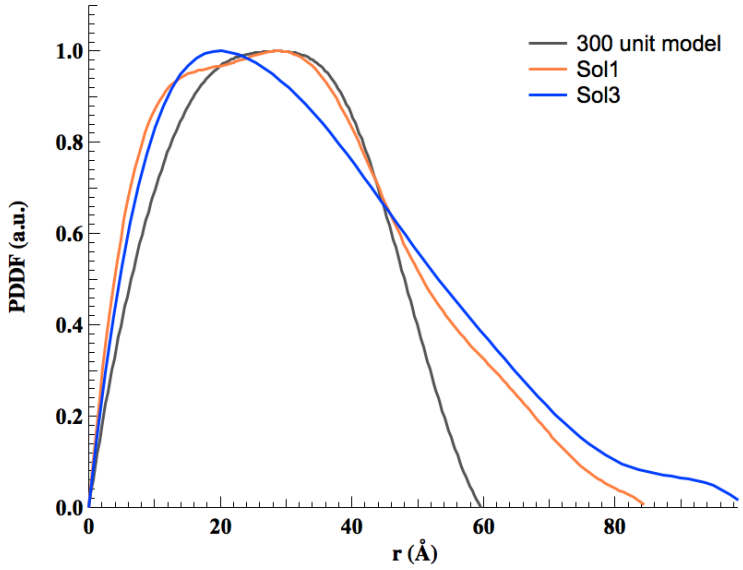


Figure 5.19: Cross-sectional Pair Distance Distribution Functions (PDDF) traces for the 300 subunit computational model (blue) in (A) compared to the experimental PDDF traces in red for So1 (A) and Sol3 (B).

a significant shoulder at higher values of *r*, which indicates that the cross section of the sample likely demonstrates a higher aspect ratio than expected for the circular cross-section of a hollow cylinder. Further, the cross-sectional PDDF for **Sol3** demonstrates a uni-modal character. A bi-modal form might intimate a hollow cross-section. The lack of bi-modal form observed may indicate the solenoidal form-factor is collapsing upon itself, creating a 'non-hollow,' ellipsoidal cross-section, as supported by the data. Notably, in contrast to **Sol3**, **Sol1** does demonstrate a bi-

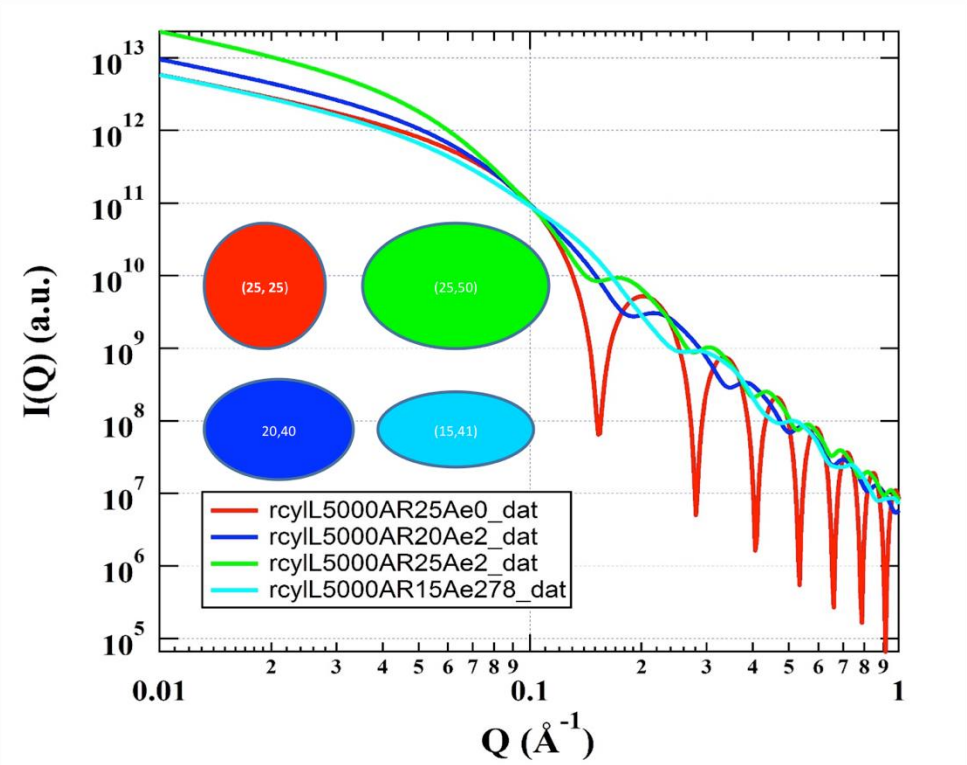


Figure 5.20: The scattering profile for a generic nanotubes with dimensions indicated as x,y wherein a uniform circular cross-section (red) with clear oscillations observed throughout the high *q* region. Modeled data indicating predicted SAXS scattering profiles for filaments with various cross-sectional dimensions, wherein more ellipsoidal cross-sections produce scattering profiles with increasingly dampened diffraction peaks (green, blue and teal respectively.

modal PDDF trace, which may be indicative of a hollow cross-section, or perhaps may reflect the results seen via TEM and STEM, wherein careful observation detects a minority of unwound fibrils. The PDDF for **Sol1** may therefore be representing the unwound minority and fully closed majority fibrillar species in solution, with each peak assigned to either of the aforementioned morphological states with differing radii. Ellipsoidal cross-sectional models created from a 300-subunit PDB model of **Sol1** are shown (Figure 5.18) in comparison to a circular fibrillar cross-section (in red). Modeled scattering curves for cross-sections that deviate from the circular demonstrate a dampening of signal in the wider-angle region. The absence of fine structure in the high q regions of the experimental scattering curves for **Sol1** and **Sol3** (Figure 5.13A and 5.13B) may indicate either that the tubes are collapsed on themselves or dynamically sampling different cross-sectional distributions on the time-scale of the measurements. This situation may arise due to the fact that the axial interfaces are likely significantly weaker than the lateral interfaces between protomers.

5.3 Ongoing Work

This research is still ongoing, with new variants being engineered in the pursuit of creating a greater library of solenoidal nanotubes which demonstrate both high designability and robust supramolecular structures. As noted, one of the primary hurdles in the self-assembly of these peptides was the somewhat common occurrence of the filaments not fully packing into fibers, but rather existing in semi-wound states. Therefore, designing changes to the sequence which decrease the likelihood of unwinding is of importance. A possible explanation for this phenomenon is an observed difference in stability between axial subunit interactions and lateral subunit interactions, with one domain overpowering the other. We aim to test the effects of increasing the strength of

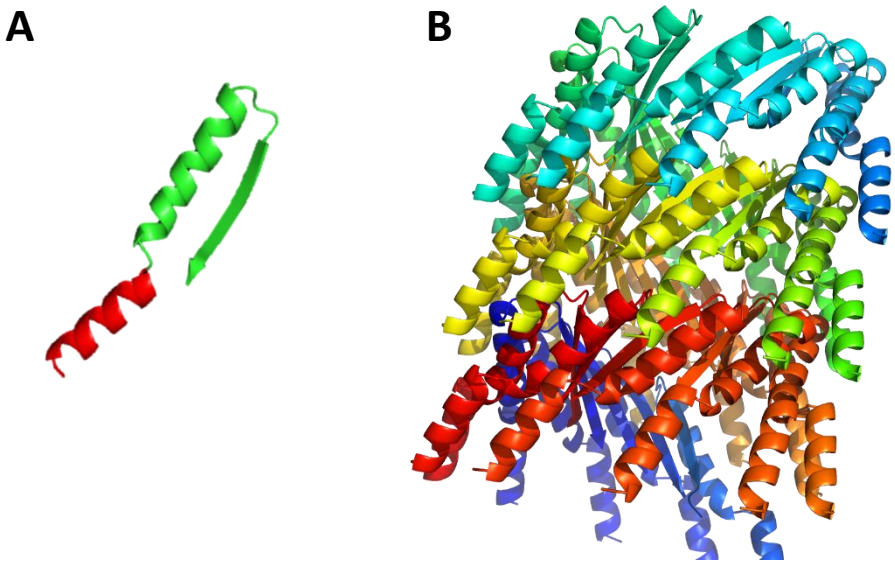


Figure 5.21: graphical representations of the computationally produced (A) newly designed Sol variant subunit and (B) the hypothesized packing morphology of this subunit.

axial interactions between subunits in comparison to lateral interactions. In this pursuit, one current design incorporates an additional alpha-helix at the solvent-facing surface of the filament. This is achieved by adding residues which demonstrate high values for helix propensity (as established via Pace et al., 1998) at the N-terminus via a short, helix-breaking loop region. By designing the helical addition in such a way, this extends the solvent-facing portion of one subunit into and between the subunits of the two most axially adjacent subunits. The loop is included to allow for the correct positioning of the added helix. It is hypothesized that this extension of the subunit into and between neighboring axial subunits will effectively increase the stability of axial interactions, leading to a more robust supramolecular structure that does not exhibit unwinding of subunits. As this work is ongoing, the precise sequence has not been shown, but the overall structural changes can be observed in Figure 5.21. As noted in 5.21A, the alpha-helix addition is shown in red, and the original sequence plus the new loop region are shown in green. The newly added helix is

smaller in length than the original. 5.19B indicates the hypothesized packing of these new alpha-loop-alpha-loop-beta subunits. The added small helix extends partially from the n solenoidal turn to the n+1 neighboring turn of the solenoidal assembly. In doing so, it rests between two of the original larger helices, potentially affording a greater hydrophobic driving force for the association of axial subunits.

5.4 Conclusion

The *de novo* creation of one-dimensional assemblies with discrete, highly predictable protomer geometries and supramolecular dimensions presents a substantial challenge to current synthetic peptide-based approaches. The plasticity of quaternary structure has emerged as a major challenge to the predictable design of extended protein assemblies [55]. Structural polymorphism is often observed for both native and designed helical protein assemblies [58-59], especially when assembled *in vitro*. The computational design of solenoid fibrils from the packing of an alpha-loop-beta structural motif represents an opportunity to test an approach based on protein designability. The **Sol** peptide series demonstrates again that relatively minute mutations in sequence-space can have measurable impacts on resultant supramolecular assemblies, whether it be fluctuations in solubility in aqueous buffer (i.e., the I24T mutation distinguishing **Sol1** from **Sol3**), or variation in packing geometry of protomers (i.e., the A20G mutation distinguishing **Sol1** from **Sol2**). Despite this, the computational approach described above still allows for a robust and reliable prediction of supramolecular morphology, with two of the three Sol variants forming filamentous structures with experimentally derived structural parameters in good agreement with those predicted by the computational modeling. STEM mass per length measurements and numerous other analytical techniques offer quantitative insight into the oligomerization of protomers within the **Sol** filaments, indicating that the protomers are self-assembling in the manner hypothesized. In the future, this approach could be applied to numerous other protomer types derived from solenoid-type tandem repeat protein motifs [60-62].

Moreover, the computational approach described above has proven to yield a greater rate of success in predicting structural outcomes for self-assembling peptide-based structures. Two of the three described computationally derived Sol variants formed filamentous assemblies with

structural parameters in agreement to what was hypothesized. Compared to peptides engineered via rational design, i.e. the Form variants, the rate of success is higher. Not described in the Form series were numerous peptides which did not self-assemble with any discernible order. Further, it is well established that minute mutations in sequence-space can produce drastic changes in quaternary structure, making the prediction of resultant supramolecular structure from primary sequence quite difficult in the absence of computational assistance. Computationally designing peptides for self-assembly aims to reduce the plasticity in the relationship between sequence and supramolecular structure. The Sol variants are good examples of the computational approach achieving a greater rate of success in designability and predictability of structure form sequence.

5.5 Methods

5.5.1 Materials. Unless otherwise stated, all chemical reagents were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO) or Anaspec, Inc. (Fremont, CA). All peptides were synthesized by and purchased from GenScript USA, Inc. (Piscataway, NJ).

5.5.2 Sequence Design. To build the assembly geometry, transformations between repeat units were identified that generated highly designable motif geometries between neighboring units and resulted in compact solenoids. The initial repeat unit was taken from PDB entry 1FSU, covering residue ranges 46-52 and 273-286, as representative of a common alpha-beta interaction geometry [DOI: 10.1016/B978-0-12-394292-0.00002-3]. Sampling of inter-unit placements was driven by a library of two-segment TERMs that was a precursor to the more complete library described in [63].The former was built in a similar fashion to the latter, but with a few key differences. As the "structural universe" a small non-redundant single-chain PISCES database was used, generated with a 1.6 Å resolution cutoff, a 0.25 R-factor cutoff, and a 20% sequence redundancy cutoff. As in MacKenzie *et al.* [63], motifs were identified by solving the set cover problem, but here only inter-residue contacts were covered (and not residues), the cutoff for defining a contact based on our contact degree metric was 0.1 [64], and all contacts were considered unique. A different RMSD cutoff function was also employed to define motif matches, one based on the Hill equation. The maximum RMSD (in Å) for a query with L residues was $\frac{A}{L^{-n}+B}$, where n was 2.5 while A and B were set so that the cutoff was 0.25 Å for a fragment of 4 residues and approached 1.1Å as the number of residues reached infinity.

For a given two-segment TERM emerging from the structural decomposition above, MASTER [65] was used to search for structural matches of each segment in the repeat unit of the

assembly, using the same RMSD cutoff function as above. Supposing that the first segment produced n matches and the second one m matches, there were a total of $n \times m$ combinations. Each of these implied a possible placement of two repeat units relative to each other. That is, aligning one copy of the repeat onto the first segment of the TERM and another copy onto the second segment (using one specific matching region in each case) placed the two copies relative to each other in a way that formed a motif structurally close to the TERM. By virtue of this, it was known that the two regions forming this interaction were highly designable. However, the remainder of the two units could still form poor interactions. Subsequently, checks for backbone clashes between the two unit copies were carried out and additional MASTER searches were performed to assure that there existed close structural matches around all other interfacial contacts. For each inter-unit geometry satisfying these filters, the transformation matrix relating the two units was deduced and applied iteratively to generate a helix. Assemblies that produced clashes between units on adjacent "turns" of the assembly solenoid were also filtered out.

Only the top 500 two-segment TERMs (by coverage) were used in the TERM library described above. The RMSD and clash filters removed most TERMs, leaving 36 two-segment motifs capable of producing non-clashing assemblies and a total of 88 assemblies (some of the TERMs produced multiple placements). These assemblies were inspected manually, discarding those with adjacent assembly turns being far apart to allow for simple engineering of axial interactions. From the remaining handful of assemblies, one was selected in which units in adjacent turns were close enough to interact and formed motifs represented in the PDB (checked using MASTER) as shown in Figure 5.22.

The next goal was to join the two-segment repeat unit into a single-chain protomer. To this end, the motif from the layer-to-layer interface was isolated and MASTER was used to identify

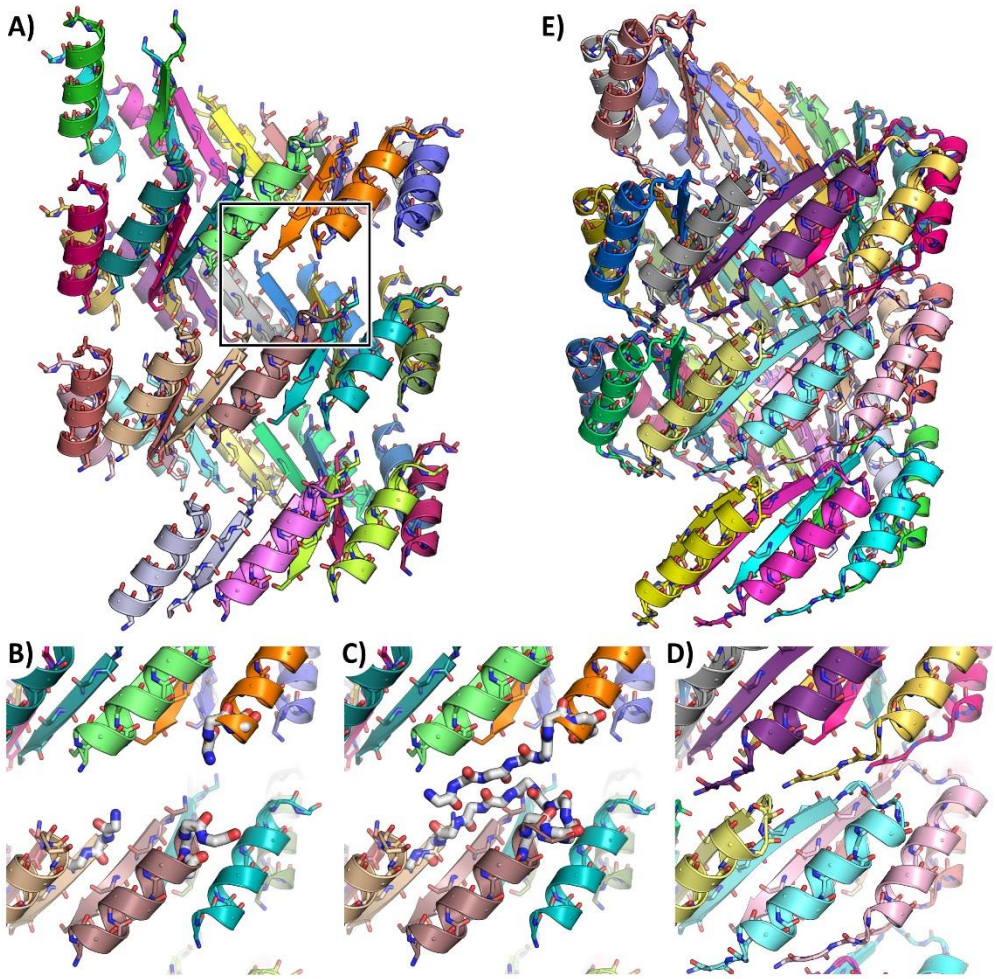


Figure 5.22: Assembly template generation procedure. Shown in A) is the designable solenoid geometry, built from our discontinuous α/β anchor motif, that was chosen for subsequent completion and design. B) To join the protomer into a single chain, relevant regions around the axial interface were selected into a single motif (shown in thick gray sticks), which was searched against the PDB. C) A match originating from entry 2AGD had both a low RMSD to the selected motif and also was remarkably appropriate for completing the unfinished interface—it simultaneously joined the two discontinuous segments of our motif into a single chain and it also showed how the N-terminus of the motif can be extended to form additional hydrogen-bond interactions across the axial interface. Using this match to fuse the assembly resulted in the geometry shown in D) and the full assembly template shown in E).

closely-matching backbone geometries in the PDB, retrieving the full structure from which each match originated. Inspecting the top 50 matches in PyMOL, a match was identified from PDB code 2AGD as illustrating both how to link the helix and the strand into a single protomer and, at the same time, extending the helix from the adjacent turn N-terminally to generate favorable beta-sheet interactions across the layers. Next, the assembly unit was "stitched" into a single protomer using the junction motif from this match to create the final assembly template. This procedure entailed finding short backbone fragments in the PDB that could seamlessly transition from the existing unit to the structure it is being joined to (e.g., from the helix to the junction motif and then to the strand). If a fragment was found that superimposed onto the two regions being stitched with very low RMSD (below 0.2 Å), the stitched structure was produced by simple coordinate averaging. If no such fragment was found, the lowest-RMSD fragment was used to optimally re-orient the two motifs being fused with respect to each other and the procedure was repeated.

