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:

_____ _____

Spencer Hughes Date

Rational and Computational Design of Superhelical Protein Nanotubes

By

Spencer A. Hughes

Doctor of Philosophy

Chemistry

_____

Dr. Vincent P. Conticello

Advisor

_____

Dr. David Lynn

Committee Member

_____

Dr. Stefan Lutz

Committee Member

Accepted:

_____

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

_____

Date

Rational and Computational Design of Superhelical Protein Nanotubes

By

Spencer A. Hughes

B.S. Chem., University of Georgia, 2014

Advisor: Vincent P. Conticello, PhD

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

2019

Abstract

Rational and Computational Design of Superhelical Protein Nanotubes

By Spencer Hughes

The design of peptide- and protein-based nanomaterials with high fidelity has long been a challenge in structural biology and materials science alike. The translation from the amino acid sequence to the folded and assembled structure is rarely facile, so design rules have been tabulated for the simplest of super-secondary structural elements (SSEs), such as coiled-coils and certain β-sheet assemblies. However, it has been shown that conservative mutations in these scaffolds has led to significant structural deviation. The search for a more predictable biomaterials scaffold has led to the characterization of Tandem Repeat Proteins (TRPs). The folding behavior of TRPs can be understood as a series of noncovalent lateral interactions between adjacent, nearly identical SSEs; in fact, the sequences of each SSE within a TRP can be aligned and statistically analyzed to derive consensus sequences for each TRP family. Consensus sequences serve as a mutagenesis guide; highly conserved positions should be held constant, whereas hypervariable positions are open for redesign. Using this strategy, we have developed the first TRP-based nanotubes, using the LRV and HEAT TRP families; we designed a single SSE from each family to self-assemble in a superhelical fashion. We used structural parameters extrapolated from the parent crystal structures to evaluate the efficacy of our design strategy. Cryo-EM was used to generate atomic models of the peptide nanotubes. In the case of LRV, the helical pitch, handedness, and number of constituent SSEs present in each superhelical turn differed from the crystal structure-based parameters. The HEAT-based design was much more effective, with the helical pitch, handedness, and number of SSEs per superhelical turn closely matching the predicted structure. A dimeric peptide derivative of each of these designs was then evaluated using the same criteria; interestingly, the LRV_dimer assembled in a similar fashion to the parent peptide and the HEAT_dimer did not.

A hexameric concatemer of the HEAT peptide sequence was bacterially expressed, purified and assembled. Low- and medium-resolution techniques (CD and TEM, SAXS and STEM, respectively) were used to compare the resultant nanotubes to those generated from the peptide constructs. At medium-resolution, the hexameric and monomeric assemblies were indistinguishable; however, the hexameric nanotube atomic model could not be solved using cryo-EM due to inherent plasticity of the structure. The outside of the HEAT nanotubes was functionalized using a SpyTag:SpyCatcher genetic fusion technique; an octameric HEAT protein was conjugated with SpyTag (ST_HEAT), and the fluorescent protein mCherry was conjugated with SpyCatcher (mCherry_SC). Post-assembly, the ST_HEAT nanotubes were introduced to the mCherry_SC construct, resulting in the functionalization of the convex surface of the HEAT nanotubes. Preliminary structural data obtained from fluorescence microscopy and CLEM indicated the successful functionalization of the nanotube.

Computational design has long been paired with rational design strategies to generate helical filaments. Three generations of computational design were applied to a helical hairpin motif, resulting in a more refined approach. Each of these design strategies was evaluated based on the solubility of the sequences, propensity to form helical filaments, and fidelity to the computational model. The most effective strategy for generating helical filaments was to redesign a known assembly, rather than building one from scratch.

Rational and Computational Design of Superhelical Protein Nanotubes

By

Spencer A. Hughes

B.S. Chem., University of Georgia, 2014

Advisor: Vincent P. Conticello, PhD

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

2019

Acknowledgments

I would like to express my respect and gratitude for my advisor, Dr. Vincent Conticello. He offered me my first research opportunity, as an undergraduate summer student. His advice and support undoubtedly helped me stay on the path towards graduate school and conducting research in the long-term. His willingness to set challenging goals and open my eyes to new opportunities has helped me to have a successful five years in graduate school. I look forward to many more years of picking his brain for new, sometimes far-fetched ideas. I am similarly indebted to my committee members, Drs. David Lynn and Stefan Lutz for encouraging me to present my research clearly and concisely, as well as providing their invaluable insights to my projects. I would also like to thank my coworkers and friends from the Conticello group during my time here: Dr. I-Lin Wu, Dr. Chunfu Xu, Dr. Tao Jiang, Dr. Elizabeth Magnotti, Dr. Charles Modlin, Dr. Rebecca Bartlett, Dr. Andre Merg, Shengyuan Wang, Avi Tuachi, Ordy Gnewou, Duong Nguyen, Jessalyn Rogers, and Gavin Touponse. I would like to highlight the contributions of Dr. Becky Bartlett, who performed the expression work from Chapter III, as well as some initial characterization experiments. Additionally, I would like to highlight Shengyuan Wang's contribution to Chapter II, as he aided with synthesis and purification of the peptides. Without my lab family, the stresses of graduate school may have been insurmountable. I would also like to thank my friends from the chemistry department, with whom I commiserated and celebrated; Tom Pickel, Matt Jenkins, Austin Baker, Sam Iamurri, Dan Wildman, Kevin Sullivan, Adam Aycock, Kevin Hannon, Jeff Schriber, and Roger Park.

I am deeply grateful to the members of the Robert P. Apkarian Integrated Electron Microscopy Core: Dr. Elizabeth Wright, Hong Yi, Art McCanna, and Jeanette Taylor, as well as

the members of Dr. Wright's research group: Dr. Cheri Hampton, Dr. Zunlong Ke, Dr. Joshua Strauss, and Dr. Rebecca Dillard. Together, all of these people helped me to learn a great deal about electron microscopy, and perhaps even more importantly, showed me how much more I have to learn. Additionally, I would like to thank Dr. Emily Weinert and Dr. Anil Mehta for very helpful conversations regarding research and navigating the murky waters of academia.

Our collaborators have been a fundamental part of every project I've worked on. Thanks to Dr. Gevorg Grigoryan (Dartmouth College), Dr. Edward Egelman (University of Virginia), Dr. Derek Woolfson (University of Bristol, United Kingdom), Dr. Fabio Parmeggiani (University of Bristol, United Kingdom), Dr. Xiaobing Zuo (Argonne National Laboratory), Dr. Elizabeth (University of Wisconsin Madison) and Dr. Joseph Wall (Brookhaven National Laboratory). Without these dedicated collaborators, my scientific worldview would have been greatly limited, as would the scope of my experiments. I cannot overstate the importance of these collaborations to the success of my projects.

Finally, I would like to thank my family. Thank you to my parents, who always emphasized the importance of education, and prioritized diligent work over raw intelligence. Thank you to my brother and sister, who kept me humble, and made me laugh. Thank you to my grandparents, Connie and Gene (Rest in Peace), for being proud of me, even when I faltered. Thank you to my nephews for making me smile. Finally, I want to express my love and gratitude for the love of my life: my wife Darrelyn. We've been together nearly the entire time I've been in graduate school, and I'm sure she looks forward to me being done!

Table of Contents

List of Figures

List of Tables

Chapter I: Introduction

1.1 The Significance of Self-Assembly

Biological self-assembly is ubiquitous in all forms of life. Nearly every biochemical process involves biological self-assembly in some fashion. Many of these self-assembly 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. The study of nature's most interesting and complex self-assembly machinery has led to the founding of the fields of virology and genetic engineering. Emerging technologies have recently been developed that have allowed us to probe these structures at or near atomic resolution; new sampling techniques for protein crystallography as well as direct electron detectors in cryo-electron microscopy. Designed helical assemblies from well-studied protein folds (α-helices and β-sheets) which focus on the interplay between *de novo* design and inspiration from native proteins have been particularly promising. Materials that comprise unnatural building blocks such as oligourea monomers and peptoid moieties exist at the interface of chemistry, biology, physics, and material science; these structures have expanded the biochemical toolbox, allowing us to design systems that are both biomimetic and bioorthogonal.

Nature uses self-assembly of helical molecules to produce complex nanomachines capable of enacting physical and chemical changes in astonishingly high yields, 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.

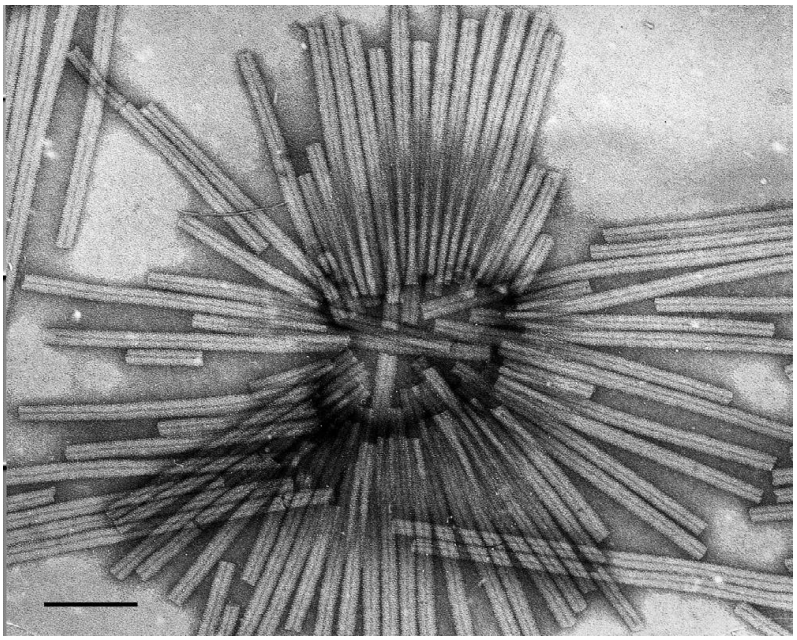


Figure 1.1 Negative-stained TEM of fully assembled TMV particles[1]. Scale bar is 100 nm. Reprinted from "Archive of Biochemical Biophysics, 581, Harris, J. R., Transmission electron microscopy in molecular structural biology: A historical survey, 3-18." Copyright (2015) with permission from Elsevier.

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.* These are the systems for which near-atomic resolution has been achieved, though increasingly complex structures are being deposited into the protein data bank (PDB) and electron microscopy data bank (EMDB) every day.

1.2 Native Helical Assemblies

One such self-assembling structure is Tobacco Mosaic Virus (TMV), the first virus ever crystallized [2]. TMV has been the subject of extensive scientific interest due to the wide range of host organisms that it infects, its lack 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 (Figure 1.1). 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)[3]. This assembly mode better accounts for the observed rapid self-assembly of TMV than a single subunit addition model.

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[4]. 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.

As it stands now, TMV is used as a mass-per-length standard in negative-stain scanning transmission electron microscopy (STEM)[5]. 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[6-9]. 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[10]. 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)[11-18]. 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[19]. Briefly, the wild type TMV sequence was point-mutated so that one exposed cysteine per coat protein was displayed on the

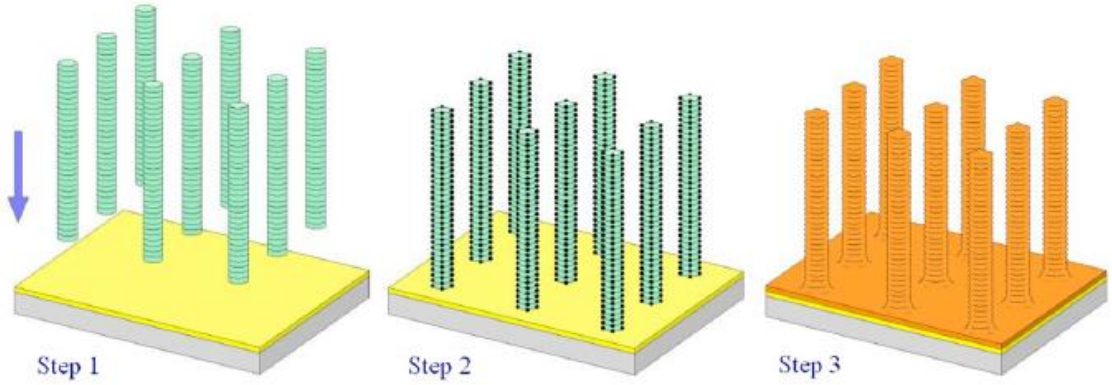


Figure 1.2 Schematic of TMV1cys incorporation into a MEMs device. Step 1 shows the conjugation of TMV1cys to a gold surface. TMV1cys is activated with palladium in Step 2 and coated with nickel in Step 3. Reprinted with permission from "Gerasopoulos, K.; McCarthy, M.; Royston, E.; Culver, J. N.; Ghodssi, R., Nanostructured nickel electrodes using the Tobacco mosaic virus for microbattery applications. *J Micromech Microeng* 2008, *18* (10). Copyright (2008) IOP Publishing.

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 (Figure 1.2).

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.

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[20]. 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 confirm this.

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* 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)[21]. 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 (Figure 1.3). 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.

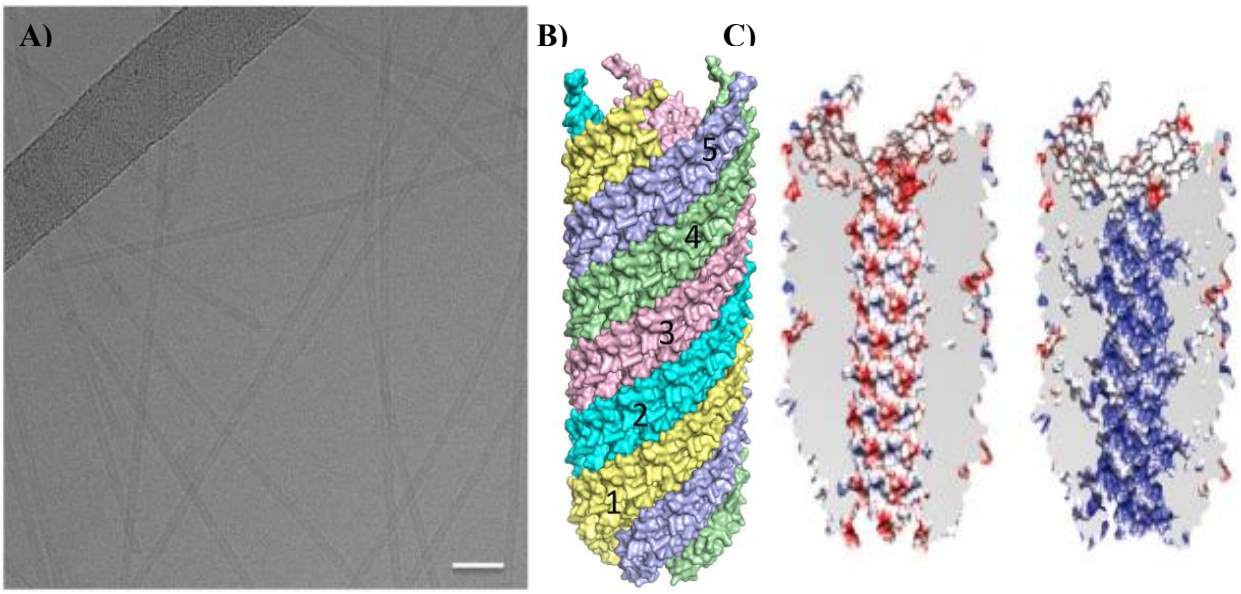


Figure 1.3. Structure of the sex pilus. a) Cryo-electron micrograph of the sex pilus (scale bar is 40 nm). b) Pilin assembly constitutes the outer surface of the pilus. c) Difference in electrostatic potential between the lumen of the pilus with and without the phospholipid core. Reprinted via the Creative Commons Attribution License from "Structure of the Bacterial Sex F Pilus Reveals an Assembly of a Stoichiometric Protein-Phospholipid Complex. Costa, T. R. D.; Ilangovan, A.; Ukleja, M.; Redzej, A.; Santini, J. M.; Smith, T. K.; Egelman, E. H.; Waksman, G. Cell, 2016¸166 (6), 1436-1444."

1.3 Non-Native Helical Assemblies of Synthesized Peptides

The Conticello group and Egelman group have engineered several alpha-helical peptide nanotube assemblies [22-24]. The first of such systems, published in 2013, is the peptide 7HSAP1.

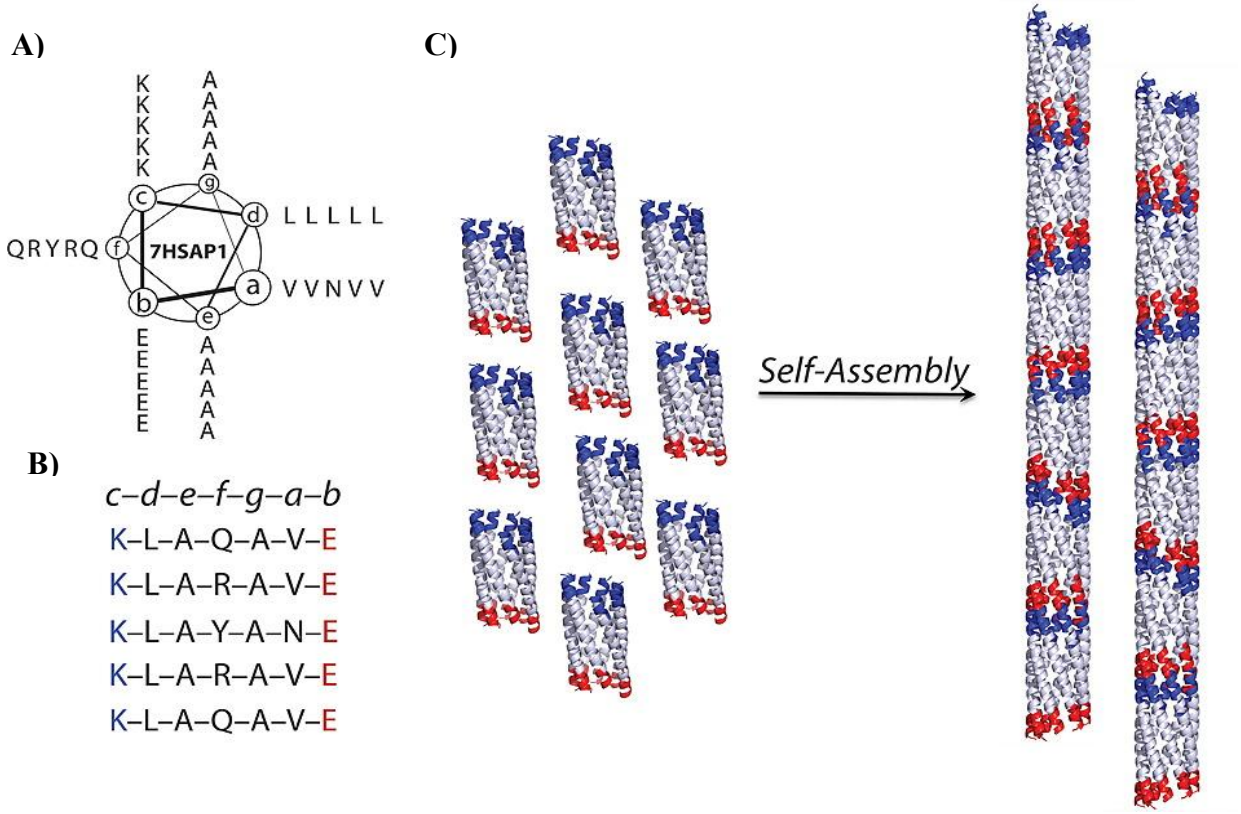


Figure 1.4. Design and assembly of 7HSAP1. (a) Helical wheel representation of the 7HSAP1 amino acid sequence derived from GCN4-pAA. (b) Linear representation of the amino acid sequence of 7HSAP1. (c) Graphical representation of the self-assembly process for 7HSAP1 wherein charge complementary termini 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. Reprinted with permission from "Rational Design of Helical Nanotubes from Self-Assembly of Coiled-Coil Lock Washers. 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. J Am Chem Soc 2013, 135 (41), 15565-15578." Copyright (2013) American Chemical Society.

Xu et al. utilized the *de novo* leucine zipper-inspired GCN4-Paa as a design scaffold for sequence manipulation. Based on the leucine zipper motif from GCN40 (a transcription factor in *S. cerevisiae)*, GN4-Paa forms a bundle of alpha-helices with an inner channel measuring 7Å 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 [24]. 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 register offset by one amino acid. This offset increases the surface area available for complementary interactions between 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 [24]. The amino acid sequence 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 1.4 shows STEM data of 7HSAP assembled in the aforementioned conditions and confirms successful fibrillar assembly. STEM analysis, further corroborated by cryo-TEM, indicates a fibril diameter of roughly 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 [25] 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.

Of particular interest to the field of helical peptide self-assembly is the design of *de novo* coiled-coil bundles and nanotubes with large lumina. In nature, such bundles (>7 subunits) are readily observed; however, *de novo* design of such structures remains a challenge. In an answer to this challenge, the Conticello group designed two *de novo* alpha-helices using prior structural principles established by Walshaw and Woolfson[26], and demonstrated that that both sequences self-assemble into structurally discrete nanotubes. The Egelman group was critical in solving the structures of both Forms at near atomic resolution. These new sequences (Form I and Form II) were based on the sequence and folding motifs of the 12-helix bundle previously witnessed in the tolC homotrimer. Tolc acts as a reasonable scaffold for sequence manipulation as it features more than seven subunits, making it a viable foundation for the creation of nanotubes with large lumina.

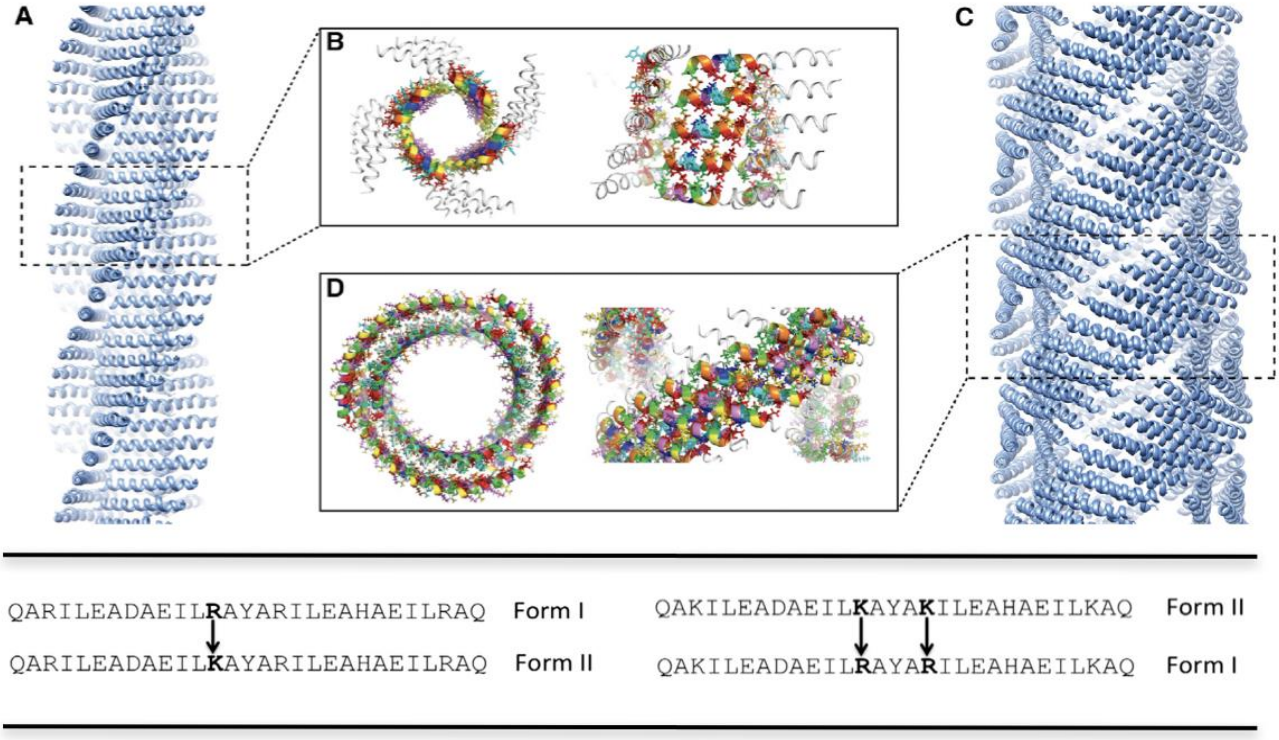


Figure 1.5. Sequences and assembly of Form I and Form II: [top] atomic models for knobs-into-holes (KIH) packing SOCKET analysis for Forms I and II. [bottom] Primary amino acid sequences for Forms I and II. Highlighted are the necessary mutations needed for conversion from Form I to II and vice versa. (a) graphical representation for Form I using SOCKET analysis. (b) graphical representation of KIH packing at the C-terminus of Form I. Both top-down and side-on views are displayed. (c) graphical representation of atomic model for Form II using SOCKET analysis (d) graphical representation of KIH packing observed in Form II, wherein such packing occurs exclusively within the small wall. No KIH packing is observed between inner and outer lumen. Both top-down and side-on views are displayed. Reprinted with permission from "Structure, 23 (2), 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, 280-289." Copyright (2015), with permission from Elsevier.

hydrophobicity into the *a/d* and *c/f* positions of the two sequences (form I and form II), creating a 'stripe' of hydrophobic residues along the helix. Alanine residues were chosen to populate the outer *a* and *f* positions with the hypothesis that Alanine's smaller sidechain (when compared to the equivalent residues in TolC) would limit steric hindrance at the hydrophobic interface between two alpha-helices, potentially affording nanotubes with larger lumen. To promote anti-parallel helix packing, charge complementary residues were engineered into the heptad of both *de novo* forms at the *b* and *e* positions. Distinguishing the two forms is a substitution of Arg residues with Lys residues in Forms I and II respectively (see Figure 1.5). Beyond these substitutions, the remainder of the sequence across both forms was held constant.

Both forms assemble in aqueous conditions at pH 4, with each form demonstrating unique morphological identities. From transmission electron microscopy and small angle x-ray scattering, diameters for Forms I and II were found to be 6 and 12 nm respectively. Direct electron detection cryo-EM coupled with an iterative helical real space reconstruction algorithm [27] produced near atomic scale structural data for Forms I and II.

As is seen in numerous self-assembling peptide systems, both forms self-assemble in a fashion that places the polar residues at the inner lumen and exterior face of the nanotubes, while burying the hydrophobic residues in a form of hydrophobic collapse to avoid being placed in a solvent-facing fashion. Interestingly, the macromolecular assemblies observed for Forms I and II are significantly dissimilar to the Tolc scaffold (an alpha-cylinder) [28]. Moreover, each is distinct from the structural models postulated by Walshaw and Woolfson [29]. Form I self-assembles into a unilaminar nanotube with four-fold symmetry, wherein individual helices pack perpendicular to the fiber axis (see Figure 1.5). Form II self-assembles into a bilaminar nanotube with three-fold symmetry, wherein three helix bilayers stack at an angle just beyond perpendicular to the fiber

axis. Therefore, when analyzing a cross-section of Form II, individual helices pack at a slight angular incline when compared to the cross-sectional plane.

The salient point in this study is that the quaternary structures of Form I and II are intimately connected with linear amino acid sequence. Indeed, Egelman *et al.* demonstrate that quaternary identity can be sensitive enough that single residue substitution can drive conversion of quaternary structure from that of Form I to that of Form II. It can be surmised then, that the resultant macromolecular nanotube assembly is not necessarily robust in sequence space, proving once again how closely connected sequence and structure can be in peptide systems.

The relatively conservative, singular residue substitution (R13K) caused a shift in resultant nanotube morphology from that of Form I to that of Form II. A second set of conservative substitutions (K13R, K17R) was sufficient to reverse nanotube morphology back to that of Form I from Form II [22]. Responsible for this morphological interchange are the critical interactions (or lack thereof) of the Arg13 and Arg17 residues with the C-terminal residues of a neighboring helix from the next stack. Importantly, this system is an excellent example of minor primary sequence modifications inducing significant shifts in quaternary structure. As such, it should be noted that future advancements in peptide-based, alpha-helical biomaterials will inevitably necessitate improved methodologies in order to create higher order structures with *increased* predictability in resultant quaternary structure from primary sequence.

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

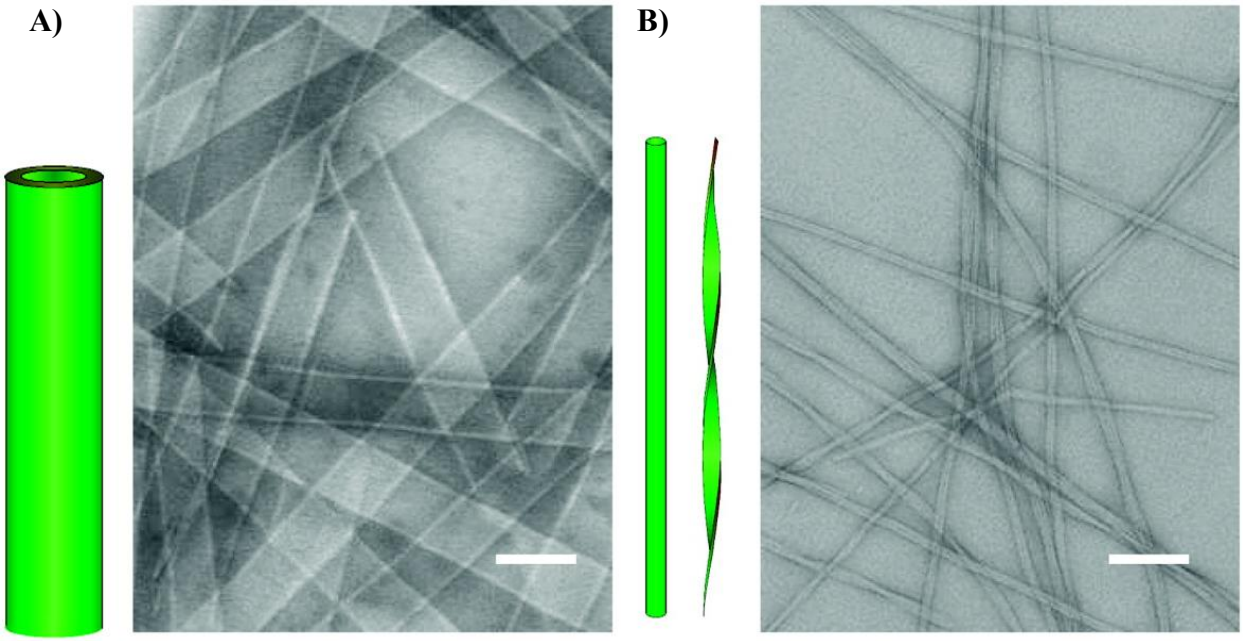


Figure 1.6. Assembly of Aβ (16-22) into tubes (a) under acidic conditions and into fibers (b) under neutral pH conditions. Scale bars = 100 nm. Reprinted with permission from "Cross-strand pairing and amyloid assembly. Liang, Y.; Pingali, S. V.; Jogalekar, A. S.; Snyder, J. P.; Thiyagarajan, P.; Lynn, D. G. Biochemistry-Us 2008, 47 (38), 10018-10026." Copyright (2008) American Chemical Society.

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[30-31]. 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 (Figure 1.6).

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[32]. 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 of iodine.

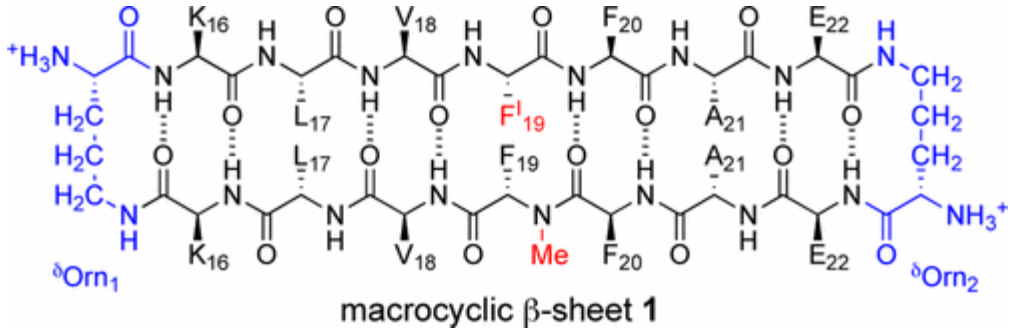


Figure 1.7. The synthesized Aβ (16-22) macrocycle. Reprinted with permission from "X-ray Crystallographic Structure of a Giant Double-Walled Peptide Nanotube Formed by a Macrocyclic beta-Sheet Containing A beta(16-22). Chen, K. H.; Corro, K. A.; Le, S. P.; Nowick, J. S. J Am Chem Soc 2017, 139 (24), 8102-8105." Copyright (2017) American Chemical Society.

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 1.7).

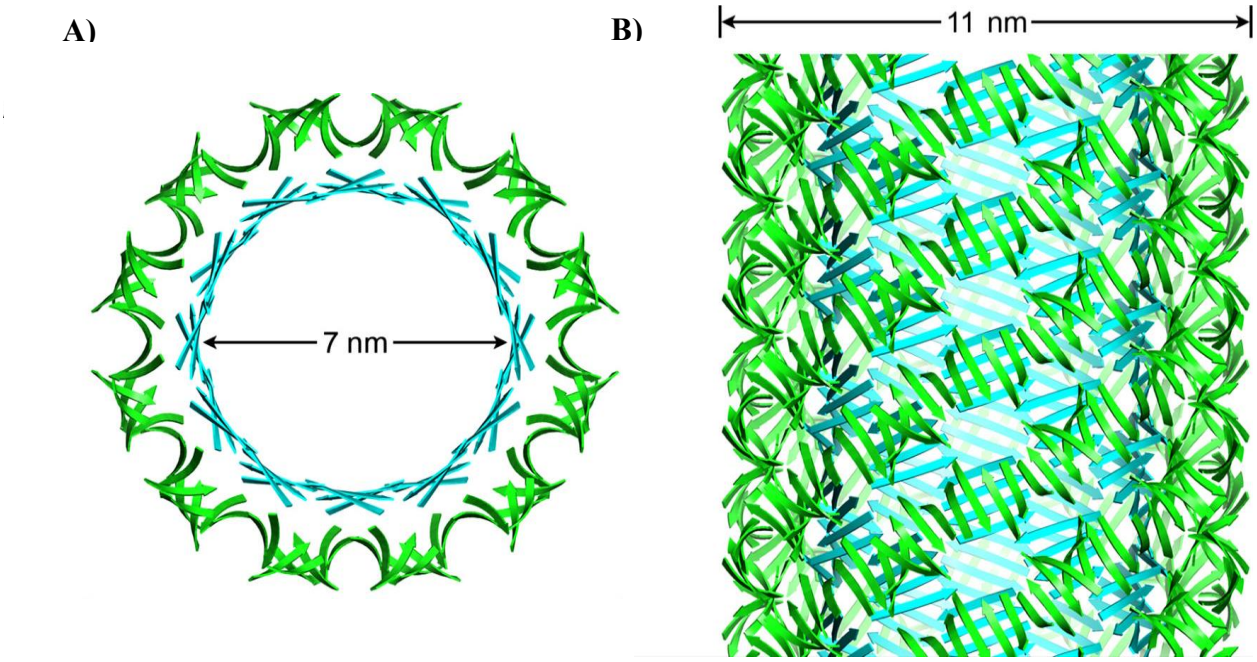


Figure 1.8 X-ray crystal structure of the macrocycle nanotube. (a) Top view highlighting the large inner lumen of the double-walled nanotube. (b) Side view of the nanotube, illustrating the packing of the dimeric and tetrameric structures against each other. Reprinted with permission from "X-ray Crystallographic Structure of a Giant Double-Walled Peptide Nanotube Formed by a Macrocyclic β-Sheet Containing Aβ(16-22). Chen, K. H.; Corro, K. A.; Le, S. P.; Nowick, J. S. J Am Chem Soc 2017, 139 (24), 8102-8105." Copyright (2017) American Chemical Society.

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 (Figure 1.8). The tetramers, on the other hand, 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.

1.4 Helical Assemblies Beyond Biomacromolecules

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 [33], (b) follow unique folding rules distinct from protein folding [34], 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 2.5

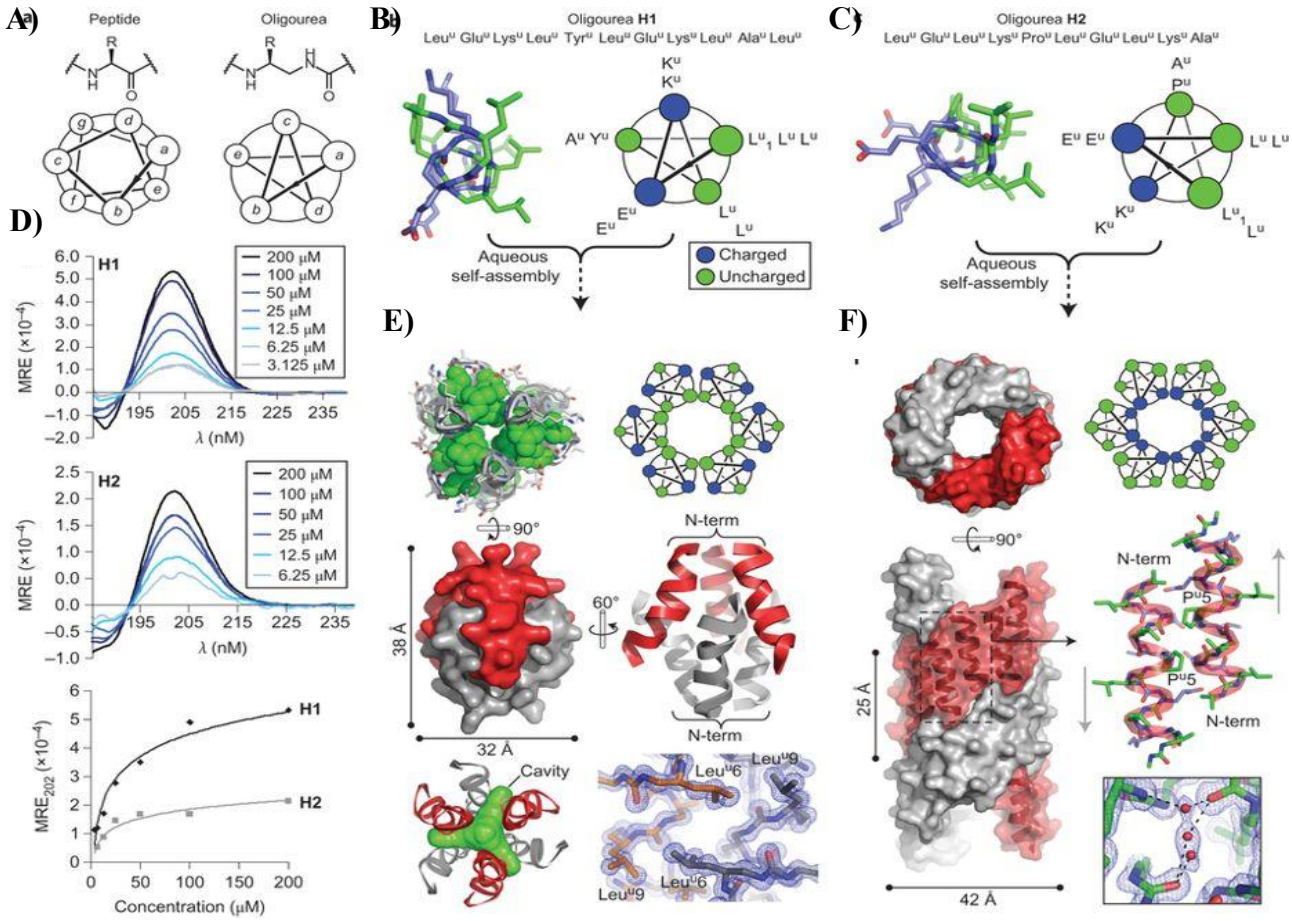


Figure 1.9. (a) Helical wheel diagrams delineating the geometric differences between alpha-peptides and aliphatic oligoureas. (b) and (c) primary amino acid sequence, top-down view of the crystal structure, and helical wheel indicating side-chain population at each position of the 2.5 helix for H1 and H2, respectively. (d) Variable concentration circular dichroism analysis of both peptides in water, shown in mean residue ellipticity. Crystal structures of the channel assemblies of H1 (e) and H2 (f) are juxtaposed to illustrate their differences. Reprinted by permission from "Shaping quaternary assemblies of water-soluble non-peptide helical foldamers by sequence manipulation, 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." Copyright (2015) Nature Chemistry.

helix, while H2 features Leu^u at the *a* and *c* positions, with Ala^u and Pro^u at the *d* position (see

Figure 1.9). H2 effectively extends the hydrophobic face of the 2.5 helix by one position, as compared to H1. 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 1.9). 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 governing their packing: (Figure 1.9a) Leu^{u}-Leu^{u} hydrophobic packing, (Figure 1.9b) reciprocal hydrogen bonding between the $Lys^{u}8$ side-chain amine and the C-terminal $Leu^{u}11$ uncapped carbonyl group, (Figure 1.9c) 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 for H1 were obtained via x-ray 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 H2 demonstrates hydrophobic Leu^{u}, Ala^{u} and Pro^{u} residues populating 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 for H2 was obtained via crystallography and resolved to 1.40 Å. H2 indicates that an extended, contiguous hydrophobic face is necessary for selective tubular formation rather than helical bundles. To corroborate this hypothesis, Collie *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.

Collie *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 sequence and structure. 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[35-37] and ion conductance across a membrane[38].

In addition to oligourea foldamers, peptoids represent one of the most promising non-protein moieties capable of forming extended helical assemblies[39]. . Whereas peptides have side-chains connected to the α-carbon, peptoids instead have side-chains attached to the amide

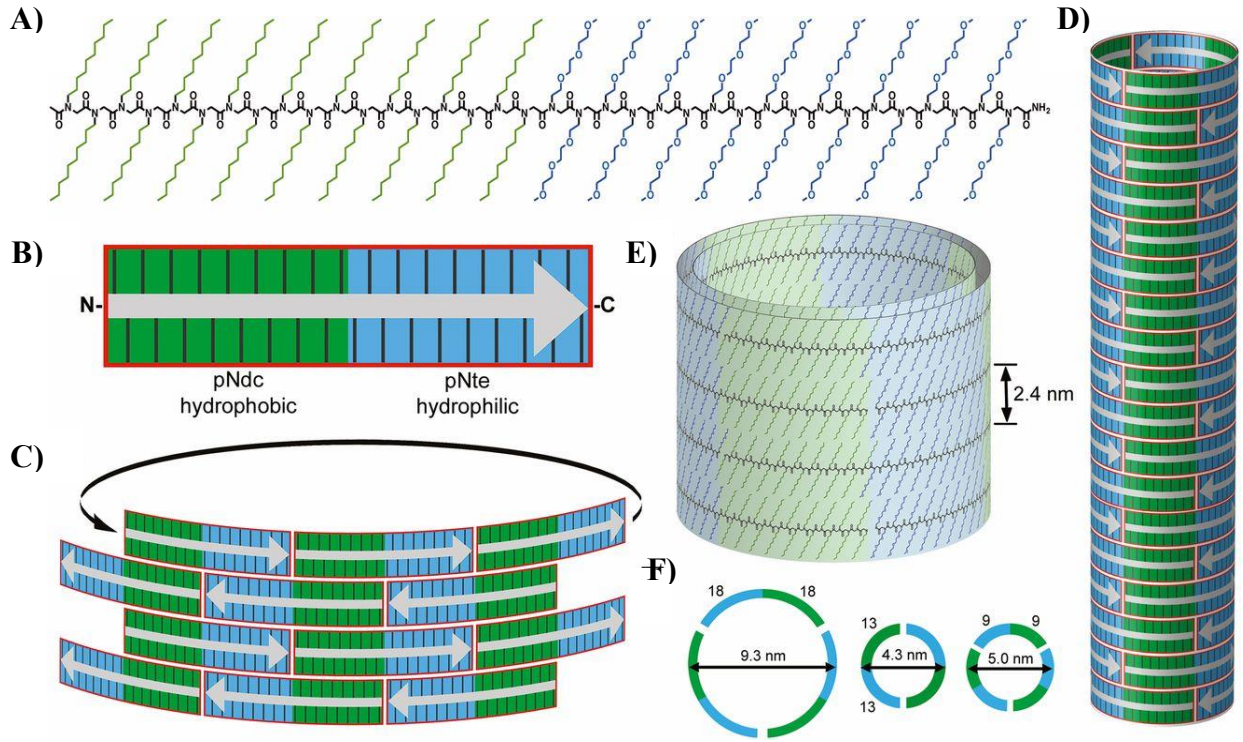


Figure 1.10. Structural schematic of the peptoid nanotubes. (a) Chemical structure of $pNdc_{18}$-b-$pNte_{18}$ (b) organization of the peptoid diblock, with the hydrophobic pNdc block at the N-terminus and the pNte hydrophilic) block at the C-terminus. (c) The peptoid chains align anti-parallel, with the hydrophobic domains and hydrophilic domains stacking on top of themselves. (d) Full side-on view of the peptoid nanotube, held together by longitudinal side-chain interactions. (e) The nanotubes comprise many stacked rings, 2.4 nm in height, with the length of the peptoid chain determining the tube diameter (f).

nitrogen[40-41]. 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[42-43], 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 (Figure 1.10)[44]. 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 neighboring subunits. The side chains of the peptoid backbone were shown to extend into the pore of the nanotubes, making these an excellent material for potential filtration applications.

