Design Strategies for Dynamic Self-assembled Protein Materials

Nathan Andrew Carter

Dissertation submitted to the faculty of Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Chemistry

Tijana Z. Grove, Chair
Alan R. Esker
Aaron S. Goldstein
Timothy E. Long

August 28, 2017
11:00 am
Kelly Hall, CafeX
Virginia Tech, Blacksburg, VA

Design Strategies for Dynamic Self-assembled Protein Materials

Nathan A. Carter

Scientific Abstract

Hierarchical structures exist throughout nature. From a materials perspective, researchers have shown great interest in mimicking biological structures and their complex structure-property relationships that have evolved from a small number of chemically similar building blocks. While extremely prevalent in research literature, bio-inspired materials and structures are even beginning to make their way into commercial products. Some of the most prevalent are adhesives that mimic the structures on gecko feet and edible food packaging that mimics the skins of fruits. While some of these studies mimic structural features with top-down manufacturing processes, they are limited in scale and struggle to match the structural complexity at the nanometer scale. To address this problem, groups use bottom-up design strategies where the resulting structural features of materials are controlled by the building blocks. As such, proteins provide a versatile building block with inherent structural and functional complexity that is unmatched by synthetic materials.

This dissertation explores more generalized strategies to influence the self-assembly of protein materials whose hierarchical structure may prove useful in a variety of applications in soft-robotics and energy storage. This work focuses on repeat proteins, specifically the 18-repeat, Consensus Tetratricopeptide Repeat protein (CTPR18). More generally, repeat proteins are a family of proteins that are formed by short-repeating sequences and structural motifs. Repeat proteins are ever present throughout the body, often controlling protein-proteins interactions and cell signal pathways. Beyond their biological functions, they are interesting from a materials perspective, because of their internal symmetry and short repeated sequences that assemble like bead on a string. Repeats occur in tandem and each motif is structurally independent of the next, where structure is dominated by interactions with neighboring residues of the tandem repeat. This is in stark contrast to globular proteins whose structure is stabilized by through-space interactions resulting frommore complex folding. These long-range interactions make rational design and genetic engineering strategies significantly more difficult, because mutations in a local area of the protein

may detrimentally affect the fold of the entire protein. In contrast, a repeat protein comprising 10 repeats of the same motif will have the same overall structure if each of the repeats are identical, if they are all different or any permutation in between.

In this work we showed CTPR18 assembles into 1D and 2D crstyalline arrays that diffract electrons. We further showed the ability to select for different epitaxial growth by changing monovalent salts, selected from the Hoffmeister series. We are able to create 2D nanosheets that maintain the aspect ratio of CTPR18, indicating equal rates of head-to-tail (HT) and side-to-side (SS) epitaxial growth. By selecting for HT epitaxial growth, CTPR18 assembles into nanowires. Whereas, SS epitaxial growth gives rise to 2D 'footballs' and low aspect-ratio tubes. These materials and more importantly assembly strategies may prove useful as molecular recognition scaffolds and further assembly of other repeat protein arrays.

We further explored the hierarchical structure and properties of the macro-scale self-assembled films of CTPR18. These films form highly oriented lamellar structures that seemingly continuous phases using polarized light microscopy. This lamellar alignment results in anisotropic mechanical properties ranging from the nano- to macro-scale. We leveraged this in the design of functional materials, by changing film casting conditions, to impart a structural gradient in the film. This proves useful in tuning the water-induced bending motion of these hygroscopic films. Herein, we showed the ability to change the bending mode (*e.g.* long axis bend, twist & short axis bend) as well as the kinematics simply by changing the underlying film morphology. Lastly, we showed that these actuators function as electroresponsive materials as well. This functionality arises from the charged character of CPTR18 films which results in one of the highest conductivities reported for biological materials to date. As actuators, electroresponsive motion is induced by ion transport through material causing asymmetric swelling in the film. These dual responsive materials may prove useful in soft robotics. Additionally we are beginning investigations into the usefulness of CTPR18 films as alternate materials for ion-transport materials like those used in lithium polymer (more commonly LiPo) and sodium-ion batteries.

Design Strategies for Dynamic Self-assembled Protein Materials

Nathan A. Carter

General Audience Abstract

Structures in nature exhibit unique and complex architectures whose order propagates from nano- (10^{-9} m) to macro-scales (mm to m). These structures give rise to a rich diversity of adaptive function that allows for life in all environments on Earth. This complex functionality has driven research into bio-inspired materials where scientists investigate the complex relationship between sequence, structure and function of these materials. A good illustrative example of the effect that hierarchical structure can have is a brick wall. Bricks are laid so that the layer on top is shifted in either direction by half of a brick. This alternating pattern is what gives the wall its strength. If a crack occurs in the mortar, it will only propagate until it hits a boundary (a neigh-boring brick). Designing nanostructures can have similar effects on materials we use every day. Some of the most prevalent are adhesives that mimic the structures on gecko feet, which allow them to stick to any surface.

This work presents bottom-up design strategies for self-assembling protein materials whose hierarchical structure may prove useful in a variety of applications in soft-robotics and energy storage. Proteins are a useful class of molecules, because they contain a level of structural complexity beyond that of synthetic materials. They are an inherently 'green' material feedstock; made in a lab using microbes like E. coli. Additionally, with the ease and availability of genetic engineering techniques we can easily modify the structure. This is especially true for the class of proteins, repeat proteins, which are the focus of this manuscript. Repeat proteins comprise small repeated sequences which are structurally independent from each other and can be strung togeth-er to create open, extended architectures. Here we explore the self-assembly emergent properties of the consensus tetratricopeptide repeat (CTPR18) . We show that this protein assembles into highly ordered 1D and 2D arrays that are shape tunable based the molecular environment (sol-vents, charge, etc). These nanomaterials may prove useful as molecular recognition scaffolds. We further explore the hierarchical self-assembled films of CTPR18. These films form highly

oriented lamellar structures that seemingly propagate the entire length of the films. These lamellae directly affect the materials mechanical properties. Accordingly, by changing the film casting conditions, we can impart a structural gradient in the film, which proves useful in tuning the water-induced bending motion of these films. Herein, we show the ability to change the speed and directionality of actuation by simply changing the underlying film morphology. Lastly, we show that these films are electroresponsive as well, owing this function to ion transport through the inherently charged character of CTPR18. These dual responsive materials may prove useful in soft robotics. Additionally we are beginning investigations into the usefulness of CTPR18 films as alternate materials for ion-transport materials like those used in lithium polymer (more com-monly LiPo) and sodium-ion batteries.

To my family, for the unwavering love, support, encouragement and occasional reality check through the years.

Signing up as a professor's first graduate student can be a bit of a black box. However, my experience in the Grove lab has been a very positive one. Aside from the occasional hiccup inherent to setting up a new lab, Tijana has always adopted a fluid approach to advising, adjusting to individuals within our group, rather than forcing all into a single mold. Her flexibility (sometimes, in other words, patience) truly allowed me to take my research in a direction that interested me, making it my own. From the bottom of my heart, thank you.