Sequence design was based on an assembly fragment comprising a single central protomer surrounded by eight of its closest neighbors (the template). The design procedure was implemented in a python script that relied on PyRosetta [66] for sequence evaluation. Starting from a random sequence, Monte Carlo simulations were performed in sequence space, making single random mutations in each iteration. Sequences were evaluated by first making the mutation in each chain of the template and side-chain repacking all chains simultaneously using the Rosetta *talaris2013* energy function. After an initial ten rounds of MC with 10,000 iterations in each (with *kT* parameter set to 3 energy units), where all 20 standard amino acids were allowed at each position, the best-scoring solution was manually inspected. As a result, of the 29 positions in the protomer, 15 were marked as needing additional optimization (positions 6, 7, 11, 14, 15, 16, 17, 19, 20, 22, 24, 26, 27, 28, 29) and a more focused library for these positions was generated using MASTER to identify

the most common amino acids in fragments matching local motifs around these positions (an average of ~7.5 amino acids per position were chosen). Using this focused library, an additional ten rounds with 1,000 iterations in each was run, this time allowing for continuous side-chain and backbone minimization upon repacking before evaluating solution energies. The best-scoring solution from this round was again manually inspected and four positions (6, 16, 17, and 29) marked for additional optimization. A final set of six MC rounds with 1,000 iterations in each were then run, varying only the above positions, and this time allowing for only side-chain minimization upon repacking. Many of the resulting solutions scored approximately equally well, so the four sequences for experimental characterization were chosen by manual inspection of top-scoring solutions, aiming to address two specific hypotheses. Position 17, which packs in the helix-helix interface, contains ALA in the best-scoring sequence, but can also accommodate a larger amino-acid (e.g., ILE) if backbone minimization is allowed. As it was difficult to determine with certainty what amount of backbone relaxation would be reasonable to expect, especially given the constraint of forming the desired solenoid, both options were tried. Position 20, which maps into the loop between the helix and the strand and packs into the axial interface between adjacent assembly turns, can accommodate either ALA or GLY, depending on the type of relaxation allowed and the specific N-terminal modification of the peptide (as the N-terminus of the unit in the adjacent turn maps closely to this residue). Both ALA and GLY were sampled at this position experimentally due the possibility of backbone adjustment, the potential importance of additional methyl-group burial at the axil interface to the stability of the assembly, and in aiming to support the possibility of different synthetic terminal modifications (all of the above optimizations were performed in the context of the acetyl N-terminal modification). The four final solutions, sol1-sol4, sampled the four possible combinations at these two sites discussed above.

5.5.3 Peptide Self-Assembly. As previously stated, peptides were purchased from GenScript, USA (Fremont, CA). Peptides were purified via reverse phase high performance liquid chromatography (RP-HPLC) using a C-18 column and a gradient of water (0.065% trifluoroacetic acid)-acetonitrile (0.05% trifluoroacetic acid). Peptides were modified at the N- and C-terminus with acetyl (CH_3CO) and methyl-amide (NHMe) capping groups respectively. 10 mM acetate buffer pH 4.0 was added to the lyophilized GenScript peptides at approximately 0.3 mg mL^{-1} peptide concentrations for all experimentation except for small angle x-ray scattering (SAXS), for which peptides were assembled at ≥3 mg mL^{-1}. Peptide concentrations were determined spectrophotometrically from measurements of the absorbance at 280 nm (A_{280}). For peptides containing Tyr, Trp, or Cys residues, the peptide concentration can be calculated from Equation 1:

$$MW \times (A_{280}/c) = 1280n_Y + 5690n_W + 120n_C \qquad \text{(Equation 1)}$$

in which c is the concentration of peptide in mg mL^{-1}, and n_Y, n_W, and n_C are the numbers of tyrosine, tryptophan and cysteine residues, respectively, in the peptide sequence (Gill and von Hippel, 1989). As each variant contains only a single tyrosine residue per sequence, c = MW x A_{280}/1280. 6M guanidinium chloride is added in a 1:9 v/v ratio to each aqueous specimen solution before CD analysis to eliminate error in determination of absorbance due to the scattering of UV light due to peptide self-assembly. All three peptides exhibited varying but limited solubility in peptide assembly buffer (as well as a library of other tested buffers at varying concentrations and pH). Subsequent 1:1 volumetric addition of 2,2,2-trifluorethanol (TFE) to the 0.3 mg mL^{-1} peptide assembly solution enabled efficient solubilization of peptide within one minute. TFE was subsequently allowed to evaporate sufficiently such that the volume of the combined peptide

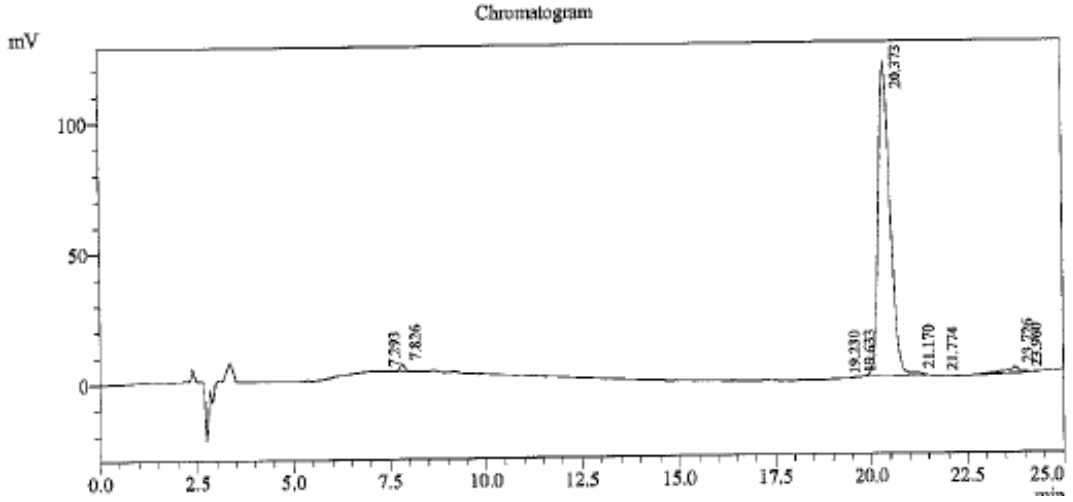
assembly buffer and TFE was reduced by half, typically taking between 1 and 5 days depending on total volume. It is assumed that the aforementioned halving of the original assembly solution volume indicates near full evaporation of TFE, and that the TFE-acetate buffer mixture did not evaporate azeotropically to a significant degree. Eppendorf microcentrifuge tubes containing the assembly solutions + TFE were covered with perforated Parafilm ® to decrease chances of environmental contamination. The Parafilm ® was pierced once using a sterile needle, allowing for controlled evaporation of TFE. For the Sol2 variant, a pre-solubilization in HPLC mimetic conditions before assembly demonstrated improved fiber formation. To achieve this, approximately one mg of Sol2 was dissolved in 0.5 mL water and 0.5 mL acetonitrile (0.1% TFA concentration), incubated at room temperature for 24 hours, frozen and then lyophilized. This lyophilized peptide powder was then assembled as described above. Following TFE addition, samples incubated at room temperature for several weeks, with fibers visible via TEM as early as five days. No evidence of sample degradation was witnessed via TEM over a time scale of several months.

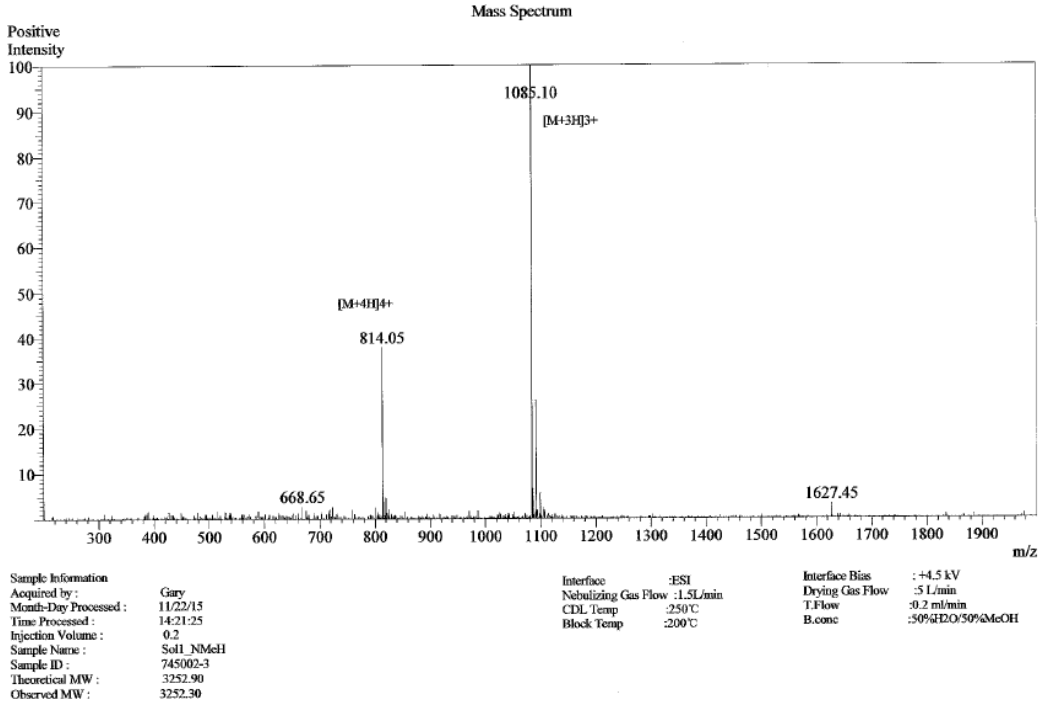
Sample Name : Sol1_NMeH
Sample ID : 745002-3

Pump A : 0.065% trifluoroacetic in 100% water (v/v)
Pump B : 0.05% trifluoroacetic in 100% acetonitrile (v/v)
Total Flow:1 ml/min
Wavelength:220 nm

Time	Unit	Command	Value	Comment
0.01	Pumps	Pump A B.Conc	35	
25.00	Pumps	Pump A B.Conc	95	
31.00	Pumps	Pump A B.Conc	95	
31.01	Pumps	Pump A B.Conc	35	
40.00	Pumps	Pump A B.Conc	35	
40.01	Controller	Stop		

<<Column Performance>>
<Detector A>
Column : AlltimaTM C18 4.6 x 250 mm





Sample Information	
Acquired by :	Gary
Month-Day Processed :	11/22/15
Time Processed :	14:21:25
Injection Volume :	0.2
Sample Name :	Sol1_NMeH
Sample ID :	745002-3
Theoretical MW :	3252.90
Observed MW :	3252.30

Interface	:ESI	Interface Bias	: +4.5 kV
Nebulizing Gas Flow	:1.5L/min	Drying Gas Flow	:5 L/min
CDL Temp	:250℃	T.Flow	:0.2 ml/min
Block Temp	:200℃	B.conc	:50%H2O/50%MeOH

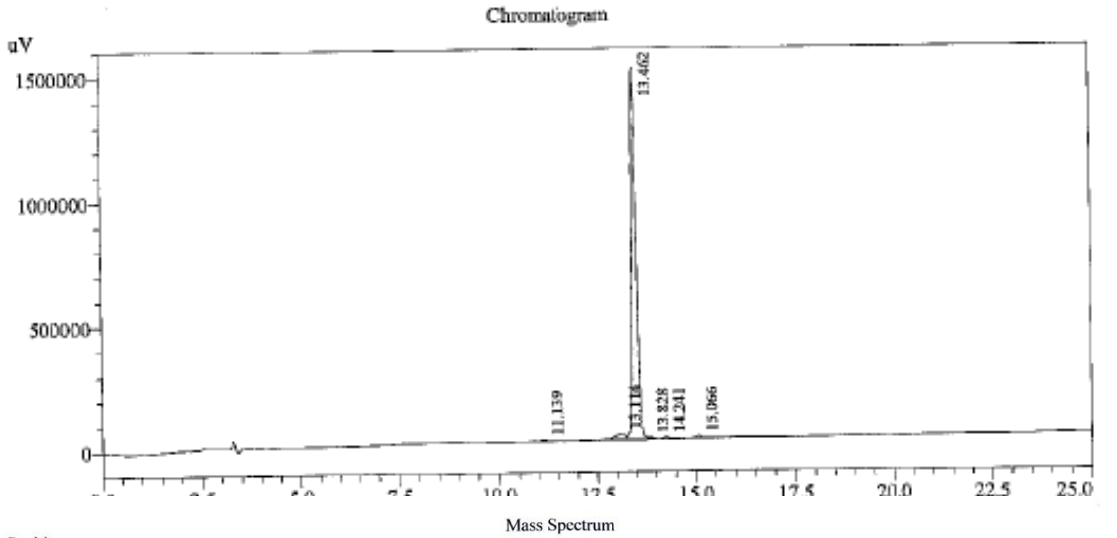
Figure 5.23: Analytical HPLC trace (top) and ESI-MS mass confirmation (bottom) of Sol1.

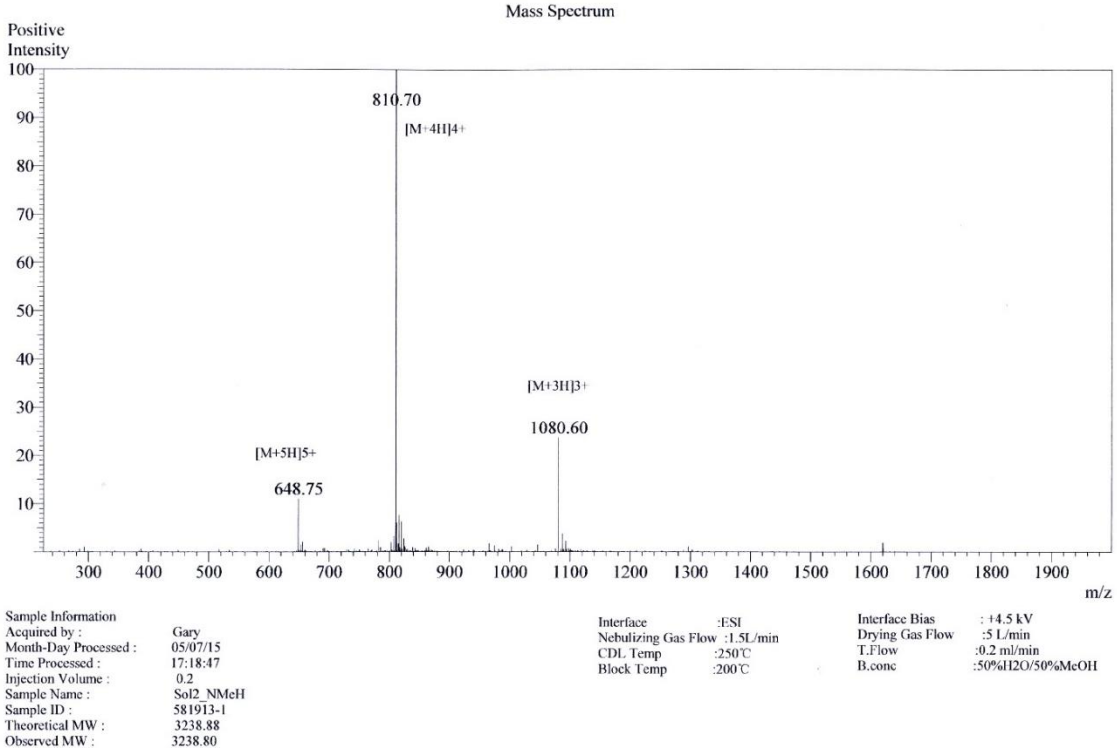
Sample Name : Sol2_NMeH
Sample ID : 581913-1

Pump A : 0.065% trifluoroacetic in 100% water (v/v)
Pump B : 0.05% trifluoroacetic in 100% acetonitrile (v/v)
Total Flow:1 ml/min
Wavelength:220 nm

Time	Unit	Command	Value	Comment
0.01	Pumps	Pump A B.Conc	5	
25.00	Pumps	Pump A B.Conc	65	
25.01	Pumps	Pump A B.Conc	95	
31.00	Pumps	Pump A B.Conc	95	
31.01	Pumps	Pump A B.Conc	5	
40.00	Pumps	Pump A B.Conc	5	
40.01	Controller	Stop		

<<Column Performance>>
<Detector A>
Column : AlltimaTM C18 4.6 x 250 mm





Sample Information

Acquired by :	Gary
Month-Day Processed :	05/07/15
Time Processed :	17:18:47
Injection Volume :	0.2
Sample Name :	Sol2_NMeH
Sample ID :	581913-1
Theoretical MW :	3238.88
Observed MW :	3238.80

Interface	:ESI
Nebulizing Gas Flow	:1.5L/min
CDL Temp	:250℃
Block Temp	:200℃

Interface Bias	: +4.5 kV
Drying Gas Flow	:5 L/min
T.Flow	:0.2 ml/min
B.conc	:50%H2O/50%MeOH

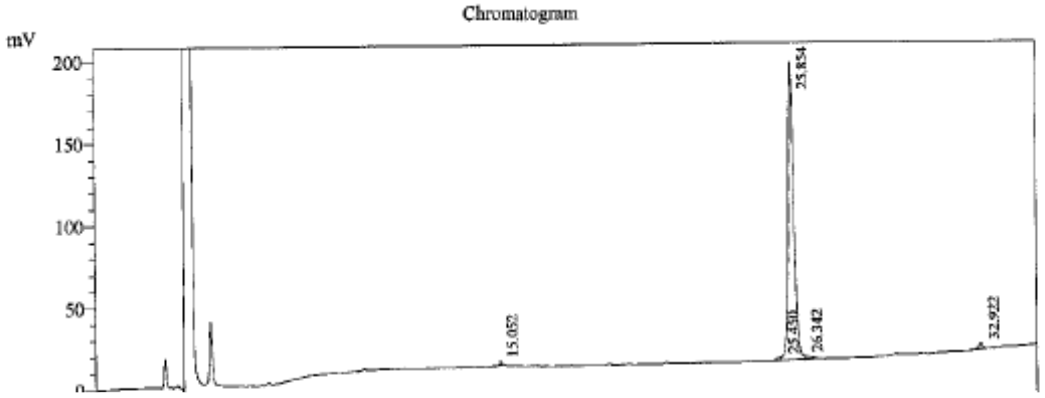
Figure 5.24: Analytical HPLC (top) and ESI-MS mass confirmation (bottom) of Sol2.