1.5 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.

1.6 References

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

2. 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.

3. 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.

4. 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.

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

6. 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.

7. 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.

8. 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.

9. 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.

10. 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.

11. 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.

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

17. 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.

18. 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.

19. 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).

20. 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.

21. 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-+.

22. 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.

23. 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.

24. 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.

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

26. 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.

27. 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.

28. V, K.; Sharff, A.; Koronakis, E.; Luisi, B.; Hughes, C., Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* **2000,** *405* (6789), 914-919.

29. Calladine, C. R.; Sharff, A.; Luisi, B., How to untwist an alpha-helix: Structural principles of an alpha-helical barrel. *J Mol Biol* **2001,** *305* (3), 603-618.

30. 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.

31. 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.

32. 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.

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

34. 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.

35. 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.

36. 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.

37. 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.

38. 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.

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

40. 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.

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

42. 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.

43. 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.

44. 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. Ambidextrous Helical Nanotubes from Self-Assembly of Designed Helical Hairpin Motifs

Published in PNAS, July 2019:

Hughes, S. A.; Wang, F.; Wang, S.; Kreutzberger, M. A. B. ; Osinski, T.; Orlova, A.; Wall, J. S.; Zuo, X.; Egelman, E. H.; Conticello, V. P. Ambidextrous Helical Nanotubes from Self-Assembly of Designed Helical Hairpin Motifs.

2.1 Self-assembly of Phage-Mimetic Peptides

Self-assembly is a ubiquitous process in biological systems. The functional macromolecular machines of living organisms commonly arise from homomeric or heteromeric self-association that involves selective recognition at the inter-molecular interfaces between protomers (i.e., structural subunits). The most structurally economical method of constructing such assemblies is based on the application of a helical symmetry operation.[1-2] In this mode of assembly, self-association between protomers occurs such that structurally adjacent subunits are related by rotation through a characteristic twist angle (ϕ) with a commensurate axial translation (z). Continuous application of this symmetry operation to successive protomers generates a supramolecular polymer with helical symmetry. Determination of the helical symmetry of biological filaments can facilitate structural determination and promote understanding of its relevance to native function. Biologically derived helical protein assemblies encompass a diversity of functional roles that would be desirable to emulate in synthetic or semi-synthetic systems, including controlled release and delivery (filamentous phage and viruses),[3-5] cargo transport (conjugative pili, type 3 secretion system needle complex),[6-7] locomotion (flagellum)[8-9], signal transduction and actuation (pyrin domains and sterile α-motifs associated with the inflammasome and the signalosome, respectively)[10-11], among other roles.

We describe herein an approach to the fabrication of synthetic helical protein nanotubes based on simple oligopeptide motifs. The peptide design incorporates structural features associated with a class of tandem repeat proteins based on helical hairpin structural motifs. These helical assemblies represent promising substrates for the construction of synthetic compartments with structural and, potentially, functional asymmetry, since the interior (concave) and exterior (convex) surfaces can be structurally distinguished based on the mode of self-association of the protomers. The resultant assemblies approximate the dimensions and helical architecture of native protein and nucleoprotein filaments, which offers the opportunity to engineer synthetic peptide-based nanomaterials that display a range of functional roles similar to those of the corresponding biological assemblies.

The reliable and predictable design of helical nanotubes depends on the ability to control the non-covalent interactions between protomers that drive self-assembly. Helical assemblies are usually characterized in terms of the number of subunits per helical turn, N (= $2\pi/\phi$), the helical pitch, P (= Nz), and the helical radius, r. In the simplest cases, these parameters can be determined from assignment of a one-start helix, i.e., a helix that passes through every subunit in the structure. These assemblies need not form closed cylindrical structures, i.e., ones in which successive turns of the helical assembly make stable physical contacts through an axial interface. However, the corresponding tubular structures may be the most suitable targets for construction of asymmetric interfaces, that is, those in which the external and internal surfaces of the assembly can be structurally distinguished and functionally addressed. Helical protein nanotubes result from cohesive lateral and axial self-association of protomers into a hollow cylinder in which a central channel (lumen) runs through the structure and is oriented parallel to the helical axis. However, considerable structural evidence suggests that the quaternary structure of helical protein assemblies may not be robust in sequence space and, therefore, may be limited in designability.[12-21] The potential plasticity of quaternary structure in sequence space represents a significant challenge to the rational design of helical nanotubes. One approach to avoid this bottleneck may be to identify peptide or protein motifs that correspond to quaternary structures that are natively designable.

Tandem repeat proteins (TRPs)[22] constitute a diverse and highly designable protein supergroup[23-25] that show considerable promise for fabrication of robust supramolecular scaffolds. TRPs comprise repetitive domains in which structural motifs of twenty to fifty amino acids are concatenated with varying levels of sequence identity that depends on the repeat protein family.[22,26] Consensus repeats have been defined for a number of solenoidal tandem repeat proteins, including TPR (tetratricopeptide repeat),[27] ankyrin,[28] armadillo,[29] HEAT (Huntingtin, Elongation Factor 3, Protein Phosphatase Subunit A, and Yeast Kinase Tor1),[30] and LRR (leucine-rich) repeats.[31] Synthetic solenoidal repeat proteins derived from consensus sequences display structures that mimic the three-dimensional structures of the respective native TRPs. Moreover, computational methods have been employed for the *de novo* design of a diverse range of synthetic TRPs.[22,32-35] In these cases, good agreement was observed between the crystallographically determined structures and the corresponding computational models for a number of TRPs based on different repeat motifs. Most native solenoid proteins form open and extended supramolecular structures with strong lateral interactions but negligible axial interactions between successive helical turns. In addition, native TRPs are often capped at the *N*- and *C*-termini to prevent uncontrolled non-covalent polymerization, which is usually an unwanted outcome from the perspective of biological function. Therefore, native TRP sequences are not necessarily amenable as substrates for the construction of synthetic helical nanotubes. Recently, Shen *et al.* reported the computational design of self-assembling filaments from self-assembly of synthetic TRPs with excellent agreement between the computational models and the cryo-EM derived structures[65]. This study treated the TRPs as generic building blocks in which the local symmetry of the repeat was not continuously maintained throughout the structure. However, we have identified two related classes of tandem repeat protein motifs based on based on helical hairpin structures derived from leucine-rich repeat variant (LRV)[36] and phycobilisome lyase HEAT-like (PBS_HEAT)[37] that may be capable of forming closed cylindrical assemblies. We report the design, synthesis, and structural characterization of two peptide sequences, LRV_M3Δ1 and HEAT_R1, based on these motifs that self-assemble into structurally robust, thermally stable helical

nanotubes, in which the local cohesive interactions between protomers corresponding to one or two repeat motifs are maintained throughout the assembly.

2.2 Materials and Methods

Chemicals and Materials

All chemical reagents were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO) or Anaspec, Inc. (Fremont, CA) unless otherwise specified. The HEAT_R1, HEAT_dimer, and LRV_dimer peptides were ordered from GenScript USA, Inc. (Piscataway, NJ). Peptide synthesis resin was ordered from Applied Biosystems (Foster City, CA).

Peptide Synthesis and Purification Methods

The LRV_M3Δ1 and HEAT_R1_W17F peptides were prepared via microwave-assisted solid phase peptide synthesis on a CEM Liberty instrument as the *N*-acetyl, *C*-amide capped derivatives. A PAL-PEG-PS resin from Applied Biosystems was used for both peptides. Standard Fmoc protection chemistry was utilized in conjunction with coupling cycles consisting of 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 cleaved from the resin by incubation at room temperature for 3 hours in a cocktail consisting of trifluoroacetic acid (TFA), distilled water, triisopropylsilane, and 2,2'-(ethylenedioxy)-diethanethiol. Cleavage was followed by filtration and subsequent precipitation in diethyl ether. The peptide/diethyl ether mixture was then centrifuged at 4 °C at 4,000 rpm for 10 min. The supernatant was then discarded, and the precipitate allowed to desiccate overnight. Following desiccation, the crude peptide gels were resolubilized in 3 mL of a 50:50 mixture of acetonitrile and water (0.1% TFA additive) and purified by reversed-phase high-pressure liquid chromatography (HPLC) on a C18 column with a water-acetonitrile (0.1% TFA-additive) gradient.

Peptide mass was confirmed using MALDI mass spectrometry. Purified HPLC fractions were then lyophilized, sealed, and stored at -30 °C.

Peptide Assembly Methods

Stock solutions of HEAT_R1 (3 mg·mL^{-1}) were prepared by solubilizing 0.6 mg of purified, lyophilized peptide in 200 μL of 10 mM MES buffer, pH 6.0. Immediately after mixture, the solution was titrated to pH 6.0 using dilute sodium hydroxide solution. Stock solutions of LRV_M3Δ1 (3 mg·mL^{-1}) were prepared by solubilizing 0.6 mg of purified, lyophilized peptide in 200 μL of 10 mM MES buffer, pH 6.0. Immediately after mixture, the solution was titrated to pH 6.0 using dilute sodium hydroxide solution. This solution was then thermally annealed using the following thermal cycler protocol: (1) rapid heating to 90° C for 30 minutes and (2) cooling to 25°C at a rate of -0.2°C/minute.

Circular Dichroism Spectropolarimetry Methods

CD measurements were performed on a Jasco J-810 CD spectropolarimeter using 0.10 mm thick quartz plates (Hellma Analytics). Spectra were collected at 50 nm/min. in the range of 190-260 nm, and a data pitch of 0.2 nm.

Transmission Electron Microscopy Methods

TEM grids were prepared using solutions of peptide (3 mg·mL^{-1}) in aqueous buffer (10 mM acetate, pH 5.0, and 10 mM MES, pH 6.0, 10 mM MOPS, pH 7.0, and 10 mM TAPS, pH 8.0). Samples were prepared by depositing 4 μL of peptide solution onto a 200-mesh carbon-coated copper grid from Electron Microscopy Services (Hatfield, PA). After 1.5 minutes of incubation on the grid, moisture was wicked away, leaving only a thin film of sample. 4 μL of negative stain (2% methylamine tungstate) was deposited onto the thin film, to allow proper mixture. After 1 minute of staining, the remaining moisture was wicked away, and the grid dried overnight in a tabletop desiccator. Electron micrographs were captured on a Hitachi HT-7700 with a tungsten filament and AMT CCD camera, at an accelerating voltage of 80 kV.

Cryo-electron Microscopy and Image Processing Methods

The HEAT_R1 peptide (4 μL of a 3 mg·mL^{-1} solution) was applied to glow-discharged lacey carbon grids and vitrified in a Vitrobot Mark IV (FEI, Inc.). Grids were imaged in a Titan Krios at 300 keV and recorded with a Falcon II direct electron detector at 1.05 Å per pixel. Images were collected using a defocus range of 0.5–3.0 μm, with a total exposure of 2 s (amounting to 70 electrons·Å^{-2}) dose-fractionated into seven chunks. All the images were first motion corrected by the MotionCorr v2, and then the CTFFIND3 program was used for determining the actual defocus of the images. Images with poor CTF estimation as well as defocus > 3 μm were discarded. A total of 386 images were selected and peptide filaments of varying lengths were boxed using the e2helixboxer program within EMAN2. The CTF was corrected by multiplying the images from the first two chunks (containing a dose of ~ 20 electrons·Å^{-2}) with the theoretical CTF.

Overlapping 256-px long boxes with a shift of 5 pixels (~ 1.5 times of the axial rise) were cut from the long filaments. The determination of the helical symmetry was by trial and error, searching for a symmetry which yielded recognizable secondary structure. Power spectra from the filaments suggested a variability in the helical parameters. A reference-based sorting procedure was used to bin the segments based on the axial rise and azimuthal rotation. Then 56,421 out of 356,100 segments were selected after this sorting and the IHRSR method implemented in in Spider was used to produce the final reconstructions at ~6 Å. The helical hand was determined by the fit of crystal structures into the cryo-EM map.

The LRV_M3Δ1 peptide (4 μL of a ~3 mg·mL^{-1} solution) was plunge-frozen in the same manner as the HEAT_R1 filament. The LRV_M3Δ1 grids were imaged using the Titan Krios at the National Cancer Institute and recorded with a K2 camera at 1.32 Å per pixel without super-resolution. A total of 296 images were selected, motion-corrected and CTF-corrected in a similar way with the HEAT_R1 peptide. Then 512-px long overlapping boxes with a shift of 8 pixels, containing a dose of circa 20 electrons·Å^{-2}, were cut out from the CTF-corrected images and were used in the initial IHRSR reconstruction. The

helical symmetry was also determined by trial and error, searching for a symmetry which yielded recognizable secondary structure. After determining the correct symmetry, segments were sorted using a reference-based procedure based on the azimuthal rotation. Then 62,616 out of 171,324 segments were selected and were further reconstructed by the IHRSR method in Spider. After the last cycle of the reconstruction, the alignment parameters were applied to an image stack containing 10 electrons·$Å^{-2}$. The final volume generated from the lower dose image stack had a resolution of ~ 4.4 Å. The correct helical hand could be directly determined from the hand of the α-helices visible at this resolution.

Model building

The initial model of the HEAT_R1 peptide was generated by Phyre2 homology modeling using residues 74-103 of 4JW3,[52] a concatemeric polypeptide derived from consensus PBS_HEAT repeats similar to 3LTJ, as the starting model (80% sequence identity). Then the map corresponding to a HEAT_R1 dimer was segmented from the experimental filament map in Chimera. At this resolution, the N/C- termini of the peptide subunits in the dimer cannot be distinguished directly from the density. There are 4 possible arrangements of the peptide dimer termini viewing from outside of the filament and from left to right: (1) N-N; (2) C-C; (3) N-C; (4) C-N. All those four possibilities were examined by docking peptide subunits into the segmented map followed by a Phenix real-space refinement. The N-N arrangement fit the map significantly better, with at least 20% higher real-space coefficient correlation than the other possibilities. A filament model was generated and further refined using Phenix, and MolProbity was used to evaluate the quality of the filament model. The refinement statistics are given in Table 2.1.

The initial model of the LRV_M3Δ1 peptide was generated by extracting the backbone from 1LRV using residues 124-146 as the starting model. The map corresponding to the model positioned in the filament was segmented from the experimental map at 6 Å radius in Chimera. The LRV_M3Δ1 model was rebuilt with the RosettaCM protocol. To obtain a well-fitting model a total of 1000 decoy models

were generated based on the segmented map. Rosetta's energy function was used to select the top 5 models from which the best fitting was chosen for further processing. Coot was used to manually edit the model to better reflect actual stereochemistry, and then it was further refined by Phenix real-space refinement protocol. Then the symmetry-related molecules were regenerated using Rosetta, and the whole assembly was further refined in Phenix real-space refinement. The quality of the filament model was assessed by MolProbity. The refinement statistics are given in Table 2.1.

Scanning Transmission Electron Microscopy

STEM data were acquired at Brookhaven National Laboratory (BNL). The STEM instrument operates at 40 keV with a scanning probe of <0.3 nm diameter produced from a cold field-emission source. Each electron emerging from the specimen is detected by one of the scintillator photomultiplier detectors collecting 0−15 mRadian (bright field), 15−40 mRadian (small-angle dark field) and 40−200 mRadian (large-angle dark field). The large-angle signal is proportional to the mass of atoms in the path of the beam. Specimen quality and mass calibration were assessed by detailed comparison of the image to the known structure of tobacco mosaic virus (TMV). For mass-per-length (M/L) measurements, TMV rafts at a theoretical M/L value of 13.1 kDa/Å were employed for calibration.

HEAT_R1 was assembled at 3 mg·mL^{-1} in 10 mM MES buffer, pH 6.0. LRV_M3Δ1 was assembled at 3 mg·mL^{-1} in 10 mM MES buffer, pH 6.0, then thermally annealed as previously described. These solutions were then diluted to 0.3 mg·mL^{-1} and 0.03 mg·mL^{-1} for the unstained STEM imaging, and to 0.1 mg·mL^{-1} for negative-stained STEM imaging. Specimens are deposited on thin carbon (circa 2 nm thick) supported on a thicker holey carbon film mounted on a titanium grid using the wet-film, hanging-drop method. TMV is added to the grid first as an internal control, followed by injection buffer, then specimen solution (in 10 mM MES buffer, pH 6.0) for 1 min, then 10 washes of 20 mM ammonium acetate pH 7.0. Excess solution is wicked from the edge with filter paper between each injection. After the last wash, the grid is wicked to a thin layer (ca. 1 mm), fast frozen by plunging into liquid nitrogen

slush and stored under liquid nitrogen. Grids are freeze-dried overnight in an ion pumped chamber with an efficient cold trap and transferred under vacuum to the STEM cold stage (−160 °C). Imaging typically uses a dose of 20 $e^{-}/Å^2$ (causing < 5% mass loss, corrected by comparison to TMV). Mass measurements were performed off-line with the customized software PCMass32.

Small- and Wide-Angle X-Ray Scattering Measurements

Synchrotron SAXS/WAXS measurements were taken at the 12-ID-B beamline of the Advanced Photon Source at Argonne National Laboratory in Lemont, Illinois. HEAT_R1 was assembled at 6 $mg{\cdot}mL^{-1}$ in 10 mM MES buffer, pH 6.0. LRV_M3Δ1 was assembled at 6 $mg{\cdot}mL^{-1}$ in 10 mM MES buffer, pH 6.0, then thermally annealed as previously described. These solutions were then dialyzed into 10 MES buffer, pH 6.0 with a 1% glycerol additive to stabilize against X-ray radiation damage. SAXS/WAXS measurements were then taken on the peptide solutions at 25 °C in a quartz capillary flow cell (1.5 mm). The solutions were raised and lowered in the flow cell to minimize radiation damage. Twenty 2D images were collected per sample, and then azimuthally averaged into 1D SAXS scattering curves after solid angle correction and normalization against the transmitted X-ray beam intensity, using the software package at beamline 12-ID-B. The 1D scattering curves were then averaged, and the averaged signal from the MES buffer was subtracted out.

2.3 Design, Synthesis, and Biophysical Characterization of the LRV and HEAT Peptide Nanotubes

Rees and co-workers first described the LRV repeat motif in 1996 from the crystallographic analysis of a protein (PDB ID:1LRV) derived from a cryptic ORF (open-reading frame) in the nitrogen-fixing bacterium *A. vinelandii* (Figure 2.1).[36] The structure of the protein comprised two domains; an Fe_4-S_4 cluster protein and a series of eight helical repeats of 24 amino acid residues described as a concatemers of an LRV domain. Presumably, the Fe_4-S_4 cluster protein is involved in redox processes and the LRV domain may be involved in substrate recognition (as is the case for many tandem repeat proteins), although the exact roles of the respective domains remain speculative. The LRV repeat motif (Pfam

database, PF01816) is based on a 3_{10} helix-loop-alpha helix structure, in which the 3_{10} helix lies at the exterior (convex) surface and the α-helix at the interior (concave) surface. The structure of LRV repeat sequence in 1LRV corresponds to a right-handed super-helix with an average twist angle between successive repeat motifs of approximately 15°. Remarkably, Rees and co-workers predicted on the basis of this crystal structure that the LRV domain could potentially form a helical nanotube with a pitch of 32 Å and approximately 24 subunits per turn of the super-helix. Subsequently, Minard and co-workers described a class of HEAT repeat proteins, PBS_HEAT (SMART database, SM000567 and Pfam database, PF03130), derived from thermostable microbes.[37] In contrast to the LRV motif, the PBS_HEAT repeat comprises an alpha helix-loop-alpha helix motif of 31 amino acids. They defined a consensus repeat sequence for the PBS_HEAT motif and synthesized and characterized a series of concatemers based on this sequence. The inner (concave) helix of the consensus sequence displays similarity to the α-helix of the LRV repeat in that specific residues that mediate interactions at the lateral interface are conserved between the two sequences (Figure 2.1A). They reported the crystal structure of a tetramer of the consensus HEAT repeat sequence, αRep-n4-a (PDB ID: 3LTJ), which displayed a right-handed helical twist with an average angle of 20° between successive repeat motifs. These results suggest that the mode of subunit packing in the folded concatemer is similar between the LRV and PBS_HEAT repeat motifs, although the internal rotational angle–analogous to the twist angle, ϕ, of a helical assembly–differs between the two structures. Neither sequence was capable of forming helical nanotubes since the termini were blocked in the full-length proteins to prevent elongation through end-to-end self-association.

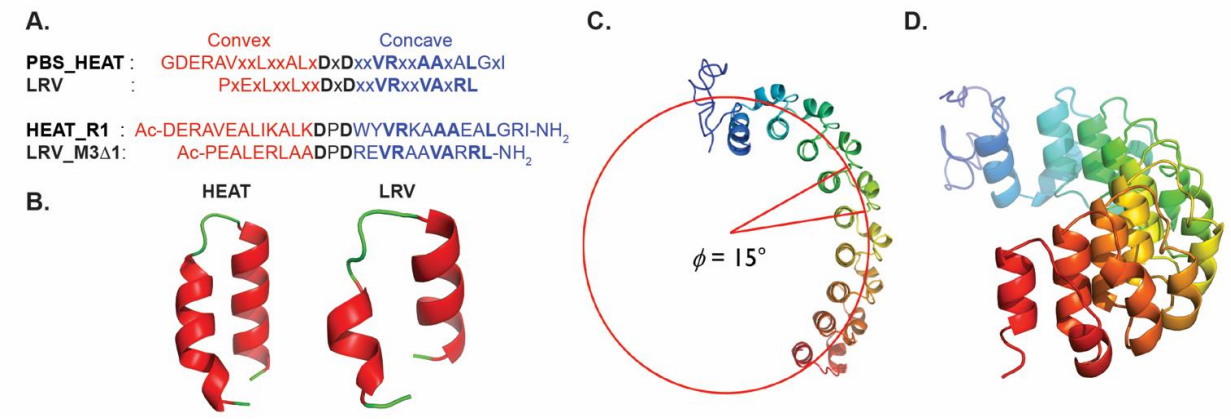


Figure 2.1. Design of synthetic helical nanotubes from tandem repeat peptide sequences. (A) Conserved sequence positions for the PBS_HEAT and LRV repeat motifs and the corresponding sequences of the synthetic peptides HEAT_R1 and LRV_M3D1. The sequence features corresponding to structural motifs located at the convex and concave surfaces are colored red and blue, respectively. Conserved and semi-conserved residues between the PBS_HEAT and LRV repeat sequences are depicted in bold and are located at the turn and concave helix. (B) Structures of representative repeat units from the crystal structures of 3LTJ and 1LRV, respectively. Concave helices are depicted on the right side of the respective images. (C) Top view of the crystal structure (PDB ID: 1LRV) of *A. vinelandii* protein (Genbank accession number M20568) comprising an *N*-terminal Fe_4S_4 cluster protein domain (line diagram in blue) fused to a series of eight LRV domains. Successive LRV repeat motifs undergo an average rotation through a twist angle of circa 15° with respect to a central axis that is perpendicular to the plane of the image. (D) Side view of the crystal structure of the same protein in which the right-handed helical twist of the LRV concatemer can be discerned.

Two peptide sequences, HEAT_R1 and LRV_M3Δ1, were designed from the consensus sequences of the PBS_HEAT and LRV repeat motifs, respectively (Figure 2.1B). The *N*-terminal amino acid was deleted from the respective consensus sequences and replaced with a smaller acetyl group to prevent steric interference between the termini of successive protomers in the helical assembly. The identities of charged

groups at surface exposed sites were chosen to minimize repulsive electrostatic interactions and maintain the pI of the resultant peptides within the near-neutral pH range. Highly conserved residues within the respective repeat sequences were maintained in the synthetic peptides, as these positions mediated structurally critical lateral interactions between protomers in the respective tandem repeat proteins. The design of each peptide was derived from a single repeat unit of the LRV and PBS_HEAT motifs based on the following considerations; ease of synthesis of traceless (i.e., tag-free) sequences, removal of additional turn sequences between concatenated repeat motifs that could inhibit axial stacking interactions, and the greater conformational freedom of short peptide sequences that may more easily accommodate the formation of large assemblies. The disadvantage of this approach is that single repeat motifs may not adopt a thermodynamically stable conformation corresponding to the native fold. The folding of tandem repeat proteins derived from consensus TPR and ankyrin motifs has been analyzed using a one-dimensional Ising model,[38-40] which suggested that the stability of the corresponding repeat proteins was related to the number of repeat units and the strength of lateral association. Individual subunits of consensus repeats may be poised on the edge of instability but still appear stable at 25 °C. Minard and co-workers described the thermal denaturation of a series of concatemers of thermostable PBS_HEAT consensus repeats (n = 1, 2, 3, 4, and 6) cloned between optimized *N*-terminal and *C*-terminal capping sequences.[37] The T_m value of the monomer, αRep-n1-a, was determined to be approximately 71-72 °C in dilute aqueous solution (100 μM protein, 50 mM phosphate buffer, pH 7). While the monomer actually consisted of three repeats due to the presence of the capping motifs, these data suggested that individual repeat motifs might be sufficiently stable at ambient temperature.

Peptides HEAT_R1 and LRV_M3Δ1 were prepared using microwave-assisted solid-phase synthesis and purified via reverse-phase HPLC. The purity and identity of the corresponding peptides were confirmed by analytical HPLC and mass spectrometry. Circular dichroism (CD) spectropolarimetry was employed to interrogate the conformational behavior of the peptides in aqueous buffer. The HEAT_R1 and LRV_M3Δ1 peptides displayed significant differences in their respective CD spectra, which could be

interpreted in terms of the relative conformational stability of the repeat motifs. The HEAT_R1 peptide displayed a classical α-helical CD signature with a positive signal at 193 nm and negative minima at 208 nm and 222 nm over a wide range of pH values (Figure 2.2). The CD signal of HEAT_R1 was monitored over the temperature range from 25 °C to 90 °C and did not display a detectable melting transition (Figure 2.3). The CD spectra of HEAT_R1 in buffered aqueous solutions strongly resemble the corresponding

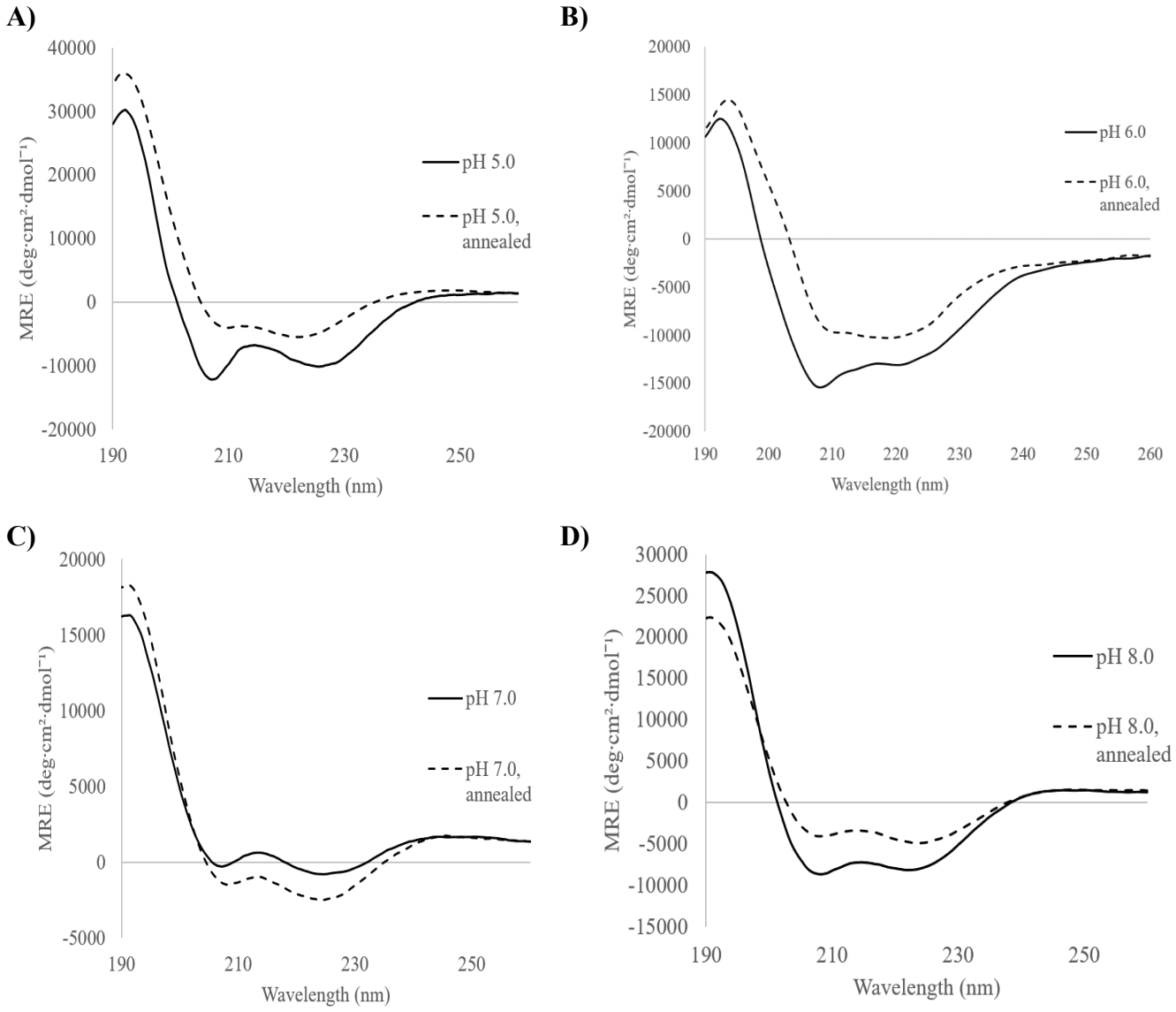


Figure 2.2 CD spectra for 250 µM HEAT_R1 at pH 5.0 (A) , pH 6.0 (B), pH 7.0 (C), and pH 8.0 (D) . Buffers were 10 mM Acetate (pH 5.0), 10 mM MES (pH 6.0), 10 mM MOPS (pH 7.0), and 10 mM TAPS (pH 8.0).

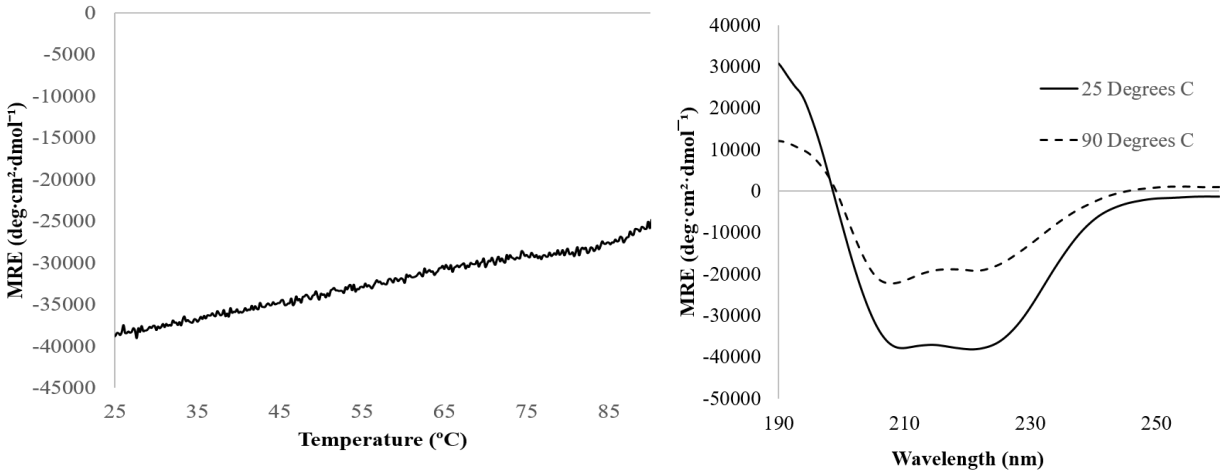


Figure 2.3 HEAT_R1 Variable Temperature CD Scan. The CD signal at 222 nm was monitored from 25-90º C (left), at a 250 µM peptide concentration. No thermal transition was observed; a standard CD spectrum was collected at 90º C (right). A helical CD signature was observed at both temperatures.

CD spectra of consensus PBS_HEAT repeats in the αRep series of capped concatemers reported by Minard and co- workers.[37] In contrast, the CD spectra of LRV_M3Δ1 displayed a spectroscopic response that evolved over time (Figure 2.4). A single strong minimum was initially observed at 226 nm with a

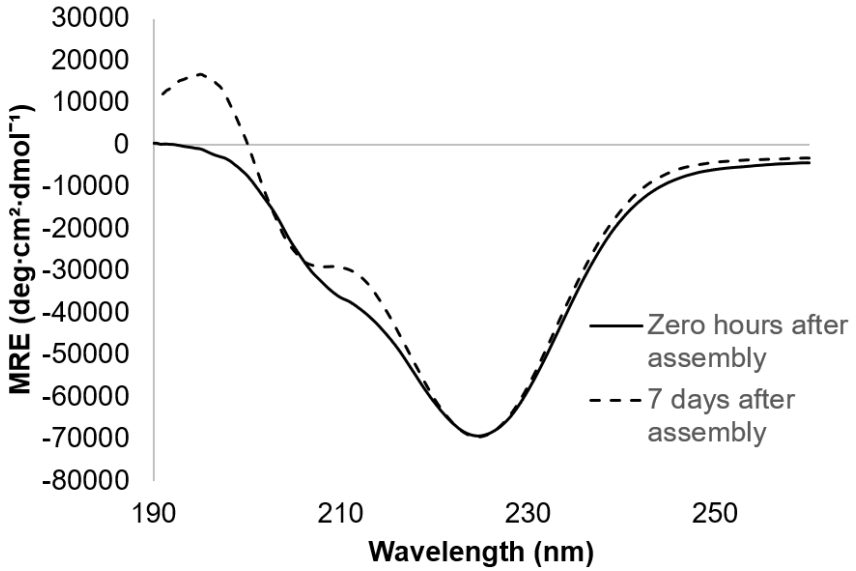


Figure 2.4 Time-dependence of circular dichroism spectra for LRV_M3Δ1 over time (250 µM peptide concentration, 10 mM MES, pH 6.0).

shoulder at shorter wavelengths. The mature CD spectrum resembled an α-helical conformation, albeit with differences in that the minima were observed at wavelengths of 204 nm and 224 nm, which were

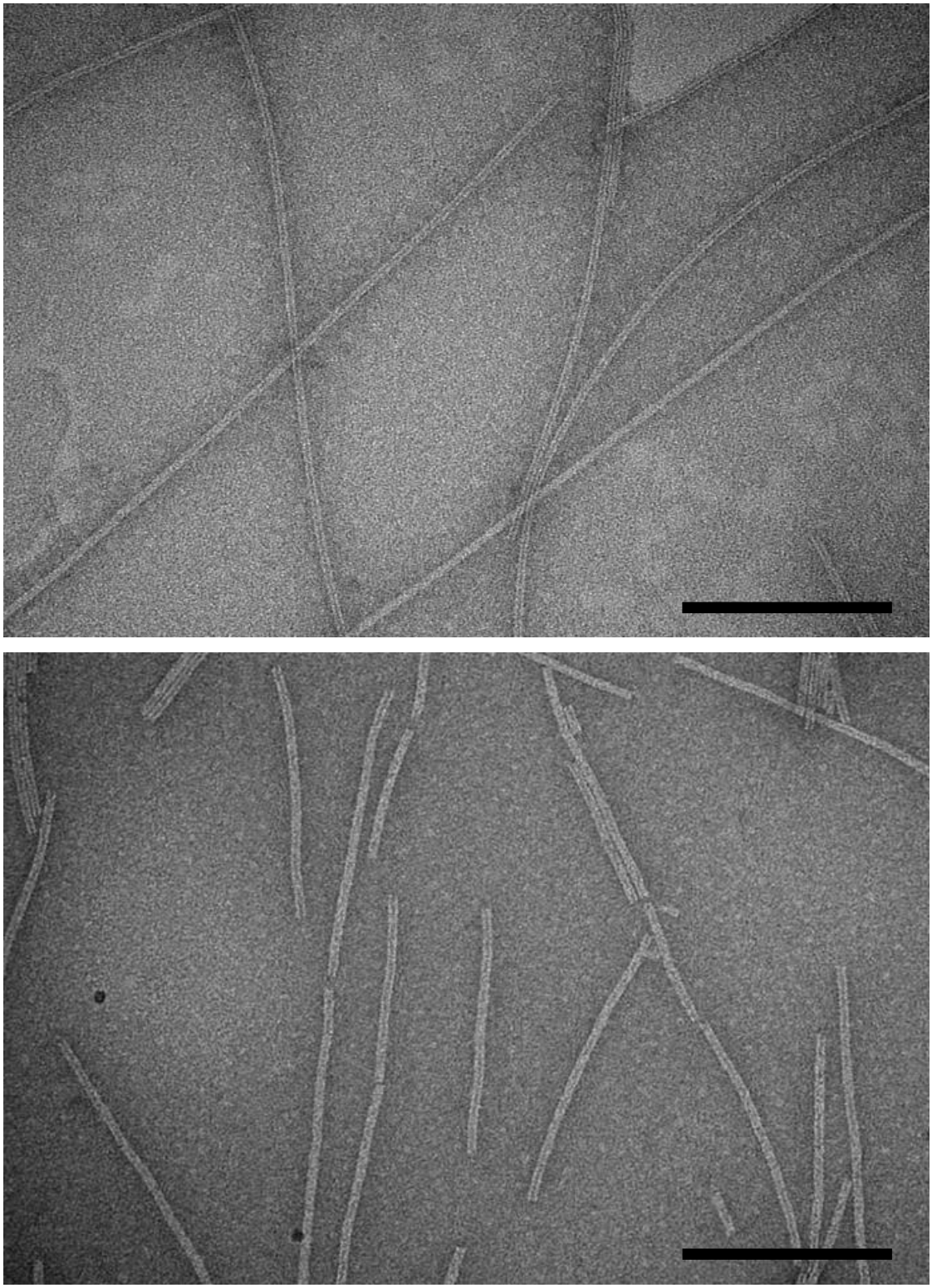


Figure 2.5 Representative negatively stained TEM images of LRV_M3Δ1 (top) and HEAT_R1 (bottom) filaments. Scale bars are 100 nm.

shifted in position with respect to a canonical a-helical conformation. The observed CD profile could not be interpreted clearly in terms of an α-helix-loop-α-helix or a 3_{10} helix-loop-α-helix.[36] The LRV repeat motif is significantly shorter than the PBS_HEAT motif and, correspondingly, may be less conformationally stable in the absence of self-association. The slow emergence of a helix-like conformation suggests that convergence to a stable conformation occurs over time as the assemblies mature.

Negative-stained transmission electron microscopy (TEM) was used to screen for self-assembly of the HEAT_R1 and LRV_M3Δ1 peptides and to determine the resultant morphology of the corresponding

assemblies (Figure 2.5). High aspect-ratio filaments were observed from self-assembly of both peptides within the pH range from 5-8 in aqueous buffer. The resultant filaments seemed to be stable in solution for a period of at least several weeks at ambient temperature. The filaments were uniform in diameter but

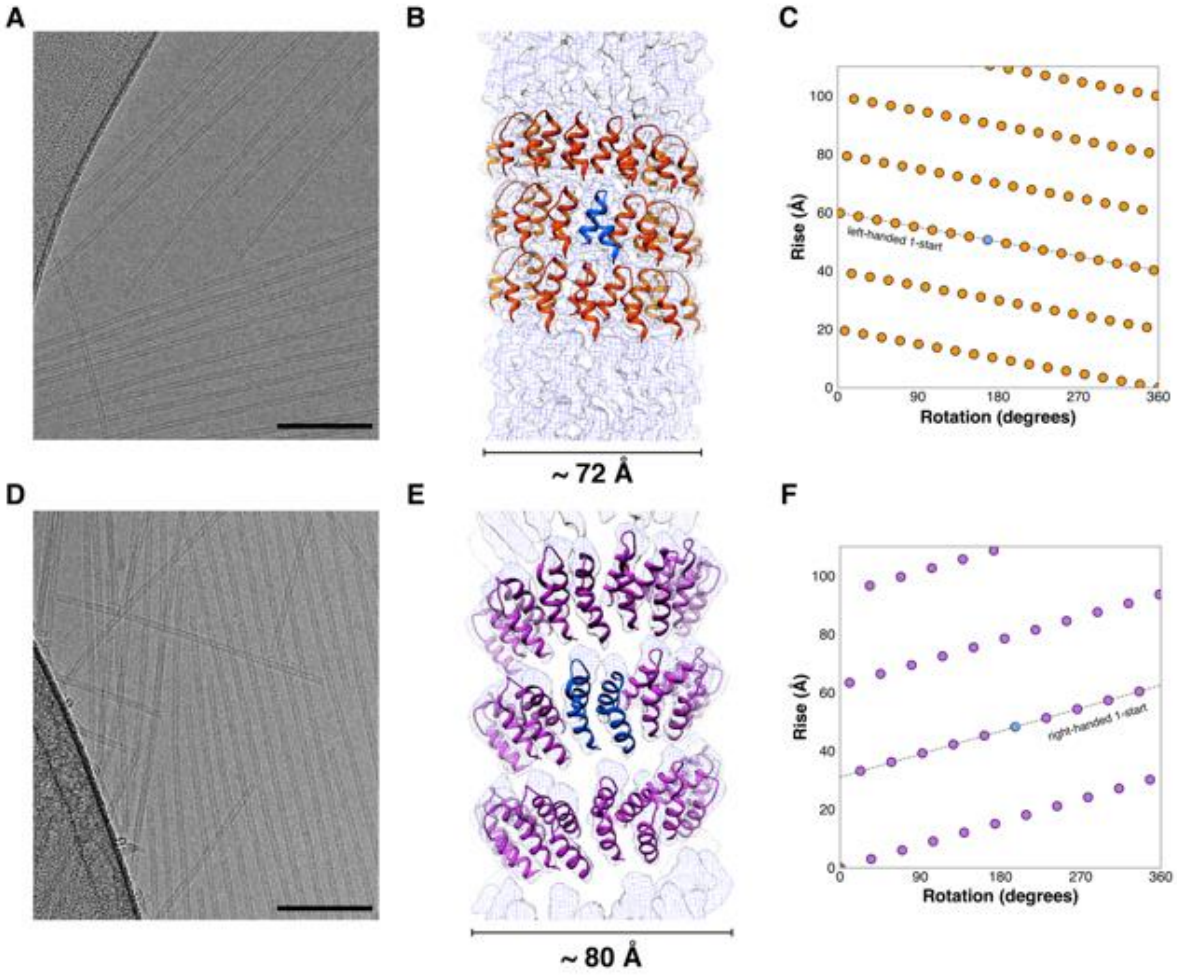


Figure 2.6 (A and D) Cryo-EM images of LRV_M3Δ1 (A) and HEAT_R1 (D) filaments (scale bar = 100 nm). (B and E) 3D reconstructions of the LRV_M3Δ1 (B) and HEAT_R1 (E) filaments fit with the corresponding atomic models. The asymmetric units are highlighted for the respective filaments in blue, which, in the case of HEAT_R1, corresponds to a dimer of peptides. (C and F) Helical nets for the LRV_M3Δ1 (C) and HEAT_R1 (F) filaments, in which the difference in helical hand is apparent. The helical nets show the unrolled surface lattice viewed from the outside of the filament.