I would also like to thank Xi Geng for the valuable and sometimes lively discussions over the past few years. Whether regarding research ideas, one of umpteen collaborations or just life I immensely enjoyed our chats. I would also like to thank my group for their support during my time here. From the birthday tradition that started by hunting for impromptu gifts in old Davidson to lending critical eyes and ears for paper drafts and presentations, I truly appreciate you all and wish you all the best in your future endeavors!

Table of Contents

Chapter 1 - Introduction

1.1. Thesis overview

Proteins are a class of biological polymers that possess a level of structural complexity that has yet proven impossible to match using synthetic materials. This ubiquitous natural complexity gives rise to functionality that academic and industrial researchers continually seek to mimic. Proteins by no means relegate synthetic materials or other biomaterials as inferior or those of the past. Rather, the discussions below are meant to address the benefits as well as drawbacks of using protein materials available with current technologies. This thesis will largely be approached from a materials perspective with the goal that its multi-disciplinary nature may help bridge the gap between the perspectives of biology and materials science. Specifically encouraging biochemists to view proteins from a materials perspective and materials scientists to think from a biological perspective.

Recent advances in protein materials such as self-assembly strategies and applications will be covered in this beginning review. Additionally, design strategies for repeat proteins, a subclass relevant to the work comprising this thesis, will be introduced. Functionalization strategies for protein-polymer materials will be reviewed in Chapter 2, again focusing more on recombinant proteins. Chapter 3 explores the hierarchical morphology in films of the consensus tetratricopeptide repeat protein (CTPR18). Furthermore, it explores the degree to which anisotropic morphology affects the resulting mechanical properties from the nano- to macro-scale. Chapter 4 builds upon the previous work, illustrating how control over casting chamber humidity can create a morphological gradient in CTPR18 films. The morphological gradient and orientation were used to tune the water-responsive actuation of CTPR18 films, controlling both kinematics and actuation mode. We also show these materials are electro-responsive. They actuate with applied electrical potential and charge/discharge in response to applied force. These materials are the first example of all protein actuators in the literature. Chapter 5 takes a step back to explore the self-assembly process of the rigid-rod CPTR18. Here we illustrate how to introduce shape control into 1D and 2D self-assemblies of CPTR18 by introducing different monovalent salts. Overall, this work lays the founda-

tion for self-assembly strategies useful for a variety of anisotropic materials, while also illustrating the utility that proteins may have from a materials perspective.

1.2. **Why protein materials?**

Natural systems of all shapes and sizes have evolved to thrive in a variety of environments on the planet. Within each individual system there is a complex network of biomaterials that are not only refined and designed for longevity, but are also adaptable and able to self-repair, in addition to many other extremely complex functions. Even seemingly straightforward systems, such as structural elements like bones or fish scales, derive their properties from complex hierarchical nanostructures that not only provide support but also protect the animal from damage. These materials are assembled through complex mineralization pathways that are still widely studied and adapted today in material design.[1] Beyond structural elements, nature has provided a great number of inspiring nanomaterials that scientists mimic today. Some prominent examples are the photonic structures seen in butterfly scales,[2] beetles,[3] sea mice[4] and chameleons.[5] Also, adhesive structures like setae on gecko feet[6] along with a host of other materials that are self-cleaning, super-hydrophobic, and antireflective.[7] This collection of remarkable functionality, comes from features assembled from a relatively simple set of building blocks.

Starting with individual atoms, nature has derived more complex materials like nucleic acids, proteins and sugars. If DNA is the universal coding language, proteins are a class of molecules that serve as the transistor, motherboard and chassis. It is proteins that are responsible for carrying out, as well as regulating the majority of cellular functions. This functional complexity arises from the unique relationship between sequence, structure and function that has evolved over time, to not only sustain life, but also to create organisms that thrive in widely different environmental conditions. Yet simply stating function derives from the structure which results from the sequence, is overly simplistic. For example, a 100 amino acid protein has 1.3×10^{130} possible sequence combinations compared with 1.6×10^{60} for the same length DNA. Within this sequence space, there are limited regions which produce a functional protein.

The pioneering work of Frederick Sanger, followed by the work that led to the 1962 Nobel Prize awarded to John Kendrew and Max Perutz, laid the foundations for protein. Sanger was the first to sequence a full peptide (bovine insulin A & B) in 1951-1952, while Kendrew and Perutz published the first low-resolution crystal structures of myoglobin (1958) and hemoglobin (1960), respectively. In the 59 years since there have been approximately 122,000 crystal structures loaded in the Protein Data Bank, not including those crystalized with nucleic acids. These structures have served as the basis for much of the protein engineering methods that are currently used. However, even with this vast structural database and the powerful computational tools at our disposal, there is still a lot unknown about protein folding. Moreover, the structure-property relationships which drive hierarchical self-assembly from nano to macro that further engender emergent properties, such as ion-transport and photonic structures, are not fully understood. To better understand and ultimately exploit these properties, scientists have worked for decades to mimic their structure and functions. From the perspective of material design, two predominating methodologies serve as the best way to codify these research efforts; top-down and bottom-up. Top-down strategies rely on material processing such as lithography, 3D printing, etc. to impart complex structures on materials. Bottom-up strategies, otherwise known as self-assembly, focus on designing building blocks that leverage any number of molecular interactions to influence the aptly named 'self-assembled' structure.

The following introduction is meant to serve as a primer to the research comprising this dissertation entitled, 'Design Strategies for Dynamic Self-assembled Protein Materials'. As one might glean from the previous discussion, this work is primarily interested in engineered protein materials. To this end, the following pages will briefly address some of the key areas of knowledge and gaps therein, as well as guide interested readers to more exhaustive reviews in each of the sub-fields. Intentionally reversed from the classic trajectory of sequence/structure/function, this introduction will guide the reader through: 1) a discussion of protein materials and their applications, including structure/property relationships (Function); 2) a discussion on protein self-assembly, classified by the driving forces that predominate, accompanied by a discussion of the dimensionality within

these systems (Structure) and finally; 3) a brief introduction on the relevant design strategies used for engineering proteins (Sequence). This order is meant to reflect the emphasis on supramolecular assembly and emergent properties that will be seen in the following chapters.

1.3. Applications of self-assembled protein materials

1.3.1 Hydrogels

Hydrogels have become stalwart materials for many biomedical applications such as targeted drug delivery, stem cell therapy, and tissue engineering. Their broad applications in the body are in large part due to their tunable mechanical and chemical properties that can be modified for a desired application. Additionally the ability to add stimuli responsive functionalities for so-called 'smart' hydrogels has proven useful. For such materials, protein-based systems afford significant advantages in lieu of synthetic polymer systems. Not the least of which is the large repository of specific sequences that provide sites for enzyme cleavage, material crosslinking, cellular adhesion or cellular growth. Additionally, with the advent of low-cost nucleic acid production and gene editing technologies such as recursive ligation,[8] scientists now have access to easy-to-design, low-cost, custom genes for extendable polypeptide sequences that are useful for biomaterials. In addition to the above mentioned functionalities, a set of polypeptides have been developed with lower critical solution temperature (LCST) transition behavior[9] that simplify recombinant purification to a simple heating step, which may further[10] validate these materials for larger scale production.[11]