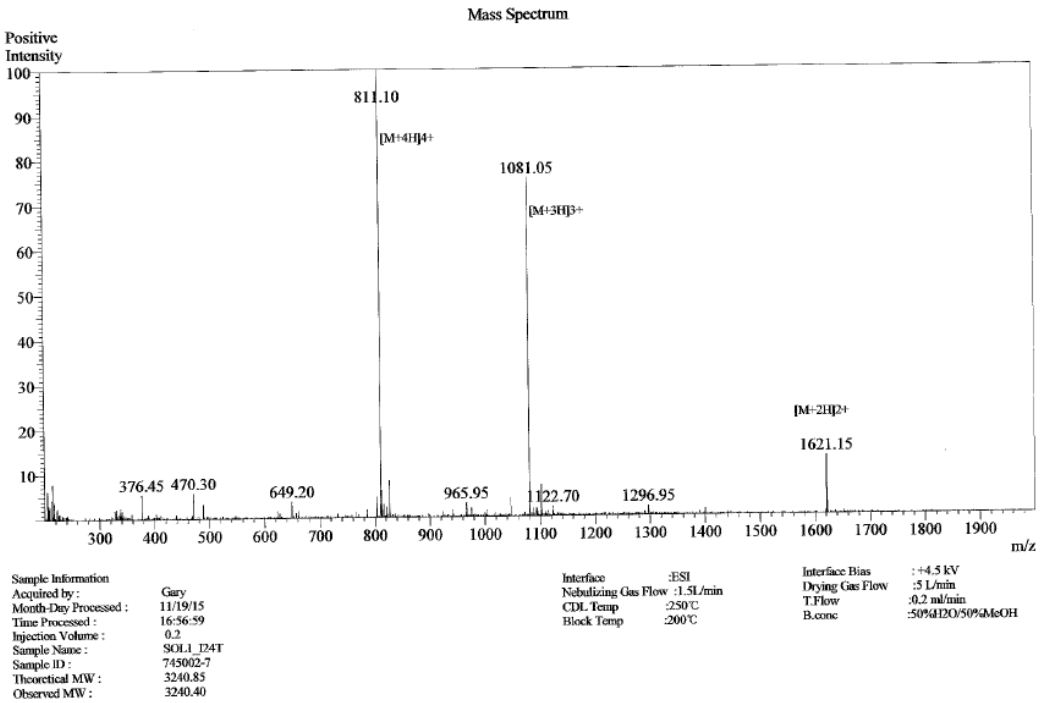
Sample Name : SOL1_I24T
Sample ID : 745002-7

Pump A : 0.065% trifluoroacetic in 100% water (v/v)
Pump B : 0.05% trifluoroacetic in 100% acetonitrile (v/v)
Total Flow:1 ml/min
Wavelength:220 nm

Time	Unit	Command	Value	Comment
0.01	Pumps	Pump A B.Conc	5	
10.00	Pumps	Pump A B.Conc	35	
35.00	Pumps	Pump A B.Conc	95	
41.00	Pumps	Pump A B.Conc	95	
41.01	Pumps	Pump A B.Conc	5	
50.00	Pumps	Pump A B.Conc	5	
50.01	Controller	Stop		

<<Column Performance>>
<Detector A>
Column : AltimaTM C18 4.6 x 250 mm





Sample Information
Acquired by : Gary
Month-Day Processed : 11/19/15
Time Processed : 16:56:59
Injection Volume : 0.2
Sample Name : SOL1_I24T
Sample ID : 745002-7
Theoretical MW : 3240.85
Observed MW : 3240.40

Interface :ESI
Nebulizing Gas Flow :1.5L/min
CDL Temp :250℃
Block Temp :200℃

Interface Bias : +4.5 kV
Drying Gas Flow :5 L/min
T.Flow :0.2 ml/min
B.conc :50%H2O/50%MeOH

Figure 5.25: Analytical HPLC (top) and ESI-MS mass confirmation (bottom) for Sol3.

5.5.4 Circular Dichroism. Circular dichroism analyses of Sol assemblies were performed using a JASCO J-810 circular dichroism spectropolarimeter. Peptide solutions were analyzed at a peptide concentration of ~100uM for both variable-wavelength scans and variable-temperature scans. For variable wavelength scans, Spectra were collected at 50nm/min from 190-260 nm with a data pitch of 0.2 nm. For variable temperature scans, spectra were collected at 222nm, with a 60°C/hour temperature progression, and a data pitch of 0.5°C. Variable temperature spectra were collected from 25-85°C. Quartz plates (0.10 mm thickness) from Hellma Analytics (Müllheim, Germany) were employed for all CD experiments.

5.5.5 Transmission Electron Microscopy. Prior to grid preparation, self-assembled peptide samples were ultra-centrifuged on a Beckman-Coulter Optima XE-90 using a Ti-45 rotor operating at 45,000G for one hour to concentrate fully formed fibers from non-fibrillar peptide aggregate and immature fibers. The supernatant was removed and discarded, and the pellet was subsequently re-suspended in peptide assembly buffer (no TFE). Grids were prepared by depositing 4 μL of peptide solution (0.3 mg mL^{-1} peptide concentration in 10 mM acetate buffer pH 4.0) to 200 mesh carbon-coated copper grids purchased from Electron Microscopy Services (Hatfield, PA). The solution was allowed to incubate on the grid for 45 seconds before being wicked away using 55mm qualitative filter paper. The peptide solution was wicked from the grid surface in such a way that a thin film of liquid remained, then 4 μL of freshly prepared, filtered (0.2 micron Whatman filter from G.E. Healthcare Services) 1% aqueous uranyl acetate solution was applied to the grid to mix with and negatively stain the sample. After 45 seconds of incubation, the stain was fully wicked from the grid using the same methodology (*vide supra*). Grids were stored in a tabletop desiccator and sufficiently dried. All transmission electron microscopy was carried out on a Hitachi HT-7700

using a tungsten filament at an accelerating voltage of 80kV. Images were captured via AMT CCD camera.

5.5.6 Scanning Transmission Electron Microscopy. STEM data were acquired at Brookhaven National Laboratory (BNL). Assembly solutions for each variant (0.3 mg mL^{-1} in 10 mM acetate buffer, pH 4.0) were deposited on thin carbon (ca. 2 nm thick) supported on a thicker holey carbon film mounted on a titanium grid using the wet-film, hanging drop method. Tobacco Mosaic Virus (TMV) is added to the grid as an internal control, then injection buffer is added, followed by specimen solution. These solutions are allowed to sit for one minute, and excess solution is wicked from the edge with filter paper. Following The grid is fast frozen by plunging into liquid nitrogen slush and stored under liquid nitrogen. All grids are freeze-dried overnight in an ion-pumped chamber with an efficient cold trap. Grids are then transferred under vacuum to the STEM cold stage, which is operated at -160 °C. Measurements were taken at 40keV with a scanning probe of <0.3 nm directed from a cold field emission source. Data was collected using bright field, small-angle dark field, and large-angle dark field scintillator photomultiplier detectors which operate at 0-15, 15-40, and 40-200 mRadian respectively. The large-angle signal is proportional to the mass of the atoms in the path of the beam. Specimen quality was tested and mass calibration was standardized using TMV. Mass per length (M/L) values were calculated for each Sol variant (380 data points for **Sol1**, 1,080 for **Sol2**, and a,093 for **Sol3**) using aforementioned TMV raft standards set at a theoretical m/L of 13.1 kDa/Å by using the software 'PCMass32' [67-68]. **Sol1** STEM analysis produced fewer viable micrographs for MPL analysis, hence the smaller data set for this variant. Data points occurring beyond three standard deviations were identified as outliers and excluded

from calculations, as these can be attributed to user error in selection of areas for PCMass32 analysis within STEM data.

5.5.7 Small- and Wide-Angle X-ray Scattering Measurements. Synchrotron SAXS/WAXS analyses were performed at the Advanced Photon Source at Argonne National Laboratory (Beamline 12-ID-B). Assembly solutions of ≥ 3 mg mL^{-1} peptide concentrations (3.6 mg mL^{-1} for Sol1, 3.0 mg mL^{-1} for Sol2, and 4.0 mg mL^{-1} for Sol3) in 10mM acetate pH 4.0 were dialyzed against pure 10mM acetate pH 4.0 such that the dialyzed sample is identical to the buffer standard in all manners except for peptide concentration. To achieve this, Slide-A-Lyzer dialysis cassettes (2,000 MWCO) from Thermo-Fisher were used to dialyze 0.5 mL of peptide assembly solution against buffer containing no peptide. SAXS/WAXS analysis was performed at 25 °C in a 1.5 mm quartz capillary flow cell (applying constant vertical agitation of the peptide assembly solution) so as to prevent radiation damage. For each Sol variant, a minimum of thirty 2-dimensional shots were gathered and azimuthally averaged into 1-dimensional scattering spectra. Identically gathered data was obtained for the dialysis buffer (10mM acetate pH 4.0, no peptide present) and data were background subtracted to produce signal correlating exclusively to peptide-based assemblies in solution. Normalization against beam intensity and solid angle correction were performed by the in-house software suite (Igor Pro) present at Beamline 12-ID-B. Radii were calculated using Equation 2:

$$R_c = R_{exp}/\sqrt{2} \quad \text{(or)} \quad R_{exp} = R_c * \sqrt{2}. \qquad \text{(Equation 2)}$$

5.6 References

1. Rothemund, P. W., Folding DNA to create nanoscale shapes and patterns. *Nature* **2006,** *440* (7082), 297-302.

2. Lin, C.; Liu, Y.; Rinker, S.; Yan, H., DNA tile based self-assembly: building complex nanoarchitectures. *Chemphyschem : a European journal of chemical physics and physical chemistry* **2006,** *7* (8), 1641-7.

3. Ke, Y.; Ong, L. L.; Shih, W. M.; Yin, P., Three-dimensional structures self-assembled from DNA bricks. *Science* **2012,** *338* (6111), 1177-83.

4. Garcia-Seisdedos, H.; Empereur-Mot, C.; Elad, N.; Levy, E. D., Proteins evolve on the edge of supramolecular self-assembly. *Nature* **2017,** *548* (7666), 244-247.

5. Egelman, E. H., Three-dimensional reconstruction of helical polymers. *Archives of biochemistry and biophysics* **2015,** *581*, 54-8.

6. Egelman, E. H., Reconstruction of helical filaments and tubes. *Methods Enzymol* **2010,** *482*, 167-83.

7. Wang, Y. A.; Yu, X.; Overman, S.; Tsuboi, M.; Thomas, G. J., Jr.; Egelman, E. H., The structure of a filamentous bacteriophage. *Journal of molecular biology* **2006,** *361* (2), 209-215.

8. Stubbs, G.; Kendall, A., Helical viruses. *Advances in experimental medicine and biology* **2012,** *726*, 631-58.

9. Straus, S. K.; Scott, W. R.; Schwieters, C. D.; Marvin, D. A., Consensus structure of Pf1 filamentous bacteriophage from X-ray fibre diffraction and solid-state NMR. *European biophysics journal : EBJ* **2011,** *40* (3), 221-34.

10. Egelman, E. H.; Francis, N.; DeRosier, D. J., F-actin is a helix with a random variable twist. *Nature* **1982,** *298*, 131-135.

11. Galkin, V. E.; Orlova, A.; Vos, M. R.; Schroder, G. F.; Egelman, E. H., Near-atomic resolution for one state of f-actin. *Structure* **2015,** *23* (1), 173-82.

12. Downing, K. H.; Nogales, E., Cryoelectron microscopy applications in the study of tubulin structure, microtubule architecture, dynamics and assemblies, and interaction of microtubules with motors. *Methods in enzymology* **2010,** *483*, 121-42.

13. Yonekura, K.; Maki-Yonekura, S.; Namba, K., Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* **2003,** *424* (6949), 643-650.

14. Galkin, V. E.; Yu, X.; Bielnicki, J.; Heuser, J.; Ewing, C. P.; Guerry, P.; Egelman, E. H., Divergence of quaternary structures among bacterial flagellar filaments. *Science* **2008,** *320* (5874), 382-385.

15. Li, J.; Egelman, E. H.; Craig, L., Structure of the Vibrio cholerae Type IVb Pilus and stability comparison with the Neisseria gonorrhoeae type IVa pilus. *Journal of molecular biology* **2012,** *418* (1-2), 47-64.

16. Craig, L.; Volkmann, N.; Arvai, A. S.; Pique, M. E.; Yeager, M.; Egelman, E. H.; Tainer, J. A., Type IV pilus structure by cryo-electron microscopy and crystallography: Implications for pilus assembly and functions. *Mol Cell* **2006,** *23* (5), 651-662.

17. Yu, X.; Goforth, C.; Meyer, C.; Rachel, R.; Wirth, R.; Schroder, G. F.; Egelman, E. H., Filaments from Ignicoccus hospitalis show diversity of packing in proteins containing N-terminal type IV pilin helices. *Journal of molecular biology* **2012,** *422* (2), 274-81.

18. Loquet, A.; Sgourakis, N. G.; Gupta, R.; Giller, K.; Riedel, D.; Goosmann, C.; Griesinger, C.; Kolbe, M.; Baker, D.; Becker, S.; Lange, A., Atomic model of the type III secretion system needle. *Nature* **2012,** *486* (7402), 276-+.

19. Loquet, A.; Habenstein, B.; Lange, A., Structural investigations of molecular machines by solid-state NMR. *Accounts of chemical research* **2013,** *46* (9), 2070-9.

20. Demers, J. P.; Habenstein, B.; Loquet, A.; Kumar Vasa, S.; Giller, K.; Becker, S.; Baker, D.; Lange, A.; Sgourakis, N. G., High-resolution structure of the Shigella type-III secretion needle by solid-state NMR and cryo-electron microscopy. *Nature communications* **2014,** *5*, 4976.

21. Wang, X.; Zhou, Y.; Ren, J. J.; Hammer, N. D.; Chapman, M. R., Gatekeeper residues in the major curlin subunit modulate bacterial amyloid fiber biogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **2010,** *107* (1), 163-8.

22. Taylor, J. D.; Zhou, Y.; Salgado, P. S.; Patwardhan, A.; McGuffie, M.; Pape, T.; Grabe, G.; Ashman, E.; Constable, S. C.; Simpson, P. J.; Lee, W. C.; Cota, E.; Chapman, M. R.; Matthews, S. J., Atomic resolution insights into curli fiber biogenesis. *Structure* **2011,** *19* (9), 1307-16.

23. Pham, C. L.; Kwan, A. H.; Sunde, M., Functional amyloid: widespread in Nature, diverse in purpose. *Essays in biochemistry* **2014,** *56*, 207-19.

24. Wasmer, C.; Lange, A.; Van Melckebeke, H.; Siemer, A. B.; Riek, R.; Meier, B. H., Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core. *Science* **2008,** *319* (5869), 1523-6.

25. Wasmer, C.; Benkemoun, L.; Sabate, R.; Steinmetz, M. O.; Coulary-Salin, B.; Wang, L.; Riek, R.; Saupe, S. J.; Meier, B. H., Solid-state NMR spectroscopy reveals that E. coli inclusion bodies of HET-s(218-289) are amyloids. *Angewandte Chemie* **2009,** *48* (26), 4858-60.

26. Wan, W.; Stubbs, G., Fungal prion HET-s as a model for structural complexity and self-propagation in prions. *Proceedings of the National Academy of Sciences of the United States of America* **2014,** *111* (14), 5201-6.

27. Lu, A.; Magupalli, V. G.; Ruan, J.; Yin, Q.; Atianand, M. K.; Vos, M. R.; Schroder, G. F.; Fitzgerald, K. A.; Wu, H.; Egelman, E. H., Unified polymerization mechanism for the assembly of ASC-dependent inflammasomes. *Cell* **2014,** *156* (6), 1193-1206.

28. Joosten, R. P.; te Beek, T. A.; Krieger, E.; Hekkelman, M. L.; Hooft, R. W.; Schneider, R.; Sander, C.; Vriend, G., A series of PDB related databases for everyday needs. *Nucleic acids research* **2011,** *39* (Database issue), D411-9.

29. Walters, R. F.; DeGrado, W. F., Helix-packing motifs in membrane proteins. *Proc Natl Acad Sci U S A* **2006,** *103* (37), 13658-63.

30. Grigoryan, G.; Degrado, W. F., Probing designability via a generalized model of helical bundle geometry. *Journal of molecular biology* **2011,** *405* (4), 1079-100.

31. Kallblad, P.; Dean, P. M., Backbone-backbone geometry of tertiary contacts between alpha-helices. *Proteins* **2004,** *56* (4), 693-703.

32. Hu, C.; Koehl, P., Helix-sheet packing in proteins. *Proteins* **2010,** *78* (7), 1736-47.

33. Ho, B. K.; Curmi, P. M., Twist and shear in beta-sheets and beta-ribbons. *Journal of molecular biology* **2002,** *317* (2), 291-308.

34. Fernandez-Fuentes, N.; Dybas, J. M.; Fiser, A., Structural characteristics of novel protein folds. *PLoS computational biology* **2010,** *6* (4), e1000750.

35. Marchler-Bauer, A.; Lu, S.; Anderson, J. B.; Chitsaz, F.; Derbyshire, M. K.; DeWeese-Scott, C.; Fong, J. H.; Geer, L. Y.; Geer, R. C.; Gonzales, N. R.; Gwadz, M.; Hurwitz, D. I.; Jackson, J. D.; Ke, Z.; Lanczycki, C. J.; Lu, F.; Marchler, G. H.; Mullokandov, M.; Omelchenko, M. V.; Robertson, C. L.; Song, J. S.; Thanki, N.; Yamashita, R. A.; Zhang, D.; Zhang, N.; Zheng, C.; Bryant, S. H., CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic acids research* **2011,** *39* (Database issue), D225-9.

36. London, N.; Movshovitz-Attias, D.; Schueler-Furman, O., The structural basis of peptide-protein binding strategies. *Structure* **2010,** *18* (2), 188-99.

37. Stein, A.; Ceol, A.; Aloy, P., 3did: identification and classification of domain-based interactions of known three-dimensional structure. *Nucleic acids research* **2011,** *39* (Database issue), D718-23.

38. England, J. L.; Shakhnovich, B. E.; Shakhnovich, E. I., Natural selection of more designable folds: a mechanism for thermophilic adaptation. *Proceedings of the National Academy of Sciences of the United States of America* **2003,** *100* (15), 8727-31.

39. Zhang, J.; Grigoryan, G., Mining tertiary structural motifs for assessment of designability. *Methods in enzymology* **2013,** *523*, 21-40.

40. Govindarajan, S.; Goldstein, R. A., Why are some proteins structures so common? *Proceedings of the National Academy of Sciences of the United States of America* **1996,** *93* (8), 3341-5.

41. Wingreen, N. S.; Li, H.; Tang, C., Designability and thermal stability of protein structures. *Polymer* **2004,** *45* (2), 699-705.

42. Li, H.; Helling, R.; Tang, C.; Wingreen, N., Emergence of preferred structures in a simple model of protein folding. *Science* **1996,** *273* (5275), 666-9.