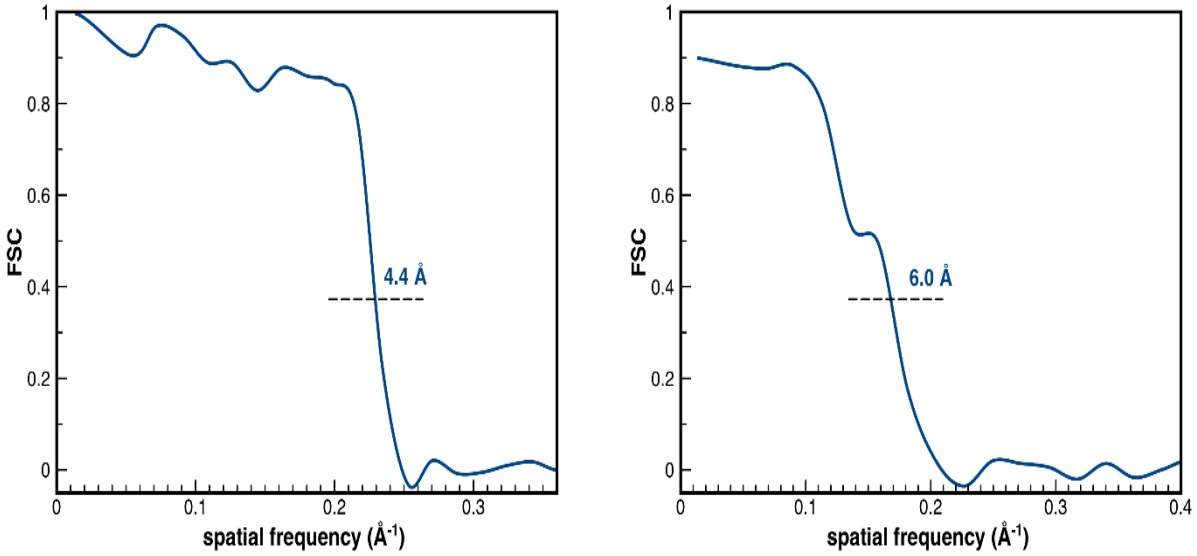


Figure 2.7 Model:Map Fourier Shell Correlation (FSC) plots for LRV_M3Δ1 (left) and HEAT_R1 (right) indicating the resolution of the two cryo-EM structures (4.4 Å and 6.0 Å) respectively. The cut-off was set at 0.38.

polymorphic in the length-ranging from 50 nm to 5 μm. The apparent widths of the filaments were measured to be circa 9 nm for HEAT_R1 and LRV_M3Δ1. The differential contrast between the edges of the structures and the internal portion confirm the formation of nanotubes. The inner lumen appears darker in negative-stained TEM images of both classes of filaments due to infiltration of negative stain into the central channel. The LRV_M3Δ1 nanotubes were generally longer and less flexible than those seen in HEAT_R1 (Figure 2.5). Despite apparent differences in peptide conformation, sequence, and length, the HEAT_R1 and LRV_M3Δ1 filaments displayed similar morphology. However, the absence of high-resolution structural information precluded a conclusive determination that the structures were equivalent to each other or to the helical nanotube structure initially proposed by Rees and coworkers.

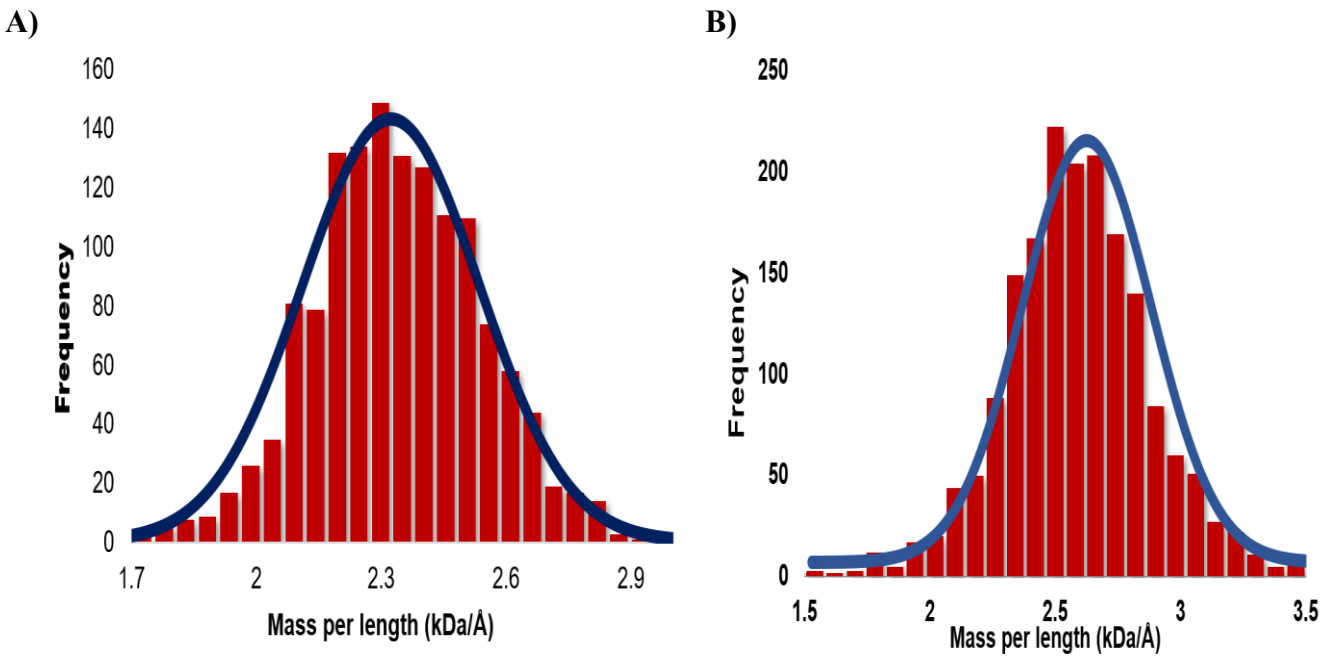


Figure 2.8 STEM-derived mass per length (MPL) distribution for HEAT_R1 (A) and LRV_M3Δ1 (B) filaments. Overlaid in navy blue is a Gaussian curve for comparison. MPL values were calculated to be 2356 ± 214 Da·Å^{-1} and 2697 ± 258 Da·Å^{-1} for HEAT_R1 and LRV_M3Δ1 respectively.

In order to resolve these issues, particularly in light of the aforementioned polymorphism of helical polymers assembled *in vitro*, the structures of the assemblies were analyzed using electron cryo-microscopy with direct electron detection (Figure 2.6). Three-dimensional reconstructions were generated using iterative helical real space reconstruction (IHRSR). Final resolutions of 4.4 Å and 6.0 Å for LRV_M3Δ1 and HEAT_R1 filaments, respectively, were determined from the Fourier shell correlation (FSC) between the maps and the models (Figure 2.7, Table 2.1). The structure of the HEAT_R1 filaments displays a right-handed helix with a pitch of approximately 31 Å and 10.3 asymmetric units per turn. Contrary to expectation, the asymmetric unit of the HEAT_R1 assembly corresponded to a dimer of helical hairpin subunits (Figure 2.6E) for an overall number of 20.7 peptides per helical turn, which was in approximate agreement with the

estimates from the STEM mass-per-length measurements (Figure 2.8). The power spectrum displayed a strong peak at 31 Å corresponding to the one-start helix, which was also observed in the synchrotron scattering curve for aqueous solutions of HEAT_R1 filaments at a *q*-value of 0.20

$Å^{-1}$ (Figure 2.9). The structures of the individual subunits resembled the structure of the HEAT repeat motifs observed in the crystal structure of aRep-n4-a (PDB ID: 3LTJ).

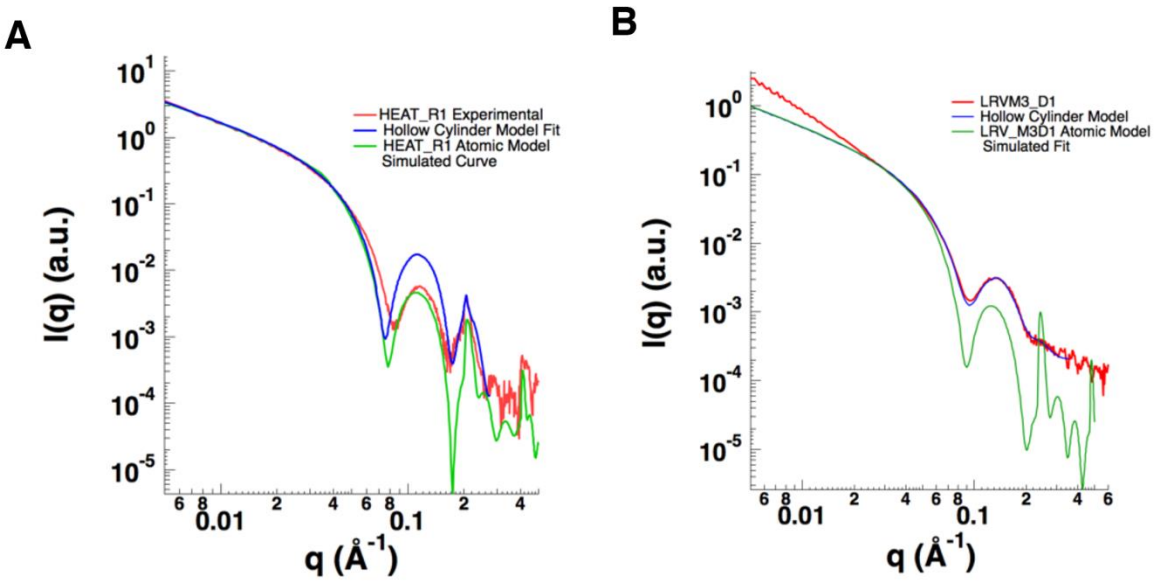


Figure 2.9 Synchrotron SAXS scattering curves for buffered aqueous solutions of HEAT_R1 (A) and LRV_M3Δ1 (B) filaments. Red curve: experimental data, blue curve: fit with hollow cylinder model, green curve: calculated from the respective atomic structural models (length ~300nm). For HEAT_R1, the low *q* (< 0.02 $Å^{-1}$) intensities follow the q^{-1} power law, indicating the assembly adopts a long rod like shape. The well-defined oscillation feature in *q* of 0.05-0.30 $Å^{-1}$indicates a highly uniformed cross-section size. The diffraction peak at ~0.206 $Å^{-1}$ arises from the ordered helical structure, with a pitch of ~ 30.5 Å. For LRV_M3Δ1, the experimental SAXS curve shows an oscillation feature, reflecting the cylindrical shape. The absence of a diffraction peak at ~0.24 $Å^{-1}$, as depicted in the simulated curve (green), indicates that the LRV_M3Δ1 filaments lack long order in the helical structure. Both peptides were assembled at 6 mg·mL^{-1} and dialyzed against 10 mM MES buffer (pH 6.0) with 1% v/v glycerol. LRV_M3Δ1 was thermally annealed (as previously described) prior to dialysis.

In comparison, the structure of the LRV_M3Δ1 filaments consisted of a left-handed helix with a pitch of 20 Å and 17.4 subunits per turn (Figure 2.6B). Notably, the observed helical parameters for the LRV_M3Δ1 filament differed significantly from the helical nanotube structure that Rees

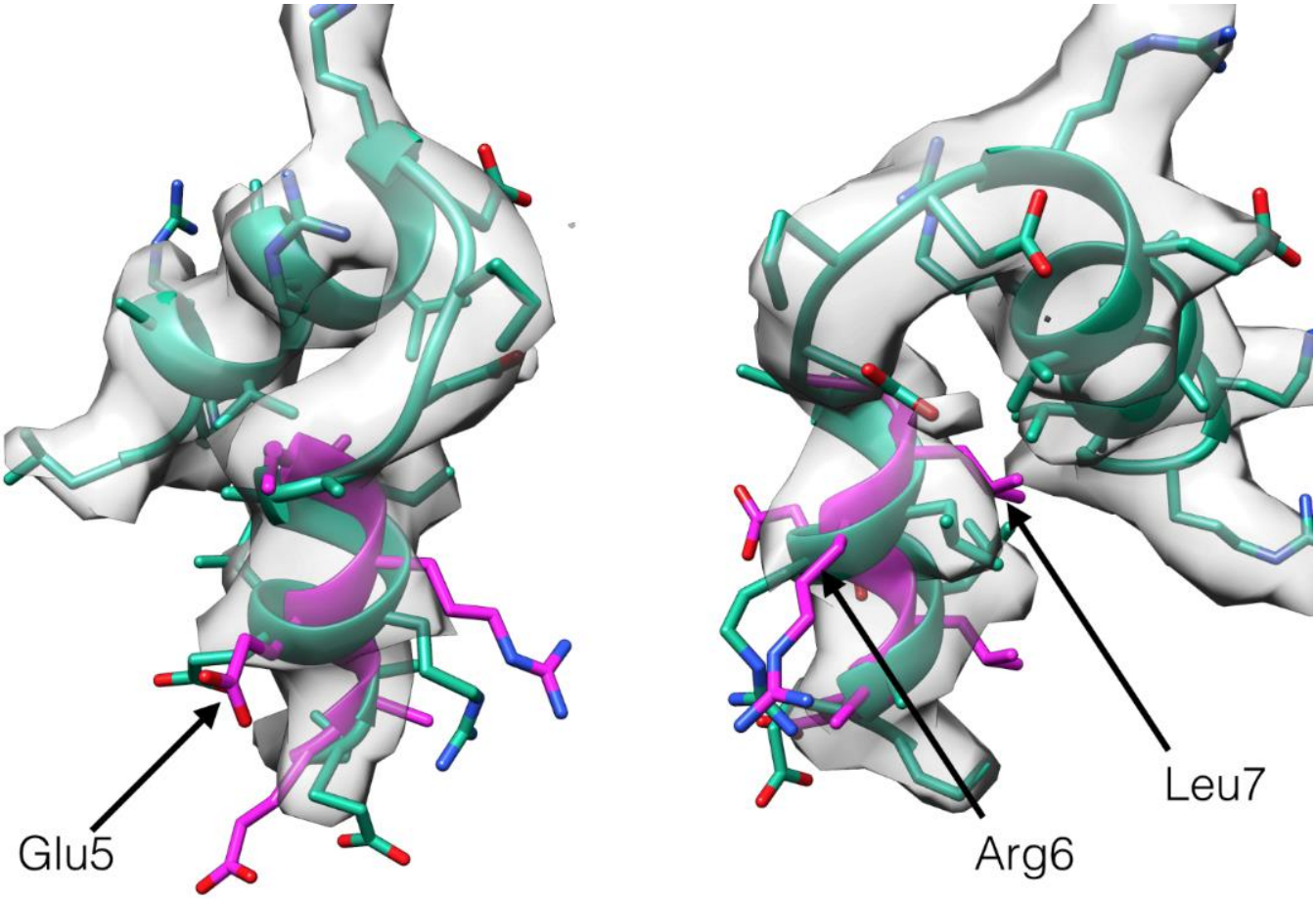


Figure 2.10 Secondary structural analysis of LRV_M3Δ1 based on the cryo-EM reconstruction. Represented in gray is the observed electron density in the cryo-EM map; fit into this density are the two candidate secondary structures. An α/α subunit is represented in green, and a single 3_{10} helix is represented in purple. Though the hypothetical subunit composition is α/3_{10}, the observed density is more consistent with an α/α subunit. Several clashing residues have been noted with black arrows.

and coworkers proposed for an assembly derived from the structure of 1LRV (*vide infra*).[36]

Moreover, the protomer structure deduced from fitting of the EM density map was consistent with an α-loop-α helical hairpin rather than the 3_{10}-loop-a helical hairpin observed in the crystal structure 1LRV (Figure 2.10). The spacing of 20 Å associated with the one-start helix of the LRV_M3Δ1 filament was not observed in the corresponding synchrotron small- angle X-ray scattering (SAXS) curve (Figure 2.9). The helical turn of a protomer within the LRV_M3Δ1 filament makes a close contact with the *N*- and *C*-termini of axially adjacent protomers of a successive helical turn (*vide infra*). Protomers within the LRV_M3Δ1 filament display a minimal tilt with respect the central helical axis of the assembly, consequently the filament appears at low

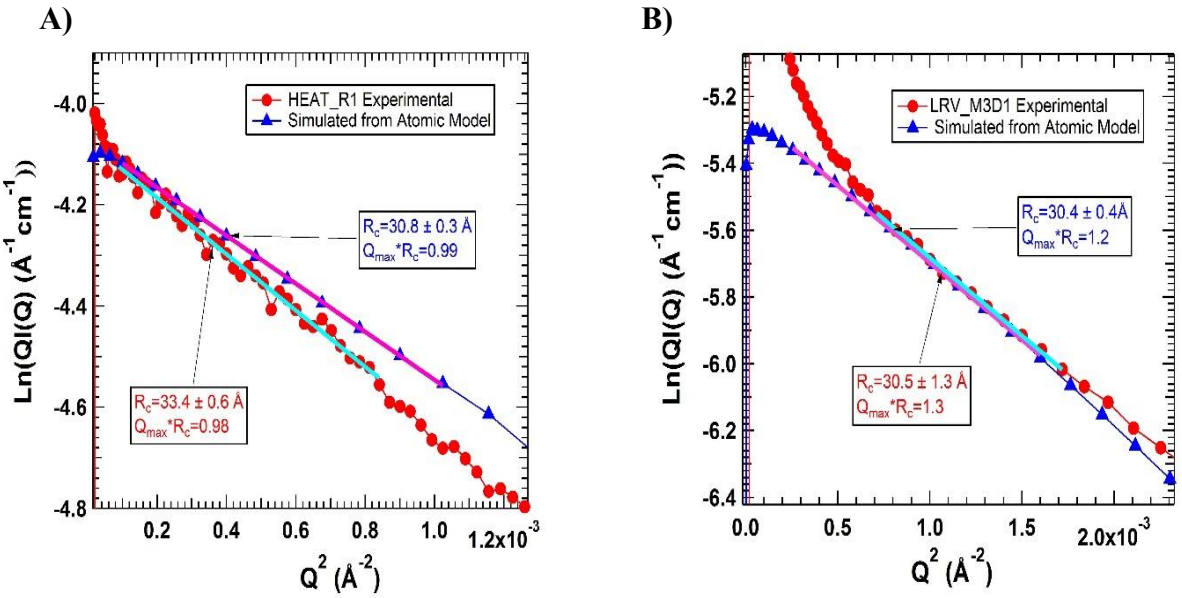


Figure 2.11 Modified Guinier fit for rod-like forms of synchrotron SAXS scattering data for buffered aqueous solutions of HEAT_R1 (A) and LRV_M3Δ1 (B) filaments. Experimental data shown in red solid circle symbols; simulated data from atomic models in solid triangle symbols. Note that for LRV_M3Δ1, the experimental SAXS data at the low *q* region deviates from q^{-1} power law, possibly caused by presence of higher order aggregates.

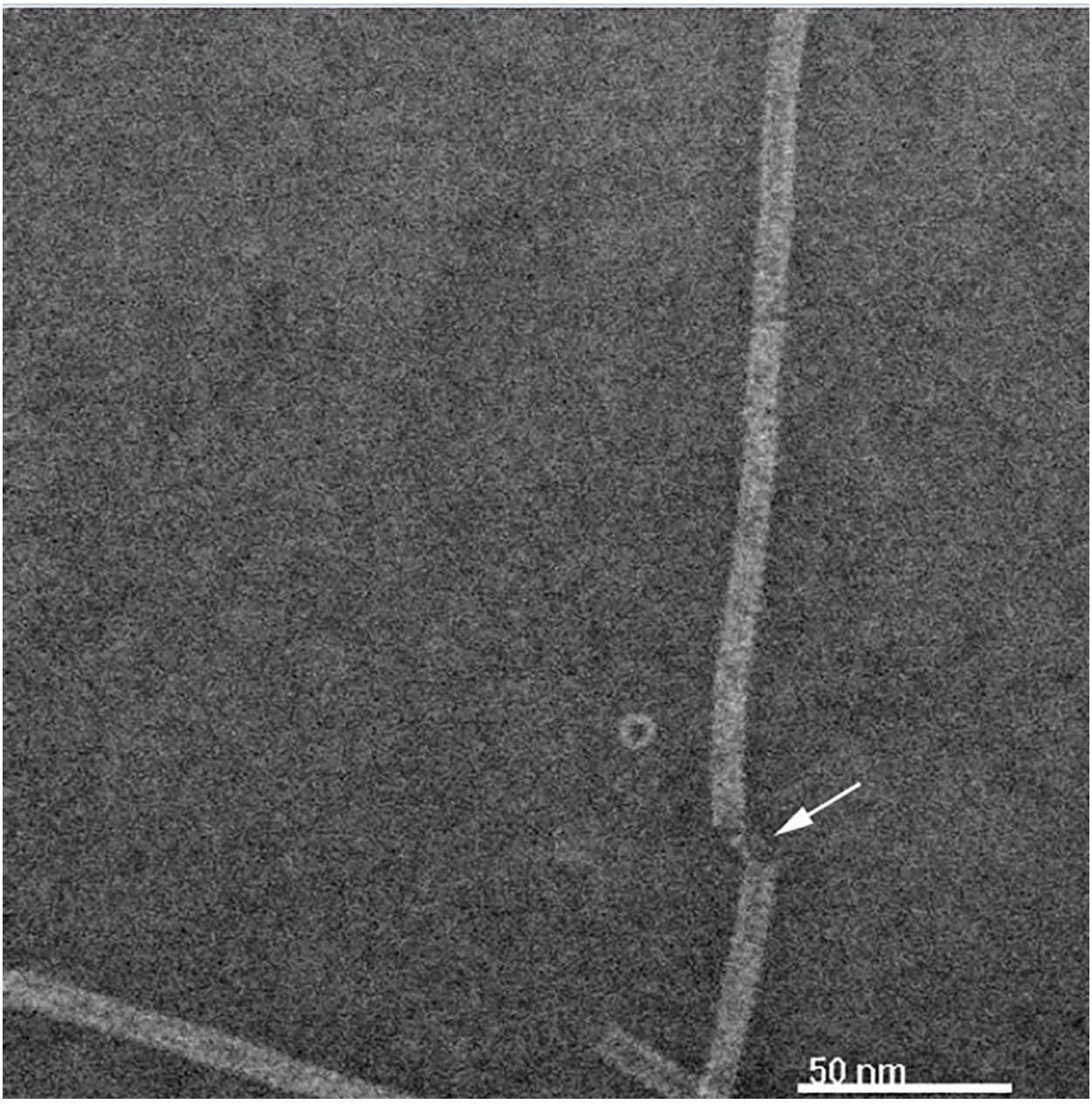


Figure 2.12 Negatively-stained STEM image of HEAT_R1 filaments, in which localized helical unwinding can be observed (white arrow). Note that the hollow cylindrical cross-section can be observed for a tubular fragment that has adventitiously adsorbed with the helical axis oriented perpendicular to the grid surface.

resolution as a smooth cylinder. In contrast, protomers within the HEAT_R1 filament are tilted away from the central helical axis of the assembly such that the concave helix at the turn surface

makes contact with the *N*-terminus of the convex helix of an axially adjacent protomer. Similarities were observed between the structures of the HEAT_R1 and LRV_M3Δ1 filaments in that the conserved concave a-helices lined the inner lumen, as expected, with the more variable *N*-terminal helix located at the convex surface, as predicted for helical nanotubes based on the structures of PBS_HEAT and LRV repeat proteins.

Synchrotron small-angle X-ray scattering measurements on aqueous solutions of the HEAT_R1 and LRV_M3Δ1 peptides confirmed the presence of uniform diameter cylindrical rods (Figure 2.9). The SAXS data for LRV_M3Δ1 and HEAT_R1 were fit to a modified Guinier fit for rod-like forms (Figure 2.11). The cross-sectional radius of gyration, R_c, was determined from the SAXS data for LRV_M3Δ1 and HEAT_R1 assemblies, which resulted in values of 30.5 ± 1.3 Å and 33.4 ± 0.6 Å, respectively. The corresponding R_c values, calculated from the atomic models derived from the 3D reconstructions were determined to be 30.4 Å and 30.8 Å for the LRV_M3Δ1 and HEAT_R1 assemblies. The experimentally determined R_c value for the LRV_M3Δ1 assemblies correlates well with that calculated from the atomic model resulting from the 3D reconstruction. In contrast, the R_c value determined from the SAXS data for the HEAT_R1 assemblies is significantly larger than that calculated from the atomic model. This discrepancy may reflect the fact that weakened axial interactions between successive helical turns in the HEAT_R1 assemblies (*vide infra*) results in greater flexibility. Local unwinding of the helical filaments is observed in STEM images of negatively stained specimens of the HEAT_R1 filaments (Figure 2.12).

The structural analyses of the HEAT_R1 and LRV_M3Δ1 filaments highlight the challenge of polymorphism in helical assemblies and the difficulties associated with prediction of quaternary structure, even within a tractable sequence context. High-resolution structural

information offers the opportunity to gain insight into the features that might be responsible for the differences between the predicted and observed structures. The HEAT_R1 filament displays

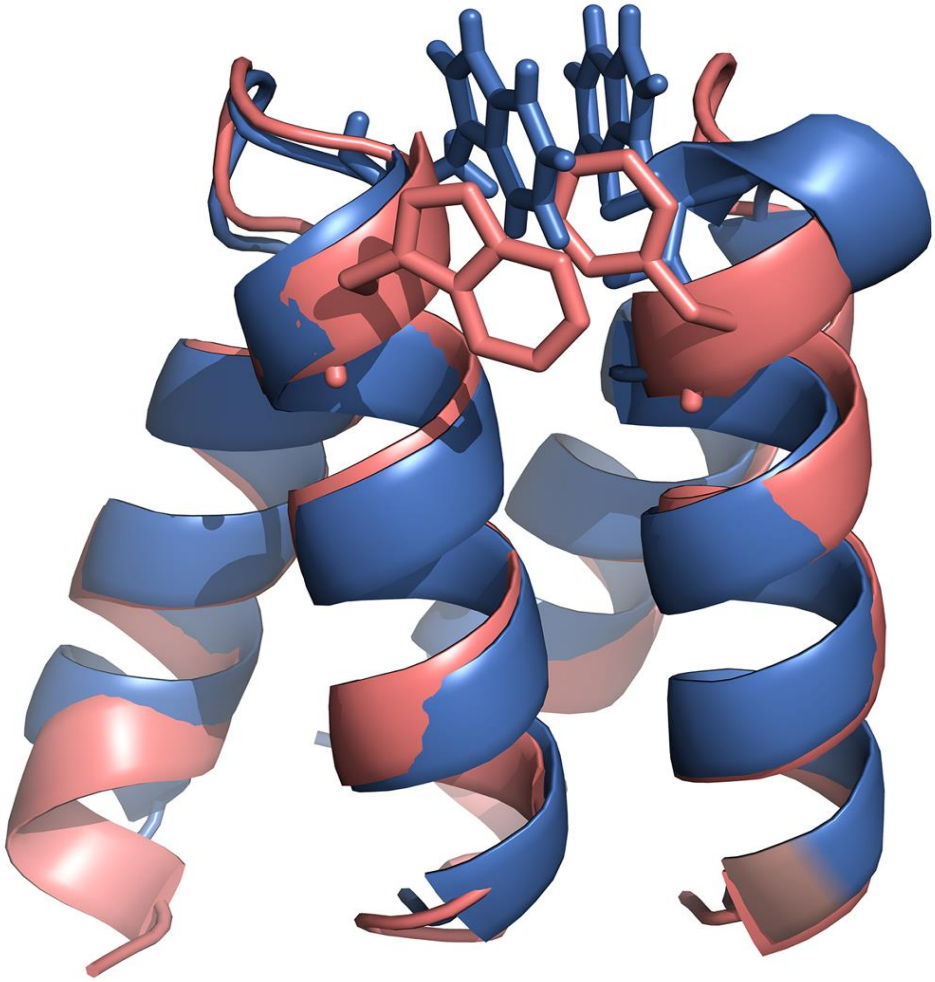


Figure 2.13 Backbone overlay of the structure of the dimeric asymmetric unit of the HEAT_R1 filament (blue) with the dimeric tandem repeat (pink) corresponding to residues 79-139 of αRep-n4-a (PDB ID: 3LTJ). The inner (concave) surfaces of the respective assemblies are oriented proximal to the viewer. Note that the formation of the Trp-Trp interaction between peptide molecules in the asymmetric unit of HEAT_R1 causes the second molecule to deviate from the expected conformation at the turn surface to accommodate the interaction. The alignment was generated using the Super command in PyMOL.

the right-handed helical twist that was observed in the crystal structures of proteins containing concatemers of LRV (PDB ID: 1LRV) and PBS_HEAT (PDB ID: 3LTJ) repeat motifs. However, a local break in symmetry occurs in the HEAT_R1 structure due to an apparent interaction between tryptophan residues within the two chains in the asymmetric unit (Figure 2.14A). In the structural model of the HEAT_R1 filament, this interaction is manifested through planar π-stacking between

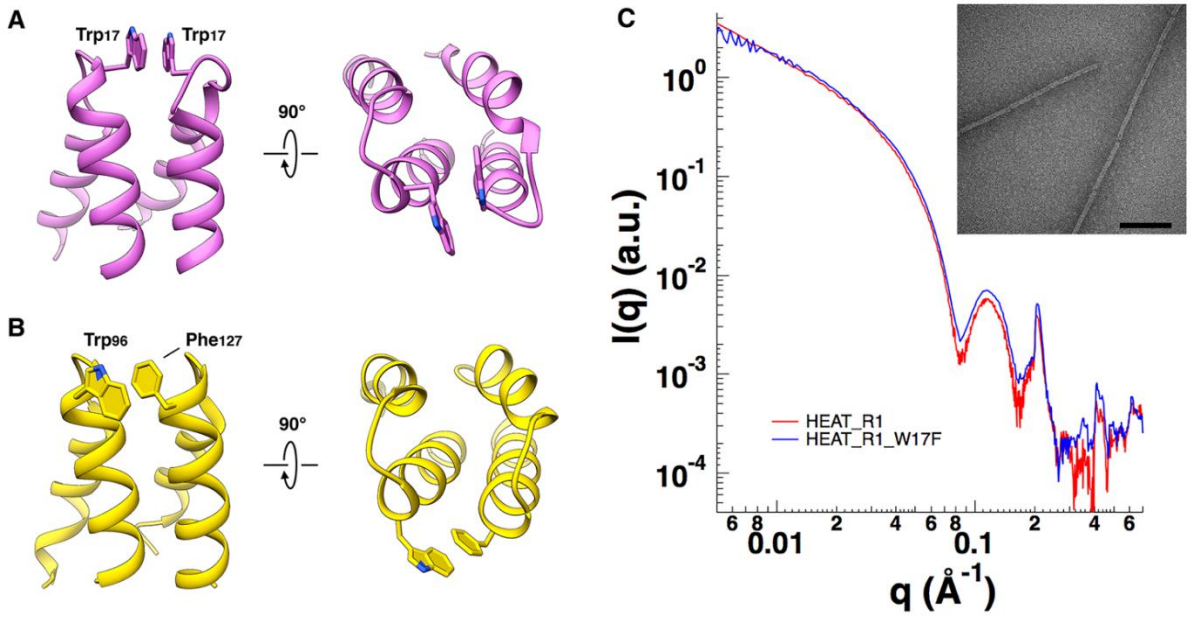


Figure 2.14. (A and B) Lateral and axial views of the p-stacking interactions between protomers in the asymmetric unit of the HEAT_R1 filament (A) and between two adjacent helical hairpin motifs ($^{79\text{-}139}$aRep-n4-a) in the crystal structure 3LTJ (B). C. Comparison of the synchrotron SAXS scattering profiles for assemblies of the HEAT_R1 peptide and the W17F mutant peptide (6 mg/mL in 10 mM MES buffer, pH 6.0, and 10 mM acetate buffer, pH 5.0, respectively). Inset: Negative stain TEM image of filament derived from self-assembly of the HEAT_R1_W17F mutant peptide (scale bar = 100 nm).

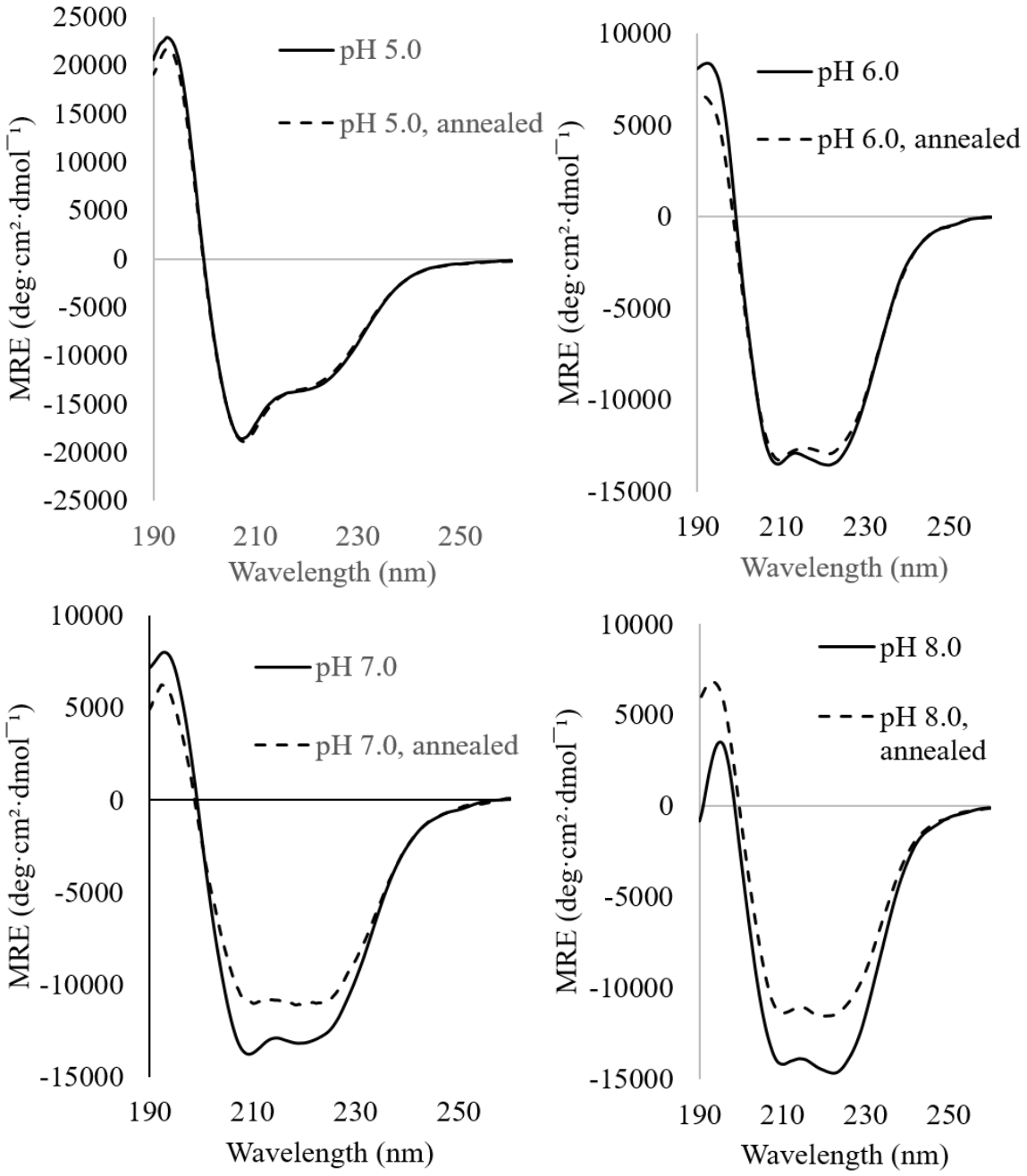


Figure 2.15 CD spectra for HEAT_R1_AW at pH 5.0, pH 7.0, and pH 8.0 (all taken at 250μM peptide concentrations). Buffers were 10 mM Acetate (pH 5.0), 10 mM MOPS (pH 7.0), and 10 mM TAPS (pH 8.0).

anti-oriented tryptophans. Consequently, only one of the two HEAT_R1 chains in the asymmetric

unit of the filament can be aligned with the consensus PBS_HEAT repeats derived from the crystal structure of 3LTJ (Figure 2.13). The geometrical requirements of the Trp-Trp interaction causes significant deviation of the other chain from the PBS_HEAT repeat structure at the turn surface, which primarily involves a large movement of the peptide backbone and a flip of the aromatic ring to promote stacking with the corresponding side-chain of a residue on the adjacent chain in the asymmetric unit. While the final resolution of the EM density map was only about 6.0 Å, the p-p stacking was adopted automatically during Phenix refinement, which suggested that the Trp-Trp interaction was energetically favorable and represented a reasonably good fit into the map. In order to probe the importance of this aromatic interaction, the tryptophan residue involved in this

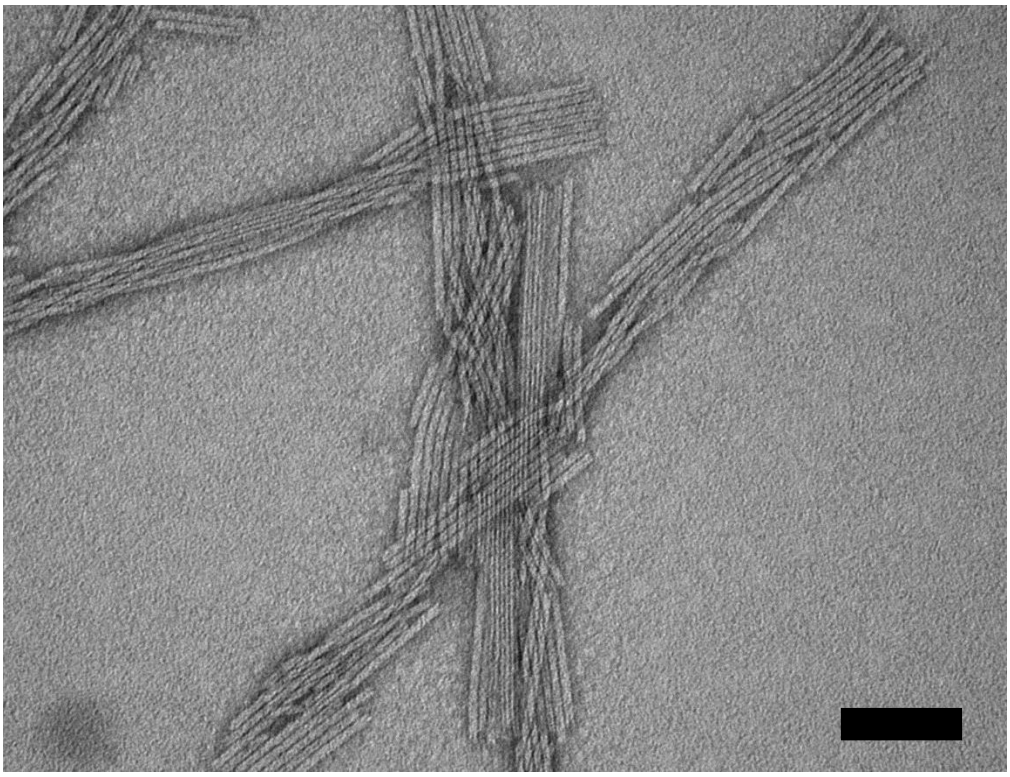


Figure 2.16 Representative negative-stain TEM of HEAT_R1_AW. Strong lateral association was seen at every tested condition. Scale bar is 100 nm.

interaction within the sequence of HEAT_R1 was mutated to a phenylalanine (Figure 2.14). Despite the potential for a weaker interaction between protomers in the asymmetric unit, the HEAT_R1_W17F mutant assembled into filaments that displayed similar properties to the wild-type sequence on the basis of negative stain TEM and SAXS measurements (Figure 2.14C). In particular, the strong diffraction peak at approximately 31 Å was maintained in the SAXS scattering curve of the HEAT_R1_W17F filaments. In the corresponding crystal structure of a more sequence-diverse PBS_HEAT tetramer, aRep-n4a (PDB ID: 3LTJ),[37] a single aromatic π-interaction occurs between Trp96 and Phe127 residues in adjacent repeat motifs at nearly equivalent positions near the *N*-terminus of the concave helix (Figure 2.14B). This interaction appears more readily accommodated within the regular geometry of the larger protein concatemer

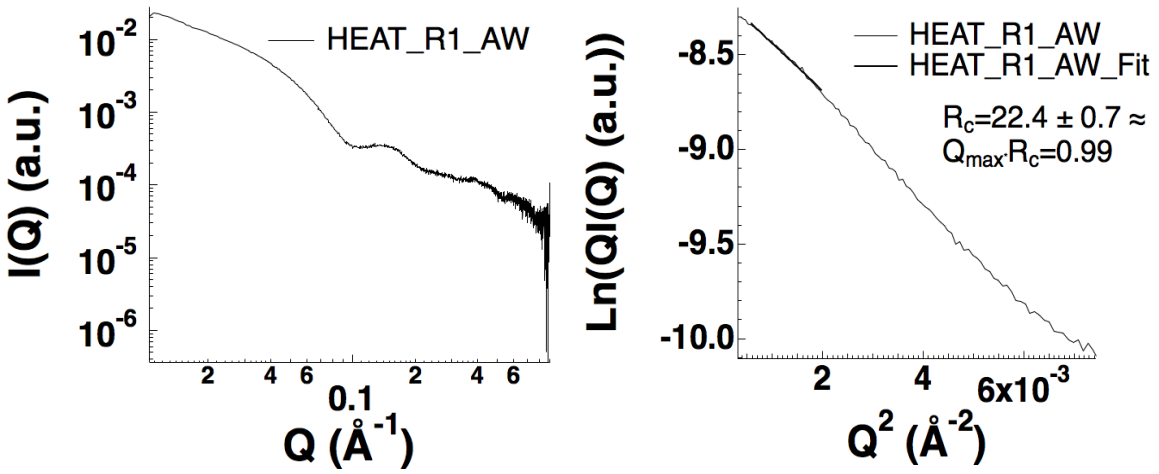


Figure 2.17 SAXS scatter (left) and Guinier (right) curves for HEAT_R1_AW. Rc is reported in Å. The observed R_c value for HEAT_R1_AW is smaller than that seen for HEAT_R1 and HEAT_R1_W17F, though this is likely caused by the intense aggregation of the filaments.

and does not result in local structural distortion, however, this protein cannot readily propagate into helical filaments due to the presence of *N*- and *C*-terminal capping motifs. A W17A,Y18W double mutant of the HEAT_R1 peptide, HEAT_R1_AW, was synthesized in which an alanine

replaced the tryptophan residue involved in the π−π interaction between peptides in the asymmetric unit. HEAT_R1_AW assembled into filaments under conditions similar to those of HEAT_R1, however, extensive lateral aggregation was observed (Figure 2.15, Figure 2.16) with concomitant loss of the 31 Å diffraction peak in the SAXS curve (Figure 2.17). These results suggest that the aromatic interaction within the protomer stabilizes the supramolecular structure and may be associated development of the axial periodicity despite the fact that the Trp residues make only one close contact (within 5 Å) with an axially adjacent protomer.

The importance of this aromatic interaction may be intuited indirectly through the absence of a chain of conserved hydrogen-bonded interactions between the main chain carbonyl group and the side chain carboxylate of Asp14 and the side chain guanidinium group of Arg20. These residues are strongly conserved at homologous positions within the sequences of PBS_HEAT and LRV motifs (Figure 2.1A) and mediate interactions between successive tandem repeats. These interactions are not observed in the structure of the HEAT_R1 filament but are preserved in the structure of the PBS_HEAT tetramer, aRep-n4-a (PDB ID: 3LTJ), even in the presence of the single aromatic interaction described above. Parmeggiani, et al., have reported the computational design of related HEAT-based tandem repeat proteins in which this Asp-Arg interaction was removed from the corresponding protein sequence.[23] However, while stable folded structures were observed from these efforts, no high-resolution structural information was reported. In the case of the HEAT_R1 filament, the potential strength of the aromatic interaction between chains in the asymmetric unit may preclude the formation of these Asp-Arg interactions. The resultant structural distortion weakens the contact between protomers located at successive turns of the helix, which introduces disorder that limits the resolution to which resultant structure of HEAT_R1 filaments can be determined. PISA analysis[41] of the HEAT_R1 filament indicated that significantly greater

surface area was buried at the lateral interfaces in comparison to the axial interfaces. Due to the presence of the dimeric asymmetric unit, two distinct lateral and axial interfaces are observed for HEAT_R1 repeats in the filament (Figure 2.14B and 2.14E). The two lateral interfaces bury 1054 $Å^2$ and 1070 $Å^2$ of surface area, which compares well to the average of 1124 $Å^2$ in buried surface area observed for lateral interactions between the consensus PBS_HEAT repeats in the crystal structure of the synthetic tetramer, 3LTJ. In contrast, the corresponding axial interfaces between protomers in the HEAT_R1 filament bury 317 $Å^2$ and 260 $Å^2$ of surface area (Figure 2.14E). The weak axial interactions may also be responsible for local unwinding of the HEAT_R1 filaments, which is frequently observed in EM image analysis (Figure 2.12). Similar local strand unwinding has been observed for chaperone-usher (CU) pili, which has been proposed as the mechanism that underlies reversible extension and retraction of the pilus under the influence of mechanical shear. Near-atomic resolution structural analysis of the corresponding pili with electron cryo-microscopy revealed that the unwinding of the helical assembly occurred at the weaker axial interface.[42-44] The cohesive interactions at the lateral interface between protomers are much stronger in the CU pili due to strand exchange between adjacent protomers derived from IgG-like domains.