One of the first advantages for protein materials are specific ligand and protein-protein interactions. Since its structure was first published in 1989 by Hendrickson *et al.,* the streptavidin/biotin binding pair has been widely used to conjugate biomolecules in a variety of applications.[12] Since it has been reviewed extensively elsewhere, it will not be addressed here. In the past 5 years a much newer peptide tag has come into vogue, the SpyCatcher/SpyTag system.[13] This system was designed by Howarth and coworkers, and was inspired by naturally occurring isopeptide bonds in some bacterial adhesion proteins. They broke apart the IgG-like collagen adhesin (CnaB2) from the fibronectin-binding protein (FbaB) of Streptococcus pyogenes, which when paired creates an isopeptide between the Asp-117 and Lys-31 respectively. This system has provided a recombinant

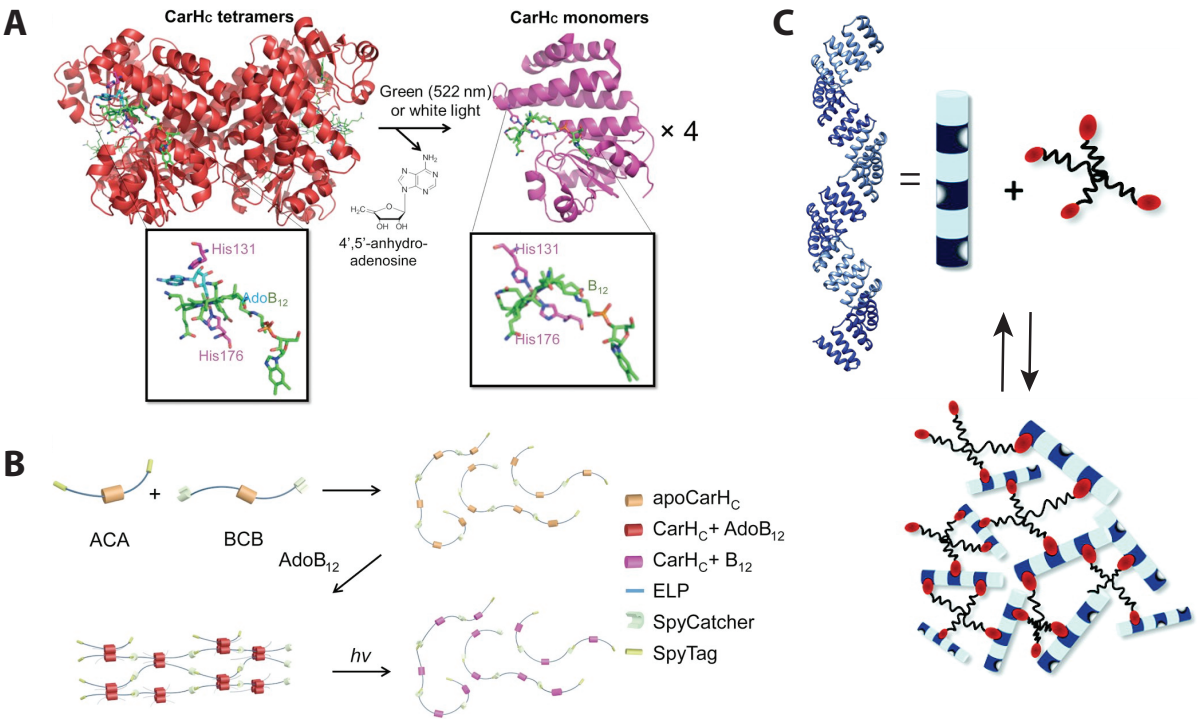


Figure 1.1. A) Light disassembles tetrameric CarHC accompanied by the degradation of AdoB12, the release of 4',5'-anhydroadenosine, and the coordination of His132 to the metal center. B) Telechelic proteins, ACA and BCB, are polymerized through SpyTag-SpyCatcher chemistry. The resulting polymers can further be assembled into a molecular network through AdoB12-induced CarHC tetramerization in the dark and disassembled on light exposure. ApoCarHC is the CarHC protein without AdoB12. C) Cartoon representation of TPR gel formation, using CTPR18 as the backbone. A PEG–peptide cross-linker is represented by black lines for the PEG component and red ovals for the peptide. The cartoon is drawn approximately to scale; each arm of the 10 kDa four-arm star PEG has a contour length of ~18 nm. A & B reproduced from Ref. [15] and C reprinted with permission from Ref. [18a]. Copyright 2010 American Chemical Society.

method to incorporate a number of topologies onto protein and polypeptide materials alike.[10, 14]

Early work by Tirrell, Arnold and coworkers showed the ability to create multiple topologies and tailorable molecular networks ('networks of spies'). Their work showed successful network formation in networks with covalent cross-linking, yet there remained a significant sol fraction of 25%. Non-covalent SpyTag-SpyCatcher networks were not stable. Wang *et al.* recently expanded on this work by incorporating the C-terminal adenosylcobalamin binding ($CarH_c$)domain of the CarH protein, a light sensitive regulatory protein controlling bacterial carotenoid synthesis.[15] $CarH_c$ forms tetramers in the dark when also in the presence of adenosycobalamin ($AdoB_{12}$) (Fig-

ure 1.1A and B). This binding is reversible using green or white light of sufficient intensity. When kept in the dark, these hydrogels were much more stable, with only 6% sol after 10 days incubation in PBS buffer. They worked sufficiently in both cellular release (fibroblasts and mesynchymal stem cells) and protein release studies (mCherry) driven by the rapid sol-gel transition caused by the light-induced $CarH_c$ disassembly. The broad utility of this system has been shown in the work of other researchers. Gao *et al.* who replaced the elastin-like peptide linkers with globular proteins to show that the Spy-binding event is not sterically hindered by large adjacent molecules.[16] Gilbert *et al.* showed the viability of the Spy-system in excretion-based expression systems (*Bacillus subtilis)* where self-cyclization dramatically increased the stability of thermo-stabile enzymes as well as illustrating the possibility of using SpyTag-SpyCatcher to colocalize enzymes in growth vessel.[17]

While covalent isopeptides have shown great promise, designable non-covalent binders have shown promise for creating tunable hydrogel systems. Grove *et al.* demonstrated self-assembling hydrogels whose driving force was an ionic interaction between a protein and specific peptide (Figure 1.1C).[18] In these works, the authors coupled tetratricopeptide repeat (TPR) domains together, incorporating peptide binding sights as handles for cross-linking with four-arm star poly(ethylene glycol) (PEG) molecules end-functionalized with their binding partner peptide. The ionic nature of the cross-link afforded stimuli responsive behavior to the material. They exhibited a reversible stimulus response where both high salt concentrations (0.5 M NaCl)[18b] and pH[18a] changes were able to break apart the gels, and gelation was recovered after dialysis or reversal of the changed pH. While the mechanical properties of these gels were within the needs for tissue engineering at ~270 Pa, the modular nature of repeat protein arrays lend them to easy modulation. Mechanisms to accomplish this might involve changing the cross-link density[19] and/or modifying the strength of the peptide-binding interaction.[20]