43. Helling, R.; Li, H.; Melin, R.; Miller, J.; Wingreen, N.; Zeng, C.; Tang, C., The designability of protein structures. *Journal of molecular graphics & modelling* **2001,** *19* (1), 157-67.

44. Kajava, A. V., Review: proteins with repeated sequence--structural prediction and modeling. *Journal of structural biology* **2001,** *134* (2-3), 132-44.

45. Bella, J.; Hindle, K. L.; McEwan, P. A.; Lovell, S. C., The leucine-rich repeat structure. *Cellular and molecular life sciences : CMLS* **2008,** *65* (15), 2307-33.

46. Papageorgiou, A. C.; Shapiro, R.; Acharya, K. R., Molecular recognition of human angiogenin by placental ribonuclease inhibitor--an X-ray crystallographic study at 2.0 A resolution. *The EMBO journal* **1997,** *16* (17), 5162-77.

47. Hrabe, T.; Godzik, A., ConSole: using modularity of contact maps to locate solenoid domains in protein structures. *BMC bioinformatics* **2014,** *15*, 119.

48. Ramisch, S.; Weininger, U.; Martinsson, J.; Akke, M.; Andre, I., Computational design of a leucine-rich repeat protein with a predefined geometry. *Proc Natl Acad Sci U S A* **2014,** *111* (50), 17875-80.

49. Zhou, J.; Grigoryan, G., Rapid search for tertiary fragments reveals protein sequence-structure relationships. *Protein science : a publication of the Protein Society* **2014**.

50. Grigoryan, G.; Kim, Y. H.; Acharya, R.; Axelrod, K.; Jain, R. M.; Willis, L.; Drndic, M.; Kikkawa, J. M.; DeGrado, W. F., Computational design of virus-like protein assemblies on carbon nanotube surfaces. *Science* **2011,** *332* (6033), 1071-6.

51. Kaltofen, S.; Li, C.; Huang, P. S.; Serpell, L. C.; Barth, A.; Andre, I., Computational de novo design of a self-assembling peptide with predefined structure. *J Mol Biol* **2015,** *427* (2), 550-62.

52. Hamley, I. W., Peptide nanotubes. *Angew Chem Int Ed Engl* **2014,** *53* (27), 6866-81.

53. Lu, K.; Jacob, J.; Thiyagarajan, P.; Conticello, V. P.; Lynn, D. G., Exploiting amyloid fibril lamination for nanotube self-assembly. *J Am Chem Soc* **2003,** *125* (21), 6391-3.

54. Magnotti, E. L.; Hughes, S. A.; Dillard, R. S.; Wang, S.; Hough, L.; Karumbamkandathil, A.; Lian, T.; Wall, J. S.; Zuo, X.; Wright, E. R.; Conticello, V. P., Self-Assembly of an alpha-

Helical Peptide into a Crystalline Two-Dimensional Nanoporous Framework. *J Am Chem Soc* **2016,** *138* (50), 16274-16282.

55. 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.

56. Xu, C.; Liu, R.; Mehta, A. K.; Guerrero-Ferreira, R. C.; Wright, E. R.; Dunin-Horkawicz, S.; Morris, K.; Serpell, L. C.; Zuo, X.; Wall, J. S.; Conticello, V. P., Rational design of helical nanotubes from self-assembly of coiled-coil lock washers. *J Am Chem Soc* **2013,** *135* (41), 15565-78.

57. Svergun, D.; Barberato, C.; Koch, M. H. J., CRYSOL– a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J Appl Crystallogr* **1995,** *28* (6), 768-773.

58. Guenther, E. L.; Ge, P.; Trinh, H.; Sawaya, M. R.; Cascio, D.; Boyer, D. R.; Gonen, T.; Zhou, Z. H.; Eisenberg, D. S., Atomic-level evidence for packing and positional amyloid polymorphism by segment from TDP-43 RRM2. *Nature structural & molecular biology* **2018,** *25* (4), 311-319.

59. Adamcik, J.; Mezzenga, R., Amyloid Polymorphism in the Protein Folding and Aggregation Energy Landscape. *Angew Chem Int Ed Engl* **2018,** *57* (28), 8370-8382.

60. Kobe, B.; Kajava, A. V., When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends in biochemical sciences* **2000,** *25* (10), 509-15.

61. Parmeggiani, F.; Huang, P. S.; Vorobiev, S.; Xiao, R.; Park, K.; Caprari, S.; Su, M.; Seetharaman, J.; Mao, L.; Janjua, H.; Montelione, G. T.; Hunt, J.; Baker, D., A general computational approach for repeat protein design. *J Mol Biol* **2015,** *427* (2), 563-75.

62. Brunette, T. J.; Parmeggiani, F.; Huang, P. S.; Bhabha, G.; Ekiert, D. C.; Tsutakawa, S. E.; Hura, G. L.; Tainer, J. A.; Baker, D., Exploring the repeat protein universe through computational protein design. *Nature* **2015,** *528* (7583), 580-4.

63. Mackenzie, C. O.; Zhou, J.; Grigoryan, G., Tertiary alphabet for the observable protein structural universe. *Proc Natl Acad Sci U S A* **2016,** *113* (47), E7438-e7447.

64. Holland, J.; Pan, Q.; Grigoryan, G., Contact prediction is hardest for the most informative contacts, but improves with the incorporation of contact potentials. *PloS one* **2018,** *13* (6), e0199585.

65. Zhou, J.; Grigoryan, G., Rapid search for tertiary fragments reveals protein sequence-structure relationships. *Protein Sci* **2015,** *24* (4), 508-24.

66. Chaudhury, S.; Lyskov, S.; Gray, J. J., PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **2010,** *26* (5), 689-91.

67. S. Wall, J.; N. Simon, M., *Scanning Transmission Electron Microscopy of DNA-Protein Complexes*. 2001; Vol. 148, p 589-601.

68. ftp.stem.bnl.gov, P. i. a. a.

Chapter VI. Increasing Structural Complexity with Self-Assembling Systems of Multiple Species: Three-Dimensional Peptide-DNA Hybrid Assemblies

6.1 Introduction:

Since its original ideation by Nadrian Seeman in the early 1980s [1], DNA origami has proven to be a promising, and rapidly expanding field of self-assembly research. Using Watson-Crick base pairing as a means to encoding self-assembly of three dimensional structures has proven to be reliable and highly open-source, with the number of possible structures seemingly restricted only by the imagination of the experimenter. Numerous examples have been documented of complex DNA assemblies across multiple size regimes, and these assemblies continue to grow in intricacy and complexity as time progresses [2-4]. The most current forays into DNA origami make use of graph theory routing algorithms and relaxation simulations, as was published by Benson et al. in 2015. This advanced methodology allows for precision engineering of larger, and more intricate DNA origami (see Figure 6.1). Until the work of Benson et al., the processes in which DNA origami was designed and coded still required significant levels of manual adjustment once a sequence was computationally generated. Benson and colleagues have optimized the automation process by employing the principles of graph theory in the computational process. Not only are the resultant structures (from this new methodology) more intricate, they also are more robust towards standard biological assays[5]. Lastly, many of the structures achieved via this new methodology are inaccessible via older protocols, allowing for the production of *de novo structures.*

Despite the advances in complexity thanks to new computational methods for sequencing, DNA origami remains limited in its complexity so long as it remains a single species system, in other words so long as the approaches remain restricted to the utilization of DNA and DNA only.

Therefore, to more closely mimic the immense degree of complexity of natural systems, adding a secondary species, such as peptide, to the system leads to a more developed overall system with greater intricacy. Currently, significant research is being performed towards this very goal: the creation of hybrid biomaterials from the co-assembly of peptide/protein and nucleic material [6-11].

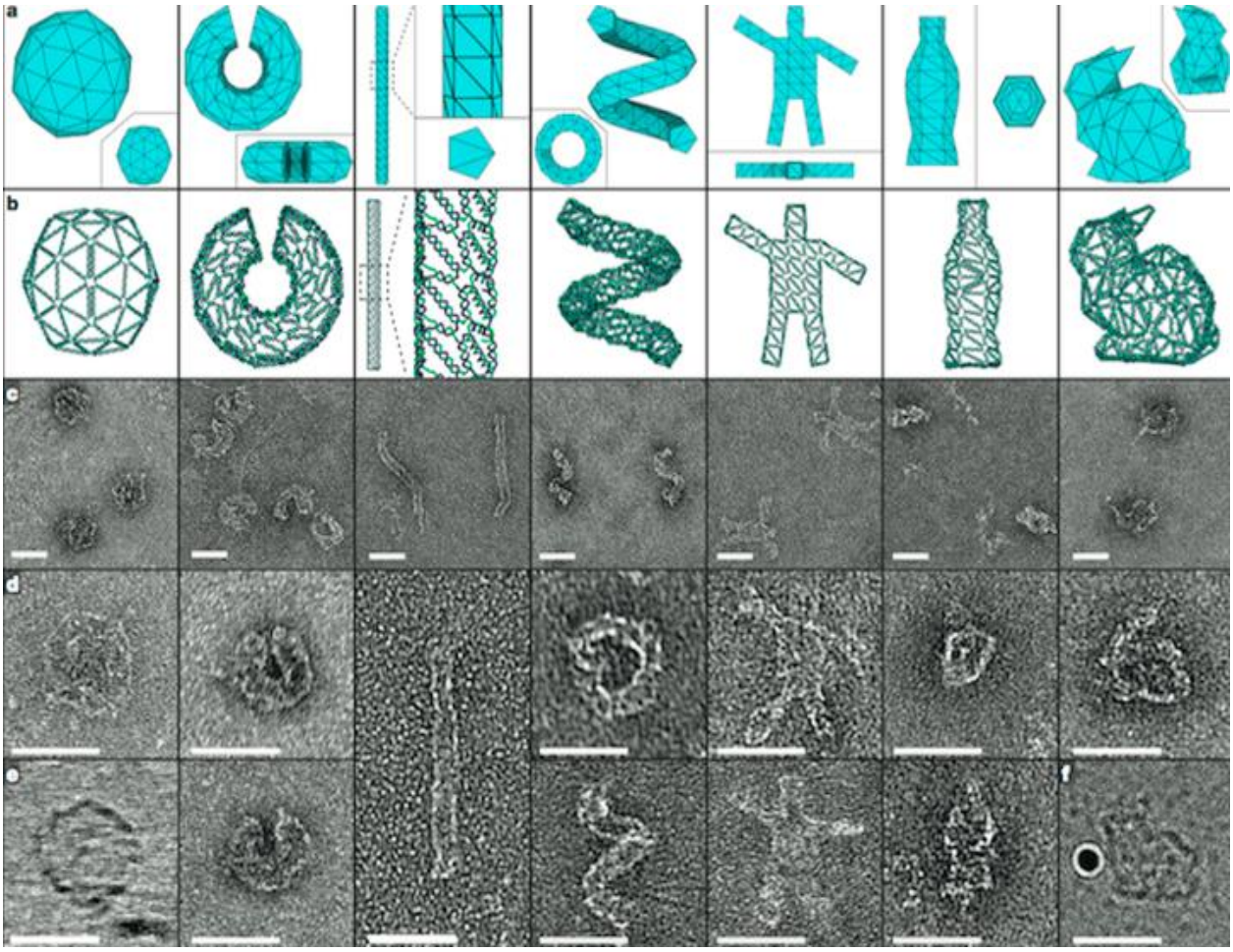


Figure 6.1: Examples of DNA origami from Benson et al. using a new graph theory based approach to computational production of nucleotide sequences. (top) graphical representations of the desired structure to be assembled and (bottom) electron microscopy analysis of actual DNA origami structures wherein their graphically represented equivalents are listed above. From Benson et al. 2015.

Thus far, peptide-specific 1-, 2- and 3-dimensional self-assembling systems have been discussed. These systems demonstrate varying degrees of programmable customization. However, as is witnessed in nature, biological systems oft exhibit a wider spectrum of structural and functional complexity, with many documented systems possessing far greater intricacy than described earlier in the scope of peptide-specific work. In other words, co-assembly of multiple structurally unique subunits can increase structural complexity, as well as imbue a greater degree of tailorable functions, within a supramolecular assembly. Tobacco mosaic virus is a superb

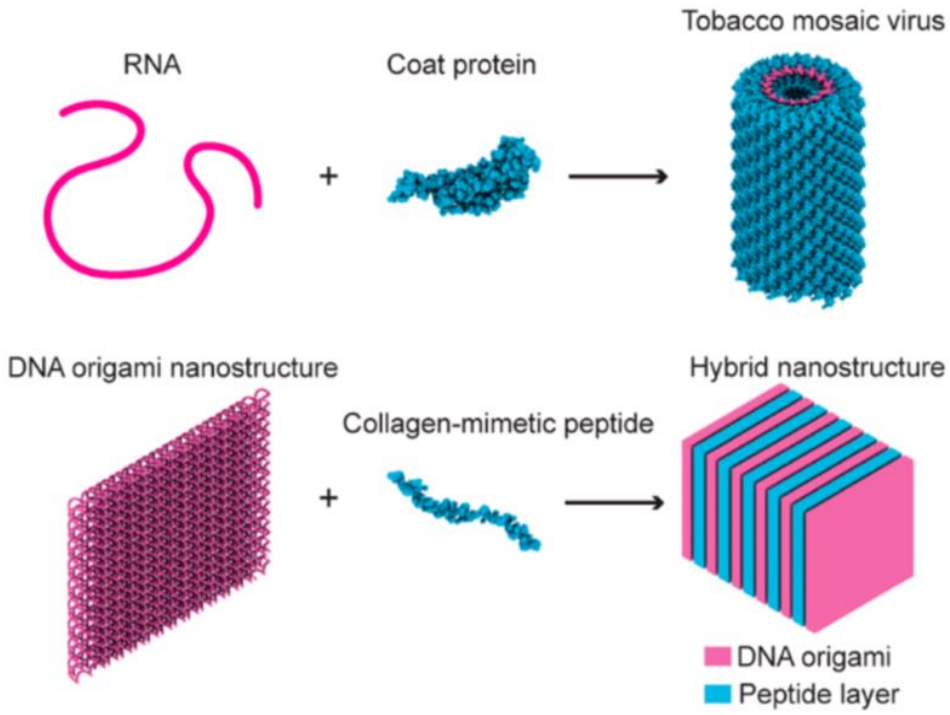


Figure 6.2: (top) tobacco mosaic virus (TMV), an example of highly complex, multi-component nanostructure found in nature. TMV displays a hybrid structure comprising RNA template and coat protein. (bottom) a *de-novo* example of a similarly complex, multi-component nanostructure comprising DNA origami and an engineered collagen-mimetic peptide. From Jiang et al. 2017.

example of such complexity: multiple coat proteins assemble via an RNA template to create an RNA-protein hybrid with greater structural and functional sophistication than would be observed for equivalent nucleic-only or proteinaceous-only assemblies. It is important to both understand and mirror the level of complexity observed in naturally occurring systems which demonstrate co-assembly of nucleic material and peptides. Critical to this understanding are many underlying factors which govern protein-DNA self-assembly: complementary electrostatic interactions [8-9, 11-12], covalent bonding [6], and molecular recognition [7, 10]. In pursuing complex multi-species systems, we describe tailorable self-assembling peptide-DNA nanowires. Three collagen mimetic peptides **CP^{+}, CP^{++}** and **sCP^{++}** are co-assembled with DNA origami (acquired from Yonggong Ke's lab at Emory University). These three sequences follow the canonical collagen triblock motif with Xaa-Yaa-Gly triads, with a neutral core block and charged blocks on either end.

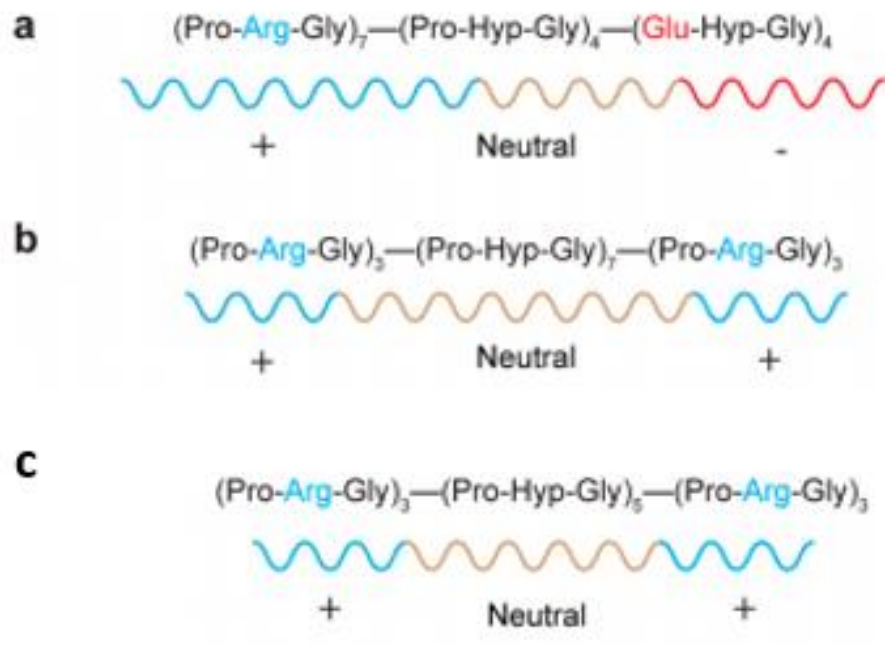


Figure 6.3: the amino acid sequences for the collagen mimetic peptides (a) CP^{+}, (b) CP^{++} and (c) sCP^{++}.

6.2 Results and Discussion

6.2.1 CP^+-TL Hybrid: A first attempt at controlled co-assembly of peptide and nucleic material

The first of the CP sequences **CP^+** self-assembles into discrete, well defined nanosheets. Self-assembly of said sheets was determined to be driven predominantly via electrostatic interactions between adjacent triple helices within the same sheet [13-14]. The positively charged (Pro-Hyp-Gly) block aligns such that it interacts with the negatively charged (Glu-Hyp-Gly) block of the neighboring triple helix. To ensure the resultant supramolecular sheet displays exclusively positive domains at the surface of the sheet, the positive (Pro-Hyp-Gly) block features seven triad repeats compared to four triad repeats for the negative block. In this manner, the constituent triple helices pack in an antiparallel fashion with the negative blocks 'buried' within the sheet, while the longer arginine-dense blocks extend from the surface of the sheet. The underlying hypothesis driving this offset negative/positive design relies on the assumption that the arginine-dense triple helix protrusions will drive association with DNA structures which inherently carry an overall net negative charge due to the sugar-phosphate backbone. It is postulated that the interface between peptide and DNA is further strengthened by hydrogen bonds formed between the guanidinium moiety of arginine and either the nucleobase or backbone of the DNA In this fashion, we hypothesize that **CP^+** may stack and bind the DNA tiles in a face-to-face manner.

Complementary to the collagen mimetic peptide sequences are numerous DNA origami supramolecular assemblies. The first of such origami is '**TL**,' a *t*wo-*l*ayer tile-like structure. A representative transmission electron micrograph of **TL** can be seen in the figure below.