Table 2.1 Refinement statistics for the peptide filament models

	HEAT_R1	LRV_M3Δ1
Helical symmetry		
Rise (Å)	3.02	1.15
Rotation (°)	34.8	-20.7
Resolution estimates (Å)		
model:map FSC (0.143/0.38/0.5)	5.1/6.0/6.3	4.0/4.4/4.5
d_{model}	5.8	4.4

d_{99}	5.5	4.3
Model vs. Data CC	0.82	0.85
Clash score, all atoms	4.15	4.61
Protein geometry		
Ramachandran favored (%)	94.6	95.2
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	0	0
Cβ deviations > 0.25 Å	0	0
RMS deviations		
Bond (Å)	0.01	0.01
Angles (°)	1.34	1.22
Molprobity score	1.69	1.57
PDB ID	6MK1	6HQE
EMDB ID	EMD-9136	EMD-0252

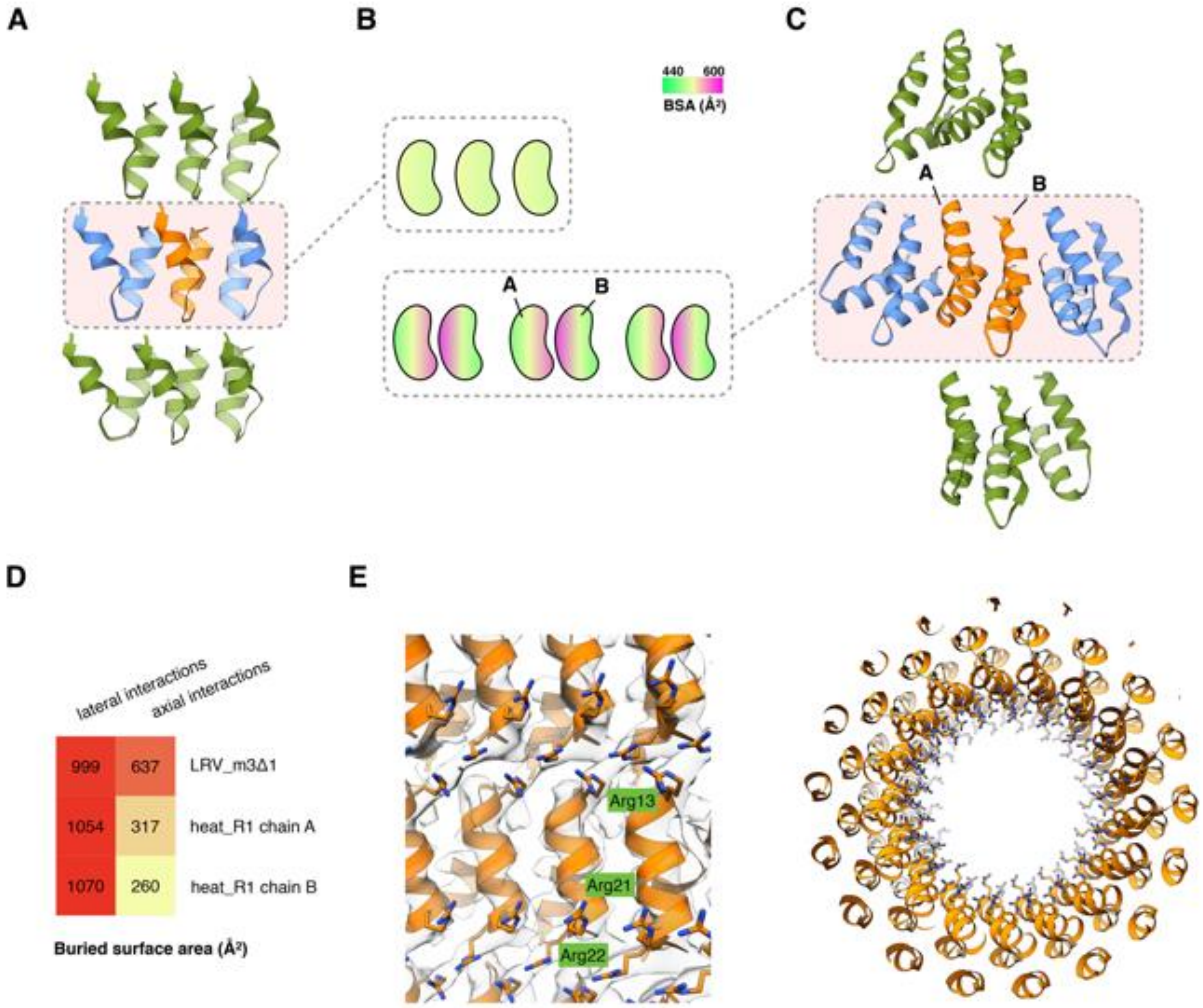


Figure 2.18. HEAT map (B) of buried surface area determined from PISA analysis (D) of lateral interactions between protomers in the 3D reconstructions of the LRV_M3Δ1 (A) and HEAT_R1 (C) filaments. (E) Lateral and axial views of the 3D reconstruction of the LRV_M3Δ1 filament in which the ordered side chains of arginine residues mediate interactions between protomers.

The nature of the cohesive interactions between protomers within the LRV_M3Δ1 filament are quite distinct from those of the HEAT_R1 filament, especially at the axial interface. The absence of aromatic residues near the *N*-terminus of the concave α-helix for LRV_M3Δ1 precludes

local pairwise de-symmetrization of adjacent subunits that would result in a structure that resembles that of the HEAT_R1 filaments. In contrast to the HEAT_R1 assembly, the position of the protomers in LRV_M3Δ1 filament is consistent with maintenance of the conserved Asp-Arg interactions between adjacent subunits in the structure, although the resolution of the corresponding structure precludes a conclusive determination. This ladder of hydrogen-bonded Asp-Arg interactions between subunits is observed between LRV domains in the crystal structure of 1LRV and between PBS_HEAT consensus motifs in the crystal structure of 3LTJ. However, neither of these proteins display axial stacking interactions that would result in the formation of helical nanotubes. Therefore, the right-handed helical twist observed in the crystal structures 1LRV and 3LTJ for concatemers of LRV and PBS_HEAT repeat motifs, respectively, must originate solely from propagation of lateral interactions between covalently linked repeat motifs. Axial contacts in the LRV_M3Δ1 filament are largely mediated through interactions between Arg13 residues and Arg22 residues within protomers located at successive turns of the helical assembly, with a contribution of Arg21 to lateral association (Figure 2.14E). Arg13 lies on the distal side of the Asp-Pro-Asp turn sequence at the *N*-terminus of the concave α-helix, while Arg22 lies near the *C*-terminus of the same helix. While Arg22 is conserved at a homologous position within the LRV domains of 1LRV, the corresponding position of Arg13 is not conserved within either LRV or PBS_HEAT motifs. These arginine residues mediate a network of non-covalent interactions that appear to be the driving force that holds together the axial interface. These axial interactions may be assisted through the pre-organization of the protomers due to the strong lateral interactions that restrict the geometry of adjacent helical hairpins.

Clusters of arginines have been observed in the form of rings, ladders, and chains at the interacting interfaces between protein oligomers,[45] and arginine residues, in general, appear to be

over-represented at protein-protein interfaces.[46,47] PISA analysis[41] of the LRV_M3Δ1 filament indicates that 999 Å^2 and 637 Å^2 of surface area is buried at the lateral and axial interfaces, respectively, between protomers. The energetics of lateral interaction for the LRV_M3Δ1 filament compares well to the corresponding interaction between lateral interfaces within the HEAT_R1 filament, as well as to the lateral interactions between LRV repeats in the crystal structure 1LRV (average buried surface area of 980 Å^2 per repeat). This lateral interaction is primarily mediated through association between concave a-helices in structurally adjacent protomers, as the convex helices are well separated as in the corresponding crystal structure. The axial interaction between LRV_M3Δ1 protomers buries nearly twice as much surface area per subunit as the corresponding set of interactions in the HEAT_R1 filament. This difference may account for the absence of filament unwinding and greater apparent persistence length for LRV_M3Δ1 filaments observed in TEM images.

The left-handed helical twist observed for the LRV_M3Δ1 filament is relatively unusual for tandem repeat proteins based on helical hairpin motifs.[35] Right-handed helical twist is predominantly observed for native proteins based on helical hairpin concatemers, although Doyle, et al., have employed computational design to afford closed oligomers of helical hairpins that display left-handed curvature.[33] Maintenance of the cohesive interactions between arginine residues at the axial interface may account for the observed differences in helical parameters for the LRV_M3Δ1 filament from those proposed by Rees and co-workers for an LRV-based nanotube.[36] The concave helices of the LRV subunits are slightly tilted rightward with respect to the super-helical axis and this packing geometry may influence the helical symmetry of the assembly through restricting the available modes in which the arginine residues can interact at the axial interface. In other words, the LRV_M3Δ1 filament may adopt the observed helical symmetry

as it can best accommodate the interactions that stabilize the axial interface under the geometrical restrictions of the lateral packing of protomers. Alternatively, the right-handed helical twist observed in the crystal structure of 1LRV may result from the presence of the *N*-terminal Fe_4-S_4 protein domain (Figure 2.1B and 2.1C), which could potentially inhibit folding of the LRV repeats with a left-handed helical twist and/or promote the formation of a right-handed helix. The attachment of a chiral endgroup has been demonstrated to induce a given helical hand in an otherwise achiral backbone to which it is attached.[48] Numerous examples of this chiral "domino" effect have been reported for peptides and synthetic foldamers.[49-51] Although the LRV repeats are chiral and may display an intrinsic preference for a specific helical geometry, a similar domino effect could be operative in the case of the 1LRV structure in which the presence of the terminal domain induces a diastereo-selection for the opposite helical hand. Elucidation of the influence of the terminal protein domain would require structural determination on the concatemer after excision of the capping domain in order to ascertain the potential relevance of these interactions to the helical hand preference.

The structural origin of the observed differences between the packing arrangements of the HEAT_R1 and LRV_M3Δ1 filaments remains an open question. One interesting observation is that the residues involved in the unique, potentially structurally determinative interactions within the respective filaments, i.e., Trp17 for HEAT_R1 and Arg13 for LRV_M3Δ1, occupy homologous positions within the corresponding sequences, i.e., immediately after the conserved Asp-Pro-Asp turn. This position at the *N*-terminus of the concave α-helix is hypervariable among native LRV and PBS_HEAT repeat sequences, which may not be surprising as residues at this site are not involved in the structurally critical lateral interactions between repeat motifs.[36-37] However, amino acid residues at the distal end of the Asp-Pro-Asp turn occupy positions that may be poised

appropriately to mediate axial interactions. The structurally similar α-helices that define the concave surface of 1LRV and 3LTJ are packed more closely together than the convex helices. Additionally, the concave helices display smaller values of the tilt angle, i.e., the angle of the helix axis with respect to a hypothetical helical axis, and a smaller and narrower range of helix crossing angles. This packing arrangement may favor not only lateral association but also axial interactions between protomers located at successive superhelical turns, if appropriately positioned amino acid residues could be introduced into the peptide sequence that are capable of mediating interactions at the concave surface. In the structure of the LRV_M3Δ1 filaments, Arg13 and Arg22 mediate these axial interactions at the concave surface through formation of a network of interactions between arginine residues (Figure 2.14E). In contrast, Trp17 cannot form a π-π interaction at the axial interface as it lacks a residue with which to interact on the *C*-terminus of the concave helix of an axially translated protomer. Instead, Trp17 interacts with the corresponding residue of a laterally translated motif, which results in a local interaction that weakens the axial interface, reduces the local symmetry, and precludes formation of the normally conserved hydrogen-bonding network involving Asp14 and Arg20 (Figure 2.13). These results suggest a potentially critical role for the residue immediately following the turn sequence in mediating axial interactions. The lack of sequence conservation at this position within the respective LRV and PBS_HEAT motifs may reflect the fact that the corresponding native proteins containing these repeat sequences have not been subjected to evolutionary pressure that would select for axial interaction. Both tryptophan and arginine residues are statistically over-represented at protein-protein interaction interfaces, especially at hot-spot sites on protein surfaces in which site-directed mutagenesis greatly diminishes protein-protein interactions.[47]

Another significant question is the relevance of the nanotube structures to the corresponding structures of the tandem repeat proteins that served as the starting point for the designs. In both the LRV and PBS_HEAT systems, similarities are observed in the structures of the respective protomers as well as the nature of the lateral interactions between subunits, while significant differences are observed in the higher order structural interactions. The LRV_M3Δ1 and HEAT_R1 peptides correspond to single repeat motifs and consequently have a greater degree of conformational freedom and, presumably, can be accommodated into a wider range of structures than repetitive proteins in which the same structural motifs are covalently linked.

In order to assess whether the structures of the synthetic nanotubes can be preserved when the repeats are covalently linked, two peptides, HEAT_dimer and LRV_dimer, were synthesized (Figure 2.19) in which the respective repeats were concatenated through use of turn linkages typically observed in the consensus sequences of the LRV and PBS_HEAT repeat motifs (Figure 2.19 A). The sequence of LRV_dimer was a direct repeat of the LRV_M3Δ1 peptide sequence,

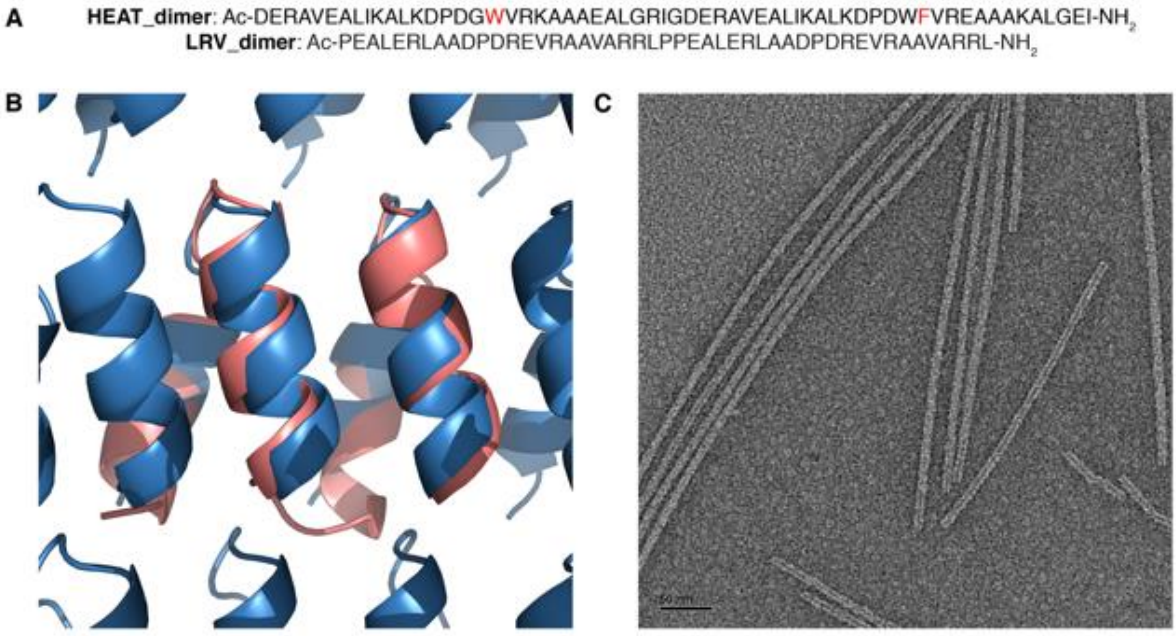


Figure 2.19 (A) Sequences of the HEAT_dimer and LRV_dimer peptides. For HEAT_dimer, aromatic residues involved in the inter-protomer interaction are highlighted in red. (B) Backbone overlay a segment of 1LRV structure (orange), $^{123\text{-}170}$1LRV, corresponding to two consecutive repeat units onto the structure of the LRV_M3Δ1 filament (blue). The backbone overlay was generated in PyMOL using the extra_fit command. (C) TEM image of assemblies of the LRV_dimer peptide.

while the sequence of HEAT_dimer was largely based on the sequence of HEAT_R1 but incorporated Trp and Phe residues at positions within the sequence that were compatible with the

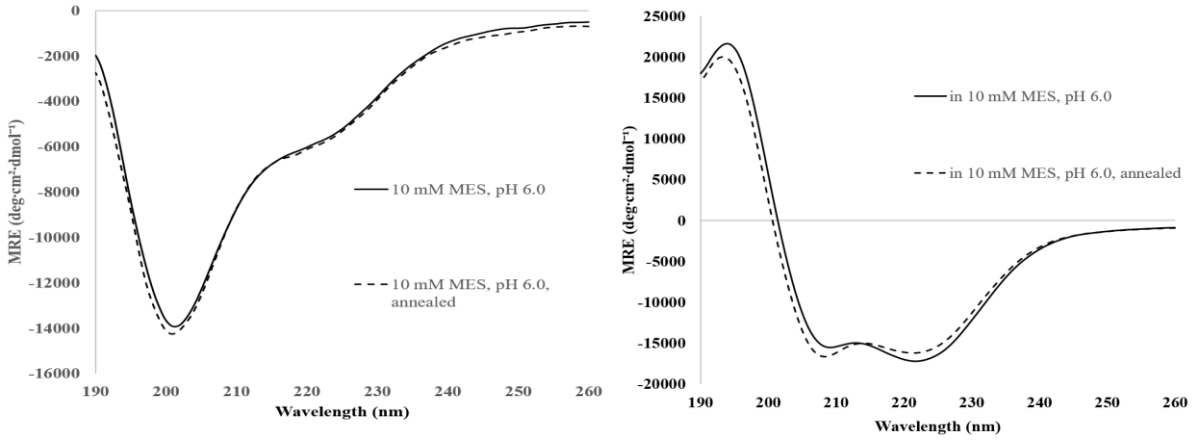


Figure 2.20 CD spectra of LRV_dimer (left) and HEAT_dimer (right) at pH 6.0 (250 µM peptide concentration, 10 mM MES).

formation of a π-interaction between adjacent repeats corresponding to amino acids ^{79}Asp-^{139}Ile in the crystal structure of αRep-n4-a (Figure 2.14B).

CD spectropolarimetric analysis of the two dimeric peptides indicated a similar conformation in solution to the corresponding LRV_M3Δ1 and HEAT_R1 peptides that contained a single repeat motif (Figure 2.20). As in the case of LRV_M3Δ1, CD spectra of the LRV_dimer peptide indicated

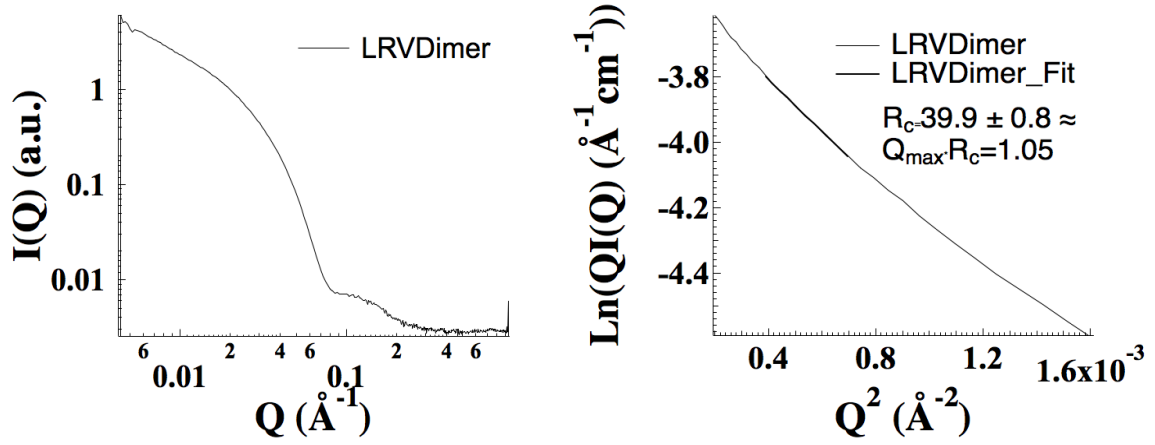


Figure 2.21 SAXS scatter (left) and Guinier (right) curves for LRV_Dimer. Rc is reported in Å.

the emergence of a persistent conformation over a period of time. The CD spectrum of HEAT_dimer peptide displayed the characteristic behavior of a classical a-helical conformation. Significant differences were observed in self-assembly behavior between the HEAT_dimer and LRV_dimer peptides. TEM imaging of the LRV_dimer indicated the presence of high aspect-ratio filaments of similar apparent diameter (~9 nm) to those observed under corresponding conditions for the LRV_M3Δ1 peptide (Figure 2.19). SAXS analysis afforded filaments with a slightly larger R_c value (39.9) than was observed for LRV_M3Δ1. It is possible that having two repeats concatenated results in less angular freedom for self-assembly. This could lead to a wider angle between repeat units, forcing the structure to have a slightly larger diameter. Surprisingly, the HEAT_dimer formed ill-defined aggregates under a range of conditions in aqueous buffered solution (Figure 2.22), including those that promoted self-assembly of the parent peptide, HEAT_R1. Notably, a segment of the 1LRV structure corresponding to two successive repeat units, $^{123\text{-}170}$1LRV, could be superimposed onto the backbone of two adjacent protomers in the LRV_M3Δ1 peptide with the an RMSD of circa 1 Å for 37 residues. The main deviations were associated with the convex helices and turn connecting the concatemers (Figure 2.19). Significant

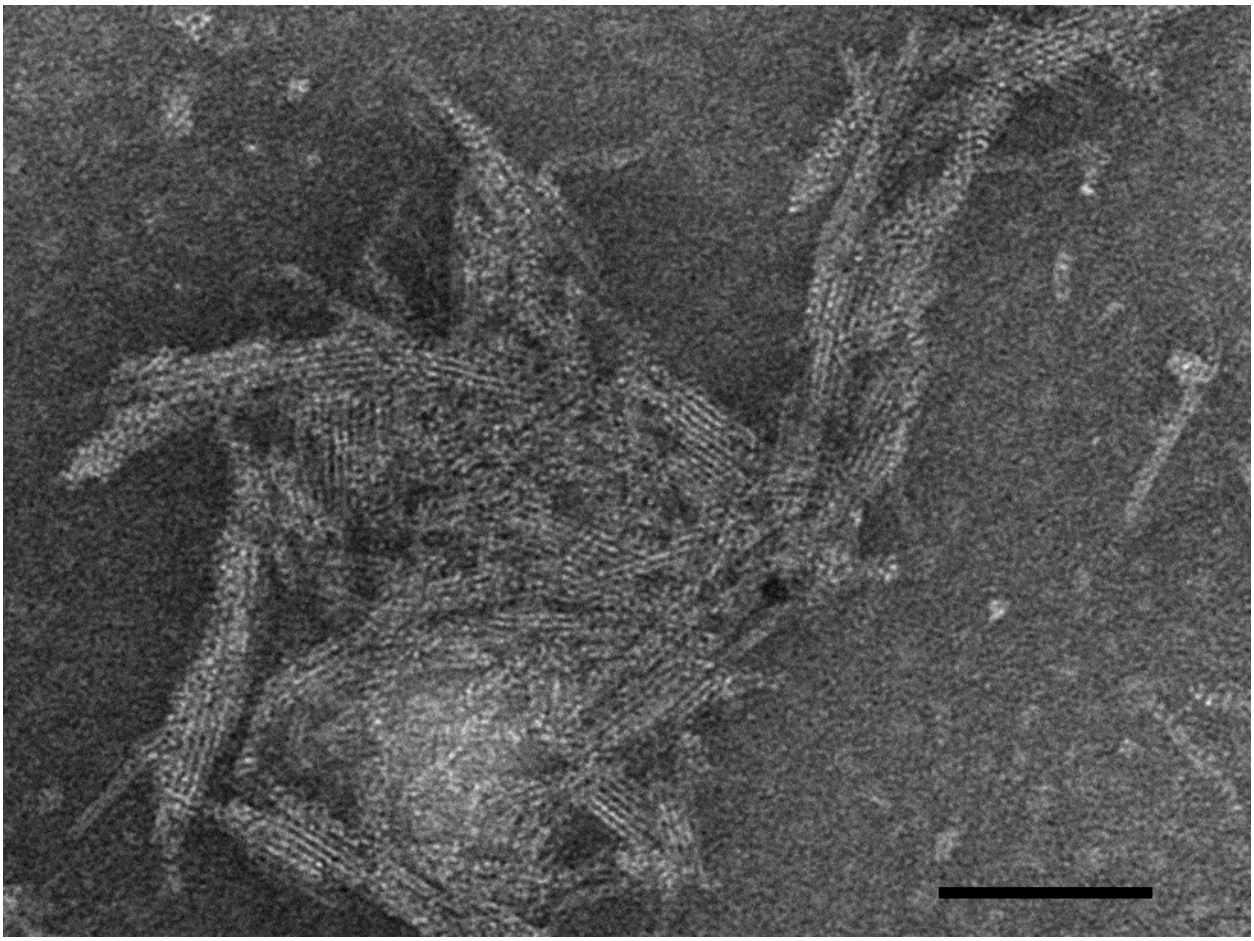


Figure 2.22 Representative TEM of assemblies of HEAT_dimer. Short nanotubes with a high propensity for lateral aggregation were observed. Scale bar is 100 nm.

deviations occurred when a similar operation was attempted for the HEAT_R1 assembly. The structure of one of the protomers with the dimeric asymmetric unit of the HEAT_R1 filament differed significantly from the structure of the repeat units in 3LTJ (Figure 2.19). The structural distortion that would result from packing of a covalently linked HEAT_dimer peptide into the HEAT_R1 filament may preclude self-assembly of the corresponding structure. The data suggest that covalently linked LRV repeats could be more easily accommodated within the structure of the LRV_M3Δ1 filament than PBS_HEAT concatemers could be accommodated within the structure of the HEAT_R1 filament. Baker and co-workers[65] recently reported that helical filaments could

be constructed from concatemers of designed tandem repeat proteins, but the protomers were not oriented such that the intrinsic symmetry of the concatemer was maintained within the filament.

2.4 Conclusion

The rich functional properties of biologically derived helical assemblies provide motivation for the *de novo* design of synthetic analogues, which, unconstrained by evolution, can be designed to perform unique functions under conditions that differ significantly from those of the biological environment. While synthetic helical assemblies have been created using nucleic acids as building blocks,[52-54] the primary substrates for *de novo* design strategies have been synthetic peptides and proteins. These material offers the opportunity for rational design over a diverse range of structural subunits. Moreover, protein- and peptide-based subunits provide the potential advantage that complex functions can be programmed into the assemblies at the sequence level, which can serve as a mechanism to couple function to assembly state. However, *de novo* design of synthetic peptide assemblies has several significant challenges that complicates the design of protein-based assemblies. Foremost among these is the apparent infinite continuum of inter-protomer interfacial geometries that are possible even for structurally simple protomers. This designability problem is most obvious in the recognition that structural polymorphism is common even among helical assemblies derived from native proteins, especially if assembled *in vitro* under non-native conditions.[15-16] Furthermore, the *de novo* design of synthetic helical assemblies necessarily requires validation of the model through structural determination at near-atomic level resolution. Currently, few structural models have been generated for helical assemblies of synthetic peptides at atomic-level detail.[17-18, 55-61,65] These structures have often revealed significant differences between the model that served as the basis for the design and the corresponding experimental structure.

The native designability of tandem repeat proteins suggested that these proteins might represent potential substrates for the construction of synthetic helical nanotubes,[62] particularly for repeat motifs in which the lateral packing permitted close axial contact between successive turns of the super-helix. Structural analysis of the helical geometry of the PBS_HEAT and LRV concatemers indicated that close contact of axial adjacent repeat motifs was feasible. We have demonstrated that the synthetic peptides HEAT_R1 and LRV_M3Δ1, which are derived from the consensus sequences of PBS_HEAT and LRV repeat motifs, respectively, do indeed form stable helical nanotubes of distinct and unique helical geometry. The results of the structural analyses of the HEAT_R1 and LRV_M3Δ1 filaments suggest that the introduction of complementary interactions between appropriately positioned residues can strengthen the axial interface, while, conversely, the absence of these interactions can weaken the interface, especially if alternative structural interactions are energetically accessible within the folding landscape. This knowledge provides new opportunities for *de novo* design of helical nanotubes from tandem repeat protein motifs through use of methods that permit site-directed introduction of stabilizing interactions at the axial interfaces between subunits. The information gleaned from the structures of the HEAT_R1 and LRV_M3Δ1 filaments provides a convenient starting point for this type of approach.

These synthetic helical nanotubes resemble the helical structures of the capsids of filamentous viruses, with respect to lateral dimension, aspect-ratio, and the presence of distinguishable concave (interior) and convex (exterior) surfaces. Filamentous viruses are exquisitely evolved macromolecular machines for controlled delivery and release, which have inspired the design of synthetic filamentous delivery vehicles, e.g., filomicelles.[66] Synthetic filamentous assemblies display promising characteristics for use in controlled delivery and release

applications, however these systems cannot fully approximate the structural and functional complexity of filamentous viruses. While synthetic helical assemblies display potential utility in these types of applications,[67] methods for the predictable design of these systems are currently limited in scope. Nevertheless, the design of peptide- and protein-based filamentous assemblies affords the opportunity to create uniquely functional and tailorable nanomaterials.

2.5 References

1. Sachse, C. Single-particle based helical reconstruction—how to make the most of real and Fourier space. *AIM Biophysics,* **2015,** *2*, 219-224.
2. Egelman, E. H. Three-dimensional reconstruction of helical polymers. *Arch Biochem Biophys,* **2015,** *581,* 54-58.
3. 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*, 209-215.
4. Wang, H.; Culver, J.N.; Stubbs, G. Structure of ribgrass mosaic virus at 2.9 A resolution: evolution and taxonomy of tobamoviruses. *J Mol Biol.,* **1997**, *269,* 769-779.
5. Morag, O.; Sgourakis, N.G.; Baker, D.; Goldbourt, A. The NMR-Rosetta capsid model of M13 bacteriophage reveals a quadrupled hydrophobic packing epitope. *Proc Natl Acad Sci U S A.* **2015,** *112*, 971-976.
6. Costa, T. R. D.; Ilangovan, A.; Ukleja, M.; Redzej, A.; Santini, J.M.; Smith, T.K.; Egelman, E.H.; Waksman, G. (2016) Structure of the bacterial sex F pilus reveals an assembly of a stoichiometric protein-phospholipid complex. *Cell,* **2016,** *166*, 1436-1444.
7. 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*, 276-279.
8. Poweleit, N.; Ge, P.; Nguyen, H. H.; Loo, R.R.; Gunsalus, R.P.; Zhou, Z. H. CryoEM structure of the Methanospirillum hungatei archaellum reveals structural features distinct from the bacterial flagellum and type IV pilus. *Nat Microbiol*, **2016,** *2*, 16222.

9. Wang, F.; Burrage, A.M.; Postel, S.; Clark, R.E.; Orlova, A.; Sundberg, E.J.; Kearns, D.B.; Egelman, E. H. A structural model of flagellar filament switching across multiple bacterial species. *Nat Commun,* **2017***, 8,* 960.
10. Lu, A.; Magupalli, V.G.; Ruan, J.; Yin, Q.; Atianand, M.K.; Vos, M.R.; Schröder, G.F.; Fitzgerald, K.A.; Wu, H.; Egelman, E.H. Unified polymerization mechanism for the assembly of ASC-dependent inflammasomes. *Cell*, **2014,** *156*, 1193-1206.
11. Knight, M.J.; Leettola, C.; Gingery, M.; Li, H.; Bowie, J.U.; A human sterile alpha motif domain polymerizome. *Protein Sci,* **2011,** *20*, 1697-706.
12. Egelman, E.H.; DeRosier, D.J. The Fourier transform of actin and other helical systems with cumulative random angular disorder. *Acta Cryst*. **1982,** *A38*, 796-799.
13. Egelman, E. H.; Francis, N.; DeRosier, D. J. F-actin is a helix with a random variable twist. *Nature,* **1982,** *298*, 131-135.
14. Wang, Y.A.; Yu, X.; Yip, C.; Strynadka, N.C.; Egelman, E.H. Structural polymorphism in bacterial EspA filaments revealed by cryo-EM and an improved approach to helical reconstruction. *Structure*. **2006,** *14,* 1189-1196.
15. Guenther, E.L.; Ge, P.; Trinh, H.; Sawaya, M.R.; Cascio, D.; Boyer, D.R.; Gonen, T.; Zhou, Z.H.; Eisenberg, D.S. (2018) Atomic-level evidence for packing and positional amyloid polymorphism by segment from TDP-43 RRM2. *Nat Struct Mol Biol.,* **2018,** *25*, 311-319.
16. Close, W.; Neumann, M.; Schmidt, A.; Hora, M.; Annamalai, K.; Schmidt, M.; Reif, B.; Schmidt, V.; Grigorieff, N.; Fändrich, M. Physical basis of amyloid fibril polymorphism. *Nat Commun.,* **2018,** *9*, 699.

17. 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*, 280-289.

18. 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-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat Methods*, **2015,** *12,* 361-365.

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

20. 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,* 15565-15578.

21. Eaton, W. A.; Hofrichter, J. Sickle cell hemoglobin polymerization. *Adv Protein Chem*., **1990***, 40,* 63-279.

22. Kajava, A. Tandem repeats in proteins: from sequence to structure. *J Struct Biol,* **2012,** *179,* 279-288.

23. 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*, 563-575.

24. Plückthun, A. Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. *Annu Rev Pharmacol Toxicol,* **2015,** *55*, 489-511.

25. Kajander, T.; Cortajarena, A.L.; Regan, L. Consensus design as a tool for engineering repeat proteins. *Methods Mol Biol,* **2006***, 340,* 151-170.

26. Kobe, B.; Kajava, A. V. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem Sci.* **2000,** *25*, 509-515.

27. Das, A.K.; Cohen, P.W.; Barford, D. The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J,* **1998,** *17*, 1192-1199.

28. Michaely, P.; Tomchick, D. R.; Machius, M.; Anderson, R. G. Crystal structure of a 12 ANK repeat stack from human ankyrinR. *EMBO J,* **2002,** *21,* 6387-6396.

29. Huber, A. H.; Nelson, W. J.; Weis, W. I. Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell*, **1997,** *90*, 871-882.

30. Groves, M. R.; Hanlon, N.; Turowski, P.; Hemmings, B. A.; Barford, D. The structure of the protein phosphatase 2A PR65/ subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. *Cell,* **1999,** *96*, 99-110.

31. Kobe, B.; Deisenhofer, J. Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature*, **1993,** *366*, 751-756.

32. Rämisch, S.; Weininger, U.; Martinsson, J.; Akke, M.; André, I. Computational design of a leucine-rich repeat protein with a predefined geometry. *Proc Natl Acad Sci U S A,* **2014,** *111,* 17875-17880.

33. Doyle, L.; Hallinan, J.; Bolduc, J.; Parmeggiani, F.; Baker, D.; Stoddard, B. L.; Bradley, P. Rational design of alpha-helical tandem repeat proteins with closed architectures. *Nature,* **2015,** *528*, 585-588.

34. Park, K.; Shen, B. W.; Parmeggiani, F.; Huang, P. S.; Stoddard, B. L.; Baker, D. Control of repeat-protein curvature by computational protein design. *Nat Struct Mol Biol*, **2015,** *22,* 167-174.

35. 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*, 580-584.

36. Peters, J. W.; Stowell, M. H.; Rees, D. C. A leucine-rich repeat variant with a novel repetitive protein structural motif. *Nat Struct Biol,* **1996,** *3*, 991-994.

37. Urvoas, A.; Guellouz, A.; Valerio-Lepiniec, M.; Graille, M.; Durand, D.; Desravines, D. C.; van Tilbeurgh, H.; Desmadril, M.; Minard, P. Design, production and molecular structure of a new family of artificial alpha-helicoidal repeat proteins (αRep) based on thermostable HEAT-like repeats. *J Mol Biol,* **2010,** *404,* 307-327.

38. Kajander, T.; Cortajarena, A. L.; Main, E. R.; Mochrie, S. G.; Regan, L. (2005) A new folding paradigm for repeat proteins. *J Am Chem Soc*, **2005,** *127*, 10188-10190.

39. Cortajarena, A. L.; Regan, L. Calorimetric study of a series of designed repeat proteins: modular structure and modular folding. *Protein Sci,* **2011,** *20*, 336-340.

40. Aksel, T.; Barrick, D. Direct observation of parallel folding pathways revealed using a symmetric repeat protein system. *Biophys J,* **2014,** *107*, 220-232.

41. Krissinel, E.; Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol,* **2007,** *372*, 774-797.

42. 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*, 269-278.

43. Spaulding, C. N.; Schreiber, H. L. 4th; Zheng, W.; Dodson, K. W.; Hazen, J. E.; Conover, M. S.; Wang, F.; Svenmarker, P.; Luna-Rico, A.; Francetic, O.; Andersson, M.; Hultgren,

S,; Egelman, E. H. Functional role of the type 1 pilus rod structure in mediating host-pathogen interactions. *Elife,* **2018,** *7,* e31662.

44. Hospenthal, M. K.; Zyla, D.; Costa, T. R. D.; Redzej, A.; Giese, C.; Lillington, J.; Glockshuber, R.; Waksman, G. The cryoelectron microscopy structure of the type 1 chaperone-usher pilus rod. *Structure,* **2017,** *25,* 1829-1838.
45. 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,* 7006-7013.
46. Kufareva, I.; Budagyan, L.; Raush, E.; Totrov, M.; Abagyan, R. PIER: protein interface recognition for structural proteomics. *Proteins,* **2007,** *67,* 400-417.
47. Moreira, I.S.; Fernandes, P. A.; Ramos, M. J. Hot spots--a review of the protein-protein interface determinant amino-acid residues. *Proteins,* **2007,** *68,* 803-812.
48. Yashima, E.; Maeda, K.; Iida, H.; Furusho, Y.; Nagai, K. Helical polymers: Synthesis, structures, and functions. *Chem Rev,* **2009,** *109,* 6102-6211.
49. Clayden, J.; Castellanos, A.; Solà, J.; Morris, G. A. Quantifying end-to-end conformational communication of chirality through an achiral peptide chain. *Angew Chem Int Ed Engl,* **2009,** *48,* 5962-5965.
50. Kamikawa, K.; Fukumoto, K.; Yoshihara, K.; Furusyo, M.; Uemura, M.; Takemoto, S.; Matsuzaka, H. Induction of one-handed helical oligo(*p*-benzamide)s by domino effect based on planar-axial-helical chirality relay. *Chem Commun,* **2009,** 1201-1203.
51. Dolain, C.; Jiang, H.; Léger, J. M.; Guionneau, P.; Huc, I. Chiral induction in quinoline-derived oligoamide foldamers: assignment of helical handedness and role of steric effects. *J Am Chem Soc,* **2005,** *127,* 12943-12951.

52. Ding, B.; Wu, H.; Xu, W.; Zhao, Z.; Liu, Y.; Yu, H.; Yan, H. Interconnecting gold islands with DNA origami nanotubes. *Nano Lett,* **2010,** *10,* 5065-5069.

53. Liu, D.; Park, S.H.; Reif, J. H.; LaBean, T. H. DNA nanotubes self-assembled from triple-crossover tiles as templates for conductive nanowires. *Proc Natl Acad Sci U S A,* **2004,** *101,* 717-722.

54. Rothemund, P. W.; Ekani-Nkodo, A.; Papadakis, N.; Kumar, A.; Fygenson, D. K.; Winfree, E. Design and characterization of programmable DNA nanotubes. *J Am Chem Soc*, **2004,** *126*, 16344-16352.

55. Zhang, S. Q.; Huang, H.; Yang, J.; Kratochvil, H.T.; Lolicato, M.; Liu, Y.; Shu, X.; Liu, L.; DeGrado, W. F. (2018) Designed peptides that assemble into cross-α amyloid-like structures. *Nat Chem Biol,* **2018,** *14,* 870-875.

56. Lee, M.; Wang, T.; Makhlynets, O. V.; Wu, Y.; Polizzi, N. F.; Wu, H.; Gosavi, P. M.; Stöhr, 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,* 6191-6196.

57. 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,* 8102-8105.

58. Nagy-Smith, K.; Beltramo, P. J.; Moore, E.; Tycko, R.; Furst, E. M.; Schneider, J. P. Molecular, Local, and Network-Level Basis for the Enhanced Stiffness of Hydrogel Networks Formed from Coassembled Racemic Peptides: Predictions from Pauling and Corey. *ACS Cent Sci,* **2017,** *3,* 586-597.

59. 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*, 9816-21.

60. 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*, 7562-72.

61. Kajander, T.; Cortajarena, A. L.; Mochrie, S.; Regan, L. Structure and stability of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR proteins. *Acta Crystallogr D Biol Crystallogr*, **2007,** *63*, 800-811.

62. Sanchez-deAlcazar, D.; Mejias, S. H.; Erazo, K.; Sot, B.; Cortajarena, A. L. Self-assembly of repeat proteins: Concepts and design of new interfaces. *J Struct Biol*, **2018,** *201*, 118-129.

63. Afonine, P. V.; Klaholz, B. P.; Moriarty, N. W.; Poon, B. K.; Sobolev, O. V.; Terwilliger, T. C.; Adams, P. D.; Urzhumtsev, A. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr D Struct Biol*, **2018,** *74*, 814-840.

64. Guellouz, A.; Valerio-Lepiniec, M.; Urvoas, A.; Chevrel, A.; Graille, M.; Fourati-Kammoun, Z.; Desmadril, M.; van Tilbeurgh, H.; Minard, P. Selection of specific protein binders for pre-defined targets from an optimized library of artificial helicoidal repeat proteins (alphaRep). *PLoS One,* **2013,** *8*, e71512.

65. Shen, H.; Fallas, J. A.; Lynch, E.; Sheffler, W.; Parry, B.; Jannetty, N.; Decarreau, J.; Wagenbach, M.; Vicente, J.; Chen, J.; Wang, L.; Dowling, Q.; Oberdorfer, G.; Stewart, L.; Wordeman, L.; de Yoreo, J.; Jacobs-Wagner, C.; Kollman, J.; Baker, D. De novo design of self-assembling helical protein filaments. *Science*, **2018,** *362*, 705-709.

66. Geng, Y.; Dalhaimer, P.; Cai, S.; Tsai, R.; Tewari, M.; Minko, T.; Discher, D. E. Shape effects of filaments versus spherical particles in flow and drug delivery. *Nat Nanotechnol,* **2007,** *2,* 249-255.

67. Wu, Y.; Norberg, P.; Reap, E.; Congdon, K.; Fries, C.; Kelly, S.; Sampson, J.; Conticello, V.P.; Collier, J. A supramolecular vaccine platform based on alpha-helical peptide nanofibers. *ACS Biomaterials Science & Engineering*, **2017**, *3,* 3128–3132.

Chapter III. Concatenation and Functionalization of Tandem-Repeat Protein-Based Helical Nanotubes

3.1 Increasing Assembly Complexity Necessitates Concatenation

While the relative speed of peptide synthesis makes it an attractive method for screening mutants of a repeat sequence, advanced designs necessitate the use of larger assembly modules. Nature often employs large tandem repeat proteins to carry out complex functions, such as binding and transporting other proteins[1-5]. Because of this, TRPs containing multiple repeats have been used in every applications-driven study to date[6-20]. One can imagine that an assembly comprising single peptide repeats could be heavily chemically functionalized with little to no steric hindrance. However, functionalizing these assemblies with other protein partners would be simply impossible, as most proteins are larger than a single tandem repeat protein subunit. That is, if we were to try functionalizing a TRP scaffold like HEAT_R1 or LRV_M3Δ1 with another protein, we cannot use a scheme that relies on 1:1 loading of the functional protein. There are methods for functionalizing at less than a 1:1 ratio, however, these often rely on limiting the availability of one of the protein partners, and lead to asymmetric loading. To achieve more precise stoichiometric control, one could use a 1:1 loading scheme, with the caveat that the TRP scaffold contains multiple TRP subunits. We hypothesized that a stable concatemer of HEAT_R1 should be able to assemble, even in the presence of a genetically fused protein. This motivated the generation of a six-repeat containing concatemer of HEAT_R1, named HEAT_6R.

3.2 HEAT_6R: A TRP Concatemer

HEAT_R1: **Ac-**DERAVEALIKALKDPDWYVRKAAAEALGRI**-NH2**

HEAT_6R: **M(**GDERAVEALIKALKDPDGWVRKAAAEALGRI

GDERAVEALIKALKDPDWFVREAAAKALGEI**)3**

GSMHHHHHH

Figure 3.1 Comparison of the sequences of HEAT_R1 and HEAT_6R. The convex helices are color-coded red, and the concave helices are color-coded blue. Capping groups are denoted in bold.