The tunability of mechanical properties of polymeric materials has been well established, whereas in contrast, protein materials have in past shown limiting mechanical properties in regards to toughness, stretchability and self-healing properties. Chapters 3 and 4 address strategies to exploit the hierarchical mechanical properties of protein self-assemblies and the affect that they can

have on materials function. There have been many efforts in recent years to mimic natural mechanisms in biomaterials to impart useful mechanical properties as a result of proteins. A classic natural example is the I-band muscle protein titin.[21] The elasticity of titin results from the series of folded IgG-like domains that are linked together by unstructured segments.[21b] The molecular mechanism of this elasticity, caused by the unfolding of individual domains has been well-studied on single molecule level using AFM.[22] Since, there have been multiple strategies that rely on the unfolding of protein domains as dissipation mechanisms.[23] Recent work by Li, Baker and co-workers illustrated this principle with a *de novo* designed protein, Di-I_5 (PDB: 2KL8), that has a ferrodoxin like structure.[23b] These designer proteins were mechanically labile, unfolding with only 5 pN of applied force, Figure 1.2A. This proved advantageous when swelled with water. Ge-

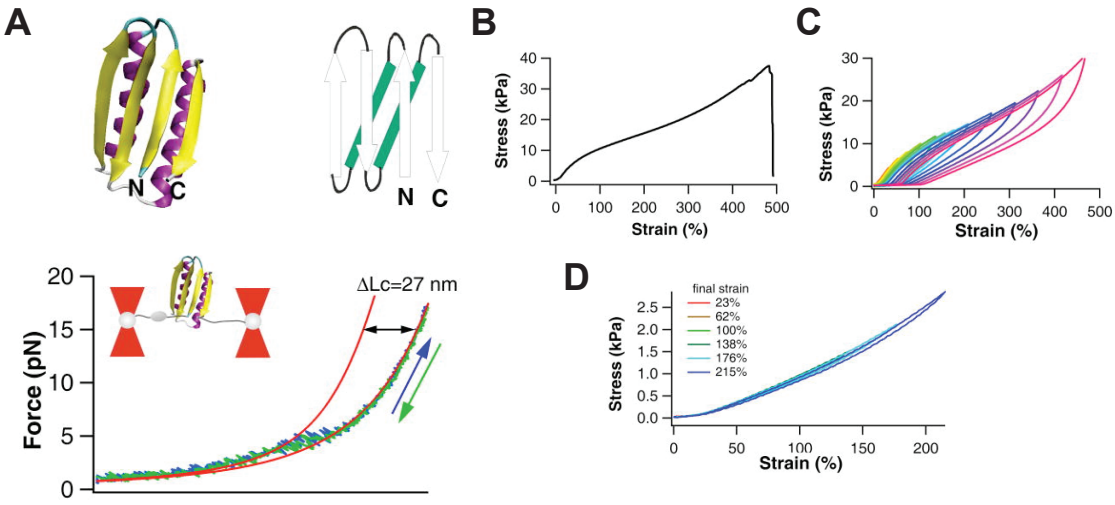


Figure 1.2. Unfolding of *de novo* designed proteins impart useful mechanical properties in hydrogels. A) The FL domain assumes the classic ferredoxin-like fold with the signature βαββαβ secondary structure (cartoon and schematic representation) arrangement along its backbone. Representative force–extension curves of a single-FL domain where the distinct unfolding force peak, occurring at ~5 pN with a contour length increment ΔLc of ~27 nm, corresponds to the unfolding of a FL domain. B) Stress-strain curve of (FL)8 hydrogels. The average breaking strain is 450±90% (n=16), Young's modulus at 15% strain is 16±3 kPa (n=27), and breaking stress is 35±6 kPa (n=16) C) Stressing–relaxation cycles from the same sample show that hydrogels built from (FL)8 exhibit massive energy dissipation. D) Consecutive stress-strain curves of (FL)8 hydrogels with different strains in 3 M GuHCl. These have a much smaller Young's modulus than native ones and do not exhibit appreciable hysteresis in stretching–relaxation cycles. Adapted by permission from Macmillan Publishers Ltd: Nature Communications, Ref. [23b], copyright 2013.

lation occurred as a result of the swelling induced unfolding of the protein, which then aggregated, giving a network with both physical and chemical cross-links. The resulting gels had remarkable mechanical properties, with an average strain-at-break of 450%, ability to recover from significant hysteresis within a few seconds, and the ability to recover mechanical stability after exposure to a denaturing environment (3.0 M guanidine hydrochloride), Figure 1.2B-D respectively.

In addition to incorporating structured proteins, there is a growing level of work, which incorporates unstructured protein materials to impart useable/tunable mechanical properties. Zhao and coworkers incorporated multifunctional unstructured proteins into gels that contained a covalent cross-link site and a labile metal coordination site linked together by an elastin-like peptide with an LCST.[24] The addition of the phase transition ELP and metal-coordination site gives rise to multiple mechanisms for force dissipation within the gels, while the covalent cross-links maintained the structure. Additionally, the labile nature of the Zn^{2+} coordination makes these materials self-adhesive and opens the door for self-healing properties. Their materials proved to have high strength (Young's moduli > 2.5 MPa), exhibited good fracture toughness (26 - 1300 J m^{-2} depending on [Zn^{2+}]) and, quite high stretchability (yield >500% strain), as well as good self-adhesion properties in Zn^{2+} solutions, capped off by high production yields of 350 mg L^{-1} of culture.

3.2 Bio-electronic devices

Electronic devices are ubiquitous in today's world. Technology is so available that most people carry around devices with more internal computing power than was used to send people to the moon. The ease of fabrication, small size and powerful computing have brought us to the point to where 'the internet of things' is a materializing reality rather than simply possibility. Thus researchers have been interested in developing materials for wearable devices that sense everything from an individual's daily steps, to more in-depth health markers. Additionally there is a large market for biomedical devices ranging from stimulating brains to powering artificial hearts. One of the issues with such materials is interfacing traditional semiconductor devices with biomaterials, where something as simple as material stiffness can wreak havoc at the tissue interface. While there have been a number of works that focus on using biomaterials as supports for traditional and or-