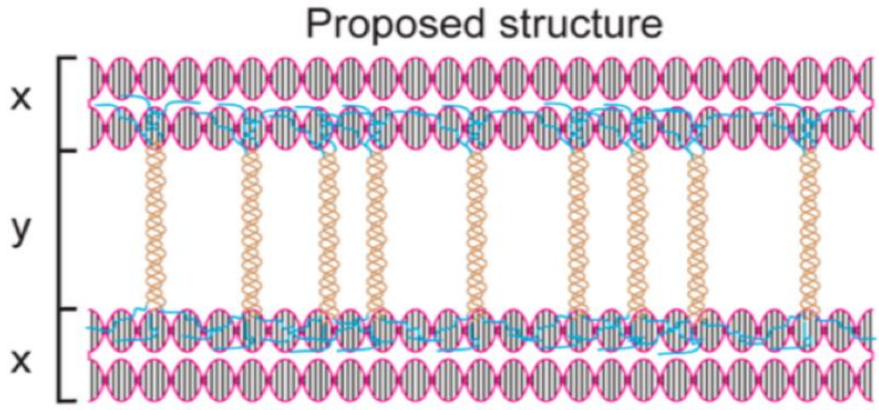


Figure 6.4: graphical representation for the hypothesized structure of the DNA-peptide hybrid nanowire assemblies. The DNA is represented as 'x' and is a bilayer. The peptide CP^{++} is represented by yellow bands. The positively charged termini of the peptide are shown in blue, and their hypothesized interaction with the DNA is shown. Of note, the peptide is predicted to associate in a perpendicular sense to the DNA tiles. From Jiang et al. 2017.

6.2.1.1 Circular Dichroism of CP^{+}

As with all previously mentioned research projects, circular dichroism is an essential method for confirming hypothesized secondary structure of the peptide. The circular dichroism spectropolarimetry data displayed below from Jiang *et al*, 2015 [14] is provided for referential purposes, as CP+ is employed in **TL-**peptide hybrid assemblies. CP+ (1 mg/mL in 20 mM MES buffer at pH 6.0) is shown in red in the left graph, and demonstrates the canonical spectrum indicative of a collagen-like triple helix, with a maximum absorption value at 222-224 nm and a minimum absorption value at 195-199 nm. Thermal denaturation studies indicate a melting transition (T_m) for **CP+** (shown again in red) around 76 °C. (An oppositely charged variant, **CP-**, is shown in black in both the left and right graphs, but is not pertinent to the experiments discussed

in this section). The data confirm that the CP+ peptide is forming the required triple helix secondary structure.

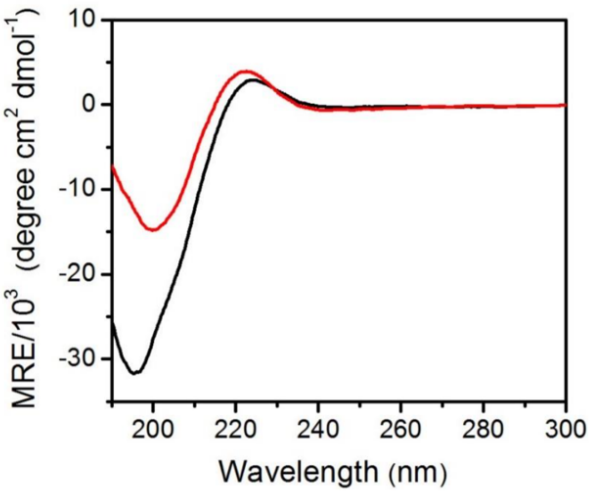


Figure 6.5: CD spectra of aqueous solutions of (red) CP^+, and (black) CP^- from 190 to 300 nm.

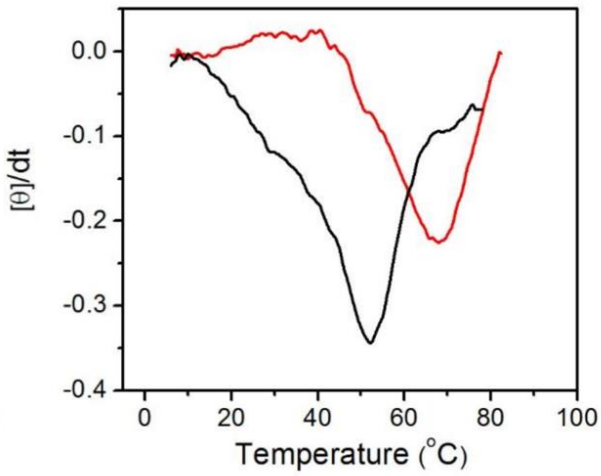


Figure 6.6: First derivative analysis of the melting curves for aqueous solutions of (red) CP^+, and (black) CP^-. First derivative curves were calculated by following the maximal signal relative to temperature.

6.2.1.2 Transmission Electron Microscopy: CP^+ and CP^+-TL

The dimensions of the DNA tile structure are approximated from TEM analysis to be 50 nm x 50 nm x 5 nm. When mixed with CP+ sheets and incubated in 5 mM Tris buffer (pH 8.0), the smaller **TL** tiles attach to the larger CP+ sheet. However, this deposition proves to progress randomly and lacks overall order. A more predictable, more ordered hybridization of DNA and

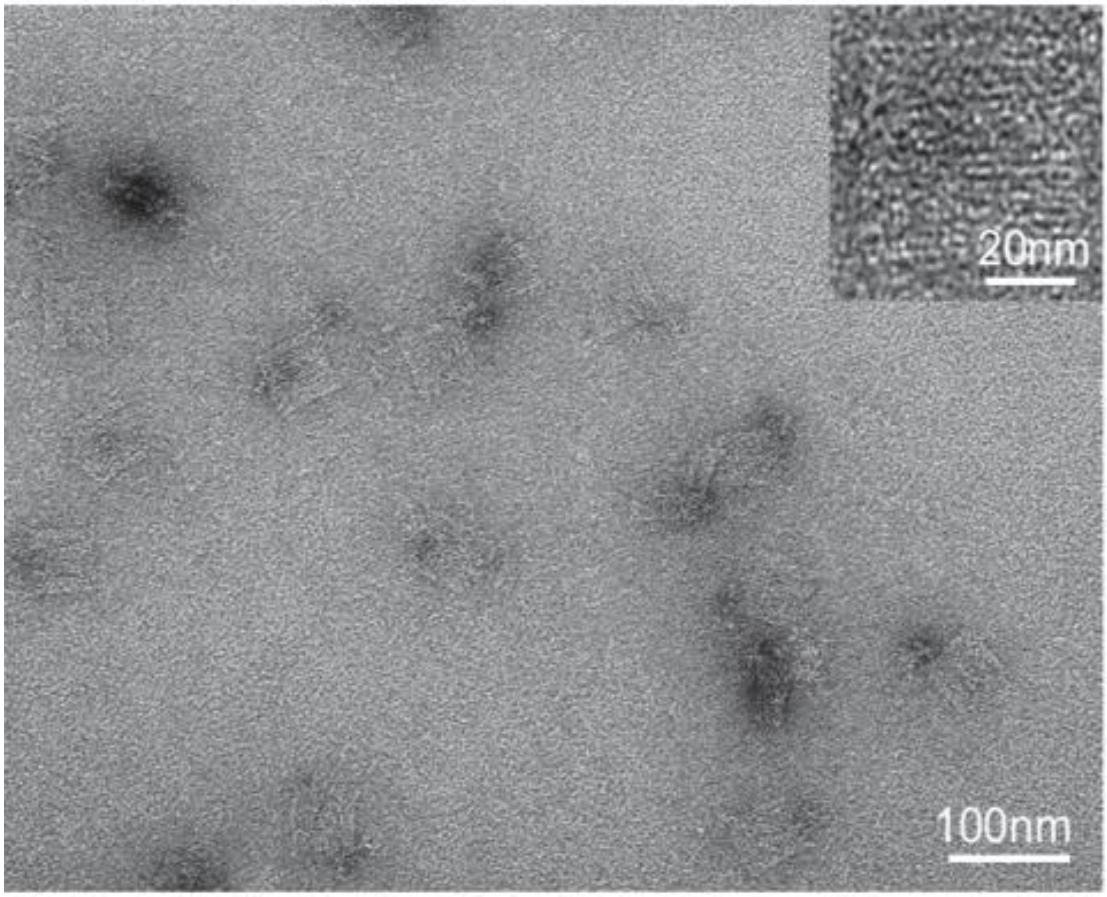


Figure 6.7: transmission electron micrograph indicating the dimensions of TL, wherein measurements for the DNA assembly are approximated at 50 x 50 x 5 nm. As the tiles are somewhat difficult to discern, a yellow outline is provided on one of the tiles for easier identification.

peptide is desired. In this pursuit, new peptide sequences were designed and will be described in subsequent sections.

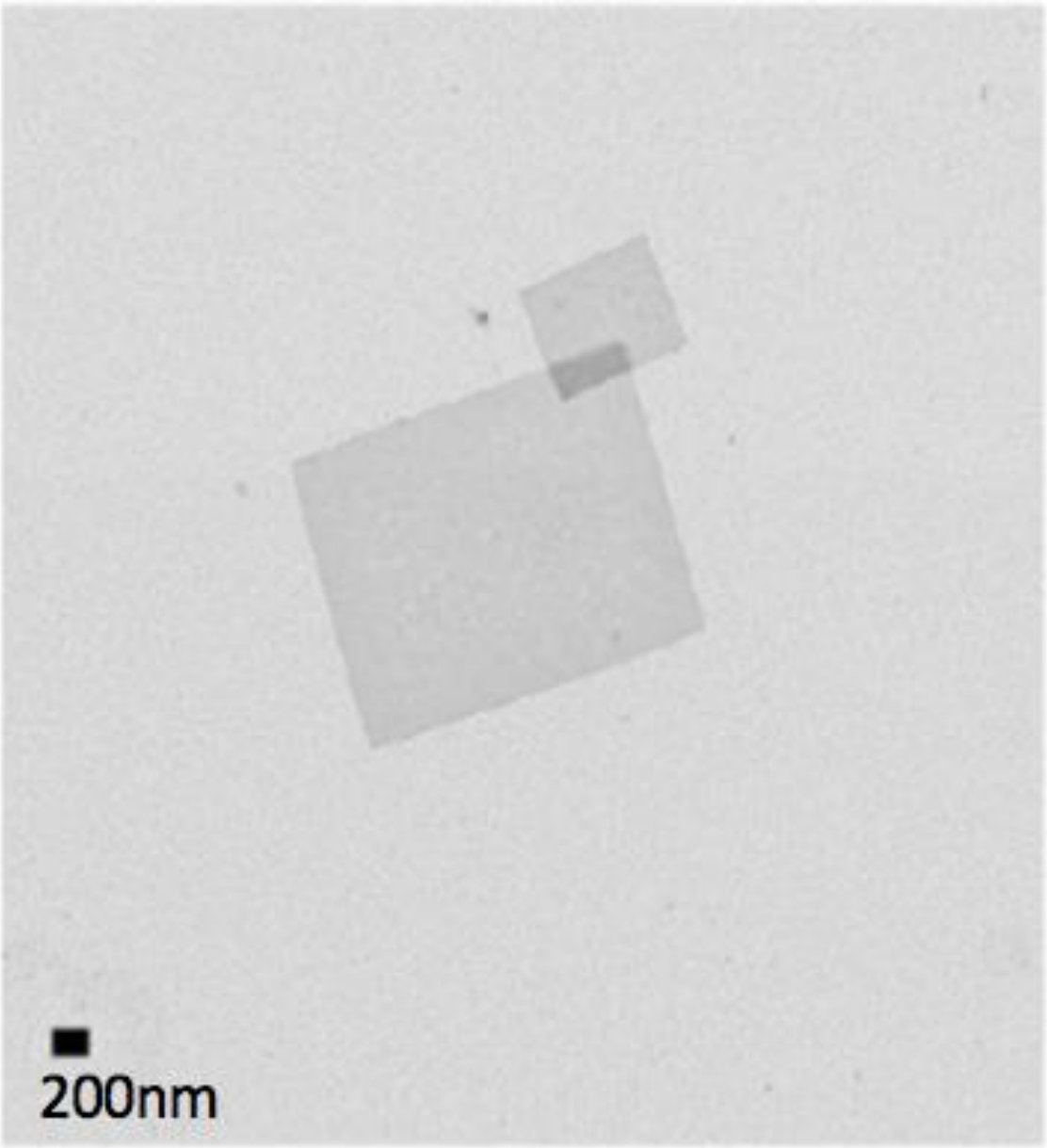


Figure 6.8: transmission electron micrograph of CP^+ indicating the approximate x,y dimensions on the scale of multiple hundreds of nanometers.

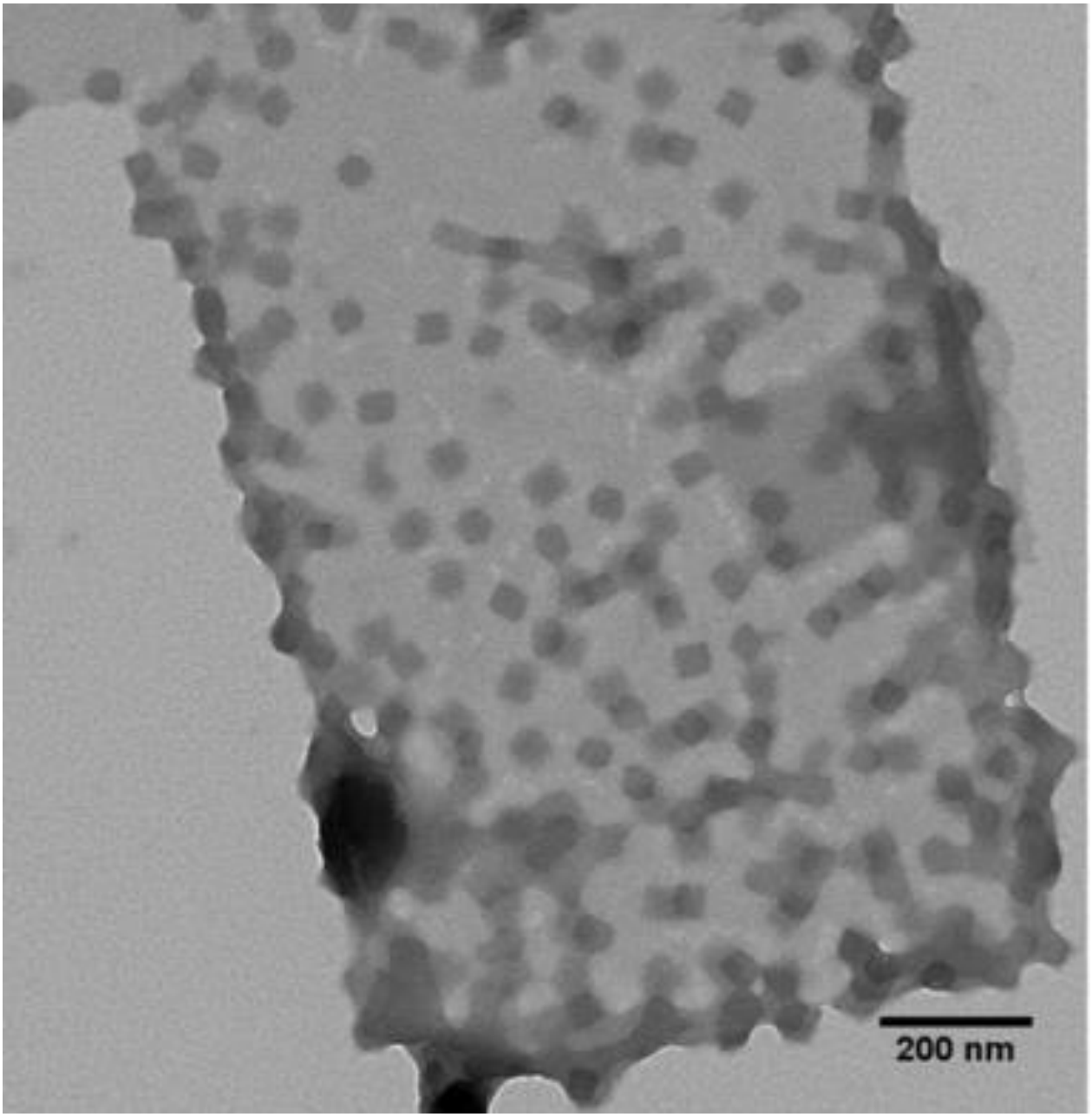


Figure 6.9: the hybrid TL/CP^+ supramolecular assembly wherein TL deposits upon the surface of the peptide sheet due to electrostatic complementarity between respective faces of each species.

6.2.2 CP^{++}-TL and sCP^{++}-TL: A new approach for creating DNA-peptide hybrid nanowires

We hypothesize that decreasing the role of the collagen mimetic CP *sheet* in the hybrid structure will enable a more ordered co-assembly of species. In other words, we propose changing the CP system from a large-scale *template* upon which **TL** deposits, to a CP system wherein the CP peptide serves as a 'looser,' or less-ordered *glue* wherein individual triple helices demonstrate greater affinity for **TL** tiles than they demonstrate for one another. In this proposed methodology, the CP peptide does not demonstrate sheet formation and remains in the form of discrete, individual helices lacking supramolecular structure (see figure). In this manner, we hope to drive the assembly process to more closely mimic the hypothesized process outlined in the 'proposed mechanism' figure above.

To achieve this transformation from *template* to *glue*, **CP^{++}** and **sCP^{++}** are employed. To abrogate self-association of CP peptide (sheet) in favor of co-assembly of CP peptide and DNA origami, **CP^{++}** features a charge-neutral central block comprising $(Pro\text{-}Hyp\text{-}Gly)_7$ and two charge-positive blocks $(Pro\text{-}Arg\text{-}Gly)_3$ at both the C- and N-termini [15]. By programming in positive charge at both termini, the new sequences will be less electrostatically stable the more they self-associate. **sCP^{++}** similarly features three positive triads at both termini but reduces the number of core triads from seven (as in **CP^{++}**) to five. See figure for graphical representations of **CP^{++}** and **sCP^{++}**.