In the Minard group's paper detailing a new class of HEAT repeat, from which we derived the HEAT_R1 sequence, it was shown that thermal stability was increased with higher numbers of concatenated HEAT repeats[21]. A great deal of flexibility was seen in the HEAT_R1 cryo-EM reconstruction, leading to a variable helical pitch and limited resolution (6.0 Å). Though the HEAT_R1 peptide was quite thermal stable, with no observed unfolding at 90°C, we reasoned that a concatemer of the HEAT_R1 peptide would have less flexibility than a single repeat. HEAT_6R is a concatemer of the HEAT_R1 sequence containing six helix-turn-helix subunits. Rather than six exact repeats of the HEAT_R1 sequence, charge-complementary dimers were

designed, and three of these were concatenated to promote solubility. A 6x His-Tag was included

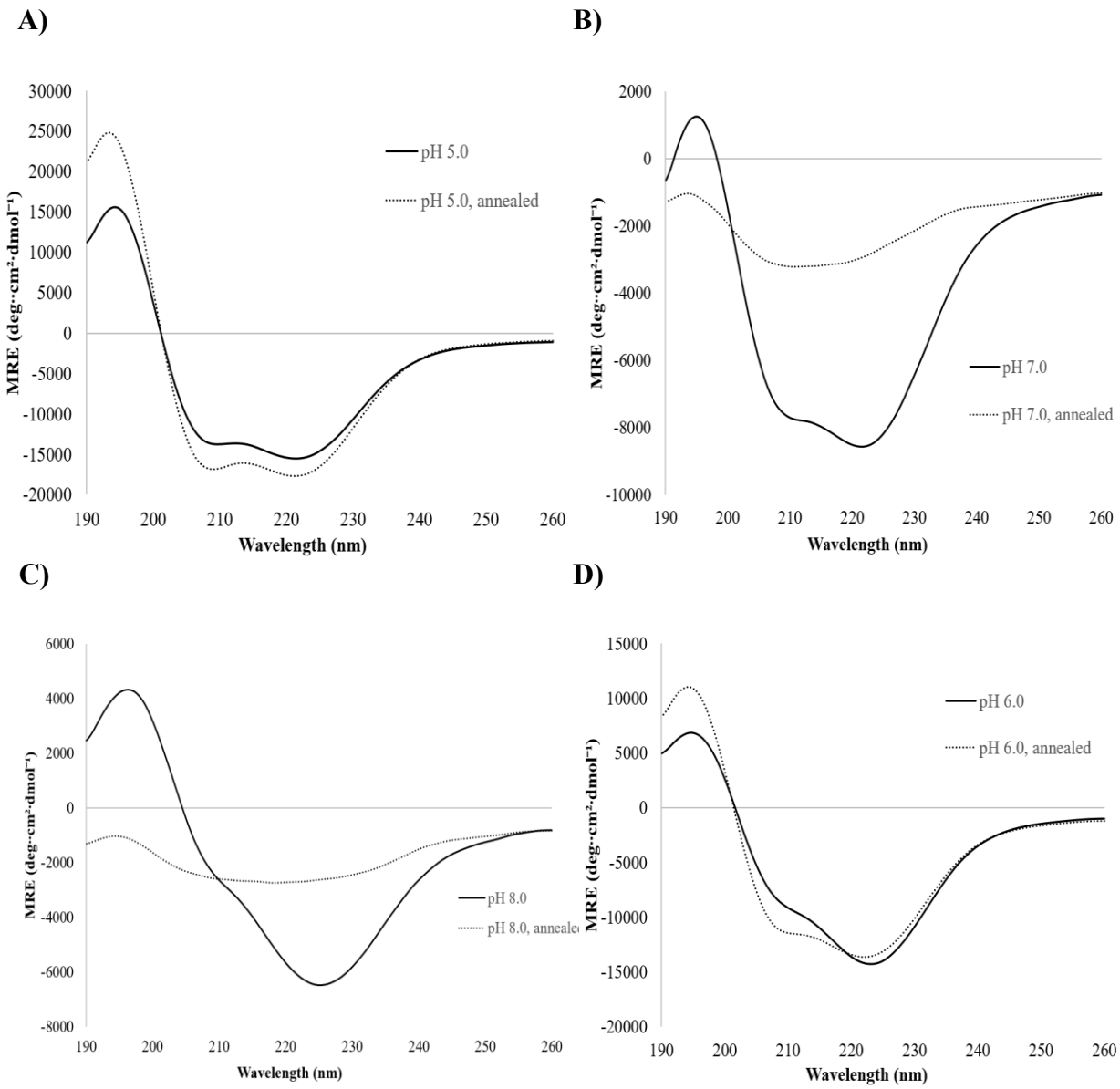


Figure 3.2 CD spectra for 250 µM HEAT_6R at pH 5.0 (a) , pH 6.0 (b), pH 7.0 (c), and pH 8.0 (d) . Buffers were 10 mM Acetate (pH 5.0), 10 mM MES (pH 6.0), 10 mM MOPS (pH 7.0), and 10 mM TAPS (pH 8.0).

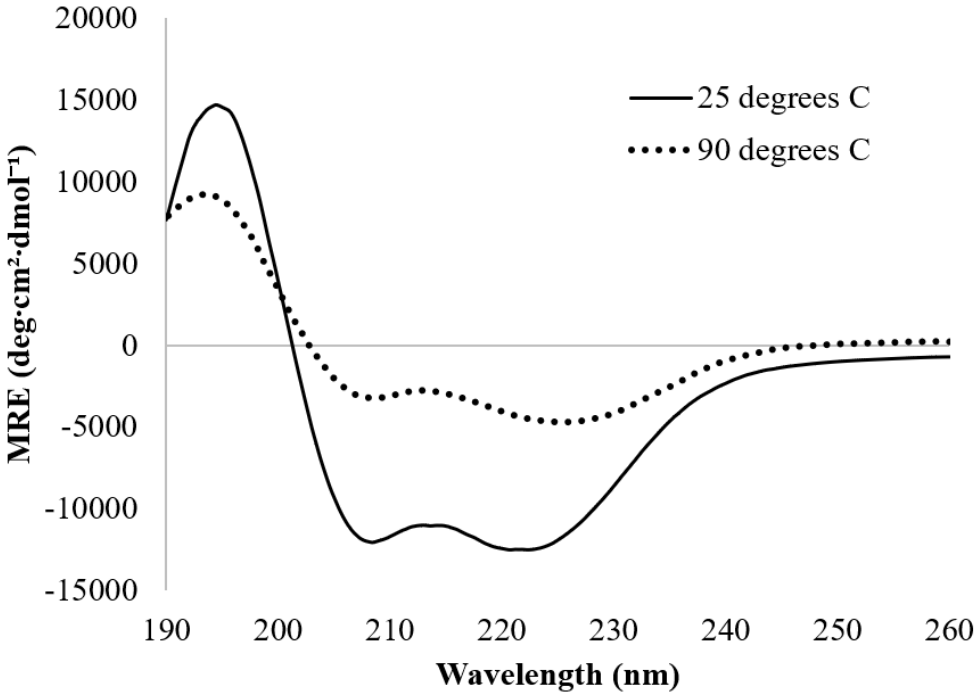


Figure 3.3 Variable Temperature CD scans for 250 µM HEAT_6R at 25 and 90 degrees C. No thermal transition was observed upon heating to 90 degrees C, and a helical signature is still seen.

at the C-terminus to provide a handle for nickel affinity column chromatography. The HEAT_6R plasmid was amplified through growth in Top10F' *E. coli* cells and then transformed into BL21 cells for IPTG-induced over-expression. After purifying HEAT_6R with a nickel column, the lyophilized protein was purified via HPLC.

The lyophilized protein was then assembled at 3 mg·mL^{-1} in buffers ranging from pH 5-8. The CD signal was consistent with that of an α-helix, as seen in Figure 3.2. The protein was completely soluble at pH 5.0, precipitated at pH 6.0 and pH 7.0, and was slightly insoluble at pH 8.0; the CD signal intensity follows this trend, with the most intense signal at pH 5.0. As predicted, HEAT_6R was resistant to thermal denaturation, with no unfolding observed with slow heating to

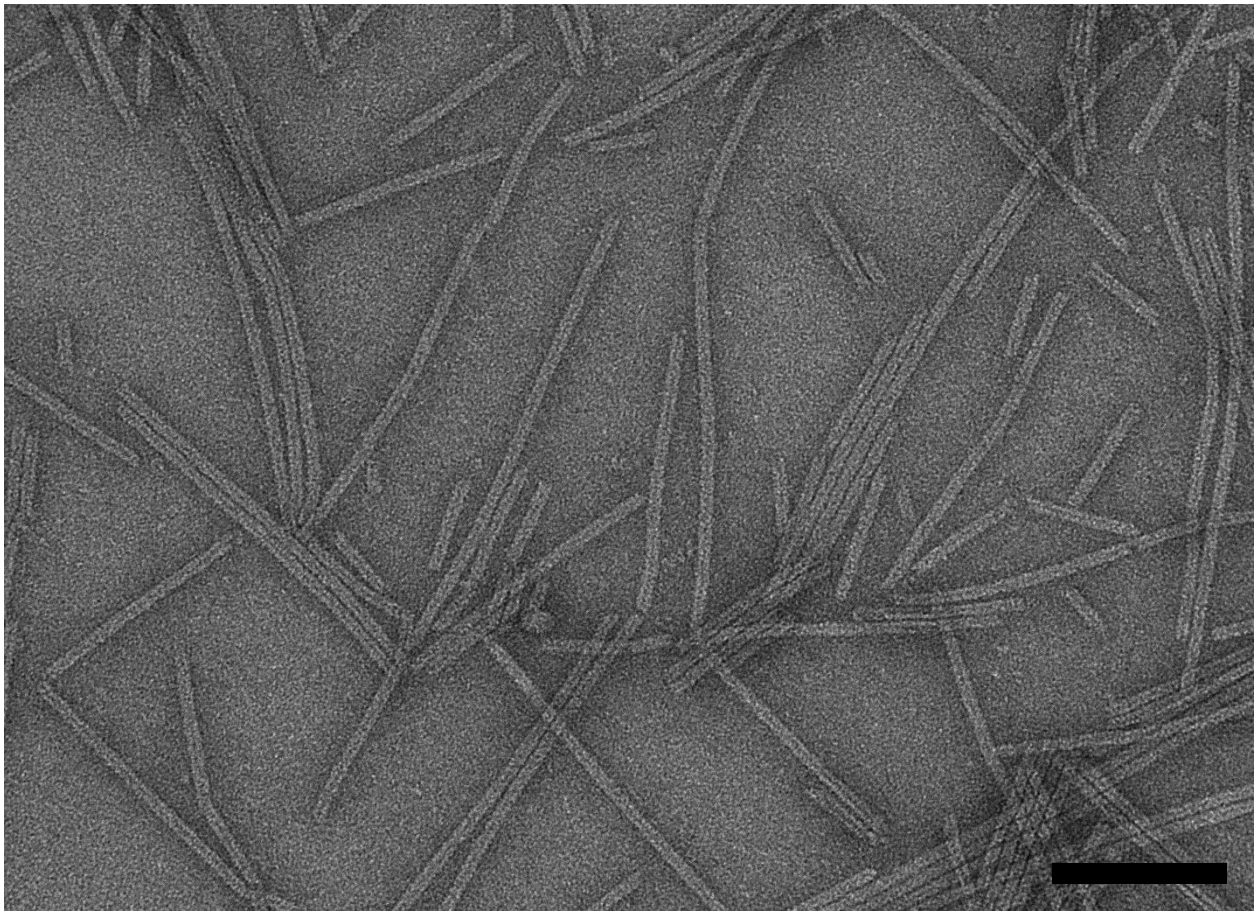


Figure 3.4 Negative-stain TEM image of HEAT_6R at pH 5.0, in 10 mM Acetate buffer. Morphology is consistent with HEAT_R1, and hollow cylinders in general. Scale bar is 100 nm.

90˚C (Figure 3.3). Via TEM, HEAT_6R was indistinguishable from HEAT_R1; high aspect ratio nanotubes were observed, which possess a smooth, featureless outer edge, along with a clearly distinguishable inner lumen (Figure 3.4). Moreover, the ImageJ measured diameter of the HEAT_6R nanotubes was identical to that of HEAT_R1 (9 nanometers). To verify that the HEAT_6R nanotubes assemble in the same fashion as the HEAT_R1 nanotubes, we analyzed HEAT_6R using SAXS. In the case of HEAT_R1, a classic hollow cylinder SAXS scattering profile was observed, with the notable addition of a Bragg-like diffraction peak, which

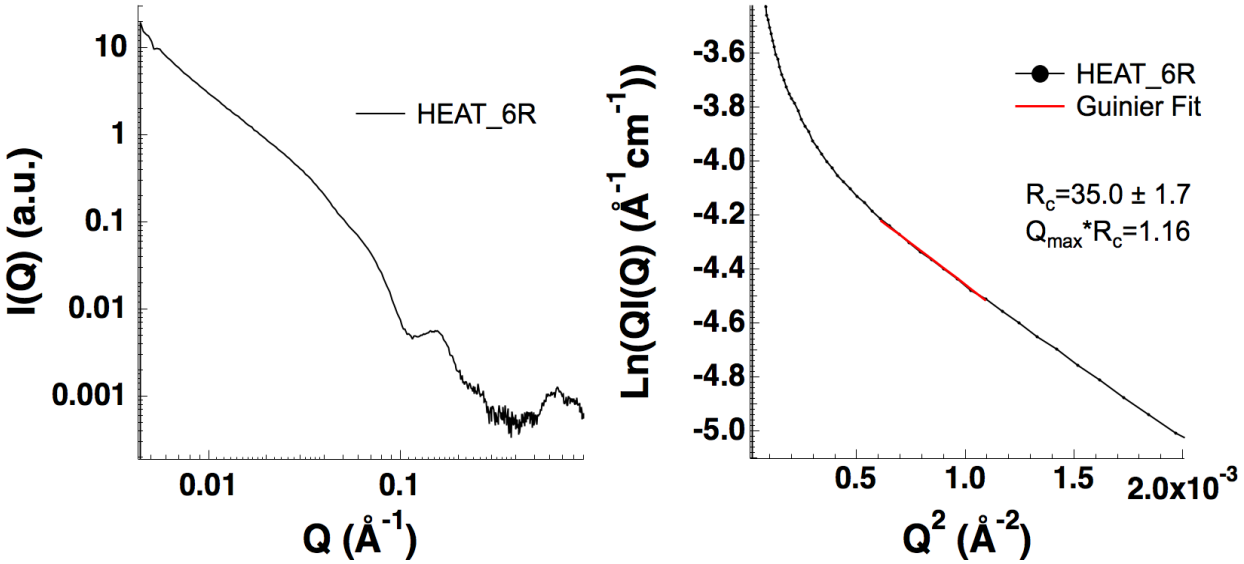


Figure 3.5 Small-angle x-Ray Scattering profile (left) and Guinier plot (right) of HEAT_6R. The strong diffraction peak observed in HEAT_R1 is not observed in HEAT_6R. However, despite this, the Guinier-derived Rc value is within error of that determined for HEAT_R1.

corresponded to a distance of 30.8 Å (the height of the helical hairpin subunit). When a similar analysis was done on the HEAT_6R nanotubes, we observed a similar scattering profile, which confirmed the hollow cylinder morphology and monodispersity in diameter. However, we did not observe a Bragg-like diffraction peak in the wide-angle scattering range. This suggests a relative lack of small-distance order in the nanotube assemblies, or variation in the helical pitch of the assembly. We initially believed this to be due to a lower protein concentration, which could lead to a lower signal to noise ratio in SAXS measurements, so we analyzed HEAT_6R at both 5 mg·mL^{-1} and 10 mg·mL^{-1} concentrations (at higher than 10 mg·mL^{-1} concentrations, the solubility of HEAT_6R is greatly decreased). However, we saw little difference in the scattering profiles of HEAT_6R for the two different concentrations, leading us to believe that the lack of diffraction is inherent to the protein itself, rather than the assembly conditions. The Guinier plot of the SAXS

data revealed that the diameter of the assemblies was very similar to what was observed for HEAT_R1.

Assemblies of HEAT_6R were incubated with TMV and freeze-dried prior to loading into a scanning transmission electron microscope. Upon freeze-drying, a large proportion of the nanotube assemblies aggregated, leading to difficulties in collecting mass-per-length measurements. Despite these complications, ~250 measurements were averaged, and a standard

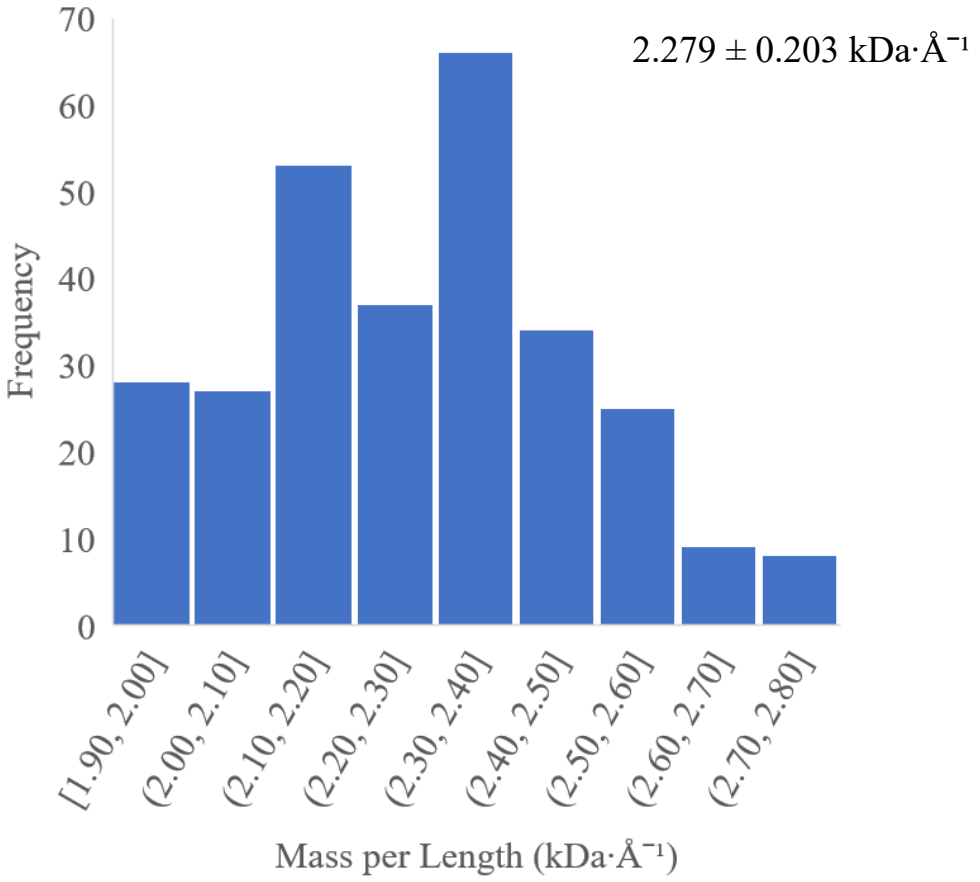


Figure 3.6 STEM-based mass per length histogram for HEAT_6R. The average MPL value for HEAT_6R was calculated to be 2.279 ± 0.203 kDa·Å⁻¹ (as compared to 2.356 ± 0.214 kDa·Å⁻¹ for HEAT_R1).

deviation was calculated. The average MPL value for HEAT_6R was calculated to be 2.279 ± 0.203 kDa·Å^{-1}. This value is well within the first standard deviation of the calculated MPL value of HEAT_R1, 2.356 ± 0.214 kDa·Å^{-1}. Based on the negative-stained EM, the SAXS Guinier plot, and the STEM MPL measurement, HEAT_R1 and HEAT_6R seem to be identical.

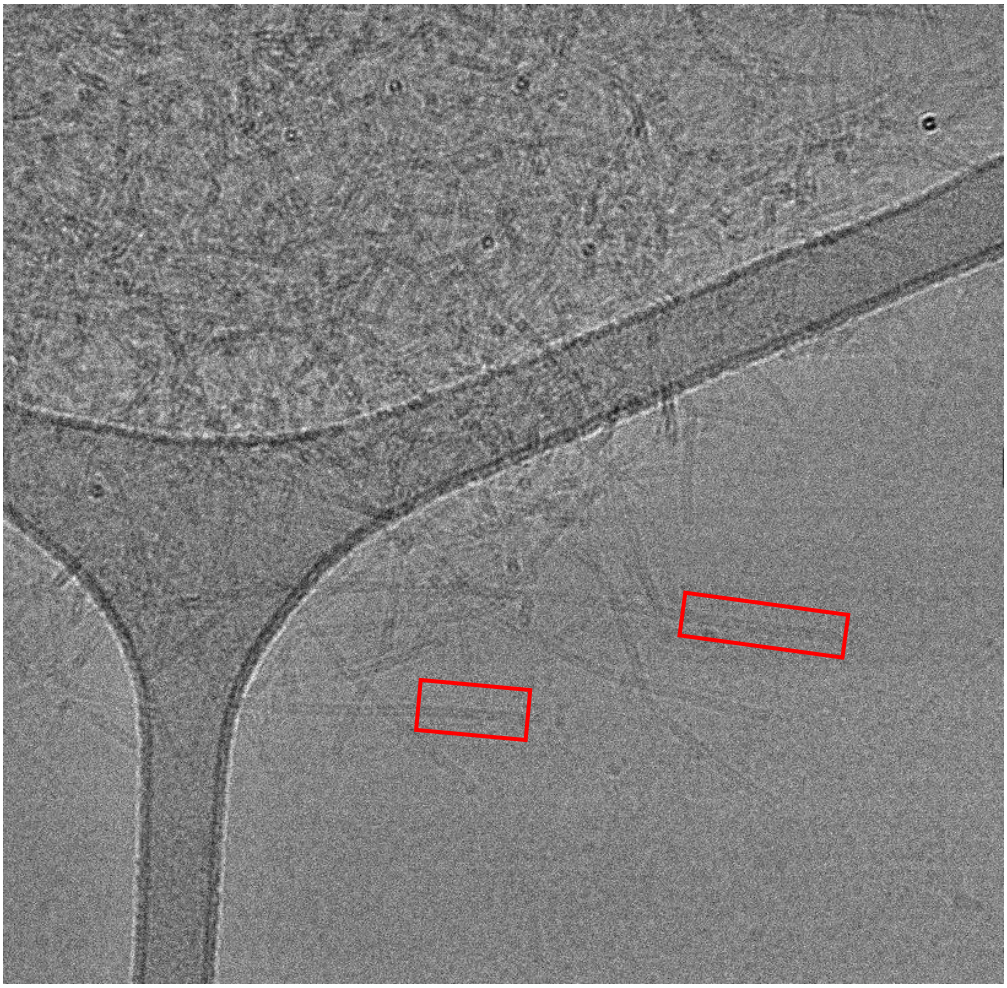


Figure 3.7 Cryo-electron microscope image of HEAT_6R in 10 mM Acetate, pH 5.0. The lighter gray background corresponds to the holes between the lacy carbon (darker gray background) of the cryo-EM grid. Some filaments are suitable for single particle analysis (red boxes), but most display an affinity for the carbon, and tend to aggregate.

In an effort to gain high resolution structural data for the HEAT_6R concatemer, we employed cryo-electron microscopy to image the assemblies. We used the same vitrification parameters to freeze grids of HEAT_6R. However, upon vitrification, the HEAT_6R assemblies tended to aggregate, and preferentially stick to the carbon film in between the holes in the grid as seen in Figure 3.7. Enough filaments remain isolated in the holes for the sample to be

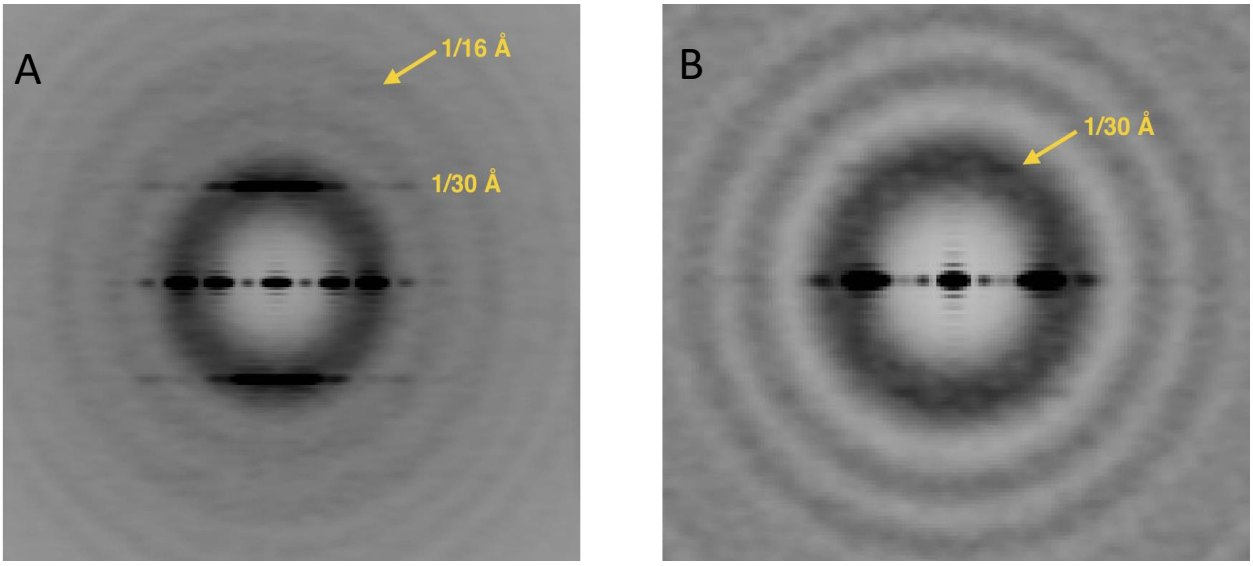


Figure 3.8 Power Spectra for HEAT_R1 (A) and HEAT_6R (B) generated via boxing of cryo-EM images. Layer lines extend out to 1/16 Å in HEAT_R1, but only out to 1/30 Å for HEAT_6R. The presence of the higher order layer lines in HEAT_R1 predicts a more well-ordered structure, suitable for high resolution reconstuction.

reconstructed, but a power spectrum analysis suggested that high-resolution reconstruction was impossible (Figure 3.8). We reasoned that the inherent flexibility of a single repeat assembly (HEAT_R1) afforded sufficient rotational and translational freedom for the subunits to pack in an optimal fashion. By covalently linking the subunits, we remove several of these degrees of freedom, and thereby generate a less stable nanotube. However, the lack of a high-resolution structure does not preclude us from making predictions or generating more complex materials; it only makes this task more challenging.

3.3 mCherry_HEAT: Design of a functional nanotube

Our ability to generate a monodisperse diameter, pH stable nanotube from HEAT concatemers led to the design of a more complex material. The Howarth group developed a general method for genetic fusion of two proteins (Figure 3.9)[22-23]. It was shown via crystallographic studies that the CnaB2 protein from *Streptococcus pyogenes* contains a single isopeptide bond, as is common in extracellular proteins in Gram positive bacteria. The likely mechanism for this isopeptide bond formation involves a catalytic triad (Figure 3.9B); the carbonyl carbon of Asp[117]

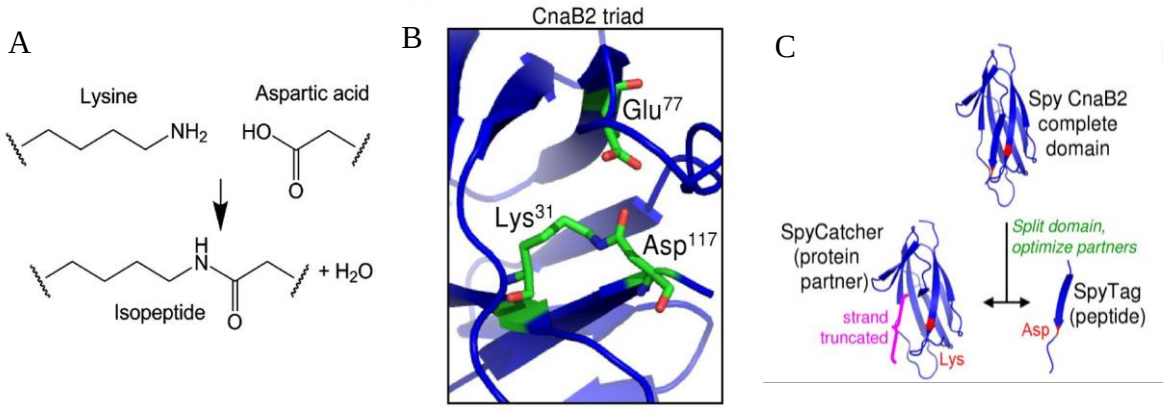
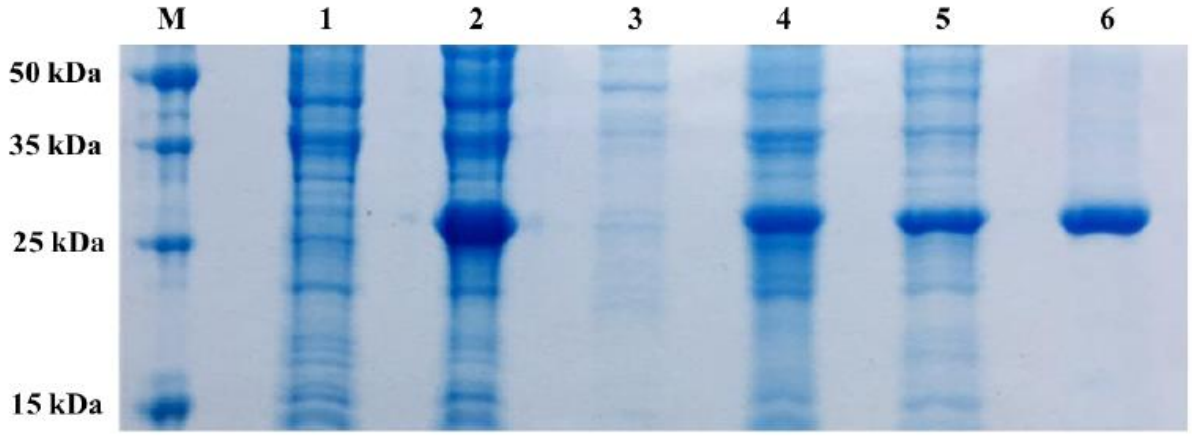


Figure 3.9 Spontaneous intermolecular amide bond formation by SpyTag. (A) Amide bond formation between Lys and Asp side chains. (B) Key residues involved in amide bond formation in CnaB2 shown in stick format (PDB accession code 2X5P). (C) Cartoon of SpyTag construction, featuring dissection of the *Streptococcus pyogenes* protein CnaB2 into a large N-terminal fragment (SpyCatcher, left) and a small C-terminal fragment (SpyTag, right). Reactive residues are highlighted in red. Reprinted by permission from PNAS. "Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M., P Natl Acad Sci USA 2012, 109, 690-697."

undergoes nucleophilic attack from the unprotonated amine of Lys^{31}, catalyzed by Glu^{77}. Howarth and coworkers reasoned that this bacterial adhesin protein from *S. pyogenes* (from which the prefix Spy is adapted) could be split into a short peptide (SpyTag) and a much longer protein partner (SpyCatcher). They redesigned the interface between these two halves to increase their affinity for

one another (the native sequence undergoes isopeptide bond formation in a few hours, whereas the redesigned sequence performs this transformation in several minutes). This covalent bond forms readily at room temperature and is stable to boiling in SDS.

MGAHIVMVDAYKPTKTSGGGSGGGAS(GDERAVEALIKALKDPDGWVRKAAAE
ALGRIGDERAVEALIKALKDPDWFVREAAAKALGEI)$_4$GSMHHHHHH



M. Perfect Protein Marker (10-225 kDa)

1. ST_HEAT, pre-induction expression

2. ST_HEAT 15 hours post-induction

3. ST_HEAT, lysed supernatant

4. ST_HEAT, lysed pellet

5. ST_HEAT, urea supernatant

6. ST_HEAT, affinity column eluent

Figure 3.10 Sequence and SDS PAGE gel of ST_HEAT. N-terminal SpyTag (purple) is followed by four identical repeats of the dimeric HEAT sequence (red), with a C-terminal 6X HisTag (green) for affinity column purification. SDS PAGE gel is shown with a legend identifying the lane contents. 6M Urea was used to solubilize the lysed pellet, which contained the vast majority of the expressed ST_HEAT protein. Lane 6 shows a very clean band at the MW expected for ST_HEAT.

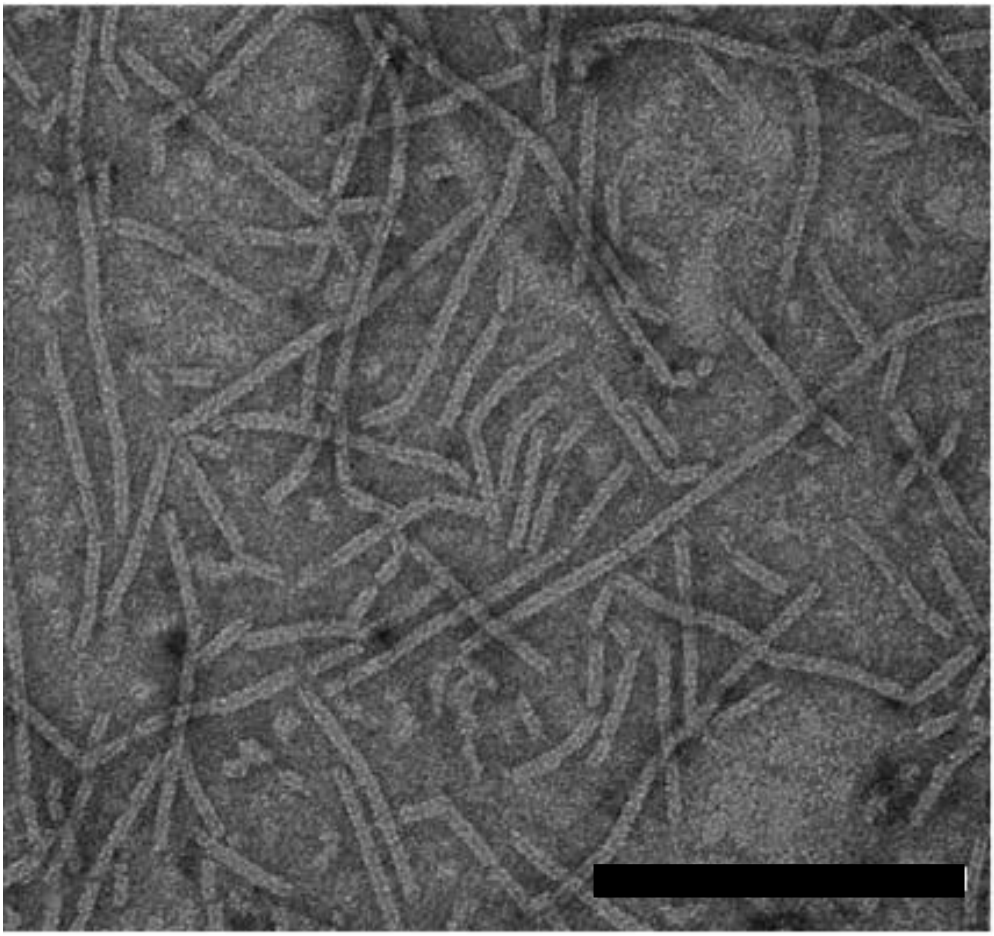


Figure 3.11 Representative negative stained TEM image of ST_HEAT. Like HEAT_R1 and HEAT_6R, the nanotubes have a smooth outer surface, and a clearly distinguishable inner lumen. Scale bar is 200 nm.

We sought to employ the SpyTag:SpyCatcher genetic fusion technique to affix a functional protein to the outer surface of the HEAT nanotubes. We initially selected super folder green fluorescent protein (sf-GFP) as a reporter protein, due to its stability and strong fluorescent activity. We designed a concatemer of HEAT containing eight repeats (essentially HEAT_6R with an additional dimer) and a SpyTag at the C-terminus. We elected to add an additional dimer to the HEAT_6R protein to space out the sf-GFPs on the HEAT nanotube surface. Simultaneously, we expressed sf-GFP with SpyCatcher at the N-terminus.

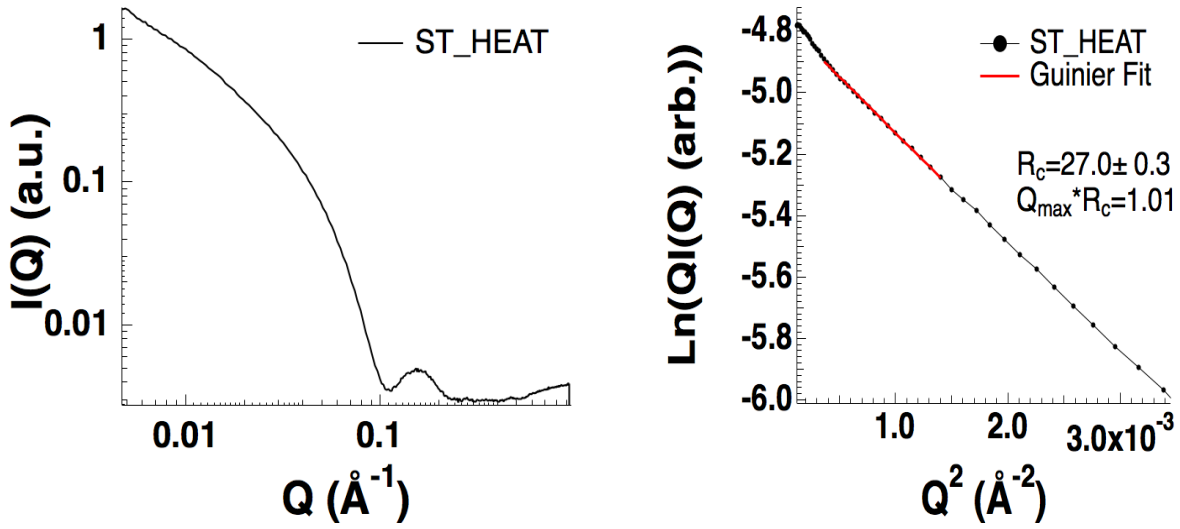


Figure 3.12 SAXS Data for ST_HEAT. The SAXS scatter curve (left) corresponds with a nanotubular assembly, as evidenced by the oscillation observed at ~0.1 Å^{-1}. No diffraction peaks were measured. A Guinier analysis (right) was applied to the SAXS data, resulting in an Rc value of 27.0 Å, which is significantly smaller than the observed Rc values for HEAT_R1 and HEAT_6R.

After the SpyTag:HEAT protein (ST_HEAT) was purified via affinity chromatography as well as HPLC, it was assembled from pH 5-8, similar to HEAT_R1 and HEAT_6R. ST_HEAT was found to fold into a helical conformation from pH 5-8. Using transmission electron microscopy, the supramolecular assemblies were determined to be high aspect ratio filaments with clearly distinguishable lumen, and featureless outer walls. SAXS scattering data were collected for the ST_HEAT nanotubes, indicating a slightly smaller diameter than was seen for HEAT_R1/HEAT_6R. Similar to HEAT_6R, analysis of the ST_HEAT WAXS scattering indicated a lack of diffraction-like peaks.

MKHHHHHHGTSENLYFQGMSSGLVPRGSHMVSKGEEDNMAIIKEFMR
FKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLPFAWDILS
PQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQD
SSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASSERMYPEDGAL
KGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITSHN
EDYTIVEQYERAEGRHSTGGMDELYKKNSGGGLVAGGSGGGSGGGTG
GGSGGGTSGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDEDGKE
LAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVA
TAITFTVNEQGQVTVNGKATKGDAHI

Figure 3.13 Sequence of mCherry_SC, which contains mCherry (red lettering), SpyCatcher (blue lettering), a 6X N-terminal HisTag (green lettering) for column affinity purification, and a TEV-protease cleavage site (burnt-orange lettering).