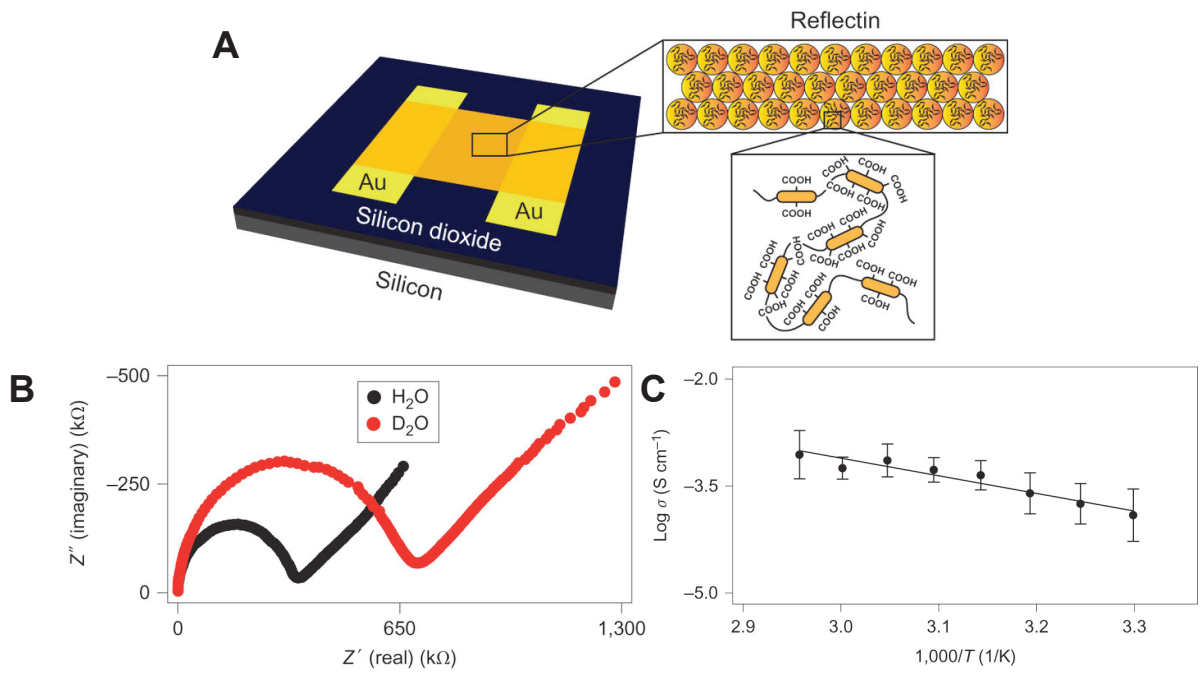


Figure 1.3. A) Illustration of a two-terminal device in which a film composed of aggregated reflectin protein bridges two gold electrodes. Wild-type reflectin contains six repeating subdomains (orange) connected by variable linker regions (black). The carboxylic acid portions of the aspartic and glutamic acid residues in the repeating subdomains are labelled as COOH. The aliphatic portions of the labelled aspartic and glutamic acid residues are omitted for simplicity. B) A typical Nyquist plot for a reflectin-bridged two-terminal device in the presence of water vapour (black) and in the presence of deuterium oxide vapour (red), both at a RH of 90%. Effective film resistance changes on switching from H_2O to D_2O. C) Arrhenius-type plot of the conductivity (σ) as a function of temperature for reflectin devices at a RH of 90%. The average value calculated form a linear fit yields an activation energy consistent with Grotthus-type proton transfer, Ea = 0.22(±0.05) eV. Adapted by permission from Macmillan Publishers Ltd: Nature Chemistry, Ref. [27a], copyright 2014.

ganic electronics[25], we are more interested in materials whose electrical characteristics are dictated by the proteins themselves. Recent work has begun to take advantage of the highly charged nature of proteins and exploiting them for proton transport (topics that will be discussed in Chapter 4 of this thesis). For example, proton-translocation is seen across nature,[26] yet has only been recently explored in bioelectronics.[27]

Pioneering work by Gorodetsky and coworkers showed bulk conductivity of protons in a drop-cast film of reflectin, a structural protein from cephalopods, (Figure 1.3A).[27a] Reflectin's highly charged character, with 30% charged residues (also 20% hydrophobic aromatic residues)

gives rise to this conductivity. This was confirmed when mutating Asp and Glu to Ala effectively knocked out conductivity. Wild-type material shared characteristics with the conductivity of a dilute acid solution and a peak conductivity of 2.6 x 10^{-3} S cm^{-1} at 90% humidity and 65 ^{o}C, the highest protein, protonic conductivity reported to date (Figure 1.3B). The authors posited that the structure is analogous to that of Nafion, in that nanoparticle ionic aggregates form the conductive channels through which protons flow, and later confirmed this supposition in a structural study on reflectin.[28] Further, the authors successfully implemented these materials into protonic transistor devices that operate in a similar manner to classic field-effect transistors. The source-drain current at different source-drain voltages is dictated by the charge carrier density and modulated by the gate voltage. These devices exhibited similar electrical properties to reported chitosan structures[29] with the added genetic engineering benefits that come with protein materials. Further work by Ordinario *et al.* showed a facile approach to doping these structures with the photoacid 8-hydroxypyrene-1,3,6-trisulfonic acid (HPTS) which has a labile proton when exposed to blue light, thus increasing the proton density and thus, the source-drain current.

Amdursky *et al.* utilized the photoacid, HPTS, to study the mechanism of short-range proton transport in electrospun bovine serum albumin (BSA) mats followed by a series of measurements to study the long-range proton conduction.[27c] BSA was selected as a cheap, commercially available alternative to designer or recombinant proteins. BSA has a pI=4.7 and therefore carries a net negative charge at pH=7, which comes from a primary sequence bias towards acidic residues. While these mats proved conductive, with conductivity values ~ 5.0 x 10^{-5} S cm^{-1}, the authors found no evidence of a kinetic isotope effect (KIE) and an activation energy (E_a) 2x higher (~0.29 eV) than expected for diffusive proton-hopping through a water network (Grotthaus mechanism) as seen in Ordinario *et al.* (KIE = 1.7, E_a = 0.2 eV). Beyond speculation, there is not much to explain these differences.

Most recently, Silberbush *et al.* used amyloid-like protein fibrils to study the sequence effects on proton conduction,[27d] a principle they first explored with electronic conductivity in the same family of peptides.[30] They showed the profound effect that charged residues have on proton

conduction, where a single insertion of an acidic or basic residue caused an order of magnitude conductivity increase. More importantly, their results matched those seen in dilute solution where acidic residues (Glu and Asp) are in fact more important than basic ones (Lys), showing a 2.7x increase in conductivity. While the acidic vs basic residue data was not a revelation in and of itself, this work was the first example to show the different effects that small changes in the sequence can have. It opened the door for sequence engineering of other protein and peptide materials, while still not addressing any possible differences between acidic moieties like Asp, Arg and His.

1.4. Self-assembled proteins structures

Each of the aforementioned applications have a slightly different spin on how to utilize protein