6.2.2.1 Circular Dichroism of CP^{++} and sCP^{++}

Circular dichroism spectropolarimetry results for both **CP^{++}** and **sCP^{++}** are shown in figure 6.9 and 6.10. Both exhibit canonical spectra indicative of a collagen-like triple helix, confirming that the mutations in both variants (when compared to **CP^{+}**) have not compromised or abrogated

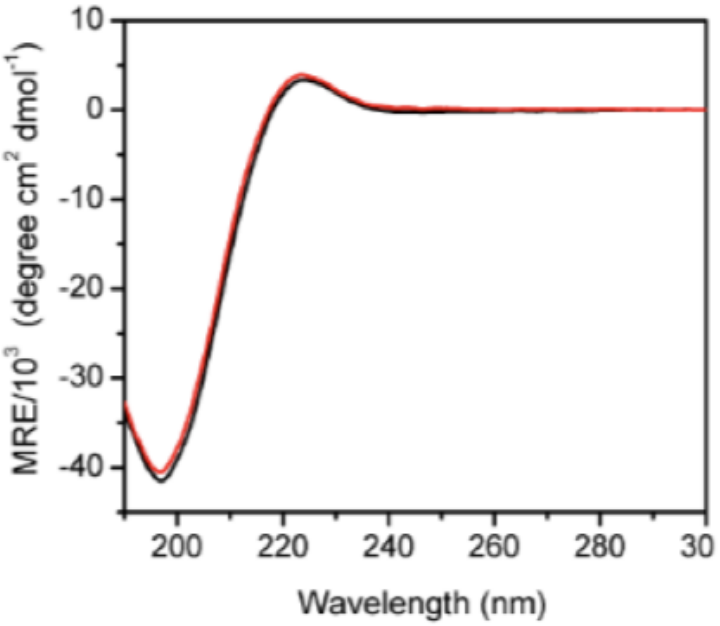


Figure 6.10: Circular Dichroism of (black) CP^{++} and (red) sCP^{++}. Both of which indicate-triple helix formation.

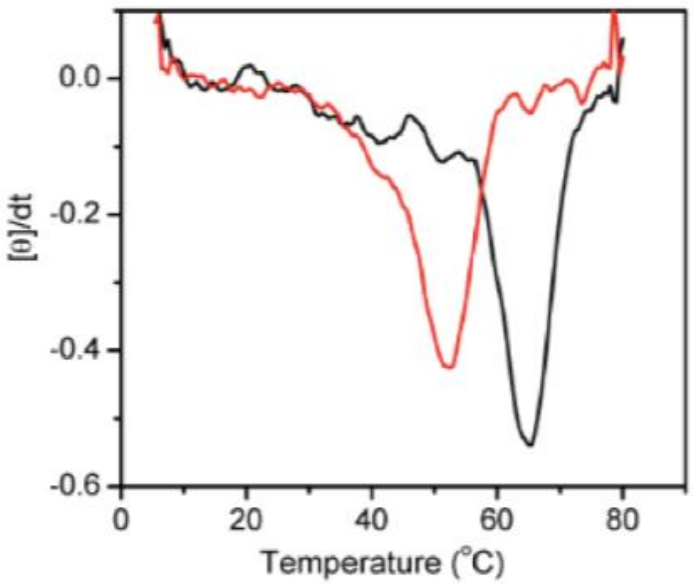


Figure 6.11: First derivative analysis of the melting curves for aqueous (black) CP^{++} and aqueous (red) s CP^{++}. First derivatives were derived by following the maximal signal relative to the temperature.

the triple-helix secondary structure which is important for subsequent co-assembly with DNA **TL**.

Of note, melting analyses of both variants produced differing melting transition temperatures (T_m): 52 and 63 °C for **sCP**$^{++}$ and **CP**$^{++}$ respectively. The variance in the T_m values can be explained by the difference in central core length for each species. It has been previously demonstrated that reducing the length of the central core of a triple helix (Pro-Hyp-Gly) correlates to a reduction in thermal stability [13]. Thus, the five-triad central core **sCP**$^{++}$ demonstrates a lower melting transition temperature than the seven-triad central core **CP**$^{++}$.

6.2.2.2 Transmission Electron Microscopy of CP^{++}-TL

Mixing **CP**$^{++}$ and **TL** at a ratio of 1600:1 (with **TL** at an initial concentration of 20 nM) produces one-dimensional nanowires. These nanowires appear (via TEM) banded in appearance due to the stacking of DNA tiles arranged with **CP**$^{++}$ spliced between each layer of DNA. As expected, the width of the banded nanowires was found to be 47 ± 3 nm, copacetic with the dimensions of the free **TL** (50 nm). Moreover, the dimension of the lighter bands in the nanowire structure corresponds reasonably with the previously measured thickness of TL sheets (4.8 ± 0.7 nm). The dark bands are 5.9 ± 0.6 nm, which corresponds to calculated predicted length of the *central* block for **CP**$^{++}$ if the rise per residue of the triple helix is assumed to be canonical (0.286 nm rise/residue * 21 residues = 6.0 nm). This inter-band value indicates that the collagen triple-helices are associating perpendicularly to the **TL** and that the positively charged terminal domains are interacting in a more parallel fashion to **TL** to drive co-assembly of DNA and peptide. Overall, this represents an improvement in the generation of multi-component nanowires when compared with the methodology described in the previous section.

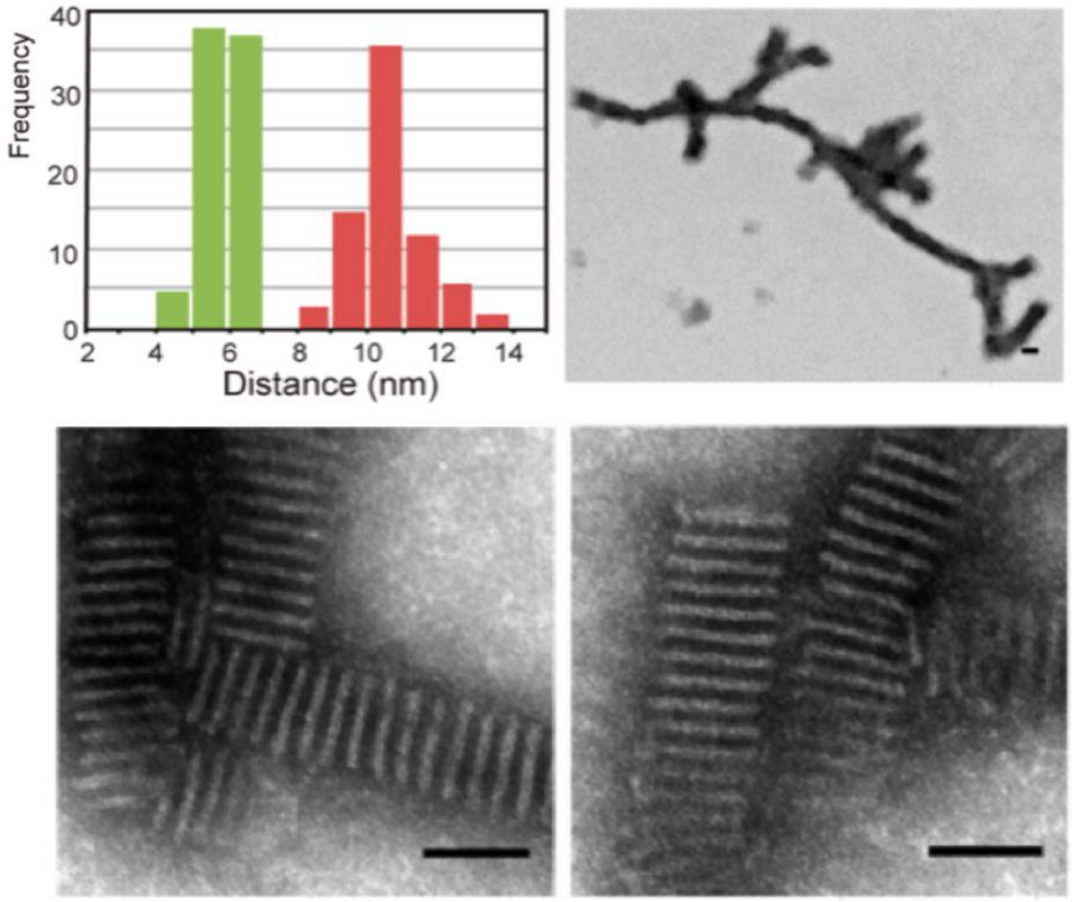


Figure 6.12: (top left) histogram representations of the measured dimensions of (red) TL plus CP^{++}, represented as x + y in the proposed structure diagram and histogram representation of the measured dimensions (red) of exclusively the peptide CP^{++} layer, represented as x in the proposed structure diagram macro view of the nanowire hybrid assemblies, demonstrating axial lengths on the scale of hundreds of nanometers. (bottom left and right) detailed transmission electron micrographs of hybrid nanowires formed from the co-assembly of DNA TL and peptide CP^{++}. Scale bars for all micrographs are 50 nm. From Jiang et al. 2017.

6.2.2.3 Transmission Electron Microscopy of Varying the Ratios of CP^{++} and TL

Similar to the previously described **CP^{+}-TL** assembly, the ratio for stable nanowire co-assembly must be equal to or greater than **CP^{++}:TL** = 1600:1. Increasing **TL** concentration yielded larger average axial nanowire lengths, while restricting the DNA concentration produced shorter axial lengths. This can be observed in the figure, wherein larger concentrations of DNA tile exhibit longer axial length values for the hybrid nanowire supramolecular assembly. It is demonstrated that smaller concentrations of TL (0.35 nM), when incubated with 32 uM **CP^{++}** peptide, produce shorter length nanowires than when mixing 20 nM TL with the equivalent concentration of peptide. This is observed via TEM (see figure). A similar effect is observed when increasing the concentration of the peptide component within the hybrid solution of DNA and collagen mimetic peptide. Increasing the amount of **CP^{++}** in comparison to available DNA **TL** leads to an increase in the number of nanowires, as well as an increase in the general order and overall length of said nanowires. This effect was maximized in the hybrid solution which featured 320 uM peptide mixed with 20 nM DNA (Figure 6.15). These nanowires formed robust and well defined structures with little globular aggregate or non-target assemblies observed.

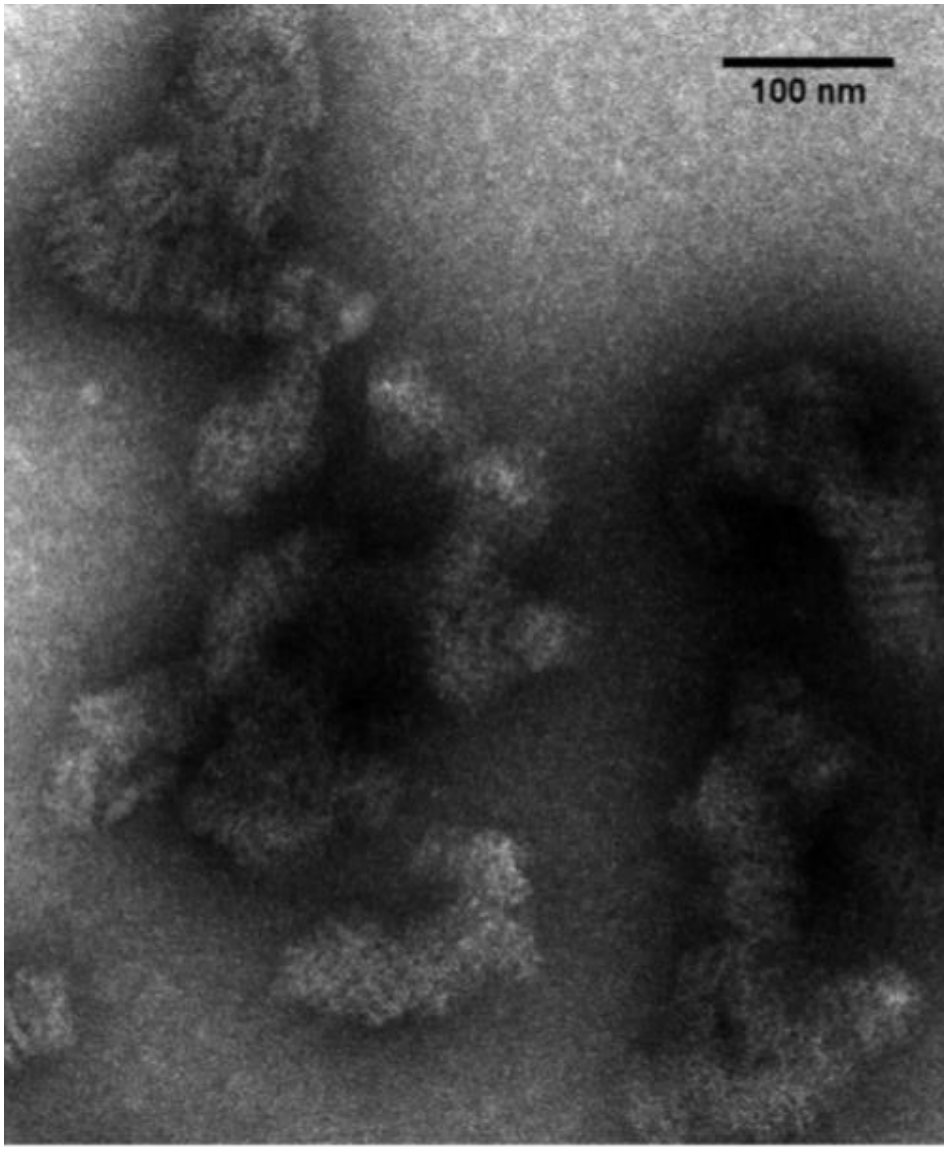


Figure 6.13: Negatively stained transmission electron micrograph of 0.35 nM DNA TL mixed with 32 uM of CP^{++}. This assembly features the lowest ratio of DNA to peptide, and demonstrates poorly ordered aggregates of TL and CP^{++} lacking discernible, patterned striations that are observed in other assemblies. This indicates optimization of the DNA:peptide ratio is required, and that this assembly is lacking enough DNA TL for proper

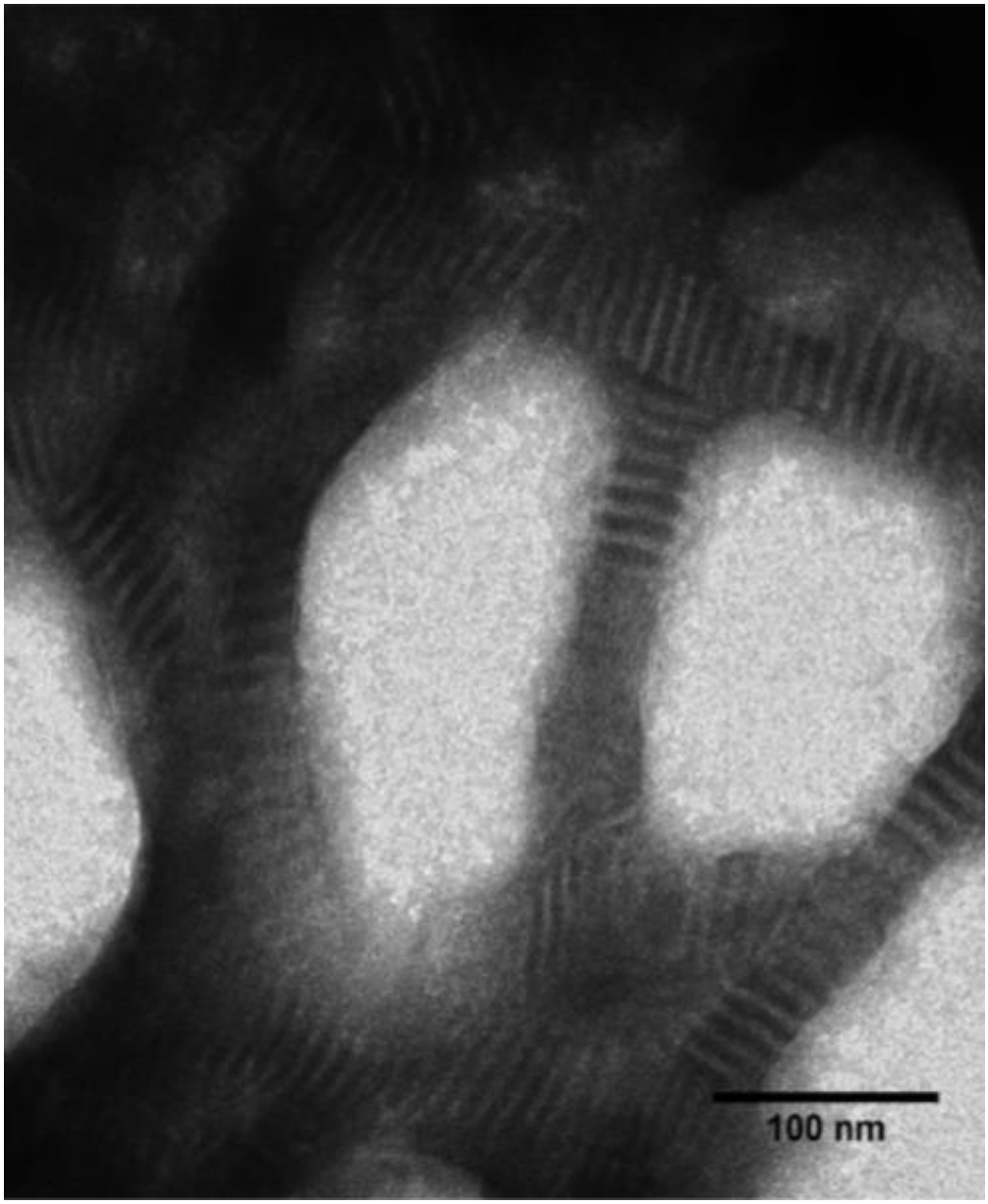


Figure 6.14: Negatively stained transmission electron micrograph of 20 nM DNA TL mixed with 32 uM of CP^{++}. This assembly features a mid-level ratio of DNA to peptide, and demonstrates comparatively well- ordered nanowires (not aggregates) of TL and CP^{++} with discernible, patterned striations between peptide and DNA. This indicates a ratio of DNA to peptide that is sufficient for self-assembly of the two species into nanowires.