Conjugation with sf-GFP however, proved to be difficult, as there was little overlap between stable assembly conditions of ST_HEAT (pH 4.5-6 optimal), and sf-GFP (pH 3-4 optimal). The tendency of sf-GFP to dimerize precluded conjugation with ST_HEAT in some cases. Because of this, we sough another fluorescent protein that better suited conjugation with ST_HEAT. One of the mFruit series proteins, mCherry, nicely fit our requirements; it is stable at higher pHs than sfGFP, expresses at high yield, and remains monomeric at low- and medium-concentrations. We expressed the mCherry-SC construct in the same fashion as we expressed the sfGFP-SC construct. The sequence for mCherry-SC is shown in Figure 3.13. Conjugation of

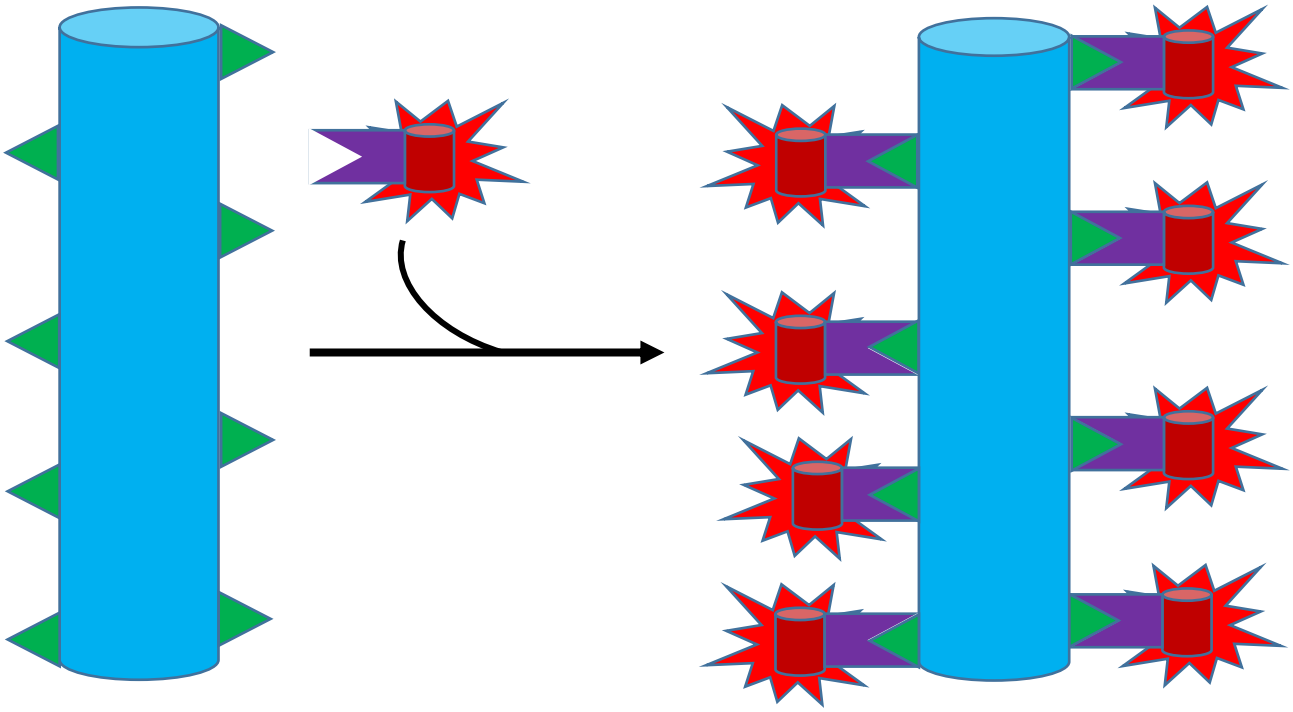


Figure 3.14 Fusion schematic for the mCherry_HEAT complex assembly. Self-assembled ST_HEAT nanotubes (blue nanotubes represent HEAT with green triangles representing the SpyTag) are incubated with mCherry_SC (red cylinders represent mCherry and the purple shapes complementary to the SpyTag triangles are SpyCatcher) at room temperature for several hours to ensure wide-spread conjugation.

mCherry-SC to ST_HEAT was accomplished by assembling the ST_HEAT nanotubes at their optimal pH (pH 5.0, thermally annealed from 90°C to 25°C at a cooling rate of -0.2°C/minute), and mixing in concentrated mCherry-SC at a 1% v/v. The resultant solution is allowed to mix overnight at room temperature, which is significantly longer than the required reaction time, in an effort to functionalize the majority of the ST_HEAT assembly units. To remove excess (unreacted) mCherry_SC from solution, the mixture is passed through 100,000 Da molecular weight cut-off centrifuge filters and washed with a large excess of the assembly buffer. Finally, the conjugated

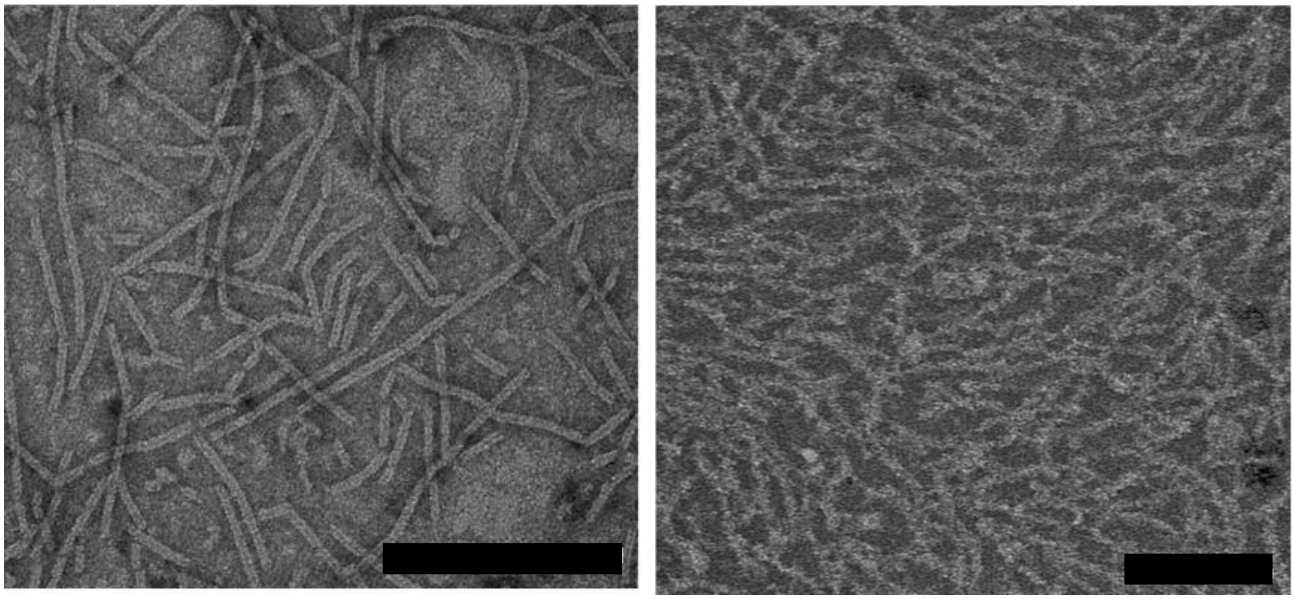


Figure 3.15 Negative-stained TEM of the (left) ST_HEAT and (right) mCherry_HEAT assemblies. The surface morphology of ST_HEAT seems to be smooth, compared with the ragged, bumpy surface morphology seen in mCherry_HEAT. Additionally, no inner lumen can be distinguished in the mCherry_HEAT filaments, likely due to obstruction by the conjugated mCherry protein. Scale bar is 200 nm for both images.

protein was collected from the filter by inversion and centrifugation, followed by dilution to the original volume (for detailed methods, see below). The conjugated protein assemblies (mCherry_HEAT) were analyzed via TEM and compared to the bare ST_HEAT assemblies. While the bare ST_HEAT assemblies displayed nanotubes with clearly distinguishable inner lumina and a featureless outer wall, the mCherry_HEAT assemblies appeared as filaments without distinguishable lumina and proteinaceous protrusions. It stands to reason that the protein sticking to the outside of the filaments is mCherry_SC, and that the lack of a distinguishable lumen in these filaments is due to obstruction by mCherry. Though the negative-stained TEM images are compelling, they are by no means diagnostic. In an effort to more clearly resolve the conjugated

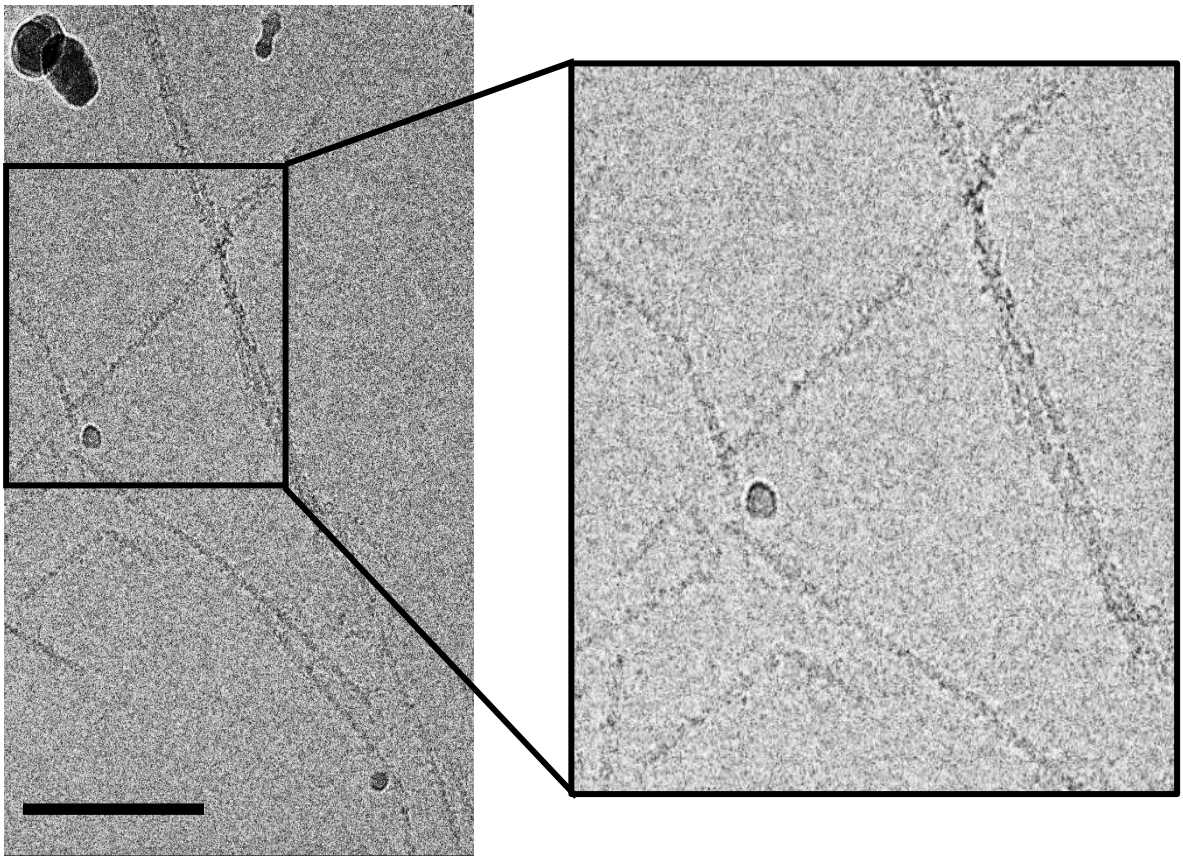


Figure 3.16 Cryo-EM image of mCherry_HEAT (left), with a zoomed-in view (right). Several filaments are observed, in which a "beads on a string" morphology is prevalent. Differential contrast is also seen between the filament and the beads on the filament. Scale bar is 200 nm.

mCherry protein on the outside of the ST_HEAT nanotubes, correlative light and electron microscopy (CLEM) was employed. The purpose of CLEM is to correlate the location of fluorescence (or other visible-light scale features) with high resolution features seen via cryo-EM. Initial CLEM imaging indicated that the incidence of mCherry fluorescence was localized to grid spots in which nanotube assemblies were found. Unfortunately, an instrument failure cut these experiments short, and the fluorescent image was never captured. However, some cryo-EM images

were captured, and offer a higher resolution look at the mCherry_HEAT construct. The cryo-EM images show rather clearly that the nanotubes have been functionalized with protein.

To show that the nanotubes have been effectively functionalized, it is important to show that the functional protein, mCherry, has retained its fluorescent function. Initial fluorescence measurements showed that after conjugation with ST_HEAT, and filtration of any unbound mCherry, fluorescence is still observed. It is important to note that during the process of thermal

annealing, mCherry loses its fluorescence (likely due to denaturation of the protein); for this reason, we anneal the ST_HEAT assemblies prior to adding the mCherry protein.

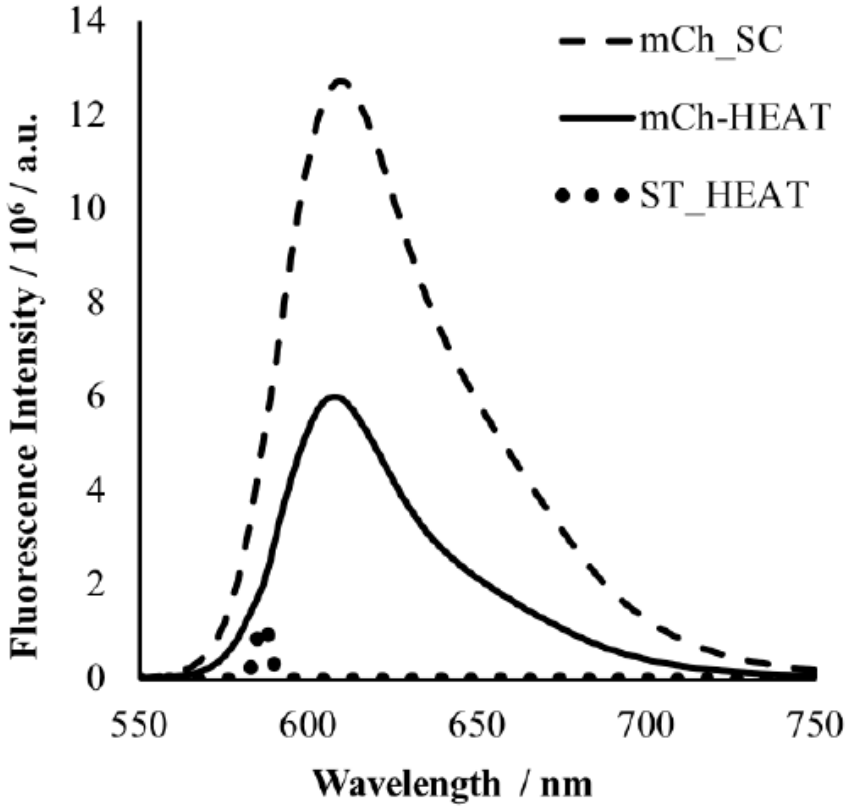


Figure 3.17 Fluorescence emission spectra of mCherry_HEAT, mCherry_SC, and ST_HEAT from 550 nm to 750 nm, with an excitation wavelength of 587 nm. Maximal fluorescence emission was observed for mCherry_HEAT and mCherry_SC at 610 nm (the emission wavelength for mCherry).

The prepared mCherry_HEAT solution was analyzed using fluorimetry, following baseline correction using the assembly buffer (10 mM Acetate, pH 5.0). The excitation wavelength for mCherry, 587 nm, was used for all samples. Unsurprisingly, ST_HEAT did not produce a significant fluorescent signal; to contrast, both mCherry_SC and mCherry_HEAT did. The

maximum emission for both of these proteins occurred at 610 nm, the emission wavelength of native mCherry. The fluorescent intensity of mCherry_SC is significantly higher than that of mCherry_HEAT, though this is likely due to the higher concentration of the mCherry_SC (20 μM versus 5 μM for mCherry_SC and mCherry_HEAT, respectively). Because the cross-linked proteins still fluoresce, it is clear that conjugation does not interfere with the function of the mCherry protein.

Finally, we sought to determine the relative loading of mCherry on the ST_HEAT nanotubes. Using epifluorescence microscopy, and calibrating using an mCherry standard, it was

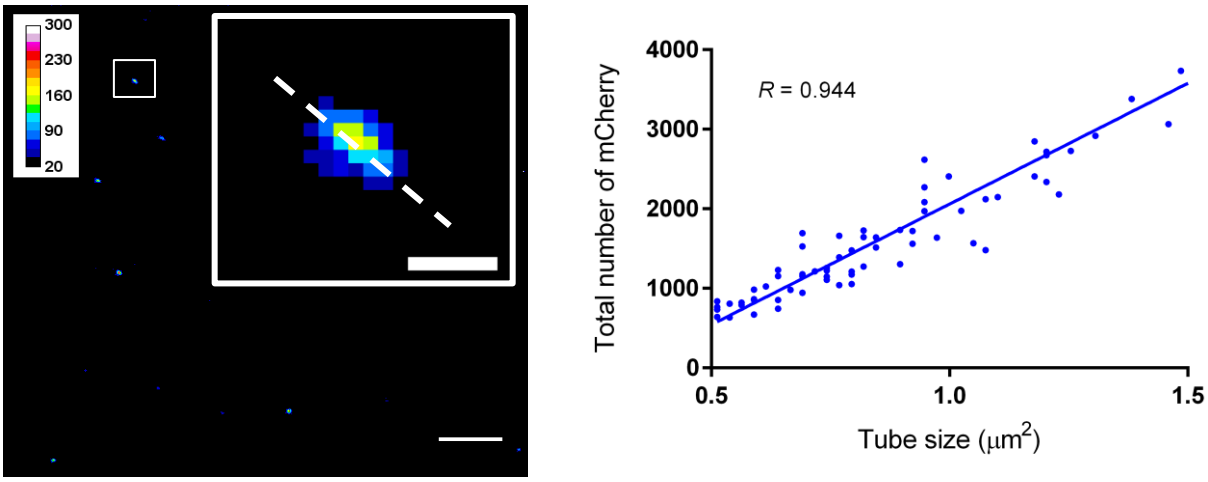


Figure 3.18 Epifluorescence microscopy image of mCherry_HEAT, depicting several mCherry_HEAT filaments (Left). In-set image is a zoomed view of a single filament. The inner scale bar is 1 μm, and the outer scale bar is 10 μm. Correlation curve of filament length versus calibrated fluorescent count (Right).

determined that the fluorescent intensity scaled linearly with the length of the mCherry_HEAT filaments (Figure 3.18). Moreover, the relative stoichiometry of mCherry to ST_HEAT can be estimated, with the assumption that ST_HEAT retains the same periodicity as the HEAT_R1

filaments; the bulk of the biophysical data supports this assumption. Prior to calculating the number of mCherry proteins per ST_HEAT copy, it is important to discuss the limitations of the epifluorescence microscopy. First, the resolution of this technique is relatively low compared to the diameter of the nanotubes; a 1 μm filament will appear as 8 linear pixels. Because of this, the error associated with the filament length distribution is very large.

From the correlation curve in Figure 3.18, we can estimate that a 1.0 μm filament has 2,000 copies of mCherry attached. The axial rise of HEAT_R1 is 3.1 nm, and in each axial turn, there are 20.7 helical hairpins. If this holds true for ST_HEAT, there should be 2.6 copies of ST_HEAT in a 3.1 nm turn (20.7 helical hairpins in a turn / 8 helical hairpins in ST_HEAT).
A filament with a length of 1 μm should have 322.6 turns (1,000 nm / 3.1 nm per turn). Thus, in a 1 μm filament, there should be 838.7 ST_HEAT proteins (322.6 turns x 2.6 ST_HEAT copies per turn). With a proposed binding stoichiometry of one mCherry per ST_HEAT, this filament should only have 838.7 mCherry proteins. Contrast this with the measured 2,000 copies in a 1 μm filament. However, due to the inherent inaccuracies in the epifluorescence microscopy, it is possible that overcounting of mCherry is occurring. While an accurate stoichiometry cannot be determined, it is important to note that a large number of functional mCherry proteins are binding to the ST_HEAT nanotube. Even if the efficiency is not near 100%, this represents a massive stride forward in developing functional nanotubes from a TRP scaffold.

3.4 Conclusion

The effective binding of functional protein to the surface of the ST_HEAT nanotubes via the SpyTag:SpyCatcher genetic fusion methodology is very encouraging. Now that the concept has been sufficiently proven, it is now possible to functionalize the ST_HEAT nanotubes with enzymes that will carry out some function. An interesting question to be addressed here is the effect of immobilizing the enzymes on their reactive capabilities. Ultimately, the ST_HEAT nanotubes are promising because the concave surface of the nanotubes is completely modular with respect to amino acid. This may extend to unnatural amino acids, which can be chemically functionalized using bioorthogonal methods such as copper free azide-alkyne cycloaddition. Coupled with the genetic fusion method employed on the outer wall, we have described a stable nanotube scaffold with two independently addressable surfaces, capable of bifunctionalization. A possible application for this system is theranostics; a reporter protein (such as mCherry) could functionalize the convex surface of the tube, while a therapeutic molecule (such as a cytotoxic anti-cancer drug) could line the concave surface of the nanotube. Combined with PEGylation for increased uptake by tumors, this system could be used for the detection and treatment of certain cancers. No matter the application it ultimately serves, the mCherry_HEAT system has brought us much closer to a functional TRP-based material.

3.5 Materials and Methods

Materials

All chemical reagents were purchased from either Thermo Fisher Scientific, Inc. (Waltham, MA) or Millipore Sigma (Burlington, MA) unless otherwise noted. Synthesis of plasmids containing codon-optimized genes encoding mCherry_SpyCatcher and

SpyTag_HEAT8R was conducted by ATUM, formerly known as DNA2.0 (Newark, CA). The Top10F′ chemically competent *E. coli* strain was obtained from Invitrogen Corp. (Carlsbad, CA), and the BL21 (DE3) chemically competent *E. coli* strain was purchased from New England Biolabs, Inc. (Ipswich, MA). The QIAprep-spin miniprep kit was purchased from QIAGEN, Inc. (Valencia, CA). Luria-Bertani broth and agar powder were purchased from Millipore Sigma (Burlington, MA), and Terrific Broth (TB) powder was purchased from Thermo Fisher Scientific, Inc. (Waltham, MA). Kanamycin monosulfate was purchased from VWR International, LLC (Suwanee, GA). Isopropyl-β-D-thiogalactopyranoside (IPTG) was purchased from Research Products International Corp. (Prospect, IL). The Perfect Protein Marker (10-225 kDa) was purchased from Millipore Sigma (Burlington, MA). Benzonase® nuclease and protease-inhibitor cocktail (EDTA-free) were purchased from Millipore Sigma (Burlington, MA). Lysozyme from chicken egg white was purchased from Research Products International Corp. (Prospect, IL). HisPurTM cobalt resin was obtained from Thermo Fisher Scientific, Inc. (Waltham, MA). TEV Protease was purchased from Millipore Sigma (Burlington, MA). SnakeSkin^{TM} Dialysis Tubing with a 10 kDa MWCO and Slide-A-LyzerTM G2 dialysis cassettes with a 20 kDa MWCO were obtained from Thermo Fisher Scientific, Inc. (Waltham, MA). Amicon® ultra centrifugal filters (10 kDa NMWL, 30 kDa NMWL, 50 kDa NMWL, and 100 kDa NMWL) for sample volumes of 15 mL and 0.5 mL were purchased from Millipore Sigma (Burlington, MA). The Coomassie (Bradford) Protein Assay Kit was purchased from Thermo Fisher Scientific, Inc. (Waltham, MA). Vacuum filters with a 0.2 μm polyether sulfone (PES) membrane were obtained from Thermo Fisher Scientific, Inc. (Waltham, MA).

General Methods

Basic molecular biology procedures were adapted from a standard molecular cloning manual23 or the protocol supplied by the manufacturer unless otherwise described in detail. All Reagents intended for use with bacteria, DNA, or recombinant proteins were sterilized by either syringe filtration through a 0.2 μm cellulose membrane, vacuum filtration through a 0.2 μm polyether sulfone (PES) membrane, or by autoclaving. All enzyme reactions were conducted in the reagent buffers provided by the manufacturer unless otherwise noted. Synthetic plasmids were ordered from ATUM, resuspended in water upon arrival, and transformed into chemically competent Top10F' or BL21 *E. coli* strains. *E. coli* strains were grown at 37 °C in Lauria-Bertani (LB) medium containing appropriate antibiotic with shaking at 200 rpm unless otherwise stated. A QIAprep-spin miniprep kit was used for isolation of plasmid DNA. All OD values were determined using a UltroSpec 3000 UV/Visible spectrophotometer at 600 nm with a 1 cm cuvette. All proteins were purified using immobilized metal affinity chromatography (IMAC) on HisPurTM cobalt resin followed by dialysis to remove imidazole. Protein electrophoresis was conducted using 12 or 16% SDS polyacrylamide gels on a Mini-PROTEAN 3 cell electrophoresis system from Bio-Rad Laboratories, Inc. (Hercules, CA). The buffer tank was filled with SDS run buffer (25 mM tris, 250 mM glycine, 0.1% SDS, pH 8.3). The Perfect Protein Marker was used as a protein standard for SDS-PAGE analysis. An initial sample volume of 10 μL was added to each well with any necessary concentration adjustments made thereafter. Gels were run at 150 V for 1-1.5 hours depending on the desired level of separation. Gels were stained in Coomassie overnight and then destained in a methanol-acetic acid buffer.

Gene and Bacterial Preparation

A gene encoding a fusion protein of SpyTag and HEAT_8R was codon optimized for expression in *E. coli* and synthesized by ATUM (formerly DNA 2.0). The lyophilized powder was resuspended in 50 μL of distilled, deionized water (ddH2O) upon arrival. Aliquots of 1 μL were used in transformations of the chemically competent *E. coli* strains BL21 (DE3) and Top10F'. The cells were recovered after heat shock in 400 μL SOC rich media for 30 minutes at 37 °C. Aliquots of 25 μL and 50 μL of the transformed suspensions were spread onto LB agar plates containing kanamycin (50 μg/mL) for antibiotic selection. The plates were incubated for 14 hours at 37 °C.

Six colonies were selected from the plate containing the Top 10 F' strain and two colonies were selected from the plate containing the BL21 strain. The selected colonies were used to inoculate eight culture tubes containing 5 mL LB media supplemented with kanamycin (50 μg/mL). The cultures were grown at 37 °C overnight on a rotator. Two of the cultures containing the Top10F' strain in addition to the two cultures containing the BL21 strain were prepared for long term storage of the plasmid contained within the *E. coli* strain. These frozen stocks were prepared by combining 200 μL of 80% glycerol with 800 μL of the culture. The frozen stocks of the BL21 and Top10F' strains containing the plasmid ST_HEAT were stored at -80 °C until needed. To make secondary stocks of the ST_HEAT plasmid DNA, a QIAprep-spin miniprep kit (QIAGEN, Inc.) was used to isolate the plasmid DNA of each of the remaining cultures of Top10F' cells. The isolated DNA was recovered in 50 μL of EB buffer (10 mM Tris-HCl, pH 8.5) and stored in solution at -30 °C when not in use.

Large-Scale Expression

The plasmid ST_HEAT was transformed into BL21 (DE3) strain and plated onto LB agar plates containing kanamycin (50 μg/mL). Single colonies of BL21 strain containing the ST_HEAT

plasmid were used to inoculate 5 mL cultures of LB broth supplemented with kanamycin (50 μg/mL). The cell cultures were grown overnight on a rotator at 37 °C. Five mL of the overnight cultures were transferred to 2,800 mL Erlenmeyer flasks containing 500 mL of Lauria-Bertani (LB) media supplemented with 50 μg/mL kanamycin. Four flasks were used for a total expression culture volume of 2000 mL (2L). The large culture flasks shook at 200 rpm and 37 °C for approximately 3 hours until cell growth reached log phase growth with an OD_{600} of ~0.6-0.8. The OD_{600} was monitored hourly. Once log phase growth was achieved, expression was induced by the addition of Isopropyl-β-D-thiogalactopyranoside (IPTG). For each 500 mL flask, 250 μL of 1M IPTG was added for a final IPTG concentration of 0.5 mM. The expression cultures were incubated at 30 °C with shaking for 15 hours. Aliquots were removed from the expression culture throughout the process for SDS-PAGE analysis. Fifteen hours after induction, the cells in the expression cultures were harvested by centrifugation at 4,000 x g for 20 minutes (4 °C). The cell pellets were resuspended in 100 mL lysis buffer (50 mM NaH2PO4, 100 mM NaCl, pH 8.0) and stored at -80 °C. One mL aliquots were removed from each flask before induction and 1.5, 3 and 15 hours after induction. The OD_{600} was determined for each aliquot and the aliquot volumes were normalized based on their OD so that the number of cells in each aliquot would be equivalent. The normalized aliquot volumes were centrifuged for 5 minutes at 4,000 x *g* to pellet the cells. The supernatants were discarded, the pellets were resuspended in 50 μL ddH2O, and the expression samples were stored at -30 °C. Aliquoted expression samples were later prepared for SDS-PAGE analysis by mixing the 50 μL samples with 50 μL of 2X SDS sample loading buffer (100 mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 20% (v/v) glycerol, 0.2% (w/v) bromophenol blue, 10% (v/v) β-mercaptoethanol) and boiling the samples at 100 °C for 5 minutes. The expression samples were run on a 16% SDS-PAGE gel according to the protocol described in the general methods section.

Cell Lysis

The cell pellets from a large-scale expression of ST_HEAT in the BL21 strain were previously frozen in 100 mL lysis buffer. Lysis of the cell pellets was initiated by three freeze/thaw cycles (-80 °C; 25 °C). After the third thaw cycle, EDTA-free protease inhibitor cocktail (1X), lysozyme (1 mg/mL), benzonase nuclease (25 units/mL), and MgCl2 (1 mM) were added to the lysate. The lysate was incubated at 30 °C for 1 hour with shaking at 200 rpm and then the temperature was dropped to 4 °C for incubation overnight. The cell lysate was sonicated (9 seconds on; 9 seconds off) at 4 °C for 15 minutes. A 50 µL aliquot was collected for SDS-PAGE analysis. An aliquot of 3 mL was removed after lysis for experiments to verify linking between ST_HEAT and mCh_SC (see section below on linking of ST_HEAT lysate with mCh_SC lysate). The remaining cell lysate was centrifuged at 10,000 rpm for 20 minutes (4 °C). The lysed supernatant was separated from the lysed pellet until analysis by SDS-PAGE determined the location of the ST_HEAT fusion protein. Meanwhile, the cell lysate aliquot was centrifuged at 10,000 rpm for 5 minutes. The aliquot supernatant was transferred into a clean microcentrifuge tube and the pellet was resuspended in 50 µL of ddH2O. The lysed supernatant and lysed pellet samples were prepared for SDS-PAGE analysis by mixing the 50 µL samples with 50 µL of 2X SDS sample loading buffer (100 mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 20% (v/v) glycerol, 0.2% (w/v) bromophenol blue, 10% (v/v) β-mercaptoethanol) and boiling the samples at 100 °C for 5 minutes. Analysis by 16% SDS-PAGE concluded that the ST_HEAT fusion protein was primarily located in the lysed pellet (insoluble fraction).

Protein Solubilization

The insoluble ST_HEAT protein in the cell lysate pellet was solubilized to allow for purification on a HisPurTM cobalt column. In order to do this, the pellets were resuspended in 120 mL total volume of urea solubilization buffer (6 M urea, 50 mM NaH2PO4, 100 mM NaCl, pH 8.0). The insoluble pellets were allowed to shake at 250 rpm for 3 days at 4 °C. A 50 µL aliquot was collected for SDS-PAGE analysis. The remaining urea solution was centrifuged at 12,000 rpm for 20 minutes (4 °C). The urea supernatant was separated from the urea pellet until analysis by SDS-PAGE determined the location of the ST_HEAT fusion protein. Meanwhile, the urea aliquot was centrifuged at 12,000 rpm for 5 minutes. The aliquot supernatant was transferred into a clean microcentrifuge tube and the pellet was resuspended in 50 µL of ddH2O. The urea supernatant and urea pellet samples were prepared for SDS-PAGE analysis by mixing the 50 µL samples with 50 µL of 2X SDS sample loading buffer (100 mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 20% (v/v) glycerol, 0.2% (w/v) bromophenol blue, 10% (v/v) β-mercaptoethanol) and boiling the samples at 100 °C for 5 minutes. Analysis by 16% SDS-PAGE concluded that the ST_HEAT fusion protein was at least partially soluble in 6M urea and primarily located in the supernatant.

Purification

The presence of a hexahistadine tag at the C-terminus of the fusion protein ST_HEAT facilitated purification by immobilized metal affinity chromatography (IMAC). The soluble fraction recovered after protein solubilization in urea was loaded directly onto 12 mL pre-equilibrated HisPurTM cobalt resin divided among four columns. The resin was then washed with 60 mL urea wash buffer (6 M urea, 50 mM NaH2PO4, 100 mM NaCl, 20 mM imidazole, pH 8.0). The target protein was isolated by the addition of 60 mL elution buffer (6 M urea, 50 mM NaH2PO4, 100 mM NaCl, 250 mM imidazole, pH 8.0). An aliquot of 100 µL of HisPurTM eluent

was removed after purification for experiments to verify linking between ST_HEAT and mCh_SC (see section below on linking of ST_HEAT HisPurTM eluent with mCh_SC HisPurTM eluent). A 50 μL aliquot of the eluent was collected and prepared for SDS-PAGE analysis by mixing the 50 μL sample with 50 μL of 2X SDS sample loading buffer (100 mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 20% (v/v) glycerol, 0.2% (w/v) bromophenol blue, 10% (v/v) β-mercaptoethanol) and boiling the sample at 100 °C for 5 minutes. The pure eluent sample was run on a 16% SDS-PAGE gel according to the protocol described in the general methods section. The purification eluent was loaded into a Slide-A-LyzerTM dialysis cassette (20,000 MWCO) and dialyzed against ddH2O for 4 days (4 L volume, switching buffer every 8 hours).

mCherry_SpyCatcher Production

After dialysis, the protein solution was transferred from the dialysis cassette to a 50 mL conical tubes and stored at -80 °C. Once frozen, the tops of the conical tubes were replaced with Kimwipes^{TM} and the tubes were placed in a lyophilizer until all that remained was the pure ST_HEAT protein in powder form. The lyophilized powder was weighed and stored at -30 °C.

Gene and Bacterial Preparation

A gene encoding a fusion protein of mCherry red fluorescent protein and SpyCatcher was codon optimized for expression in *E. coli* and synthesized by ATUM. The lyophilized powder was resuspended in 50 μL of distilled, deionized water (ddH2O) upon arrival. Aliquots of 1 μL were used in transformations of the chemically competent *E. coli* strains BL21 (DE3) and Top10F'. The cells were recovered after heat shock in 400 μL SOC rich media for 30 minutes at 37 °C. Aliquots of 50 μL of the transformed suspensions were spread onto LB agar plates containing kanamycin

(50 μg/mL) for antibiotic selection. The plates were incubated for 14 hours at 37 °C. Six colonies were selected from the plate containing the Top 10 F' strain and two colonies were selected from the plate containing the BL21 strain. The selected colonies were used to inoculate eight culture tubes containing 5 mL LB media supplemented with kanamycin (50 μg/mL). The cultures were grown at 37 °C overnight on a rotator. Two of the cultures containing the Top10F' strain in addition to the two cultures containing the BL21 strain were prepared for long term storage of the plasmid contained within the *E. coli* strain. These frozen stocks were prepared by combining 200 μL of 80% glycerol with 800 μL of the culture. The frozen stocks of the BL21 and Top10F' strains containing the plasmid mCh_SC were stored at -80 °C until needed. To make secondary stocks of the mCh_SC plasmid DNA, a QIAprep-spin miniprep kit (QIAGEN, Inc.) was used to isolate the plasmid DNA of each of the remaining cultures of Top10F' cells. The isolated DNA was recovered in 50 μL of EB buffer (10 mM Tris-HCl, pH 8.5) and stored in solution at -30 °C when not in use.

Large-Scale Expression

mCh_SC plasmid was transformed into BL21 (DE3) strain and plated onto LB agar plates containing kanamycin (50 μg/mL). Single colonies of BL21 strain containing the mCh_SC plasmid were used to inoculate 5 mL of LB broth supplemented with kanamycin (50 μg/mL). The cell cultures were grown overnight on a rotator at 37 °C. Five mL of the overnight cultures were transferred to 2,800 mL Erlenmeyer flasks containing 500 mL of Terrific Broth media supplemented with 50 μg/mL kanamycin. Four flasks were used for a total expression culture volume of 2000 mL (2L).

The large culture flasks shook at 200 rpm and 37 °C for 2-3 hours until cell growth reached log phase growth with an OD_{600} of ~0.6-0.8. The OD_{600} was monitored hourly. Once log phase

growth was achieved, expression was induced by the addition of Isopropyl-β-D-thiogalactopyranoside (IPTG). For each 500 mL flask, 250 µL of 1M IPTG was added for a final IPTG concentration of 0.5 mM. The expression cultures were incubated at 30 °C with shaking for 4 hours. Aliquots were removed from the expression culture throughout the process for SDS-PAGE analysis. Four hours after induction, the cells in the expression cultures were harvested by centrifugation at 4,000 x g for 20 minutes (4 °C). The supernatants were transferred into a clean 2,800 mL Erlenmeyer flask since there appeared to be a large amount of mCh_SC present in the media. A 50 µL aliquot of the expression media was collected for SDS-PAGE analysis. The cell pellets were resuspended in 100 mL lysis buffer (50 mM NaH2PO4, 100 mM NaCl, pH 8.0) and stored at -80 °C. One mL aliquots were removed from each flask before induction and each hour after induction. The OD_{600} was determined for each aliquot and the aliquot volumes were normalized based on their OD so that the number of cells in each aliquot would be equivalent. The normalized aliquot volumes were centrifuged for 5 minutes at 4,000 x *g* to pellet the cells. The supernatants were discarded, the pellets were resuspended in 50 µL ddH2O, and the expression samples were stored at -30 °C. Aliquoted expression samples were later prepared for SDS-PAGE analysis by mixing the 50 µL samples with 50 µL of 2X SDS sample loading buffer (100 mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 20% (v/v) glycerol, 0.2% (w/v) bromophenol blue, 10% (v/v) β-mercaptoethanol) and boiling the samples at 100 °C for 5 minutes.

Cell Lysis

The cell pellets from a large-scale expression of mCh_SC in the BL21 strain were previously frozen in 100 mL lysis buffer. Lysis of the cell pellets was initiated by three freeze/thaw cycles (-80 °C; 25 °C). After the third thaw cycle, EDTA-free protease inhibitor cocktail (1X), lysozyme (1 mg/mL), benzonase nuclease (25 units/mL), and MgCl2 (1 mM) were added to the

lysate. The lysate was incubated at 30 °C for 1 hour with shaking at 200 rpm and then the temperature was dropped to 4 °C for incubation overnight. The cell lysate was sonicated (9 seconds on; 9 seconds off) at 4 °C for 15 minutes. A 50 μL aliquot was collected for SDS-PAGE analysis. A 3 mL aliquot was removed after lysis for experiments to verify linking between mCh_SC and ST_HEAT (see section below on linking of ST_HEAT lysate with mCh_SC lysate). The remaining cell lysate was centrifuged at 15,000 rpm for 20 minutes (4 °C). The lysed supernatant was separated from the lysed pellet until analysis by SDS-PAGE determined the location of the mCh_SC fusion protein. Meanwhile, the cell lysate aliquot was centrifuged at 15,000 rpm for 5 minutes. The aliquot supernatant was transferred into a clean microcentrifuge tube and the pellet was resuspended in 50 μL of ddH2O. The lysed supernatant and lysed pellet samples were prepared for SDS-PAGE analysis by mixing the 50 μL samples with 50 μL of 2X SDS sample loading buffer (100 mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 20% (v/v) glycerol, 0.2% (w/v) bromophenol blue, 10% (v/v) β-mercaptoethanol) and boiling the samples at 100 °C for 5 minutes. Analysis by 12% SDS-PAGE concluded that the mCh_SC fusion protein was primarily located in the lysed supernatant (soluble fraction) with some remaining in the lysed pellet (insoluble fraction).

Purification

The presence of a hexahistadine tag at the N-terminus of the fusion protein mCh_SC facilitated purification by immobilized metal affinity chromatography (IMAC). The soluble fraction recovered after lysis was loaded directly onto 12 mL pre-equilibrated HisPurTM cobalt resin distributed across four columns. The resin was then washed with 60 mL wash buffer (50 mM NaH2PO4, 100 mM NaCl, 20 mM imidazole, pH 8.0). The target protein was isolated by the addition of 40 mL elution buffer (50 mM NaH2PO4, 100 mM NaCl, 250 mM imidazole, pH 8.0). An aliquot of 100 μL of HisPurTM eluent was removed after purification for experiments to verify

linking between mCh_SC and ST_HEAT (see section below on linking of ST_HEAT HisPurTM eluent with mCh_SC HisPurTM eluent). The remaining eluent was loaded into a Slide-A-LyzerTM dialysis cassette (20,000 MWCO) and dialyzed against 50 mM NaH_2PO_4, 100 mM NaCl, pH 8.0 for 12 hours (4 L volume, switching buffer every 4 hours). After dialysis, the protein solution was transferred from the dialysis cassette to a 50 mL conical tube and stored at 4 °C. A 50 μL aliquot of the eluent was collected and prepared for SDS-PAGE analysis by mixing the 50 μL sample with 50 μL of 2X SDS sample loading buffer (100 mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 20% (v/v) glycerol, 0.2% (w/v) bromophenol blue, 10% (v/v) β-mercaptoethanol) and boiling the sample at 100 °C for 5 minutes.

As noted previously, the expression media present after harvesting the cells contained a significant amount of mCh_SC. The expression media was centrifuged at 15,000 rpm for 20 minutes to ensure the removal of all cells. The supernatant was saved for purification. The expression media shook with 20 mL of HisPurTM cobalt resin for 1 hour and the resin was added to four columns. The loaded resin was washed with 100 mL wash buffer (50 mM NaH_2PO_4, 100 mM NaCl, 20 mM imidazole, pH 8.0). The target protein was isolated by the addition of 40 mL elution buffer (50 mM NaH2PO4, 100 mM NaCl, 250 mM imidazole, pH 8.0). The eluent was loaded into a Slide-A-LyzerTM dialysis cassette (20,000 MWCO) and dialyzed against 50 mM NaH_2PO_4, 100 mM NaCl, pH 8.0 for 12 hours (4 L volume, switching buffer every 4 hours). After dialysis, the protein solution was transferred from the dialysis cassette to a 50 mL conical tube and stored at 4 °C.

TEV Protease Cleavage

A TEV protease cleavage site was included to facilitate the removal of the His-tag from the mCh_SC fusion protein after purification. TEV protease was added to half of the purified mCh_SC protein to a final concentration of 1:100 (10,000 units TEV protease for every 100 mg of target protein). The solution was transferred to a Slide-A-LyzerTM dialysis cassette (20,000 MWCO) and dialyzed against 50 mM NaH2PO4, 100 mM NaCl, pH 8.0 at 4 °C for 3 days (4 L volume). A stir bar was used to agitate the dialysis buffer, and the dialysis buffer was switched out every 8 hours.

To remove the cleaved His-tag-TEV protease unit, the dialyzed protein was removed from the dialysis cassette and loaded onto a column containing pre-equilibrated HisPurTM cobalt resin. Flow through was collected to obtain the TEV-cleaved pure mCh_SC. The cleaved His-tag and the TEV protease enzyme remained attached to the cobalt resin. Dialysis buffer (50 mM NaH2PO4, 100 mM NaCl, pH 8.0) was added until all mCh_SC left the column. A total volume of 50 mL of eluent was collected. A 50 μL aliquot of the eluent was collected and prepared for SDS-PAGE analysis by mixing the 50 μL sample with 50 μL of 2X SDS sample loading buffer (100 mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 20% (v/v) glycerol, 0.2% (w/v) bromophenol blue, 10% (v/v) β-mercaptoethanol) and boiling the sample at 100 °C for 5 minutes. A small portion of the cleaved mCh_SC was dialyzed using a Slide-A-LyzerTM dialysis cassette (20,000 MWCO) using step-wise dialysis from pH 8.0 to pH 3.5. The pH was dropped incrementally by 0.5 while the solubility and fluorescence were monitored. The remaining cleaved pure mCh_SC protein was filtered through a 0.2 μm filter and concentrated using Amicon® Ultra-15 centrifugal filters with a 30 kDa NMWL. The volume was decreased from 50 mL down to approximately 6 mL of concentrated eluent. The concentrated mCh_SC was dialyzed using a Slide-A-LyzerTM dialysis

cassette (20,000 MWCO) using step-wise dialysis to 10 mM acetate pH 5.0. After dialysis, the protein solution was removed from the dialysis cassette and stored at 4 °C.

Concentration Calculation

A Coomassie (Bradford) Protein Assay Kit from Thermo Fischer Scientific was used to determine the concentration of the mCh_SC solution. Diluted albumin (BSA) standards were prepared with a final BSA concentration range of 0-2000 μg/mL using 10 mM Acetate pH 4.5 as a diluent. Dilutions were prepared using a sample that was a 100% solution of the concentrated mCh_SC and a sample that was a 2% solution of the concentrated mCh_SC. The 100% mCh_SC and 2% mCh_SC samples were each diluted using serial dilution to a dilution factor of 1, 2, 4, 8, 16, and 32. A 30 μL volume of each standard and unknown sample was combined with 1.5 mL Coomassie (G-250) reagent. The samples sat at room temperature for 10 minutes to equilibrate. The UV/Visible spectrophotometer was zeroed with ddH2O in a 1 cm cuvette at 595 nm. The absorbances of all standards and unknown samples were then measured at 595 nm in triplicate. The BSA standards and unknown samples were blank-corrected by subtracting the average of the absorbances of the 0 μg/mL BSA standard (containing only ddH2O and Coomassie) from the average of the absorbances of all standards and unknown samples. A BSA standard curve was prepared by graphing the blank-corrected BSA standard absorbances vs. the concentration of the BSA standards in μg/mL. A trendline and equation were produced using the points that allowed for the best linear trendline. The blank-corrected absorbances of the mCh_SC dilutions were plotted against the BSA standard curve. The mCh_SC dilutions within the concentration range of the linear potion of the BSA standard curve were used to determine the concentration of the original concentrated mCh_SC solution. The linear trendline equation was used to calculate the

concentrations of the unknown dilutions in μg/mL. The concentrations were multiplied by the dilution factors to determine the concentration of the original stock mCh_SC. The samples that originated from the 2% stock concentration were adjusted to reflect the concentration of the 100% stock solution. The concentrations were then converted to mg/mL by dividing by 1000 and then to molarity (M) by dividing by the molecular weight of mCh_SC (42406.74 g/mol). The final concentration of the stock mCh_SC solution was recorded in μM after multiplying by 106. After the concentration was determined, 50 μL volumes of pure concentrated mCh_SC in 10 mM Acetate pH 5.0 was aliquoted in 1.5 mL microcentrifuge tubes. The protein was stored at -80 °C until needed.

Linking of ST_HEAT Lysate with mCh_SC Lysate

The 3 mL aliquot of ST_HEAT lysate was combined with the 3 mL aliquot of mCh_SC lysate. The solution was incubated at room temperature with shaking for 1 hour. Aliquots of 100 μL were removed throughout the 1 hour linking process. The time points collected were 1 minute, 5 minutes, 10 minutes, 15 minutes, 30 minutes, 45 minutes, and 60 minutes after combination of the two samples. The aliquots were collected and immediately prepared for SDS-PAGE analysis by mixing the 100 μL sample with 100 μL of 2X SDS sample loading buffer (100 mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 20% (v/v) glycerol, 0.2% (w/v) bromophenol blue, 10% (v/v) β-mercaptoethanol) and boiling the sample at 100 °C for 5 minutes.

Linking of ST_HEAT HisPurTM Eluent with mCh_SC HisPurTM Eluent

The 100 μL of ST_HEAT HisPurTM eluent was combined with the 100 μL of mCh_SC HisPurTM eluent. The solution was incubated at room temperature with shaking for 24 hours.

Aliquots of 15 μL were removed throughout the 24-hour linking process. The time points collected were 2 minutes, 10 minutes, 30 minutes, 1 hour, 5 hours, 9 hours, and 24 hours after combination of the two samples. The aliquots were collected and immediately prepared for SDS-PAGE analysis by mixing the 15 μL sample with 15 μL of 2X SDS sample loading buffer (100 mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 20% (v/v) glycerol, 0.2% (w/v) bromophenol blue, 10% (v/v) β-mercaptoethanol) and boiling the sample at 100 °C for 5 minutes.

Conjugating ST_HEAT and mCherry_SC

0.1 mg of HPLC purified, lyophilized ST_HEAT was dissolved in 300 μL of filter-sterilized 10 mM Acetate, pH 5.0. The concentration of ST-HEAT was determined using UV-Vis spectroscopy, and then diluted to 5 μM with 10 mM Acetate, pH 5.0. This solution was then thermally annealed from 90˚C to 25˚C at a rate of -0.2˚C/minute. The solution was allowed to sit overnight at room temperature to encourage elongation of the ST_HEAT nanotubes. The ST_HEAT solution was filtered through 100 kiloDalton molecular weight cut-off centrifuge filters as described in the following passage.