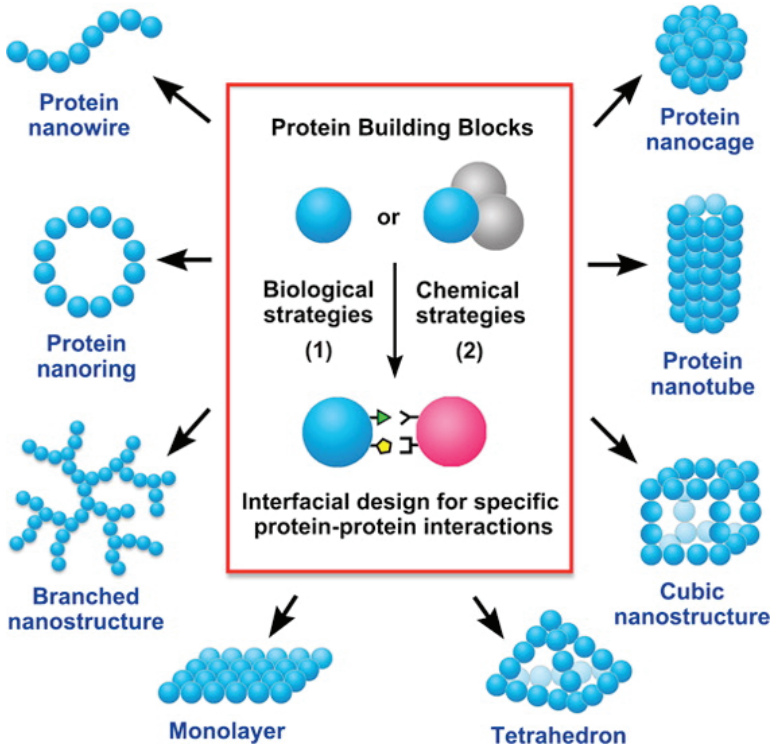


Figure 1.4. There are many strategies to drive protein self-assembly that take advantage of both existing protein-protein interactions as well as chemical strategies. Chemical strategies may include covalent bond formation, cross-linking through metal cross-linking or assembly modulated through electrostatic interactions. Each of these strategies can be implemented to influence protein self-assembly. Shape control comes from directional interactions that propagate preferentially in one or more directions. Reprinted with permission from Ref. [40]. Copyright 2016 American Chemical Society.

components, whether purely as structural elements, as conductive matrices for ion conduction or through imparted dynamic motion as a result of ligand-binding events. To design these functions, scientists have explored a variety of self-assembly strategies to promote self-assembly of protein 'building blocks'. Some of these strategies borrow from the supramolecular toolbox seen in polymer and other biomaterials, with the addition of an expansive repertoire of specific binding interactions to exploit. While the ability to rationally design complex architectures is in its infancy relative to nucleic acid self-assembly, protein materials have been shown to assemble into multidimensional arrays in 0D, 1D, 2D and 3D, (Figure 1.4). From a design standpoint, the most important thing to consider is the self-assembly mechanism, which all boil down the atomic level intermolecular interactions. From here it becomes possible for scientists to manipulate the position, density or amount of interacting moieties on a molecule to change the structure of the resulting assembly. As previously discussed, material assembly may have a great effect on mechanics and function. There are a variety of driving forces that have been shown in the literature. They are grouped in the following manner: 1) electrostatic interactions, 2) metal coordination, and 3) engineered protein-protein interfacial assembly.

1.4.1 Electrostatic assembly

In the realm of protein materials electrostatic interactions are the 'lowest hanging fruit' for tuning assembly. Proteins are inherently highly charged molecules, with surfaces decorated with charged amino acids. Across the entire UniProt Knowledgebase, charged amino acids account for 23.6% of all amino acids in proteins. The positive and negative charges come from the weakly basic and acidic residues, respectively, that are generally distributed in a heterogeneous manner across the surface of the protein. They normally exist in charged pockets that lend themselves as easy targets to exploit for protein assemblies if geometric constraints are met. Further, these interactions are inherently and easily tunable by changing the ionic strength and/or pH of the buffer used. These advantages have a downside as well. Given the current limitations of protein engineering technologies, it may be extremely difficult to change the charge distribution across the surface. Additionally, there is no reliable methodology to quantify the electrostatic interactions through space

in a highly charged environment, meaning assembly conditions come through an iterative process, if at all.

Such assembly strategies have been used heavily in the assembly of patchy protein cages. The seminal work of Kostiainen *et al.* showed that electrostatic interactions guide the assembly of protein cages comprising the cowpea chlorotic mottle viruses (CCMVs, pI = 3.8).[31] They chose the CCMV because it was an isocohedral assembly that formed from 180 identical coat proteins, which exhibited a reversible assembly in relation to pH. They induced hierarchical assemblies by incorporating a cationic dendron functionalized with spermine, a molecule with both primary and secondary amines that would bind to the negative surface of the CCMV (Figure 1.5A). Additional-