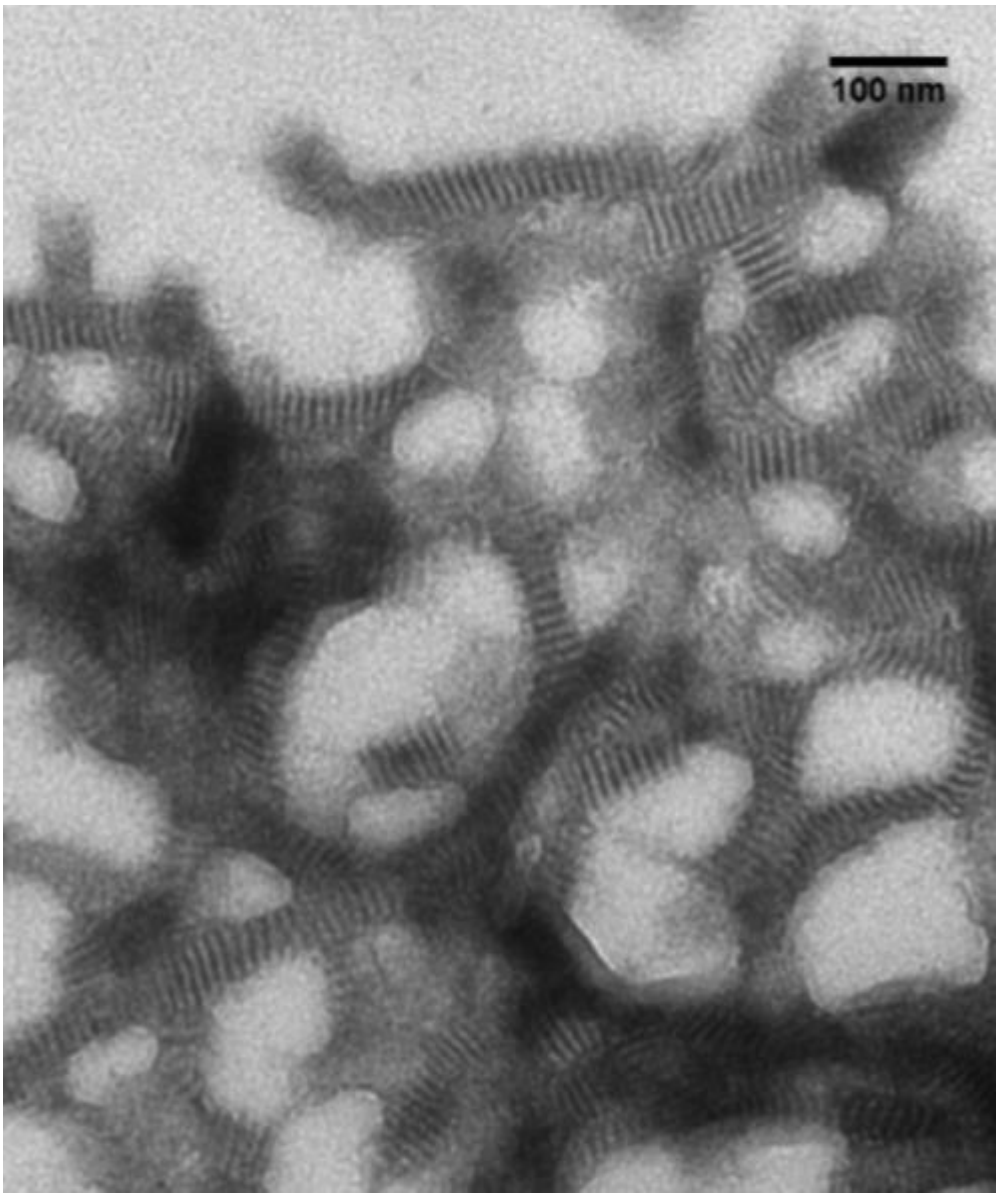


Figure 6.15: Negatively stained transmission electron micrograph of 20 nM DNA TL mixed with 320 uM of CP^{++}. This assembly features a large ratio of peptide to DNA, and demonstrates very well-ordered nanowires (not aggregates) of TL and CP^{++} with discernible, patterned striations between peptide and DNA. This indicates a highly optimized ratio of DNA to peptide that is sufficient for self-assembly of the two species into nanowires.

6.2.2.4 Small- and Wide-Angle X-Ray Scattering of CP^{++}-TL

Small- and wide-angle x-ray scattering was performed on **CP^{++}-TL** co-assemblies in 5 mM Tris buffer at pH 8.0 with 5% glycerol addition to dampen the effects of beam degradation to the sample. The data indicate a Bragg diffraction peak at q = 0.061 ± 0.004 $Å^{-1}$ (see figure 6.15). This Bragg reflection yields a *d*-spacing of 10.3 ± 0.7 nm ($2\pi/0.061$ $Å^{-1}$ = 103 Å). This *d*-spacing reasonably matches and confirms the inter-band distance in the nanowires (measured center-to-center of neighboring light bands). This value is represented graphically as x + y in the proposed structure figure (Figure 6.2). When subjected to Scherrer analysis [16] (see equation below), the aforementioned diffraction data yield an averaged nanowire length of 109 nm. This value assumes that peak broadening within the Scherrer analysis was primarily due to the nanowires' finite lengths (see Figures 6.13 and 6.14).

Scherrer equation [17]***:***

$$\tau = \frac{K\lambda}{\beta \cos\theta}$$

In the equation above, τ represents the average size of the crystalline fields ($\tau \leq$ grain size), *K* represents a (dimensionless) shape factor, λ represents the wavelength of the X-ray used, β represents the line broadening at full width half maximum (FWHM) minus the experimental line broadening, and θ represents the Bragg angle. The Scherrer equation yields insight into the dimensions of sub-micrometer crystallites. The degree of broadening of an X-ray diffraction peak is correlated to the size of the analyte particle. The Porod region yields a power law slope of approximately -2.6, indicating the supramolecular assembly adopts a conformation between that of a two-dimensional disc (power law slope of -2) and a large three-dimensional sphere (power

law slope of -3). This data could indicate that, despite addition of glycerol as a protecting agent, radiation damage is occurring (creating 2D spheres), followed by the formation of higher order aggregates (clumping of aggregates into 3D spheres).

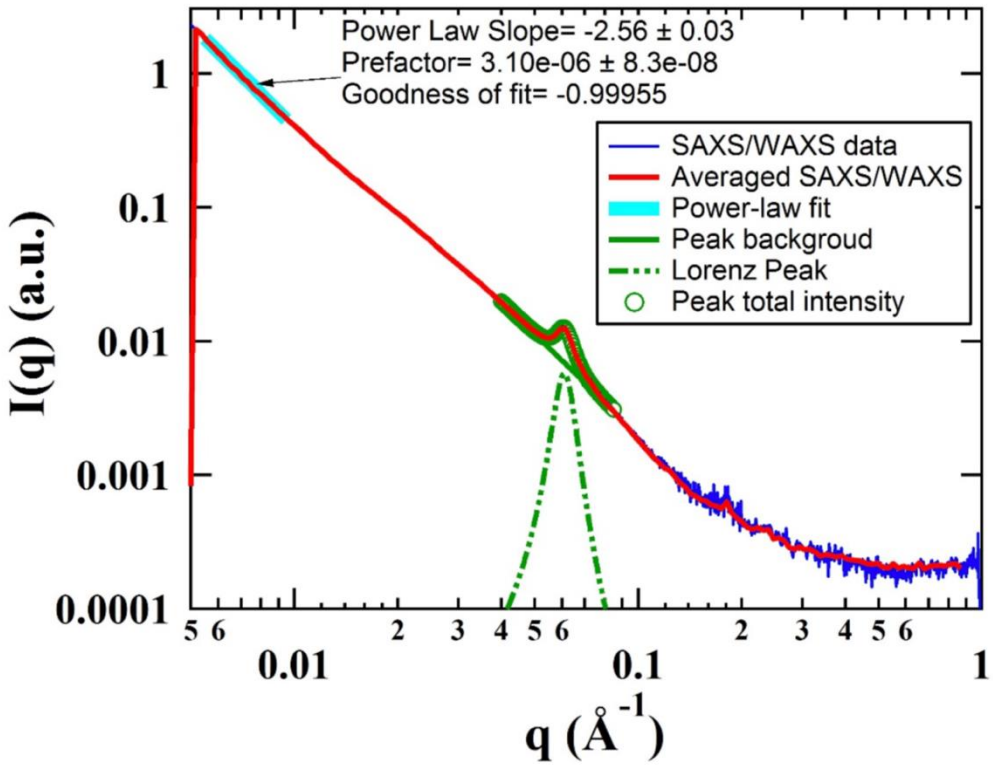


Figure 6.16: SAXS/WAXS profiles for CP^{++}-TL co-assemblies in 5 mM Tris buffer at pH 8.0 with 5% glycerol addition by volume. (Red) SAXS/WAXS averaged scattering profiles including Guinier region. (Green dash) Lorenz peak fitting showing a major diffraction peak at q = 0.061 ± 0.004. (Cyan) power law slope in the Porod region indicating a power law slope of -2.6.

6.1.3.6 Transmission Electron Microscopy: sCP^{++}-TL

When mixing **sCP^{++}** with **TL**, the resultant nanowire assemblies exhibit a morphology different than **CP^{++}-TL** nanowires. Briefly, **sCP^{++}** features a shorter neutral core block than does **CP^{++}** *(vide supra).* As anticipated, the inter-tile distance is decreased to match the theoretical collagen triple-helix length with two fewer triads than **CP^{++}** (0.286 nm rise/residue * 15 residues = 4.3 nm). The experimentally measured distance was found to be 3.8 ± 0.9 nm. As a possible result of decreased thermal stability due to the smaller neutral core block of **sCP^{++}**, the observed average length for the hybrid nanowire assembly is significantly smaller than observed for **CP^{++}**. The **CP^{++}-TL** average nanowire length was measured to be 133 ± 92 nm, whereas the **sCP^{++}-TL** average nanowire length was measured to be 46 ± 19 nm.

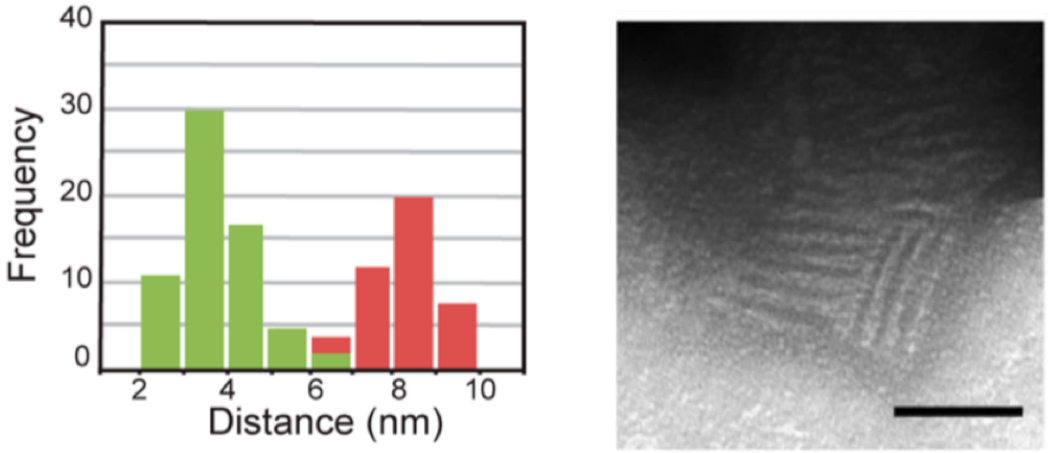


Figure 6.17: (left) histogram of measured distances for (red) TL-layer plus sCP^{++} peptide-layer (x + y in the proposed structure diagram and for (green) the distances of sCP^{++} peptide-layer (y in the proposed structure diagram. (right) negatively stained transmission electron micrograph of sCP^{++} and TL nanowires wherein the lighter bands are the TL layer and the darker bands the peptide layer. Scale bar is 50 nm. From Jiang et al. 2017.

Of note, the aforementioned **TL** sheets were not observed to associate in an edge-to-edge manner when incubated with **CP**$^{++}$. The **TL** sheets measure 5 nm in thickness, and it is hypothesized that the **TL** edges are not sufficiently large enough to provide a thermodynamically stable interface for association with **CP**$^{++}$ and **sCP**$^{++}$. It is assumed that in the 'extended-state,' the $(Pro\text{-}Arg\text{-}Gly)_3$ terminal blocks demonstrate a radius of 3 nm. This value is obtained by multiplying the contour length per residue value (0.33 nm [18]) by the number of residues (9 residues).

6.2.3 A DNA design for ribbon formation: the DNA-brick

A third design is proposed to further test the aforementioned size and thermodynamic stability quandary. In other words, DNA-bricks [19] are incubated with **CP**$^{++}$. Two DNA-bricks

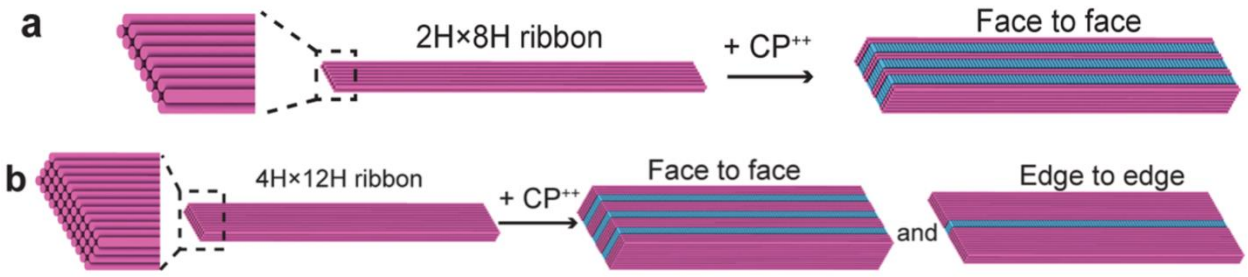


Figure 6.18: Graphical models of the DNA ribbons, indicating their 'edge' and 'face' dimensions as well as their hypothesized modes self-association based on available surface area. In both cases the ribbons associate via CP^{++} and its positively charged terminal domains. (a) the 2Hx8H is hypothesized to lack sufficient edge surface area to associate in an edge-to-edge manner, packing exclusively face-to-face. (b) the 4Hx12H is hypothesized to afford sufficient edge surface area, as well as face surface area, for both face-to-face and edge-to-edge packing.

are tested: **2H x 8H**, and **4H x 12H**, wherein the overall morphology resembles a ribbon-like structure and the H dimensions describe the width and height of the ribbon (Figure 6.17). **2H x 8H** has cross-sectional dimensions measuring 5 nm by 20 nm, and **4H x 12H** has cross-sectional dimensions measuring 10 nm by 30 nm. It is hypothesized that, based on prior data regarding surface area and its relation to thermodynamic stability (vide supra), the **2H x 8H** ribbon will associate exclusively face-to-face, as it lacks sufficient lateral surface on the 'edge' of the brick required for stable association with **CP**$^{++}$ (see Figure). Alternatively, the **4H x 12H** ribbon is hypothesized to associate in both an edge-to-edge, and face-to-face, as it features a larger 'edge' surface area. Both ribbon designs are co-assembled with **CP**$^{++}$ in order to test these hypotheses. Importantly, the **2Hx8H** ribbon has a thickness equivalent to that of the DNA tile, making the **4H x 12H** ribbons twice the thickness of the tile.

6.2.3.1 Transmission Electron Microscopy

Following self-assembly of the hybrid system (using a similar concentration ratio as prior assemblies, i.e. nM concentrations of DNA and μM concentrations of collagen mimetic peptide), transmission electron microscopy was performed to confirm hybrid self-assembly of the two species. Transmission electron microscopy data confirms the hypothesis that the **4H x 12H** DNA ribbon, when co-assembled with **CP**$^{++}$, associates in both face-to-face and edge-to-edge fashion (Figure 6.20). Of note, it was observed that the DNA ribbons were associating via both of aforementioned stacking patterns, even within the same supramolecular assembly. Lastly, a high degree of edge-to-edge packing was not observed in this system, and it is presumed that the relatively low persistence length of the ribbons allows for structural flexibility and twisting, which may preclude edge-to-edge association.

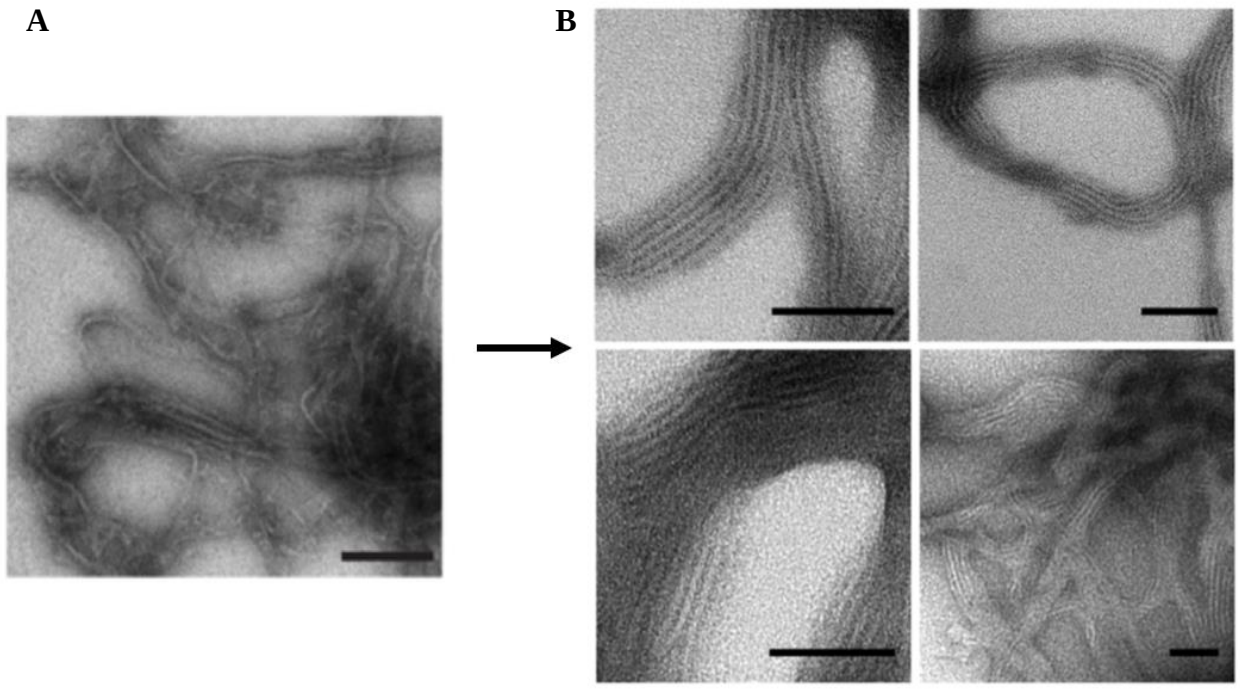


Figure 6.19: Negative stained electron micrographs of 2H x 8H DNA ribbon hybridized with CP^{++}, wherein the ribbons lack sufficient edge surface area to associate in an edge-to-edge manner. EM confirms that these ribbons, when co-assembled with $C\backslash P^{++}$ assemble exclusively in a face-to-face fashion. (A) micrograph indicating singular DNA ribbons in the absence of peptide. (B) four micrographs showing DNA ribbon in the presence of peptide, and subsequent stacking of ribbons as a result of co-assembly with CP^{++}.

The **2H x 8H** DNA ribbon, when analyzed via negative stain transmission electron microscopy, adopts a face-to-face stacking pattern. Importantly, as hypothesized, it lacks sufficient edge surface area to adopt an edge-to-edge stacking pattern. The differences between the ability for face-to-face packing (the only packing manner observed for the hybrid peptide/DNA sample with smaller DNA bricks) and a mixture of both face-to-face *and* edge-to-edge (as observed in the hybrid peptide/DNA samples with larger DNA bricks) can be visualized in the TEM data (Figures 6.19 and 6.20). The larger DNA bricks can be seen forming ribbons with larger striations of low-

electron density. Specifically, the mixture of face-to-face and edge-to-edge packing morphologies can be observed in the bottom right transmission electron micrograph of Figure 6.20B, where low electron density striations demonstrate heterogeneity in width, concurrent with a mix of packing styles.