The filter was primed with 500 μL of 10 mM Acetate buffer, pH 5.0, and spun at 14,000 x for 10 minutes. The flow through was discarded, and the ST_HEAT solution was added, followed by centrifugation at 14,000 x g for 7 minutes. The flow through was again discarded, and the protein was then washed with three aliquots of 500 μL of 10 mM Acetate buffer, pH 5.0. Each of these was spun at 14,000 x g for 7 minutes. The protein was collected by inverting the centrifuge filter in a clean collection tube., and centrifuging at 1,000 x g for 3 minutes. The filtered protein was then diluted to its original volume of 300 μL with 10 mM Acetate buffer, pH 5.0.

The proteins were linked using the SpyTag:SpyCatcher genetic fusion technique. 1 % v/v concentrated mCherry_SC was added to the filtered ST_HEAT solution, and the two were mixed overnight at room temperature. The resultant mCherry_HEAT solution was filtered similarly to the ST_HEAT solution; the sole difference in the filtration methodology is that the number of buffer washes is increased from three to ten.

3.6 References:

1. Kajava, A.V. Tandem repeats in proteins: From sequence to structure. *J Struct Biol*, **2011**, *179*, 279-288.

2. Kobe, B.; Kajava, A.V. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends in Biochem Sci* **2000**, *25*, 509-515.

3. Fournier, D.; Palidwor, G.A.; Shcherbinin, S.; Szengel, A.; Schaefer, M.H.; Perez-Iratxera, C.; Andrade-Navarro, M.A. Functional and Genomic Analyses of Alpha-Solenoid Proteins. *PLOS One.* **2013**, *8*, 1-13.

4. Andrade, M.A.; Bork, P. HEAT repeats in the Huntington's disease protein. *Nat. Genet.***1995**, *11*, 115-116.

5. Kobe, B.; Gleichmann, T.; Horne, J.; Jennings, I.G.; Scotney, P.D.; Teh, T. Turn up the HEAT. *Structure.* **1999,** *7*, 91-97.

6. Plückthun A. Designed ankyrin repeat proteins (DARPins): binding proteins for research, diagnostics, and therapy. *Annu Rev Pharmacol Toxicol*, **2015**, *55*, 489-511.

7. Kaspar Binz, H.; Stumpp, M. T.; Forrer, P.; Amstutz, P.; Plückthun, A. Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Bio.* **2003**, *332*, 489-503.

8. Kaspar Binz, H.; Amstutz, P.; Kohl, A.; Stumpp, M. T.; Briand, C.; Forrer, P.; Grütter, M. G.; Plückthun, A. High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotech.* **2004**, *22*, 575-582.

9. Steiner, D.; Forrer, P.; Plückthun, A. Efficient Selection of DARPins with Sub-nanomolar Affinities using SRP Phage Display. *J. Mol. Bio.* **2008**, *382*, 1211-1227.

10. Simon, M.; Frey, R.; Zandemeister-Wittke, U.; Plückthun, A. Orthogonal Assembly of a Designed Ankyrin Repeat Protein–Cytotoxin Conjugate with a Clickable Serum Albumin Module for Half-Life Extension. *Bioconjugate Chem.* **2013**, *24*, 1955-1966.

11. Martin-Killias, P.; Stefan, N.; Rothschild, S.; Plückthun, A.; Zangemeister-Wittke, U. A novel fusion toxin derived from an EpCAM-specific designed ankyrin repeat protein has potent antitumor activity. *Clin Cancer Res.*, **2011,** *17*, 100-110.

12. Zahnd, C.; Kawe, M.; Stumpp, M. T.; de Pasquale, C.; Tamaskovic,R.; Nagy-Davidescu, G.; Dreier, B.; Schibli, R.; Kaspar Binz, H.; Waibel, R.; Plückthun, A. Efficient Tumor Targeting with High-Affinity Designed Ankyrin Repeat Proteins: Effects of Affinity and Molecular Size. *Cancer Res.* **2010,** *70*, 1595-1605.

13. Léger, C.; Di Meo, T.; Aumont-Nicaise, M.; Velours, C.; Durand, D.; de la Sierra-Gallay, I.L.; van Tilbeurgh, H.; Hildebrandt, N.; Desmadril, M.; Urvoas, A.; Valerio-Lepiniec, M.; Minard, P. Ligand-induced conformational switch in an artificial bidomain protein scaffold. *Sci Reports*, **2019,** *9*, 1178-1191.

14. Hadpech, S.; Nangola, S.; Chupradit, K.; Fanhchaksai, K.; Furnon, W.; Urvoas, A.; Valerio-Lepiniec, M.; Minard, P.; Boulanger, P.; Hong, S.-S.; Tayapiwatana, C. Alpha-helicoidal HEAT-like Repeat Proteins (αRep) Selected as Interactors of HIV-1 Nucleocapsid Negatively Interfere with Viral Genome Packaging and Virus Maturation. *Sci Reports,* **2017,** *7*, 16335-16354.

15. Valerio-Lepiniec, M.; Urvoas, A.; Chevrel, A.; Guellouz, A.; Ferrandez, Y.; Mesneau, A; de la Sierra-Gallay, I.L.; Aumont-Nicaise, M.; Desmadril, M.; van Tilbeurgh, H.; Minard, P. The

αRep artificial repeat protein scaffold: a new tool for crystallization and live cell applications**.** *Biochem Soc Trans,* **2015,** *43,* 819-824.

16. Di Meo, T.; Ghattas, W.; Herrero, C.; Velours, C.; Minard, P.; Mahy, J.-P.; Ricoux, R.; Urvoas, A. αRep A3: A Versatile Artificial Scaffold for Metalloenzyme Design. *Chem Eur J*, **2017,** *23*, 10156 – 10166.

17. Chevrel, A.; Mesneau, A.; Sanchez, D.; Celma, L.; Quevillon-Cheruel, S.;Cavagnino, A.; Nessler, S.; de la Sierra-Gallay, I.L.; van Tilbeurgh, H.; Minard, P.; Valerio-Lepiniec, M.; Urvoas, A. Alpha repeat proteins (αRep) as expression and crystallization helpers *J Struct Bio* **2018**, *201*, 88–99.

18. Chevrel, A.; Urvoas, A.; de la Sierra-Gallay, I.L.; Aumont-Nicaise, M.; Moutel, S.; Desmadril, M.; Perez, F.; Gautreau, A.; van Tilbeurgh, H.; Minard, P.; Valerio-Lepiniec, M. Specific GFP-binding artificial proteins (*α*Rep): a new tool for in vitro to live cell applications. *Biosci Rep,* **2015**, *35*, 1-15.

19. Guellouz, A.; Valerio-Lepiniec, M.; Urvoas, A.; Chevrel, A.; Fourati-Kammoun M. G. Z.; Desmadril, M.; van Tilbeurgh, H.; Minard, P. Selection of Specific Protein Binders for Pre-Defined Targets from an Optimized Library of Artificial Helicoidal Repeat Proteins (alphaRep). *PLOS ONE*, **2013**, *8*, 1-20.

20. Gurunatha, K.L.; Fournier, A.C.; Urvoas, A.; Valerio-Lepiniec, M.; Marchi, V.; Minard, P.; Dujardin, E. Nanoparticles Self-Assembly Driven by High Affinity Repeat Protein Pairing. *ACS Nano* **2016**, *10*, 3176−3185.

21. Urvoas, A.; Guellouz, A.; Valerio-Lepiniec, M.; Graille, M.; Durand, D.; Desravines, D.C.; van Tilbeurgh, H.; Desmadril, M.; Minard, P. Design, production and molecular structure of a new

family of artificial alpha-helicoidal repeat proteins (αRep) based on thermostable HEAT-like repeats. *J Mol Biol,* **2010,** *404*, 307-327.

22. Zakeri, B.; Howarth, M. Spontaneous Intermolecular Amide Bond Formation between Side Chains for Irreversible Peptide Targeting. *J Am Chem Soc*, **2010**, *132*, 4526-4527.

23. Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M., Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *PNAS,* **2012**, *109*, 690-697.

Chapter IV. Computational Design of Helical Filaments

4.1 Computational Versus Rational Versus *De Novo* Design

Native proteins perform the most complex tasks in biochemistry; systemic repair, light harvesting, and all other facets of metabolism rely on the near-perfect function of globularly folded proteins. The incredible specificity with which native proteins interact with other

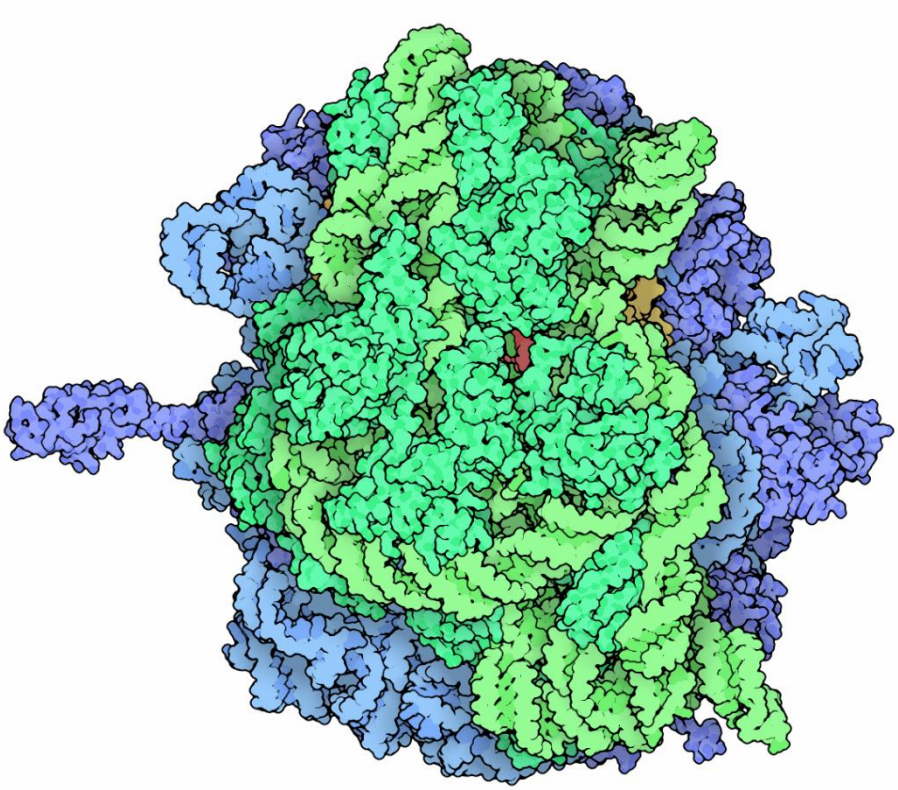


Figure 4.1 Illustration of a bacterial ribosome, with the large subunit in blue, and the small subunit in green.

biomacromolecules and small molecules is encoded by their complexity. Perhaps the most impressive example of this complexity is the ribosome; comprising two massive protein-RNA complexes, the ribosome is the piece of supramolecular machinery that synthesizes new protein

(Figure 4.1). Protein design aims to eventually emulate and even surpass this complexity. There are three general strategies for designing protein complexes: *de novo* design, rational design, and computational design. *De novo* design relies on using established rules to design protein assemblies from scratch; so far, rules have been established for single α-helices, β-sheets, and

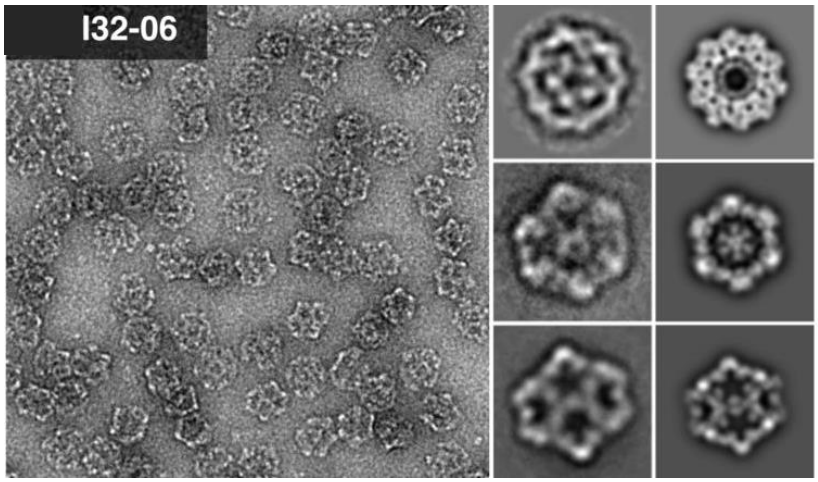


Figure 4.2 Negative-stained TEM of one of the designed icosahedral particles (left) with computationally averaged particles (left inset) and back-projections calculated from the models (right inset). From "Accurate design of megadalton-scale two-component icosahedral protein complexes. Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P.; Baker, D. Science, 2016, 353, 389-394." Reprinted with permission from AAAS.

coiled-coil systems. *De novo* design has not yet successfully been applied to assemblies containing more complex subunits, and the design of globular proteins is especially elusive. Rational design starts with a known structure and employs advantageous mutations to key positions to produce novel assemblies. Rational design has been used for complex motifs, such as helical hairpins (see Chapter 2) and ankyrin repeats, which contain two helices and two β-sheets. Computational design is the newest of these strategies, and the continuous increase in computing power has led to massive increases in the complexity of designed systems. Recent work in the Baker group showed

the design of several icosahedral, virus-like particles (Fig 4.2.)[1]. These massive complexes comprise 120 subunits (60 each of two proteins), and represent the largest nanostructure

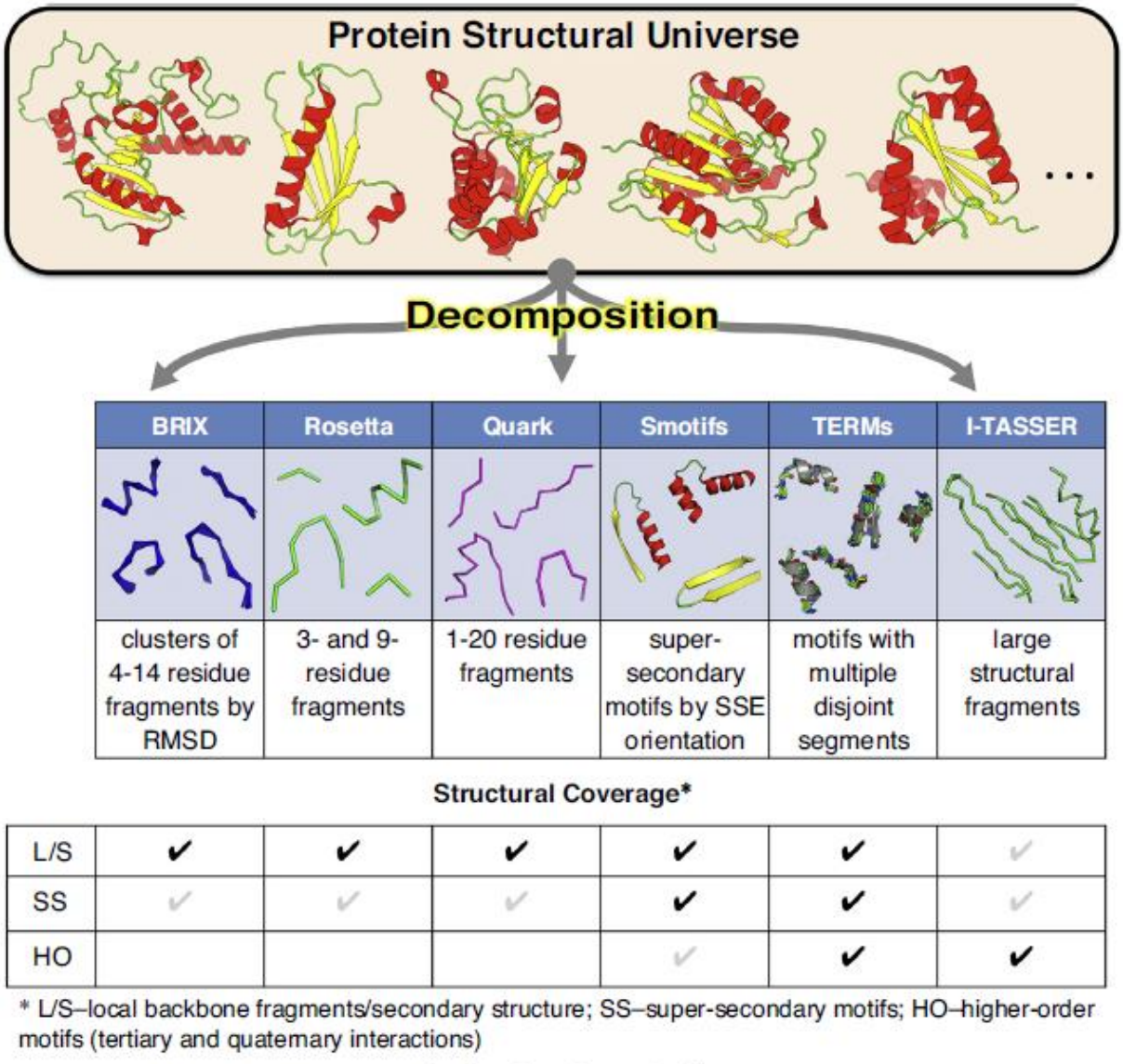


Figure 4.3 Scheme comparing multiple methods of parsing the protein structural universe. Effectiveness in covering different ranges of structural features is assessed using checkmarks, where a bold check denotes highest level of coverage, and a light grey check denotes poor coverage. Reprinted from "Current Opinions in Structural Biology, 44, Mackenzie, C. O.; Grigoryan, G. , Protein structural motifs in prediction and design, 161-167." Copyright (2017), with permission from Elsevier.

characterized at high resolution. This strategy is the most likely to produce designed assemblies with similar complexity to those found in nature. Additionally, computational design allows us to probe structures outside of the "protein universe", which describes all of the natively designed protein folds. Due to the evolutionary origin of native proteins, they are highly degenerate in structure, and sample a small portion of the physically possible protein folds. Some fascinating work (also from the Baker group) was done in which a left-handed α-helical tandem-repeat protein-based solenoid was designed to high accuracy[2]; natively, all α-helical tandem-repeat proteins fold with a right-handed superhelical structure.

4.2 TERM-based Designs and the Sol Series

The ultimate test of our current understanding of protein folding is to computationally design a structure, synthesize the corresponding sequence, and show with high-resolution biophysical analysis that we have successfully produced that same structure. A great deal of progress has been made in recent years regarding this goal; the use of Rosetta to design protein folds has led to an explosion of designed protein crystal structures. However, producing meso-scale structures from simple building blocks using computational design has remained quite challenging, with no symmetry-conserved 1-, 2-, or 3-dimensional nanomaterials reported. Recently, Shen et al. described some computationally designed filaments that closely match their predicted structures; however, the local symmetry of the repeats is not conserved in the filamentous structure[23]. In collaboration with the Grigoryan Laboratory at Dartmouth College, we developed a general approach to designing solenoidal repeat proteins to form superhelical filaments.

The first successful instance of this is the Sol series of peptides (short for Solenoidal). These were designed using tertiary motifs, or TERMs. TERMs are a general classification for

structural motifs that describe an amino acid in a protein structure, and every inter- or intra-strand contact that it makes within that protein structure. The Grigoryan lab has previously reported the effectiveness of describing protein folding using TERMs; amazingly, it took only 600 TERMs to describe 50% of the known protein structural universe[3-4]. Other attempts at parsing the protein structural universe have relied on either small fragments (BRIX[5-8], Rosetta[15-19], and Quark[12-14]) or large subunits (Smotifs[20-21] and I-TASSER[9-11]); in all of these cases, only continuous segments are used (Figure 4.3). TERMs describe interactions including those between disjoint segments; this allows for a more accurate representation of real protein folding, using the minimal number of amino acids (which helps to minimize computational cost).

The Sol series of peptides were based on an α/β hairpin, like that of the Leucine Rich Repeat, and were designed to form helical nanotubes, albeit with very small lumena. Excellent agreement between the designed structure, and the actual structure (Modlin *et al.,* unpublished work) was very encouraging. However, the approach used for the generation of the Sol assemblies was limited; it was discovered that the solubility of the peptides was very low, even with the replacement of a key hydrophobic residue with a threonine. Additionally, the persistence lengths of the assemblies were very low, making them poor candidates for cryo-EM based helical reconstruction.

4.3 Helical Assembly Builder

A useful starting point for the design of helical nanotubes is identifying the self-assembly subunit. In order to more easily compare the computationally designed peptide assemblies to the natively designable motifs discussed in Chapters II and III, the selected subunit for Helical Assembly Builder design was a helical hairpin comprising fourteen amino acids in each α-helix, and a disordered turn varying in length from one to five amino acids; HEAT_R1 is a helical hairpin comprising two fourteen amino acid α-helices joined by a three amino acid turn. Using bespoke software developed in the Grigoryan lab, every instance of a helical hairpin fitting these parameters was extracted from the Protein Data Bank, and these structures were organized by the number of

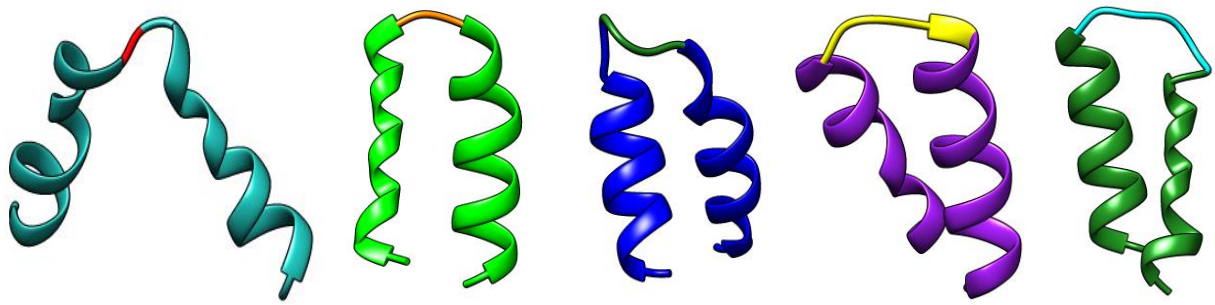


Figure 4.4 Examples of the helical hairpins mined from the PDB. From left to right, 4ty0, 4mpq, 2qqy, 1wy0, and 3l9w. Structures are named based on their PDB accession code.

amino acids in the turn sequence. Once clustered, the structures were analyzed using the Helical Assembly Builder program, which varies the relative subunit orientation and number of subunits in a superhelical turn to generate single turn assemblies. These structures are then scored, with the most energetically favorable assembly from each parent structure advancing to the second round of scoring. Here, the scoring is relatively simple; it considers the potential for geometric clashing, as well as the favorability of the inter-subunit contacts that are predicted. Notably, the inter-subunit contacts are mainly lateral in nature; axial interactions in these structures were absent or negligible

in most cases. Some user input was required for evaluating structures, as the most energetically favorable structure was not always physically reasonable. In some cases, the proposed assembly was too small in diameter, which could lead to hydrophobic collapse into a helical fiber. Conversely, some proposed assemblies were much too large; in this case, it was more likely that the assembly would be a straight chain that would collapse into a globular aggregate (Figure 4.5). The optimal geometric assemblies from each parent structure were then compared, and two or three of the best-scoring assemblies from each cluster were selected for sequence-based design in PyRosetta; this was done to ensure adequate representation of helical hairpins with varying turn sequence length. PyRosetta is a Python-compatible iteration of the Rosetta protein design software package which is highly customizable. For each of the sixteen candidate structures that were

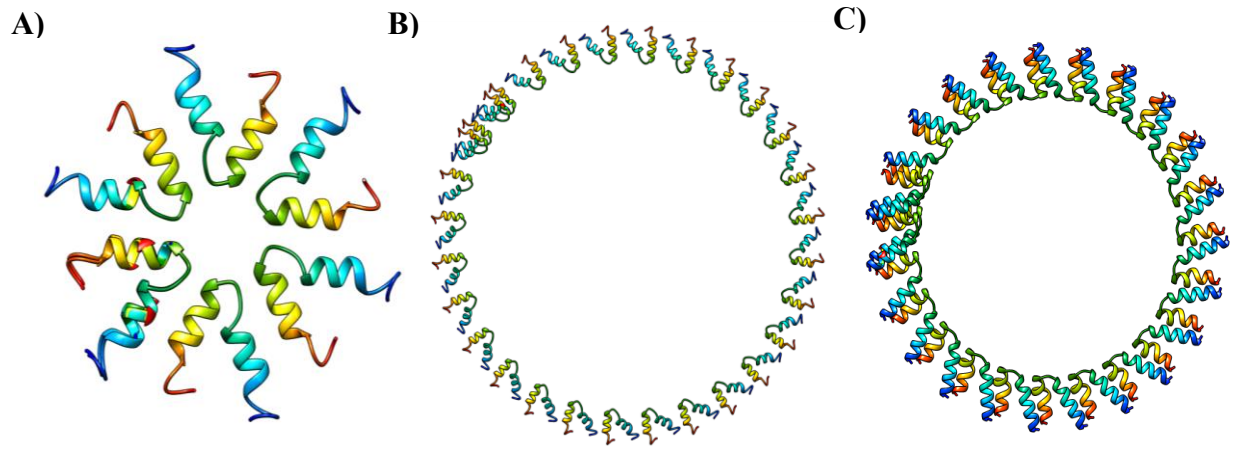


Figure 4.5 Three highly scoring assemblies generated from the helical hairpin in PDB: 1b3u. (A) Assembly featuring just six subunits in a superhelical turn. (B) Assembly featuring 24 subunits in a superhelical turn. (C) Assembly featuring ten subunits in a superhelical turn. (A) and (B) were discarded as non-physical solutions as described above.

selected, the single-turn assembly was lengthened to five or six full turns to provide the complete structural context for the subunit. Using this lengthened assembly, a single subunit would be selected, and every adjacent subunit that could possibly contact it would be included with it to

form a “template” structure. Usually this contained nine subunits, with the initial hairpin located in the center, as shown in Figure 4.6B. Owing to helical symmetry, a design template that incorporates a single subunit and all of its possible contacts completely describes the superhelical structure. Thus, an assembly containing ten thousand residues can be computationally designed from a chunk containing no more than 300 amino acids. Once generated, the template structure was fed into the PyRosetta design pipeline, which preserves its amide backbone, and optimizes the sidechains of each residue in an iterative fashion. At the first step in the PyRosetta design, the amino acid sequence for the template structure is mutated to a polyglycine chain. The iterative design proceeds using a Monte Carlo design principle; essentially, the structure is scored after

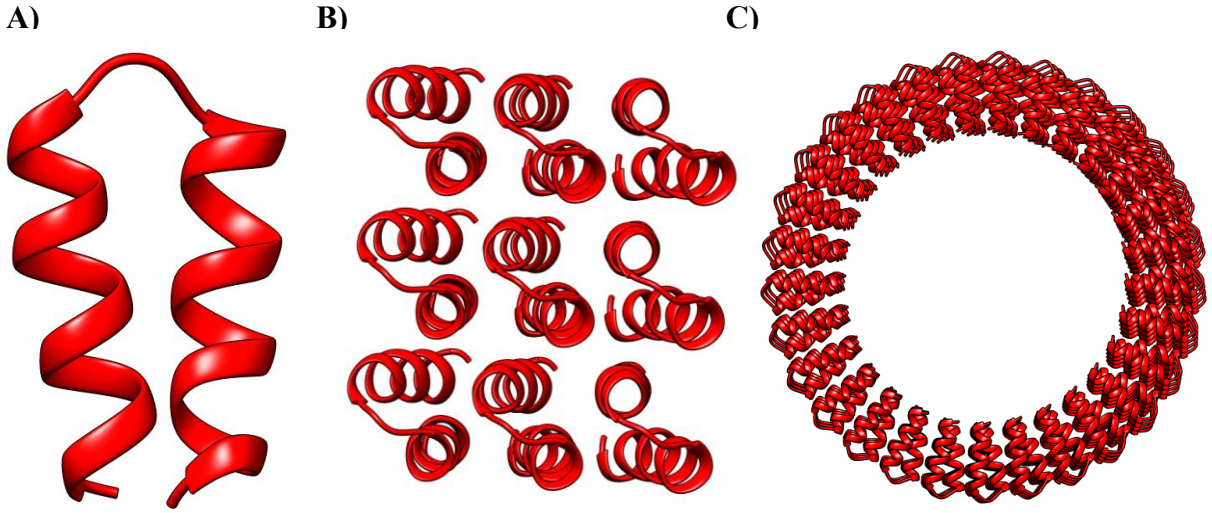


Figure 4.6 Subunit (a), template (b), and full assembly (c) of the helical hairpin found in PDB Structure 1wy0.

every mutation; mutations that provide a decrease in energy (lower energy correlates with higher stability) are automatically accepted, while mutations that *slightly* raise the energy (some cutoff is established here) are temporarily accepted for the next iteration, and mutations that raise the energy beyond the cutoff are thrown out. For example, in a given structure, if a Glycine → Alanine mutation results in a decrease in energy from -205 to -208 (energy units here are arbitrary), future

iterations will incorporate this mutation. If, however, we set an energy threshold of +2, and a Glycine → Methionine mutation results in an increase of energy from -205 to -204, this mutation is temporarily accepted for future iterations. This scheme of energy minimization is commonly used to prevent structures from falling into a thermodynamic sink (local minima on an energy landscape). A residue file is taken as one of the inputs for the optimization command; one can limit the potential amino acids to be mutated at each position. This functionality is quite useful in cases where one wants to limit hydrophobicity for the sake of solubility, or, more likely, avoid cysteines to negate concerns of disulfide-bridging. Some patterns emerge after several thousand optimization steps (Figure 4.7). A sequence that starts as a polyglycine chain will often retain some

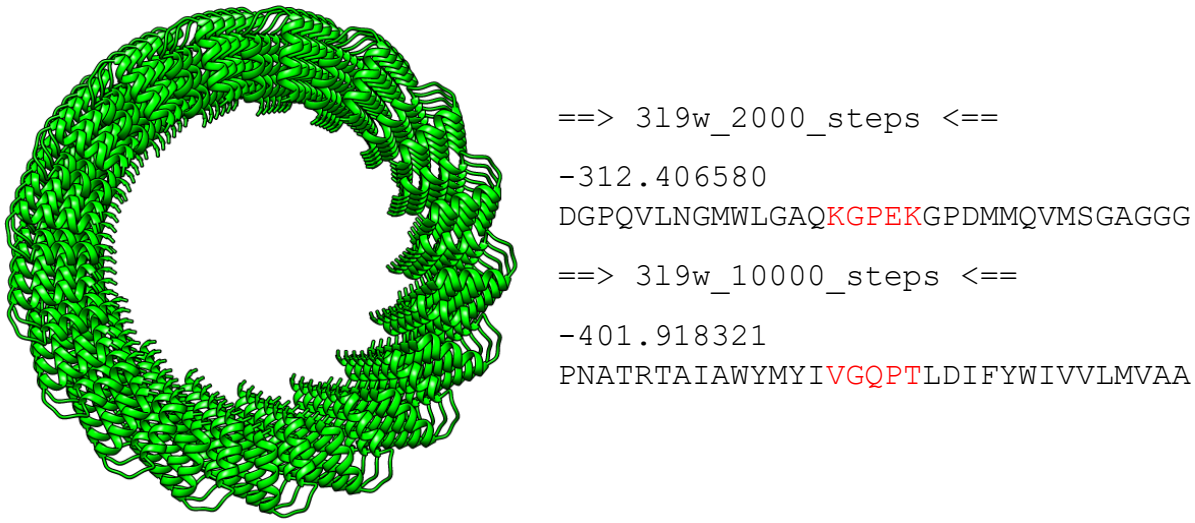


Figure 4.7 Full assembly of the helical hairpin from PDB structure 3l9w, with the PyRosetta optimized sequences at (top) 2000 and (bottom) 10000 iterations. Scores at arbitrary scale are provided as a comparison.

glycine residues at non-turn positions; this is significant because glycine tends to destabilize helices because of its flexibility. However, after ten thousand optimization steps, the only glycine

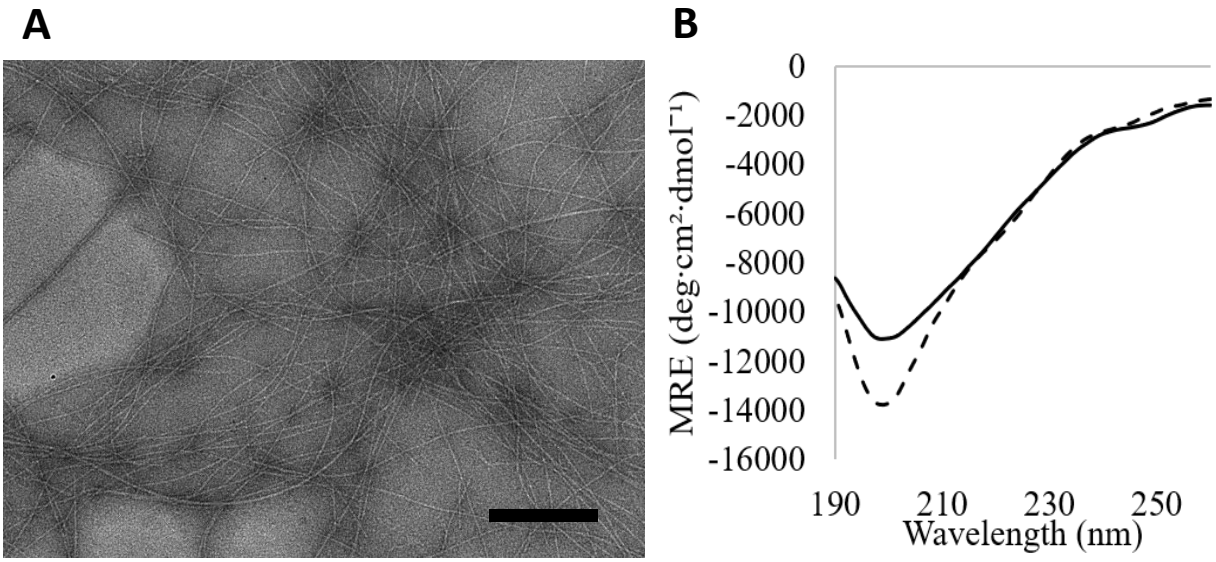


Figure 4.8 Negative-stain EM of 1wy0_F assembled at pH 8.0, with TFE (A) shows high-aspect ratio filaments of monodisperse diameter (scale bar is 500 nm), suggesting ordered assemblies. Representative CD spectra (B) of 1wy0_F under ideal assembly conditions. The solid line denotes the unannealed pH 8.0 sample, and the dashed line denotes the annealed pH 8.0 sample. A strong minimum is observed at ~195 nm, suggesting a random coil secondary structure.

residue is found inside the flexible turn, and proline is only found in the turn or at one of the sequence termini. To evaluate the effectiveness of helical assembly builder, we synthesized a total of twenty-four peptides. Of these, fourteen were soluble enough to be purified via HPLC. The Rosetta design software tends to favor structures with strong interfaces; this can lead to "overpacked" designs with little room for solvation. Each of the fourteen purified peptides was then assembled from pH 5.0 to 8.0. CD spectroscopy was used as an initial screening of assemblies and revealed that the majority of the designed peptide sequences did not fold into alpha-helices.

This is despite the ready formation of filaments in most of the assemblies (Figure 4.8). Due to this deviation from the theoretical assembly models, the design method was re-evaluated.

TERM-based design[22] was employed to create assemblies with native-like sequences, in order to promote proper folding and solubility. Two peptide sequences were designed using TERMs, and preserving the helical hairpin subunit structure: 4ty0_3.5a and 3dhi_4.5a. These sequences were synthesized and purified, and assembled from pH 5-8 as previously described. Interestingly, both peptides showed more stable CD signatures than the helical assembly builder designs, with little fluctuation based on pH. Moreover, 4ty0_3.5a displayed a helical CD signature under certain conditions (Figure 4.9), and a blue-shifted helical signature under others. 3dhi_4.5a displayed a CD signature that is commonly seen in short α-helices and 3_{10} helices. The computational models for 4ty0_3.5a and 3dhi_4.5a predicted outer diameters of 6.9 nm and 8.4 nm respectively. The filaments assembled from 4ty0_3.5a appeared to wind and unwind, making diameter estimation unreasonable. Conversely, the filaments observed in the 3dhi_4.5a assemblies were monodisperse in diameter, and ImageJ analysis determined that the measured and theoretical diameters were in close agreement (8.4 nm theoretical, 8.4 ± 0.3 nm measured). Despite this close agreement, attempts to obtain SAXS scattering data were unsuccessful; heavy beam degradation and low apparent concentration caused the sample signal to match that of the buffer (SAXS measurement of 4ty0_3.5a was not attempted due to the lack of homogeneity, and prevalence of fibers over tubes). In any case, the fact that neither of these advanced designs displayed both a strong helical signature and monodisperse filaments inspired us to move to a new design scaffold. We believe that the α-helical hairpin subunit may have been too complicated of a scaffold to test early version design software.

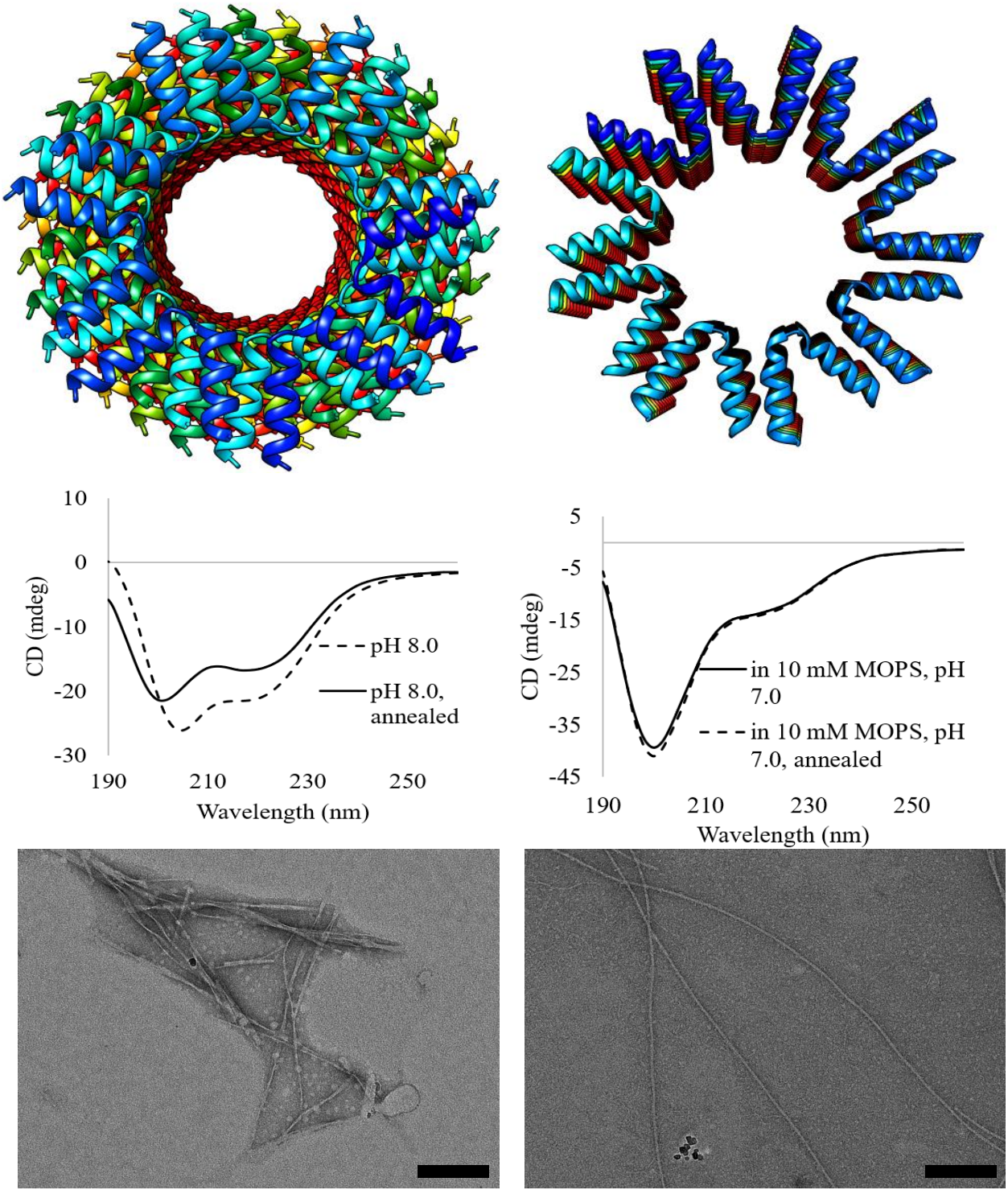


Figure 4.9 Computational models (top) of the 4ty0_3.5a (left) and 3dhi_4.5a (right) assemblies, with corresponding CD spectra (middle) and negatively-stained EM images (bottom, scale bar is 200 nm).