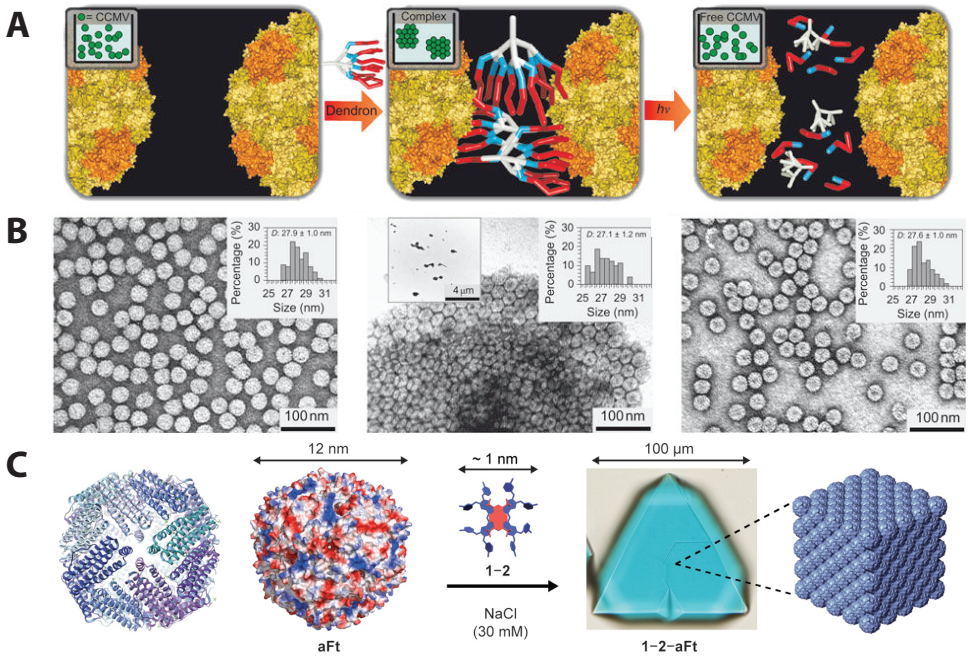


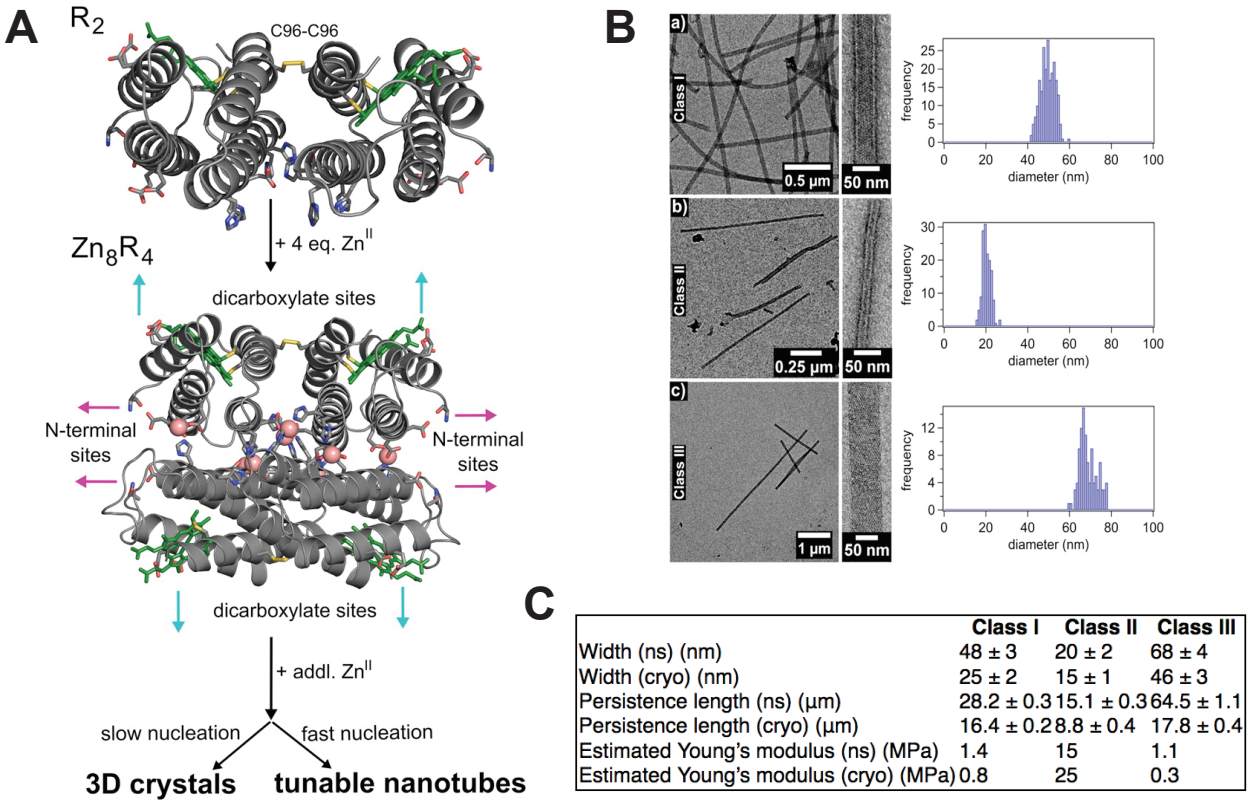
Figure 1.5. A) On addition of a cationic dendron to the negatively charged viral particles (middle), viral particles are 'glued' together by electrostatic interactions, which results in a large hierarchical assembly. Optical irradiation destroys the multivalent binding interaction and leads to the release of the virus particles. B) Representative TEM images matching each of the assembly/disassembly steps in A. C) Illustrations of the 12 nm sized aFt cage and its further cocrystallization with ZnPc-PTSA complex driven by electrostatic interactions. Illustrations of aFt show the folded protein subunits (left) and the calculated electrostatic surface potential (right) with negatively charged patches (in red) covering the cage surface. A & B adapted by permission from Macmillan Publishers Ltd: Nature Chemistry, Ref. [31], copyright 2010. C adapted with permission from Ref. [34]. Open access through ACSAuthorChoice.

ly they incorporated an optical trigger to release the assembly that incorporated a UV-photocleavable o-nitrobenzyl moiety as the spermine/dendron core linker. These particles were tunable by changing ionic content as well as the ratio of dendron:protein. They easily assembled into a hexagonal morphology as seen in Figure 1.5B. They have expanded this work to incorporate metallic nanoparticles[32] and shown the ability to tune the lattice parameters of similar binary systems comprising apoferitin (aFt, pI = 4.5) and poly(amidoamine) dendrons.[33] More recently Mikkila and coworkers created ternary co-crystals of aFt, octanoic Zn-phthalocyanine (ZnPc) and tetraaninonic pyrene (PTSA). PTSA and ZnPc assemble through electrostatic and pi-pi interactions, which then further stick to the anionic patches on the surface of aFt. These resulting face-centered cubic (fcc) materials were fluorescent and able to generate $^1O_2$, all the while still tunable with ionic concentration (Figure 1.5C).[34]

1.4.2 Metal coordination driven assembly

Metal coordination in proteins has a huge role in biology. Metals are vital to the function of redox active enzymes. Their strong coordination strength stabilizes protein structures[35] and even proves integral to macro-scale properties like the adhesion of marine mussels.[36] In fact, roughly one-third of all proteins are metalloproteins.[37] Metals bring with them a series of advantages in that: 1) ligand-metal bonds are stronger than other non-covalent bonds, 2) metal valencies make the bonds highly directional, thus imparting geometrical control on interfaces, 3) metal-ligand bonds are kinetically labile and 4) lastly can be easily controlled using external factors like pH since, with the exception of Cys, the ligands in proteins are weak acids.[38]

Using metal coordination to drive protein self-assembly was largely pioneered by Tezcan and coworkers. While Tezcan's earlier work has been well addressed in reviews: Yang *et al.*,[39] Luo *et al.*,[340] and Slagado *et al.*,[38] this discussion will address some of their more recent studies. Previous work from the group showed the construction of D_2 symmetric assemblies of the monomeric protein cytochrome cb_{562} using Zn^{2+} coordination.[41] They further designed a variant of this, RIDC3, which was further engineered to create Zn-mediated dimerization exhibited assembly in 1D, 2D and 3D arrays. In recent work Brodin *et al.* explored the possibility of creating multiple



C

	Class I	Class II	Class III
Width (ns) (nm)	48 ± 3	20 ± 2	68 ± 4
Width (cryo) (nm)	25 ± 2	15 ± 1	46 ± 3
Persistence length (ns) (µm)	28.2 ± 0.3	15.1 ± 0.3	64.5 ± 1.1
Persistence length (cryo) (µm)	16.4 ± 0.2	8.8 ± 0.4	17.8 ± 0.4
Estimated Young's modulus (ns) (MPa)	1.4	15	1.1
Estimated Young's modulus (cryo) (MPa)	0.8	25	0.3

Figure 1.6. A) Proposed Zn-mediated of two disulfide dimers (R2) into a closed, D_2 symmetric tetramer (Zn_8R_4). This tetrameric subunit assembles into larger supramolecular architectures with further Zn coordination. Heme cofactors are shown as green sticks. The two competing assembly motifs are N-terminal sites and dicarboxylate sites. B) Negative-stained TEM images of the three classes of Zn_8R_4 nanotubes C) This table illustrates the different assemblies that were possible from a single building block molecule. It shows physical parameters and derived mechanical properties of the Zn8R4 nano-tubes (both cryoTEM and negatively-stained). Adapted with permission from Ref. [42]. Copyright 2015 American Chemical Society.

morphologies from the same subunit.[42] Using the RIDC3 as a starting point because of its symmetry, they engineered a Cys at position 96 to form a disulfide dimer, and added a His at position 59 to ensure tetrameric assembly through 8 Zn^{2+} ions (Zn_8R_4). The proposed coordination model is shown in Figure 1.6A,. Zn8R4 was able to assemble into 2D arrays that were not flat due to 6_1 screw axis running along the 2D *bc* plane. The screw axis provided the necessary twist to yield a helical tube, which can be thought of as a 2D ribbon, wrapped around a column. 1D growth is favored by fast nucleation, accomplished by increasing the available Zn^{2+}. Figure 1.6B shows the varied morphologies that were available from a singular building block.