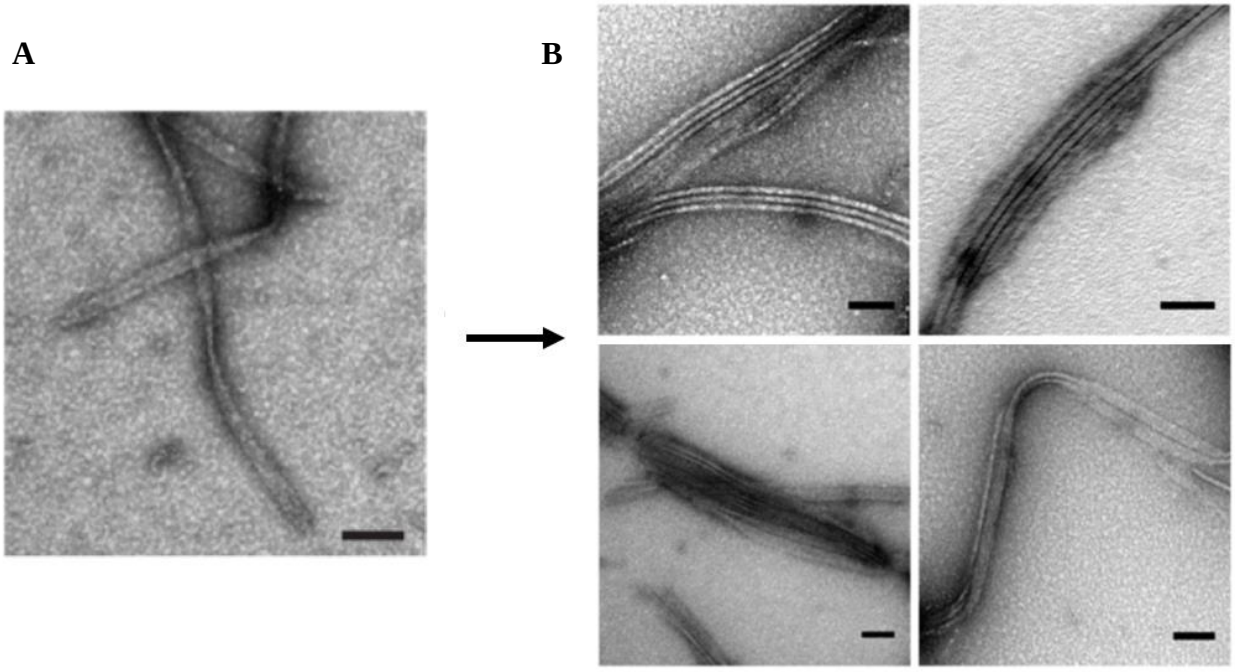


Figure 6.20: Negative stained electron micrographs of 4H x 12H DNA ribbon hybridized with CP^{++}, wherein the ribbons provide sufficient edge surface-area to associate in an edge-to-edge manner as well as a face-to-face manner. EM confirms that these ribbons, when co-assembled with CP^{++} assemble in both fashions. (A) micrograph showing DNA ribbons in the absence of peptide. (B) four micrographs showing DNA ribbon in the presence of peptide, and subsequent stacking of ribbons as a result of co-assembly with CP^{++}, wherein the top row shows face-to-face packing and the bottom row shows a mixture of face-to-face and edge-to-edge packing.

6.3 Conclusion

The pursuit of highly complex, self-assembling biomaterials has been a key goal of scientific research over the past several decades. We have thus far proven the capacity for self-assembling peptide biomaterials (primary helical nanotubes), but there is benefit to increasing the complexity of a self-assembling system beyond what can be achieved with only one subunit or species (i.e. peptide-only). The addition of DNA is a facile and efficient method for expanding the complexity of self-assembling biomaterials. Given DNA's immensely well-understood structure and function, as well as the vast library of DNA origami structures that have been produced and characterized, it presents a viable foundation for the building of multi-species systems in which DNA is partnered with peptide. This project establishes the validity, and ease, of using DNA and peptide in tandem to create highly complex, unique biomaterials with predictable size and structure.

6.4 Methods.

6.4.1 Materials.

Chemical reagents were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO) unless otherwise noted. Amino acids were purchased from Anaspec. Inc. (Fremont, CA). Fmoc-Gly-PEG-PS resin for solid-phase synthesis was purchased from Applied Biosystems, Inc. (Waltham, MA). DNA staple strands were purchased from Integrated DNA Technology, Inc. (Coralville, IA).

6.4.2 Peptide Synthesis.

Peptides CP++ and sCP++ were prepared on a Fmoc-Gly-PEG-PS resin using a CEM Liberty microwave-assisted solid-phase peptide synthesizer. Standard Fmoc protection chemistry was used with HBTU/DIEA-mediated activation and piperidine - induced deprotection of the Fmoc group. Peptide was purified via RP-HPLC on a C18 column using a gradient of water-acetonitrile with 0.1% trifluoroacetic acid. The mass and purity were analyzed by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry and analytical HPLC. Pure fractions of peptides were collected and dialyzed against deionized water 3x8 hours to remove residual trifluoroacetate (MWCO = 2,000 Da), and lyophilized. Lyophilized peptides CP++ and sCP++ were dissolved in Tris buffer (5 mM, pH 8.0) for following experiments. Solution concentrations were determined from the measured mass of peptide. Peptide solutions were thermally annealed at 85 °C for 45 min, and slowly cooled down to ambient temperature.

6.4.3 TL Nanosheet Preparation.

The single-stranded scaffold (p7560) was prepared using a previously reported method1. The staples were mixed with p7560 scaffold in a molar stoichiometric ratio of 10:1 in the 1×TE buffer (5 mM Tris, 1 mM EDTA, pH 8) supplemented with 14 mM MgCl2. The final concentration of the scaffold was adjusted to 20 nM. The strand mixture was then annealed in a thermal cycler at 65oC for 10 min and then cooled with a linear ramp from 60 °C to 25 °C over 72 hours. Annealed DNA origami samples were subjected to agarose gel electrophoresis for purification. Samples were run in 1% agarose gel in a 0.5×TBE-Mg2+ buffer (45 mM Tris, 45 mM Boric acid, 1 mM EDTA, 10 mM Mg2+) at 60 V for 3 h within an ice-water bath. The migrating bands corresponding to the correctly folded structures were then visualized with ultraviolet light and cut out from the gel. Excised bands were crushed and transferred into DNA gel extraction spin column (BIO-RAD,

Catalog number: 732-6166). The DNA structure solution was recovered by centrifugation of the loaded column for 10 min at 16,000g.

6.4.4 2Hx8H and 4Hx8H Ribbon Preparation.

The strands were mixed together at equal molar ratio, in the 1×TE buffer (5 mM Tris, 1 mM EDTA, pH 8) supplemented with 10 mM MgCl2. The final concentration was 100 nM per strand. The strand mixture was then annealed in a thermal cycler at 65oC for 10 min and then cooled with a linear ramp from 60 °C to 25 °C over 18 hours.

6.4.5 Co-assembly of Peptide and DNA.

2 μl of 2 mg/ml annealed CP++ or sCP++ solutions in Tris buffer (5 mM, pH 8.0) was added to 30 μl of 20 nM TL solutions diluted in Tris buffer (5 mM, pH 8.0) or unpurified 100nM ribbon solutions at 4 °C. The peptide-DNA mixtures were then incubated at room temperature for CP++, and 4 °C for sCP++ for more than three days before investigation. The mixed solution turn cloudy after peptide-DNA co-assembly.

6.4.6 Circular Dichroism.

CD spectra were recorded on a Jasco J- 810 CD spectropolarimeter in 0.10 mm quartz cells in Tris buffer (5 mM, pH 8.0). Spectra were recorded from 300 to 190 nm at a scanning rate of 100 nm/min and a resolution of 0.5 nm. CD melting experiments were performed in the temperature range from 5 °C to 85 °C at a heating rate of 1 °C/min. The intensity of the CD signal at 224 nm was monitored as a function of temperature. Melting temperatures were obtained from the first derivative of the melting curves.

6.4.7 Transmission Electron Microscopy:

TEM specimens were prepared from solutions of CP++ (2 mg/ml), sCP++ (2 mg/ml), or peptide-DNA co-assemblies (0.125 mg/ml of peptides with 20 nM of TL nanosheet) in Tris buffer (5 mM, pH 8.0). 5 μl of the samples were deposited onto 200 mesh carbon coated copper grids from Electron Microscopy Sciences (Hatfield, PA). After a 1 minute incubation period, excess liquid was wicked away. The specimens were stained with 5 μl of fresh solution of uranyl acetate (1 %). Excess stain was wicked away after incubation on the grid for 5-10 s. The sample grids were dried under vacuum and stored in a desiccator before investigation. TEM measurements were acquired on a Hitachi HT7700 transmission electron microscope at an accelerating voltage of 80 kV.

6.4.8 Small- and Wide-Angle X-Ray Scattering:

Synchrotron SAXS/WAXS measurements were performed at the 12-ID-B beamline of Advanced Photon Source at Argonne National Laboratory. A simultaneous SAXS/WAXS setup was used and the sample-to-detector distances were set such that the overall scattering momentum transfer, q, range was achieved from 0.005 to 1 Å-1, where $q = 4\pi\sin\theta/\lambda$, 2θ denoting the scattering angle and λ the x-ray wavelength. The wavelength was set at 1.033 Å during the measurements. Scattered X-ray intensities were measured using a Pilatus 2M (DECTRIS Ltd) detector for SAXS and Pilatus 300K for WAXS. SAXS/WAXS measurements were performed on solutions of 1D banded nanowires from co-assembly of CP++ and TL nanosheets in Tris buffer (5 mM, pH 8.0) at room temperature. 5% glycerol was added to reduce the radiation damage. A quartz capillary flow cell (1.5 mm diameter) was employed to prevent radiation damage to samples. Twenty images

were collected for each sample and buffer. The 2-D scattering images were converted to 1-D SAXS curves through azimuthally averaging after solid angle correction and then normalizing with the intensity of the transmitted x-ray beam, using the software package at beamline 12-ID-B. The 1-D scattering curves of the samples were averaged and subtracted with the background measured from the Tris buffers with 5 % glycerol.

6.4.9 DNA Sequences for TL and Brick Structures.

Full oligonucleotide sequences for all DNA origami structures can be found in the supplemental information addendum for Jiang et al. 2017.

6.5 References

1. Seeman, N. C., Nucleic-Acid Junctions and Lattices. *J Theor Biol* **1982,** *99* (2), 237-247.

2. Ke, Y. G.; Sharma, J.; Liu, M. H.; Jahn, K.; Liu, Y.; Yan, H., Scaffolded DNA Origami of a DNA Tetrahedron Molecular Container. *Nano Lett* **2009,** *9* (6), 2445-2447.

3. Liedl, T.; Hogberg, B.; Tytell, J.; Ingber, D. E.; Shih, W. M., Self-assembly of three-dimensional prestressed tensegrity structures from DNA. *Nat Nanotechnol* **2010,** *5* (7), 520-524.

4. He, Y.; Tian, Y.; Ribbe, A. E.; Mao, C. D., Highly connected two-dimensional crystals of DNA six-point-stars. *J Am Chem Soc* **2006,** *128* (50), 15978-15979.

5. Benson, E.; Mohammed, A.; Gardell, J.; Masich, S.; Czeizler, E.; Orponen, P.; Hogberg, B., DNA rendering of polyhedral meshes at the nanoscale. *Nature* **2015,** *523* (7561), 441-U139.

6. Kye, M.; Lim, Y. B., Reciprocal Self-Assembly of Peptide-DNA Conjugates into a Programmable Sub-10-nm Supramolecular Deoxyribonucleoprotein. *Angew Chem Int Edit* **2016,** *55* (39), 12003-12007.

7. Mou, Y.; Yu, J. Y.; Wannier, T. M.; Guo, C. L.; Mayo, S. L., Computational design of co-assembling protein-DNA nanowires. *Nature* **2015,** *525* (7568), 230-+.

8. Ni, R.; Chau, Y., Tuning the Inter-nanofibril Interaction To Regulate the Morphology and Function of Peptide/DNA Co-assembled Viral Mimics. *Angew Chem Int Edit* **2017,** *56* (32), 9356-9360.

9. Noble, J. E.; De Santis, E.; Ravi, J.; Lamarre, B.; Castelletto, V.; Mantell, J.; Ray, S.; Ryadnov, M. G., A De Novo Virus-Like Topology for Synthetic Virions. *J Am Chem Soc* **2016,** *138* (37), 12202-12210.

10. Praetorius, F.; Dietz, H., DNA NANOTECHNOLOGY Self-assembly of genetically encoded DNA-protein hybrid nanoscale shapes. *Science* **2017,** *355* (6331).

11. Ruff, Y.; Moyer, T.; Newcomb, C. J.; Demeler, B.; Stupp, S. I., Precision Templating with DNA of a Virus-like Particle with Peptide Nanostructures. *J Am Chem Soc* **2013,** *135* (16), 6211-6219.

12. de la Escosura, A.; Janssen, P. G. A.; Schenning, A. P.; Nolte, R. J. M.; Cornelissen, J. J. L. M., Encapsulation of DNA-Templated Chromophore Assemblies within Virus Protein Nanotubes. *Angew Chem Int Edit* **2010,** *49* (31), 5335-5338.

13. Jiang, T.; Xu, C. F.; Liu, Y.; Liu, Z.; Wall, J. S.; Zuo, X. B.; Lian, T. Q.; Salaita, K.; Ni, C. Y.; Pochan, D.; Conticello, V. P., Structurally Defined Nanoscale Sheets from Self-Assembly of Collagen-Mimetic Peptides. *J Am Chem Soc* **2014,** *136* (11), 4300-4308.

14. Jiang, T.; Vail, O. A.; Jiang, Z. G.; Zuo, X. B.; Conticello, V. P., Rational Design of Multilayer Collagen Nanosheets with Compositional and Structural Control. *J Am Chem Soc* **2015,** *137* (24), 7793-7802.

15. Jiang, T.; Meyer, T. A.; Modlin, C.; Zuo, X. B.; Conticello, V. P.; Ke, Y. G., Structurally Ordered Nanowire Formation from Co-Assembly of DNA Origami and Collagen-Mimetic Peptides. *J Am Chem Soc* **2017,** *139* (40), 14025-14028.

16. Cullity, B. D. S., S. R., *Elements of X-Ray Diffraction*. Prentice Hall: Upper Saddle River, NJ, 2001; Vol. 3rd Edition.

17. Patterson, A. L., The Scherrer formula for x-ray particle size determination. *Phys Rev* **1939,** *56* (10), 978-982.

18. Terao, K.; Mizuno, K.; Murashima, M.; Kita, Y.; Hongo, C.; Okuyama, K.; Norisuye, T.; Bachinger, H. P., Chain Dimensions and Hydration Behavior of Collagen Model Peptides in Aqueous Solution: [Glycyl-4(R)-hydroxyprolyl-4(R)-hydroxyproline](n), [Glycylprolyl-4(R)-

hydroxyproline](n), and Some Related Model Peptides. *Macromolecules* **2008,** *41* (19), 7203-7210.

19. Ke, Y. G.; Ong, L. L.; Shih, W. M.; Yin, P., Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science* **2012,** *338* (6111), 1177-1183.

Chapter VII. Conclusion and Outlook

7.1 Conclusion

Peptide self-assembly remains an efficient, yet variable and unpredictable, method for the engineering and construction of complex supramolecular systems. The Form projects described in this thesis portray the inherent labile and fluctuant nature of many self-assembling peptide systems. The Form I and II peptides[1] demonstrate the delicate and sensitive response of helical nanotubes to both minute amino acid substitutions, as well as environmental changes, as these peptides only self-assemble over a small pH range centrally focused around pH 4.0. Form IA corroborates the highly sensitive quality of the superhelical Form assemblies. Form IA takes the level of plasticity inherent to these systems even further, demonstrating a novel concentration-dependent switch from hypothesized superhelical nanotubes to complex, crystalline structures. Importantly, the Sol series demonstrates an increase in the predictability and designabilty of superhelical nanotubes compared to the Form series. While the Form series represent unique *de novo* systems, the resultant structures deviated significantly from what was hypothesized. The computational approach for the Sol peptides improves the designability of self-assembling peptide nanotubes, with the resultant supramolecular structures demonstrating structural parameters in accordance with what is hypothesized. Lastly, it is important to push for engineered systems which rival the structural complexity that is often observed in nature. To this end, the combination of collagen mimetic peptide with DNA origami has proven to be an efficient and reliable method for creating complex self-assembling systems of multiple species.[2]

7.2 Outlook

Research into self-assembling peptide systems is poised for great advancement in the near future. The studies contained within this thesis add to an ever-increasing library of self-assembling peptide systems. This growth helps in the universal understanding of the highly complex, and often unpredictable, principles which govern how peptide and protein interfaces interact. Consistent improvements in cryo-electron microscopy,[3-5] combined with the ever-growing computational capabilities in the field,[6-7] greatly increases the number of tools available for the engineering and production of novel, *de novo* self-assembling peptide materials. Indeed, given the documented level of variability in the relationship between primary amino acid sequence and resultant quaternary (and often tertiary) structure, computational modeling and computational design will likely become a critical cornerstone in the future design and creation of peptide systems (and biomaterials in general), especially as researchers strive to push the limits of complexity in the attempts of rivaling that which is nearly ubiquitous in nature. Logical next steps in the field will surely include not only programming highly complex structures, but also highly complex functions. Already, researchers are seeing the immense benefit of employing peptides in the pursuit of biological 'nanomachines,'[8] and as the ability to design and select for unique and complicated peptide scaffolds increases, so will the capacity for new and novel functions. It is truly an exciting time to be active in the field of peptide research.

7.3 References

1. 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.

2. Jiang, T.; Meyer, T. A.; Modlin, C.; Zuo, X. B.; Conticello, V. P.; Ke, Y. G., Structurally Ordered Nanowire Formation from Co-Assembly of DNA Origami and Collagen-Mimetic Peptides. *J Am Chem Soc* **2017,** *139* (40), 14025-14028.

3. Murata, K.; Wolf, M., Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochim Biophys Acta* **2018,** *1862* (2), 324-334.

4. Razi, A.; Britton, R. A.; Ortega, J., The impact of recent improvements in cryo-electron microscopy technology on the understanding of bacterial ribosome assembly. *Nucleic acids research* **2017,** *45* (3), 1027-1040.

5. Frank, J., Advances in the field of single-particle cryo-electron microscopy over the last decade. *Nature protocols* **2017,** *12* (2), 209-212.

6. Zheng, F.; Zhang, J.; Grigoryan, G., Tertiary structural propensities reveal fundamental sequence/structure relationships. *Structure* **2015,** *23* (5), 961-971.

7. Zheng, F.; Grigoryan, G., Simplifying the Design of Protein-Peptide Interaction Specificity with Sequence-Based Representations of Atomistic Models. *Methods in molecular biology (Clifton, N.J.)* **2017,** *1561*, 189-200.

8. Nagamune, T., Biomolecular engineering for nanobio/bionanotechnology. *Nano convergence* **2017,** *4* (1), 9.