4.4 PDCC and GGHEAT

Through our biophysical characterization of both naturally-derived and computationally-designed peptides, we seemed to always generate assemblies with strong, lateral interactions between subunits, and relatively weak axial interactions between superhelical turns. We believe

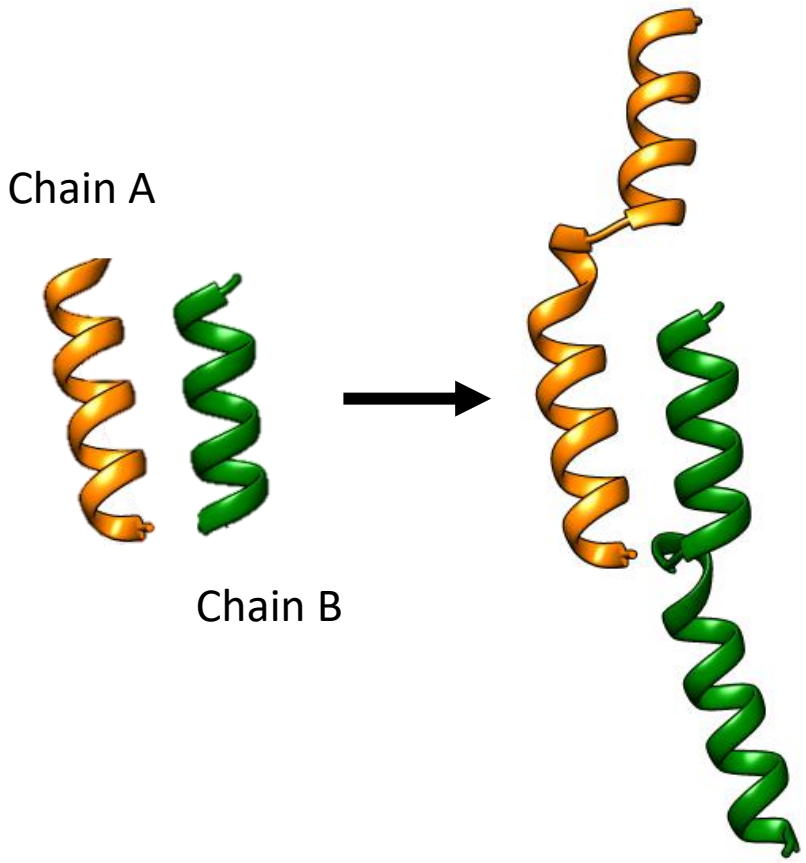


Figure 4.10 Design of the PDCC assembly. (Left) Non-covalently linked helices which undergo dimerization in solution. (Right) Depiction of two subunits of PDCC, wherein the inter-layer covalent interactions are shown.

this flexibility is what led to poorer resolution in the HEAT_R1 cryo-EM structure, and caused the apparent unwinding of several computationally-designed peptides. Naturally, we determined that an explicit design for axial interaction would be the most facile method for remedying this unwanted flexibility. Our first design target was a simplistic parallel, dimeric coiled-coil (PDCC)

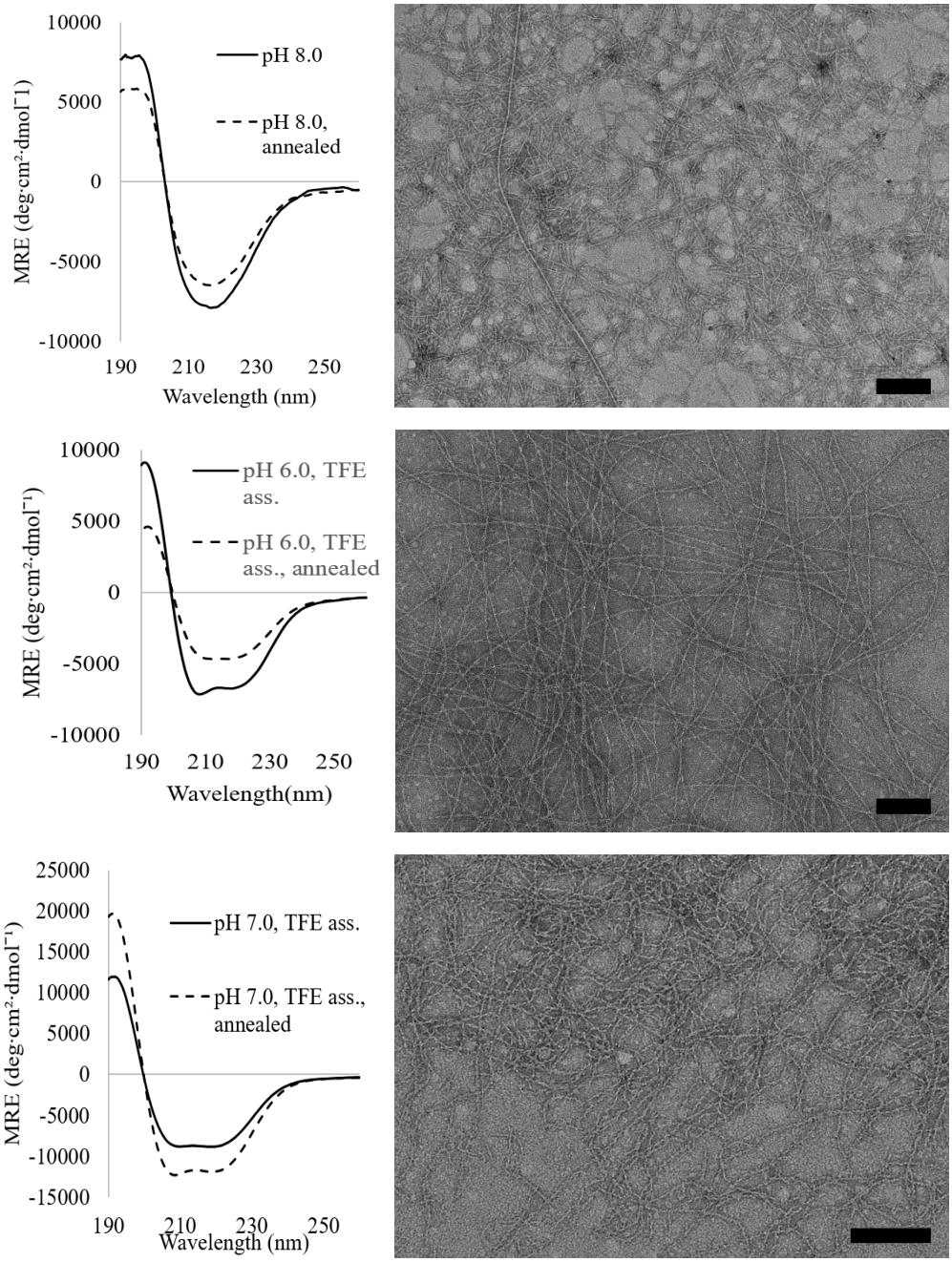


Figure 4.11 Assembly data for PDCC_1.1 (top), PDCC_1.2 (middle), and PDCC_1.3 (bottom). Circular dichroism spectra (left) show clear α-helical signals for PDCC_1.2 and PDCC_1.3, and a mixture of α-helix and β-sheet for PDCC_1.1. Negative-stain EM (right) show high-aspect ratio filaments with no visible lumen. Scale bars are 500 nm (top, middle) and 200 nm (bottom).

motif. The basic idea was to connect the two helices that would normally dimerize to form a coiled-

A)

Ac-ESIGRAAVEAARSATTKELGEQARDAAREARE-NH_2 | GGHEAT_1

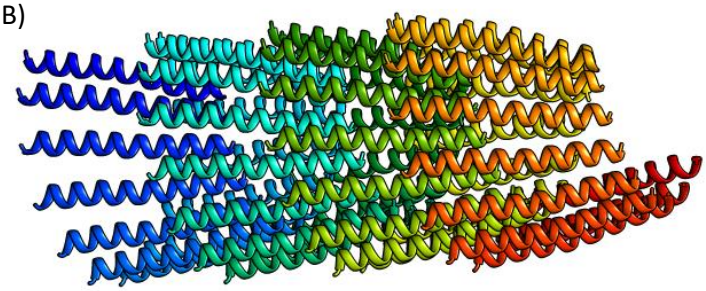
Ac-ESIGRAAAEAARSATTKELGEQARDAAREARE-NH_2 | GGHEAT_1.1

Ac-ESIGRAAAEAARSATTKELGEQARDAAREGRE-NH_2 | GGHEAT_1.2

Ac-EGIGRAAAEAARSATTKELGEQARDAAREGRE-NH_2 | GGHEAT_1.3

Ac-EAVARAAVEAARSATLKELLKEVKELAREARE-NH_2 | GGHEAT_2

B)



C)

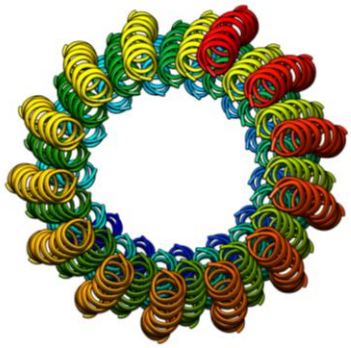


Figure 4.12 (A) Sequences of the five GGHEAT peptides; the differences between the different sequences containing 18 hydrophilic residues (GGHEAT_1-GGHEAT_1.3) are highlighted in red. The peptide sequence containing only 16 hydrophobic residues (GGHEAT_2) is also displayed. (B) Side view of the tubular assembly in which one conical helical turn nests into the turn above it. (C) Top-down view of the assembly model, with a prominent lumen.

coil via an axial turn (Figure 4). The three top-scoring designs were selected, and named PDCC_1.1, PDCC_1.2, and PDCC_1.3, in order of least to most optimization steps. TEM of PDCC_1.1 showed high-aspect ratio filaments with no clearly distinguished inner lumen, and CD revealed the secondary structure to be a mixture of α-helix and β-sheet character. PDCC_1.2 and PDCC_1.3 both displayed a strong α-helical CD signature, and filaments that looked like those found in PDCC_1.1. The inner lumen of these assemblies is not visible by negative-stain TEM, though, due to the flexibility and small diameter of the assemblies, it is possible that the stain

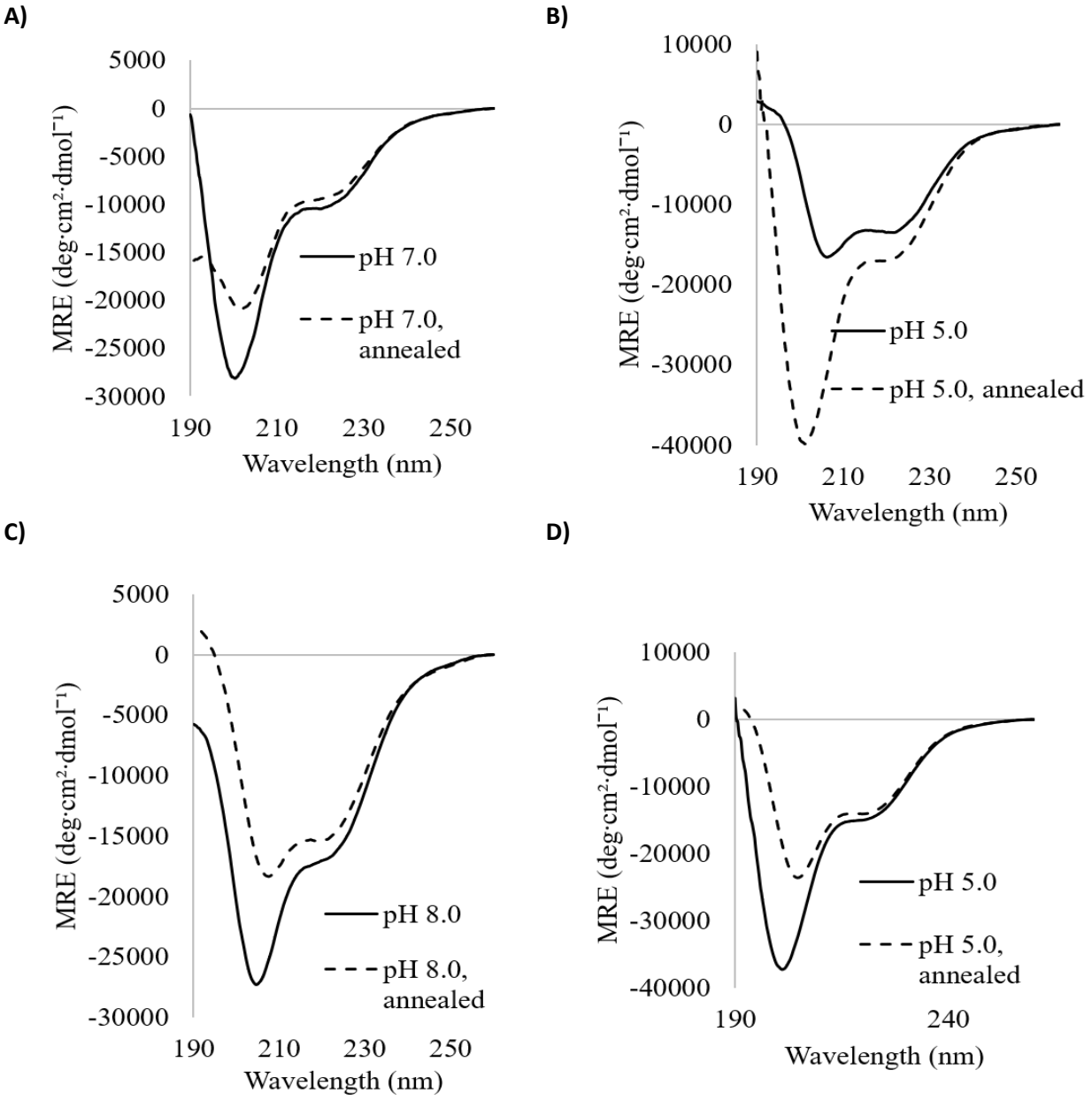


Figure 4.13 CD spectra of (A) GGHEAT_1, (B) GGHEAT_1.1, (C) GGHEAT_1.2, and (D) GGHEAT_1.3 at their optimal assembly conditions. Each peptide showed a CD signature consistent with a mixture of α-helix and random coil.

would not permeate through the structure. The most promising assemblies of PDCC_1.2 were analyzed via SAXS, but no significant signal was observed. This uncertainty led us to pursue a

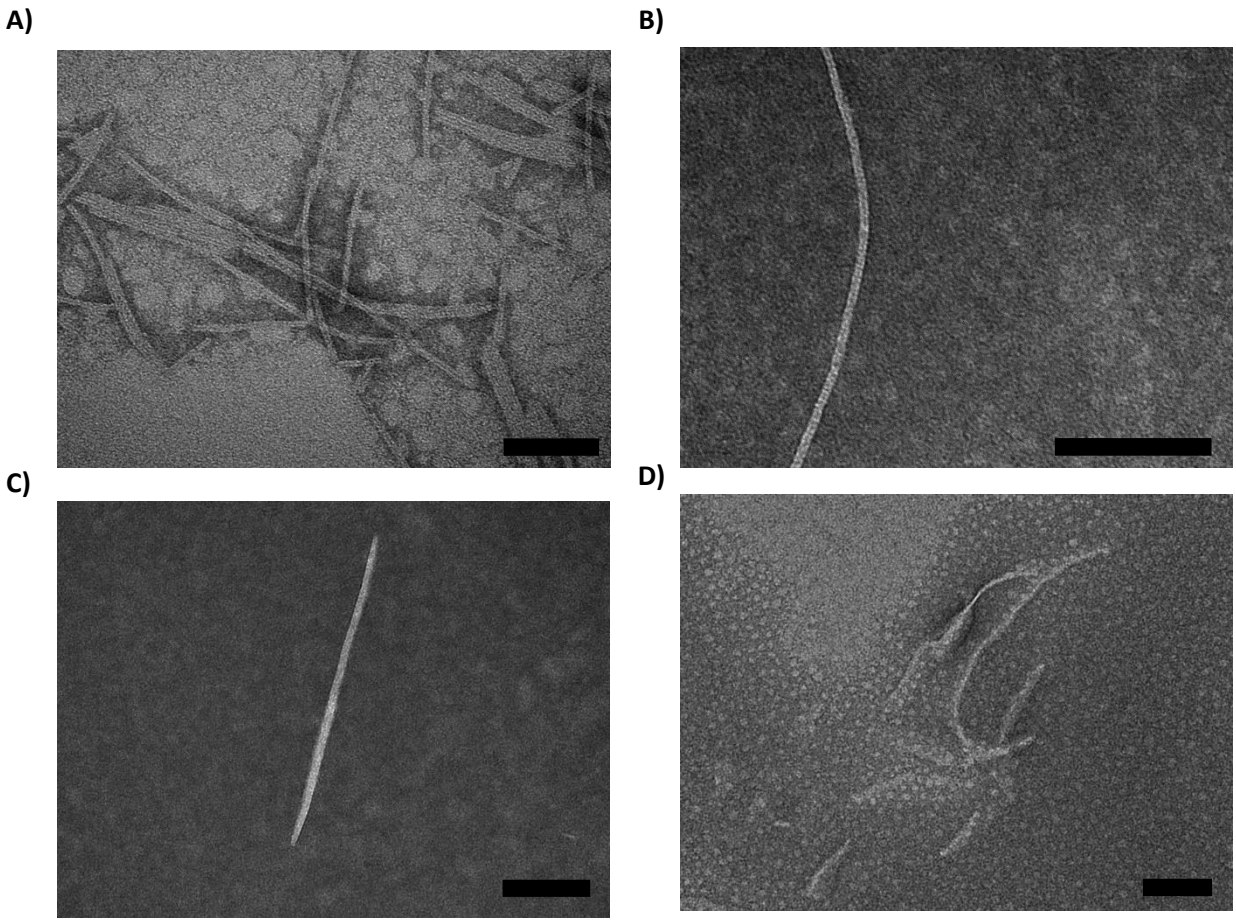


Figure 4.14 Representative negative-stain TEM images of (A) GGHEAT_1, (B) GGHEAT_1.1, (C) GGHEAT_1.2, and (D) GGHEAT_1.3, in the same assembly conditions shown in Figure 4.13

more promising design scaffold, which combined the axial finger motif with the HEAT peptide structure, known as GGHEAT. The helix-turn-helix subunit of 3LTJ (the parent crystal structure of the HEAT_R1 peptide) associates with other subunits through non-covalent interactions. Thus, the turn between helices in the subunit is not aiding in the self-assembly process. We reasoned then that this turn could be leveraged if it were designed to connect two helices in an axial fashion, rather than a lateral one (similar to the concept of PDCC). This transforms the helix-turn-helix

subunit into a single, long α-helix, which is predicted to self-assemble in a superhelical fashion (Figure 4.12B).

All five peptides were assembled at pH 5.0, pH 6.0, pH 7.0, and pH 8.0, under room temperature and thermal annealing conditions, as previously described. The CD signatures observed for all of the peptides was a mixture of α-helical and random coil signature. The assemblies were visualized using negative-stain TEM. For GGHEAT_1, GGHEAT_1.1, GGHEAT_1.2, and GGHEAT_1.3, tubular filaments were observed, with large degrees of unwinding. This is evident in Figure 4.14D, where a filament is shown in transition from a nanotube with a clearly defined inner lumen to a flattened sheet of peptide. Some buffer conditions promoted, and some buffer conditions inhibited, this unwinding effect, but ultimately every assembly condition for these peptides resulted in a mixture of tubular filaments, and various stages of unwound, associated peptide. In initial assemblies of GGHEAT_2, very few filaments were observed, and assembly with trifluoroethanol was employed to promote helicity. The helical content via CD was increased, but no filaments were observed. Serendipitously, one of the TFE-assisted assemblies produced needle-like crystals of ~ 4 mm length (visible to the naked eye). These crystals were mounted in an X-ray diffractometer, and high-quality diffraction data were collected. Ultimately, the unit cell of GGHEAT_2 was determined from the diffraction data and is shown in Figure 4.15. Not surprisingly, the unit cell consisted of interacting α-helices bearing the sequence of the peptide we synthesized; however, the predicted symmetry was not observed. Rather, four helices interact in a staggered fashion to create a flat surface which propagates at will to form the long, thin crystals we observed. Because no filaments were observed for the crystal-forming sample, but some are found in the non-crystal forming samples, we reasoned that the

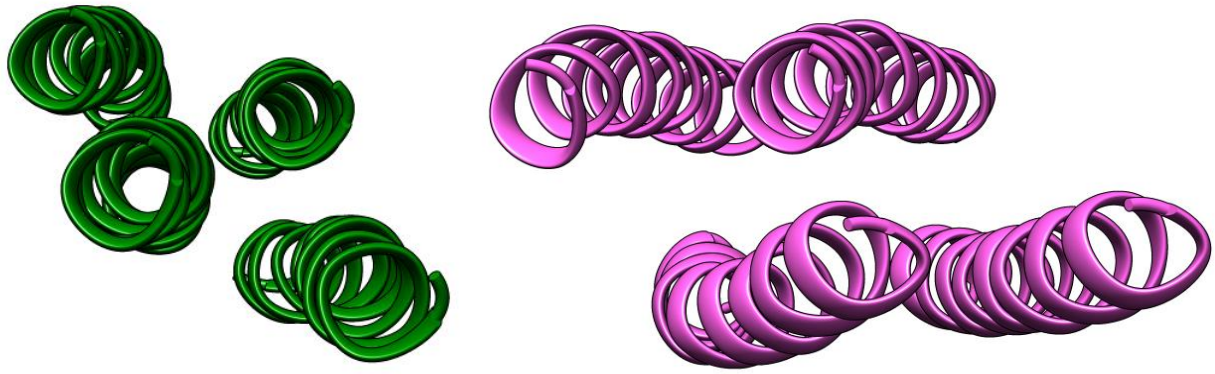


Figure 4.15 View of four helices in the GGHEAT computational model (green) and in the GGHEAT_2 crystal structure (pink). The helices in the computational model are all in the same angular orientation relative to one another, whereas those in the crystal structure adopt two distinct angular orientations (the ends of the top row of helices are pointed down, and the ends of the bottom row of helices are pointed up).

packing of subunits in the crystalline form is likely different from the packing observed in the helical assemblies. Unfortunately, due to low reproducibility of the filament-forming samples, it was impossible to collect higher resolution solution-phase data. Ultimately, GGHEAT represented a promising step towards generating self-assembling helical peptides using computation, but reproducibility issues precluded a more meaningful analysis of the designed assemblies.

4.5 Conclusion

Through multiple generations of computational design, we were able to fine-tune an approach to predicting solenoidal protein assembly structures. While we have not been able to experimentally produce a structure that matched its computational assembly at high resolution yet, we have gained a great deal of insight into which design strategies are effective and which still need work. In a few steps, we were able to transition from generating insoluble or unfolded proteins to generating assemblies roughly matching the diameter and secondary structure that was

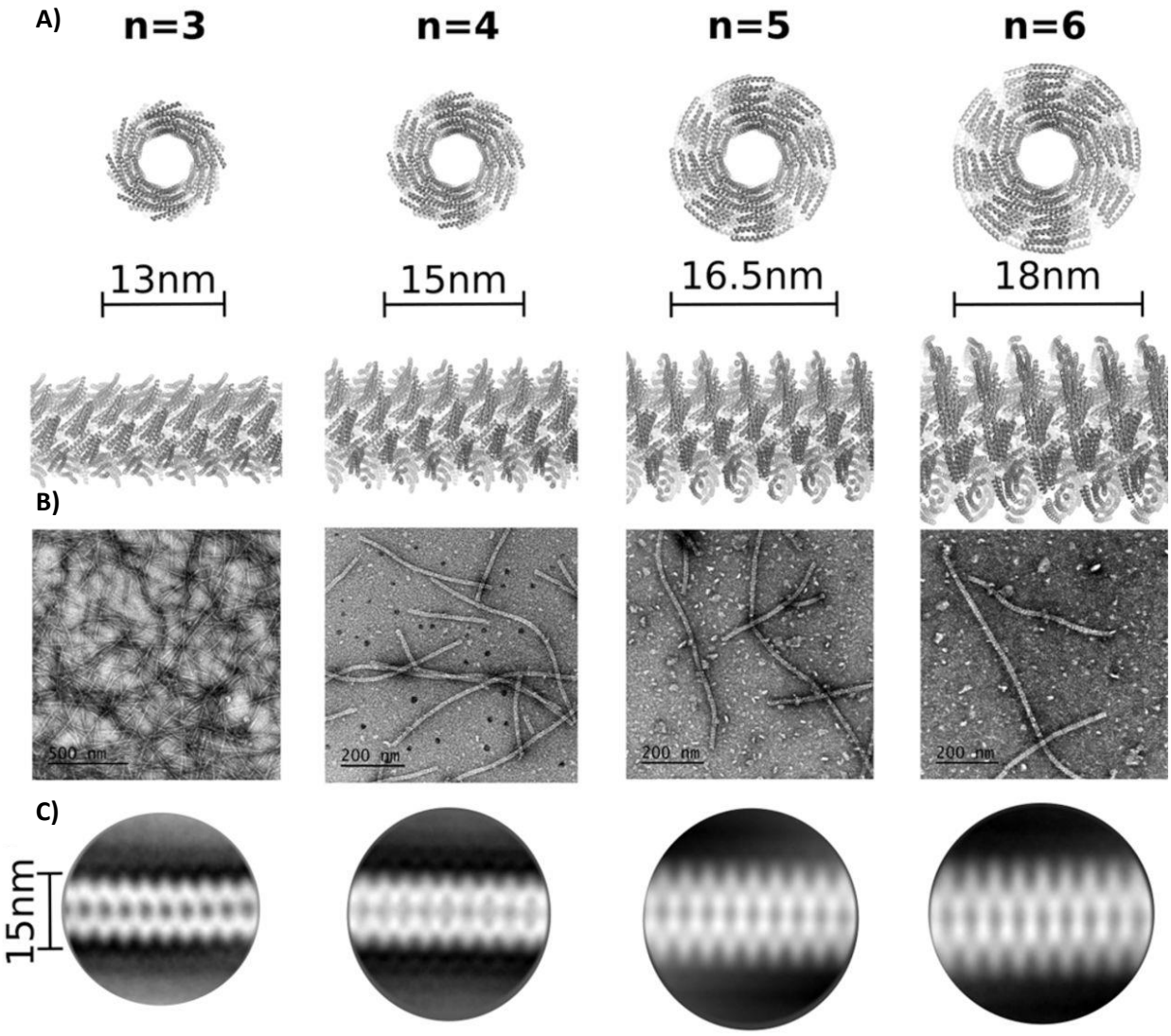


Figure 4.16 (A) Computational model of DHF58 with increasing numbers of repeats, and thus increasing diameter. (B) Negative-stained TEM of the DHF58 filamentous assemblies. (C) 2D-Class Averages of the filaments depicted in (A) and (B). A fifteen nanometer scalebar is provided to highlight the increasing diameter of the filament series, as a function of increased number of repeats in the protein. From "De novo design of self-assembling helical protein filaments. Shen, H.; Fallas, J. A.; Lynch, E. Sheffler, W.; Parry, B.; Jannetty, N.; Decarrea, J.; Wagenbach, M.; Vicente, J. J.; Chen, J.; Wang, L.; Dowling, Q.; Oberdorfer, G.; Stewart, L.; Wordeman, L.; de Yoreo, J.; Jacobs-Wagner, C.; Kollman, J.; Baker, D.. Science, 2018, 362, 705-709." Reprinted with permission from AAAS.

predicted. Computational design is rapidly becoming a powerful tool in the production of biomaterials. While computation is incredibly powerful, it is also only as effective as the assumptions and shortcuts that are input by protein designers; in this way, computational design can be used as a measurement of our understanding of the protein folding problem. As processing becomes more facile and efficient, we will likely reach a point where restraints and mathematical assumptions no longer need to be placed on the computations that we set up.

Perhaps the most promising development in computational design of helical filaments was published last year in *Science* by the Baker lab[23]. Expanding on previous work from their lab in which they generated designed helical repeat (DHR) proteins[24-26], they described designs in which assembly and disassembly can be triggered, and the diameter of the filaments can be fine-tuned by increasing or decreasing the number of repeats in the expressed protein (Figure 4.16). The design scheme used for these filaments relied on breaking the native symmetry of the scaffold protein, unlike the work we reported on the HEAT and LRV filaments. However, this approach allows for greater flexibility of scaffold choice, and may ultimately represent the path forward in the bespoke design of helical filaments for specific applications.

4.6 Materials and Methods

Chemicals and Materials

All chemical reagents were purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO) or Anaspec, Inc. (Fremont, CA) unless otherwise specified. Peptides not synthesized in house were ordered from GenScript USA, Inc. (Piscataway, NJ). Peptide synthesis resin was ordered from Applied Biosystems (Foster City, CA).

Peptide Synthesis and Purification Methods

In house synthesized peptides were prepared via microwave-assisted solid phase peptide synthesis on a CEM Liberty instrument as the *N*-acetyl, *C*-amide capped derivatives. A PAL-PEG-PS resin from Applied Biosystems was used for both peptides. Standard Fmoc protection chemistry was utilized in conjunction with coupling cycles consisting of 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 cleaved from the resin by incubation at room temperature for 3 hours in a cocktail consisting of trifluoroacetic acid (TFA), distilled water, triisopropylsilane, and 2,2'-(ethylenedioxy)-diethanethiol. Cleavage was followed by filtration and subsequent precipitation in diethyl ether. The peptide/diethyl ether mixture was then centrifuged at 4 °C at 4,000 rpm for 10 min. The supernatant was then discarded, and the precipitate allowed to desiccate overnight. Following desiccation, the crude peptide gels were resolubilized in 3 mL of a 50:50 mixture of acetonitrile and water (0.1% TFA additive) and purified by reversed-phase high-pressure liquid chromatography (HPLC) on a C18 column with a water-acetonitrile (0.1% TFA-additive) gradient. Peptide mass was confirmed using MALDI mass spectrometry. Purified HPLC fractions were then lyophilized, sealed, and stored at -30 °C.

Peptide Assembly Methods

Stock solutions of peptide (3 $mg \cdot mL^{-1}$) were prepared by solubilizing 3.0 mg of purified, lyophilized peptide in 500 μL of HPLC-grade water. 125 μL aliquots were then added to 20 mM buffers, resulting in four pH conditions: 10 mM Acetate (pH 5.0), 10 mM MES (pH 6.0), 10 mM MOPS (pH 7.0), and 10 mM TAPS (pH 8.0). Immediately after mixture, the solutions were titrated to their respective pH using dilute sodium hydroxide solution or hydrochloric acid. 125 μL of each solution was then thermally annealed using the following thermal cycler protocol: (1) rapid heating to 90° C for 30 minutes and (2) cooling to 25°C at a rate of -0.2°C/minute. TFE assemblies were

made in the same fashion, though prior to thermal annealing, an equal amount of TFE was added to the solution (125 μL TFE added to 125 μL of peptide/buffer mixture). Over 48 hours, TFE was completely evaporated inside a laminar flow hood.

Circular Dichroism Spectropolarimetry Methods

CD measurements were performed on a Jasco J-810 CD spectropolarimeter using 0.10 mm thick quartz plates (Hellma Analytics). Spectra were collected at 50 nm/min. in the range of 190-260 nm, and a data pitch of 0.2 nm.

Transmission Electron Microscopy Methods

TEM grids were prepared using solutions of peptide (3 $mg \cdot mL^{-1}$) in aqueous buffer (10 mM Acetate, pH 5.0, and 10 mM MES, pH 6.0, 10 mM MOPS, pH 7.0, and 10 mM TAPS, pH 8.0). Samples were prepared by depositing 4 μL of peptide solution onto a 200-mesh carbon-coated copper grid from Electron Microscopy Services (Hatfield, PA). After 1.5 minutes of incubation on the grid, moisture was wicked away, leaving only a thin film of sample. 4 μL of negative stain (2% methylamine tungstate) was deposited onto the thin film, to allow proper mixture. After 1 minute of staining, the remaining moisture was wicked away, and the grid dried overnight in a tabletop desiccator. Electron micrographs were captured on a Hitachi HT-7700 with a tungsten filament and AMT CCD camera, at an accelerating voltage of 80 kV.

Computational Methods

All generations of design were carried out with the support of the Computer Science Department at Dartmouth College. Design jobs were submitted to the computer cluster at Dartmouth from a local node, using the SSH client, PuTTY. For PyRosetta submissions, 10,000

steps was deemed to be an appropriate standard. TERM generation was carried out as described in Mackenzie *et al.* (2016).

4.7 References:

1. Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P.; Baker, D. Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science,* **2016,** *353*, 389-394.
2. Park, K.; Shen, B.W.; Parmeggiani, F.; Huang, P.S.; Stoddard, B.L.; Baker, D. Control of repeat-protein curvature by computational protein design. *Nat Struct Mol Biol*, **2015,** 22, 167-74.
3. Mackenzie, C. O.; Grigoryan, G. Protein structural motifs in prediction and design. *Curr. Opin. Struct. Biol.* **2017,** *44*, 161-167.
4. Mackenzie, C. O.; Zhou, J.; Grigoryan, G. Tertiary alphabet for the observable protein structural universe. *PNAS*, **2016,** *47*, 438-447.
5. Baeten, L.; Reumers, J.; Tur, V.; Stricher, F.; Lenaerts, T.; Serrano, L.; Rousseau, F.; Schymkowitz, J. Reconstruction of protein backbones from the BriX collection of canonical protein fragments. *PLoS Comput. Biol.* **2008**, *4*, 1-11.
6. Vanhee, P.; Verschueren, E.; Baeten, L.; Stricher, F.; Serrano, L.; Rousseau, F.; Schymkowitz, J. BriX: a database of protein building blocks for structural analysis, modeling and design. *Nucleic Acids Res.* **2011**, *39*, 435-442.
7. Verschueren, E.; Vanhee, P.; Rousseau, F.; Schymkowitz, J.; Serrano, L. Protein–peptide complex prediction through fragment interaction patterns. *Structure* **2013**, *21*, 789-797.
8. Verschueren, E.; Vanhee, P.; van der Sloot, A. M.; Serrano, L.; Rousseau, F.; Schymkowitz, J. Protein design with fragment databases. *Curr. Opin. Struct. Biol.* **2011**, *21*, 452-459.

9. Fernandez-Fuentes, N.; Fiser, A. A modular perspective of protein structures: application to fragment based loop modeling. In Protein Supersecondary Structures, vol. 932. Edited by Kister AE . Humana Press; **2012**, 141-158.
10. Menon, V.; Vallat, B.; Dybas, J. M.; Fiser, A. Modeling proteins using a super-secondary structure library and NMR chemical shift information. *Structure* **2013**, *21*, 891-899.
11. Vallat B.; Madrid-Aliste, C.; Fiser, A. Modularity of protein folds as a tool for template-free modeling of structures. *PLoS Comput. Biol*. **2015**, *11*, 1-16.
12. Xu, D.; Zhang, Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins: Struct. Funct. Bioinform.* **2012**, *81*, 1715-1735.
13. Xu, D.; Zhang, Y. Toward optimal fragment generations for ab initio protein structure assembly. *Proteins: Struct. Funct. Bioinform.* **2013**, *81*, 229-239.
14. Zhang, Y. Interplay of I-TASSER and QUARK for template-based and ab initio protein structure prediction in CASP10: composite protein structure prediction in CASP10. *Proteins: Struct. Funct. Bioinform.* **2014**, *82*, 175-187.
15. Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. *J. Mol. Biol.* **1997**, *268*, 209-225.
16. Gront, D.; Kulp, D.W.; Vernon, R.M.; Strauss, C.E.M.; Baker, D. Generalized fragment picking in Rosetta: design, protocols and applications. *PLoS One* **2011**, *6*, 1-10.
17. Brunette, T.; 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*, 580-584.

18. Koga, N.; Tatsumi-Koga, R.; Liu, G.; Xiao, R.; Acton, T.B.; Montelione, G.T.; Baker, D. Principles for designing ideal protein structures. *Nature* **2012**, *491*, 222-227.

19. Lin, Y-R.; Koga, N.; Tatsumi-Koga, R.; Liu, G.; Clouser, A.F.; Montelione, G.T.; Baker, D. Control over overall shape and size in de novo designed proteins. *Proc. Natl. Acad. Sci.* **2015**, *112*, 5478-5485.

20. Zhang, Y.; Skolnick, J. Automated structure prediction of weakly homologous proteins on a genomic scale. *Proc. Natl. Acad. Sci.* **2004**, *101*, 7594-7599.

21. Wu, S.; Zhang, Y. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res*. **2007**, *35*, 3375-3382.

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

23. Shen, H.; Fallas, J. A.; Lynch, E. Sheffler, W.; Parry, B.; Jannetty, N.; Decarrea, J.; Wagenbach, M.; Vicente, J. J.; Chen, J.; Wang, L.; Dowling, Q.; Oberdorfer, G.; Stewart, L.; Wordeman, L.; de Yoreo, J.; Jacobs-Wagner, C.; Kollman, J.; Baker, D. De novo design of self-assembling helical protein filaments. *Science*, **2018,** *362*, 705-709.

24. 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*, 580-584.

25. Parmeggiani, F.; Huang, PS.; 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*, 563-575.

26. Doyle, L.; Hallinan, J.; Bolduc, J.; Parmeggiani, F.; Baker, D.; Stoddard, B. L.; Bradley, P. Rational design of alpha-helical tandem repeat proteins with closed architectures. *Nature.,* **2015,** *528*, 585-588.

Chapter V. Conclusions

Throughout my thesis projects, a central theme has emerged: addressing the protein folding problem through fundamental exploration. To this end, we attempted to predict the structure of Tandem Repeat Protein-based nanotubes, the existence of which had been theorized, but never experimentally realized. Using a crystal structure comprising six nearly identical LRV repeats as a structural benchmark, we extrapolated a short protein concatemer into a full assembly model. With a consensus sequence-based design strategy we designed the sequence of a single LRV repeat (LRV_M3Δ1) such that it would self-assemble at physiologically relevant pH into high aspect ratio peptide nanotubes. Medium resolution techniques such as SAXS and STEM revealed apparent disparities between the assembly model and the experimental structure. Cryo-EM followed by iterative real space helical reconstruction was employed to solve the structure at 4.4 Å resolution. This resolution was sufficient to disprove some key structural features of the assembly model. First, the secondary structure of the assembly subunit developed into an α/α over time instead of a 3_{10}/ α subunit. Second, the actual helical pitch was 20 Å, with 17.4 subunits per helical turn; compare this to the predicted helical pitch of 32 Å and 24 subunits per helical turn. Finally, we had predicted from the 1LRV crystal structure that an LRV nanotube would be right-handed, as all reported helical solenoids have been; instead a left-handed solenoid was observed. However, due to the relatively small helical pitch, and the presence of a large Fe_4:S_4 domain at the N-terminus of the 1LRV crystal structure, we propose that this disparity in handedness is justified. Simply put, we believe that the Fe_4:S_4 domain precludes the adjacent LRV subunit from adopting the left-handed position, and each LRV subunit forces the subsequent LRV subunit to switch handedness, akin to one domino felling its neighbor until all the dominoes have collapsed.

In an effort to more reliably produce TRP-based nanotubes, we employed a similar strategy to a more well-studied system: αRep. Each αRep HEAT-like repeat was nearly superimposable on the LRV repeats shown in the 1LRV crystal structure. The robust consensus sequence for this family of HEAT proteins was developed by aligning and averaging the sequences of 100 proteins from several different organisms. This analysis revealed regions of high sequence conservation, and regions of extreme sequence variation. Interestingly, these hypervariable regions all point towards the concave surface of the solenoidal scaffold. In fact, this unique property bestows the αRep proteins their ability to bind a wealth of diverse protein partners. We again used the consensus sequence to design a single repeat that would be capable of self-assembly into superhelical nanotubes (HEAT_R1). Medium resolution characterization techniques pointed to the assembly model being very accurate, with SAXS and STEM revealing a 30.8 Å helical pitch (within error of the predicted 32 Å), and 21.3 subunits per superhelical turn (quite close to the predicted 24). Again, we employed cryo-EM followed by iterative helical real space reconstruction to solve the nanotube structure to 6.0 Å resolution. The reconstruction largely agreed with the structural model; a right-handed superhelical structure with a 31 Å helical pitch, and 20.7 subunits per superhelical turn. The one surprising result from the cryo-EM analysis was the discovery that the asymmetric unit consisted of two, not one, helical hairpins. This dimerization event is caused by a strong, π-π stacking interaction between two Tryptophan residues in adjacent helical hairpins; this interaction is sufficiently strong to break the local symmetry of the system, and form a dimer. A similar Trp-Phe π-π stacking interaction is observed in the crystal structure of one of the αRep proteins, but does not cause a symmetry break. The importance of this interaction was explored by synthesizing a Trp to Phe mutation in the original peptide sequence, as well as a Trp to Ala mutation. The Trp to Phe mutation led to a peptide nanotube with similar assembly morphology

to the parent peptide. To contrast, a large degree of lateral association was observed in the Trp to Ala mutation. In both cases, helical nanotubes were formed, but the Trp to Ala mutation formed less ordered tubes.

Inspired by our ability to reliably produce nanotubes using the HEAT consensus sequence, we designed a concatemer of the HEAT_R1 peptide, containing six repeats (HEAT_6R). The idea behind concatemerization was to provide a larger subunit onto which we could affix functional molecules (small molecules could be used to functionalize HEAT_R1, but proteins could not). Using low- and medium-resolution characterization techniques, we showed that the concatemer assembled in much the same fashion as HEAT_R1. Thus, we pursued a SpyTag:SpyCatcher genetic fusion technique to affix fluorescent mCherry onto the convex surface of HEAT nanotubes (we employed an even larger HEAT concatemer, with 8 helical hairpins). The mCherry_HEAT nanotubes were then analyzed using medium-resolution techniques, with initial results pointing towards success. Ultimately, CLEM will be used to show colocalization of fluorescent signal with the occurrence of nanotubes. Once we have definitive proof of our concept, we will design a new system employing the SpyTag:SpyCatcher system, in which the SpyCatcher is affixed to a functional protein, like an enzyme. Of particular interest is whether the immobilized enzyme will have a slower or faster turnover rate than the free enzyme. It may even be possible in the future to combine this functionalization scheme with a chemical modification of the concave surface of the nanotube to achieve a dual-function TRP-based material.

Finally, we employed three generations of computational protein design to develop superhelical filaments from helical hairpin motifs (though our design strategies are sufficiently robust to incorporate other assembly modules). Initially, we combed the Protein Data Bank to find any structure containing a helical hairpin similar in size to the HEAT subunit, and evaluated each

motif's ability to form a helical turn. Elongated assemblies were then modeled computationally, from which a template structure was cut; a template would contain a single helical hairpin, along with every other helical hairpin it could possibly touch in the assembly. In this way we captured every lateral and axial interface, and redesigned the sequence using PyRosetta. Unfortunately, this methodology produced many insoluble sequences, because PyRosetta has a tendency to overpack interfaces, leaving very little room for solvation by water. The soluble sequences were assembled at physiological pH and analyzed using TEM and CD. Most of these sequences formed filaments, but the agreement between model and experimental diameter was erratic; the CD signatures of the filamentous assemblies similarly did not match the expected α-helical spectrum.

A second generation of designs was employed, using TERMs (tertiary motifs, which describe a single amino acid, and any tertiary contacts it may make). TERMs have the advantage of sampling real protein sequences, which are inherently soluble (only well characterized structures are deposited in the PDB). Initial characterization of the TERM-based designs was encouraging, though still problematic. Two peptides were synthesized from this generation; one which displayed a helical CD signature, but disordered filaments, and one with a random coil CD signature, with beautiful filaments exactly matching their predicted diameter.

Finally, we realized the most facile way to generate novel helical structures using computation would be to redesign a known assembly, rather than create one from scratch. This led to the redesign of the 3LTJ crystal structure, such that the turn in each helical hairpin was used to connect two axial layers of the filament (GGHEAT). This design strategy produced several soluble sequences, with nearly helical CD signatures, and filaments which possessed a distinguishable inner lumen. However, a large degree of helical unwinding was observed in the GGHEAT filaments, leading us to believe that we sacrificed lateral contact strength for axial contact strength.

A successful design will have to possess a balance between strong lateral and strong axial contacts, lest one dominate the other.

Ultimately, the two protein design approaches presented here should be combined to most effectively generate predictable helical filaments. As computation becomes more rapid and more user-friendly, it will likely supplant rational design. At the very least, it will revolutionize the way that peptide- and protein-based nanomaterials are designed and evaluated, and lead to an explosion in the number of TRP-based nanomaterials.

Appendix: Re-use Acknowledgements

Figure 1.1 Negative-stained TEM of fully assembled TMV particles:

Reprinted from "Archive of Biochemical Biophysics, 581, Harris, J. R., Transmission electron microscopy in molecular structural biology: A historical survey, 3-18." Copyright (2015) with permission from Elsevier.

Figure 1.2 Schematic of TMV1cys incorporation into a MEMs device:

Reprinted with permission from "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). Copyright (2008) IOP Publishing.

Figure 1.3 Structure of the bacterial sex pilus:

Reprinted via the Creative Commons Attribution License from "Structure of the Bacterial Sex F Pilus Reveals an Assembly of a Stoichiometric Protein-Phospholipid Complex. Costa, T. R. D.; Ilangovan, A.; Ukleja, M.; Redzej, A.; Santini, J. M.; Smith, T. K.; Egelman, E. H.; Waksman, G. *Cell*, **2016**¸*166 (6),* 1436-1444."

Figure 1.4 Design and assembly of 7HSAP1:

Reprinted with permission from "Rational Design of Helical Nanotubes from Self-Assembly of Coiled-Coil Lock Washers. 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. *J Am Chem Soc* **2013,** *135* (41), 15565-15578." Copyright (2013) American Chemical Society.

Figure 1.5 Sequences and assembly of Form I and Form II:

Reprinted with permission from "Structure, 23 (2), 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, 280-289." Copyright (2015), with permission from Elsevier.

Figure 1.6 Assembly of Aβ (16-22) into tubes under acidic conditions and into fibers under neutral pH conditions:

Reprinted with permission from "Cross-strand pairing and amyloid assembly. Liang, Y.; Pingali, S. V.; Jogalekar, A. S.; Snyder, J. P.; Thiyagarajan, P.; Lynn, D. G. *Biochemistry-Us* **2008,** *47* (38), 10018-10026." Copyright (2008) American Chemical Society.

Figure 1.7 The Kekulé structure of the Aβ (16-22) macrocycle:

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

Figure 1.8 Atomic model of the macrocycle nanotube:

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

Figure 1.9 Comparison of α-helical peptides and oligoureas, with biophysical characterization of the latter:

Reprinted by permission from Nature Chemistry, "Shaping quaternary assemblies of water-soluble non-peptide helical foldamers by sequence manipulation, 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." Copyright (2015) Nature Chemistry.

Figure 1.10 Schematic of the peptoid nanotube assembly from primary to quaternary structure:

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

Figure 3.9 Spontaneous intermolecular amide bond formation by SpyTag:

Reprinted by permission from PNAS. “Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. Zakeri, B.; Fierer, J. O.; Celik, E.; Chittock, E. C.; Schwarz-Linek, U.; Moy, V. T.; Howarth, M., P Natl Acad Sci USA 2012, 109, 690-697.”

Figure 4.2 Negative-stained TEM of one of the designed icosahedral particles with computationally averaged particles and back-projections calculated from the models:

From “Accurate design of megadalton-scale two-component icosahedral protein complexes. Bale, J. B.; Gonen, S.; Liu, Y.; Sheffler, W.; Ellis, D.; Thomas, C.; Cascio, D.; Yeates, T. O.; Gonen, T.; King, N. P.; Baker, D. Science, 2016, 353, 389-394.” Reprinted with permission from AAAS.

Figure 4.3 Scheme comparing multiple methods of parsing the protein structural universe:

Reprinted from “Current Opinions in Structural Biology, 44, Mackenzie, C. O.; Grigoryan, G. , Protein structural motifs in prediction and design, 161-167.” Copyright (2017), with permission from Elsevier.

Figure 4.16 Computational model of DHF58 with increasing numbers of repeats paired with negative-stained TEM of the assemblies, and 2D-Class Averages of the filaments:

From “De novo design of self-assembling helical protein filaments. Shen, H.; Fallas, J. A.; Lynch, E. Sheffler, W.; Parry, B.; Jannetty, N.; Decarrea, J.; Wagenbach, M.; Vicente, J. J.; Chen, J.;

Wang, L.; Dowling, Q.; Oberdorfer, G.; Stewart, L.; Wordeman, L.; de Yoreo, J.; Jacobs-Wagner, C.; Kollman, J.; Baker, D.. Science, 2018, 362, 705-709." Reprinted with permission from AAAS.