Other examples of coordination driven assemblies, at much lower protein concentrations, have been explored by Liu and coworkers. Here they use a C2-symmetric, homodimeric enzyme, glutathione transferase (sjGST) as the building block for assembly.[43] To direct GST self-assembly they used a recombinant GST with an N-terminal His6 -tag (sjGST-6His), a standard purification tag, which when dimerized, gives a dual-handled molecule that produced nanowires in the presence of Ni^{2+} (Figure 1.7A and B). Expanding on this work, Bai *et al.* sought to tune the assembly of the sjGST-6His by removing the His_6-tag, and incorporating a bishistidine (bis-His) clamp on the surface of the sjGST (sjGST-2His) specifically moving it from the C2-axis.[44] In their design, they ensured these bis-His clamps remained on the solvent exposed surface and further oriented in such

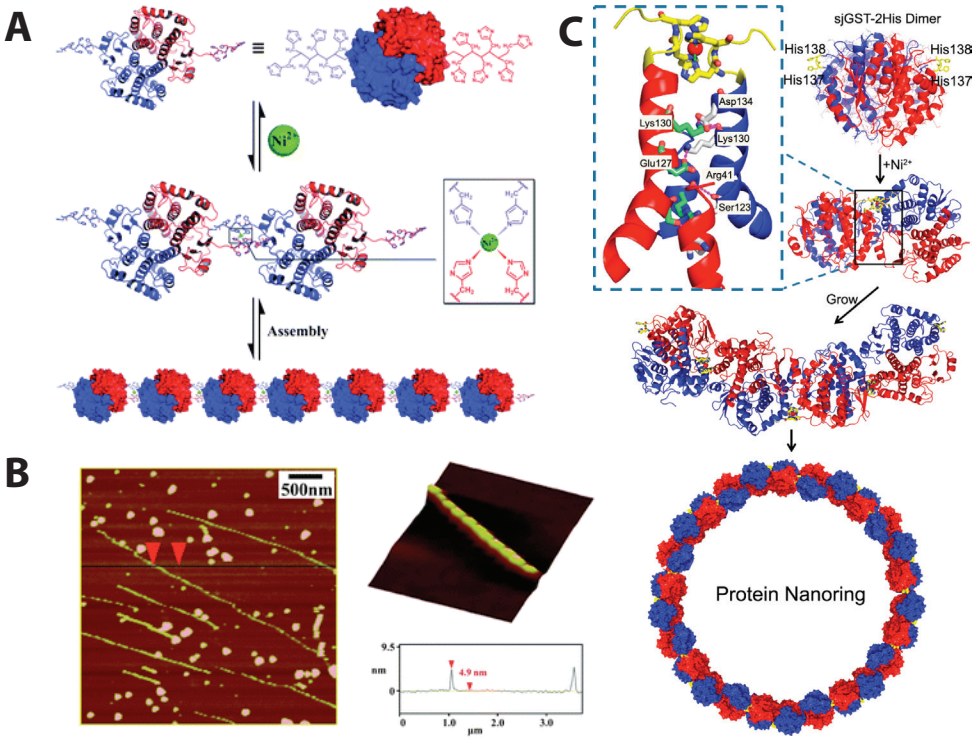


Figure 1.7. A) GST is a homodimeric enzyme with C2-symmetry. By incorporating N-terminal His6-tags on each monomer, Liu and coworkers were able to grow 1D nanowires in the presence of Ni^{2+} and disassemble them by chelating off the Ni ions B) AFM height images of the 1D GST nanowires. C) Schematic representation of the formation of protein nanorings through Ni^{2+}–His coordination. The ring-like assembly is driven by metal chelation sites off of the C2-axis, allowing the supramolecular assembly to twist. A& B adapted with permission from Ref. [43]. Copyright 2015 Royal Society of Chemistry. C adapted with permission from Ref. [44]. Copyright 2013 American Chemical Society.

a way that there was an acute angle between the respective bis-his clamps on sjGST-2His dimers. Additionally, they accounted for electrostatic dimer binding interactions to find the optimal bis-His placement. The combination of acutely angled chelation sites accompanied by local electrostatic interactions promoted growth into nanorings (Figure 1.7C). Coincidentally the electrostatic dimer interface proved useful in tuning the diameters of the protein nanorings, by simply changing the ionic strength.

1.4.3 Engineered protein-protein interfacial assembly

Self-assembly based on protein-protein interactions is a largely electrostatic process that is stabilized by larger interaction surfaces. With the recent development of protein-oriented, energy minimizing computational suites such as Rosetta; inexpensive, high-powered computing; and low-cost

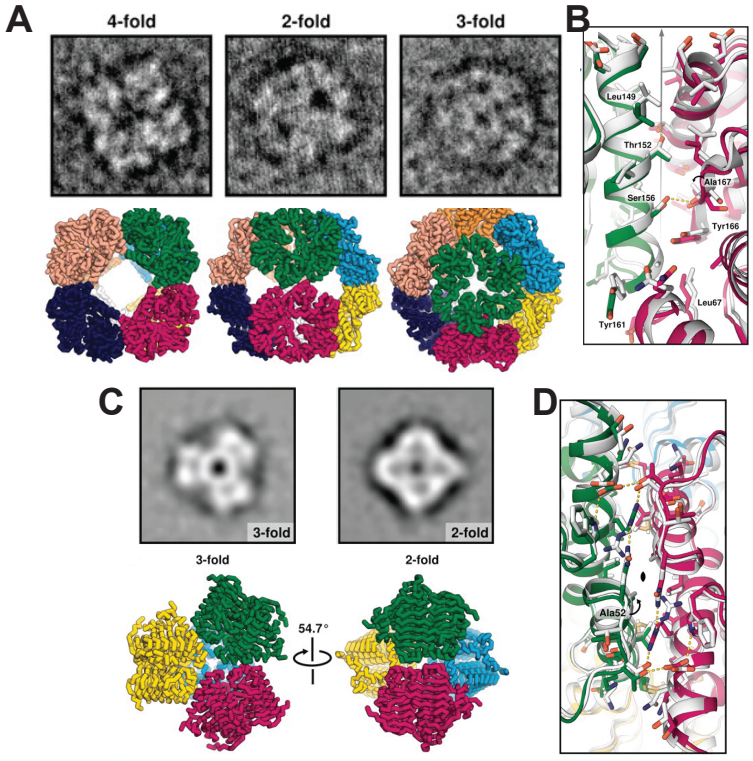


Figure 1.8. A) A negative stained TEM image of the multi-fold symmetry of the O3-33, octagonal design showing the 4-fold, 2-fold and 3-fold rotational axes. The oriented crystal structures are shown below. B) The designed interface in O3-33, which fit well with the crystal structure. C) A negative stained TEM image of the multi-fold symmetry of the T3-10, octagonal design showing the 3-fold and 2-fold rotational axes. The oriented crystal structures are shown below. B) The designed interface in T3-10, which fit well with the crystal structure. Adapted with permission from Ref. [45]. Copyright 2012 the American Association for the Advancement of Science.

synthetic genes, groups like Baker and coworkers were able to explore the sequence space of both computationally engineered and *de novo* proteins. In 2012, King *et al.* introduced a computational strategy for self-assembling proteins, using the energy minimizing Rosetta suite as the computational backbone.[45] Here they present a two-part approach in which they design a desired geometrical assembly around symmetrical protein building blocks that match the desired orientation, and

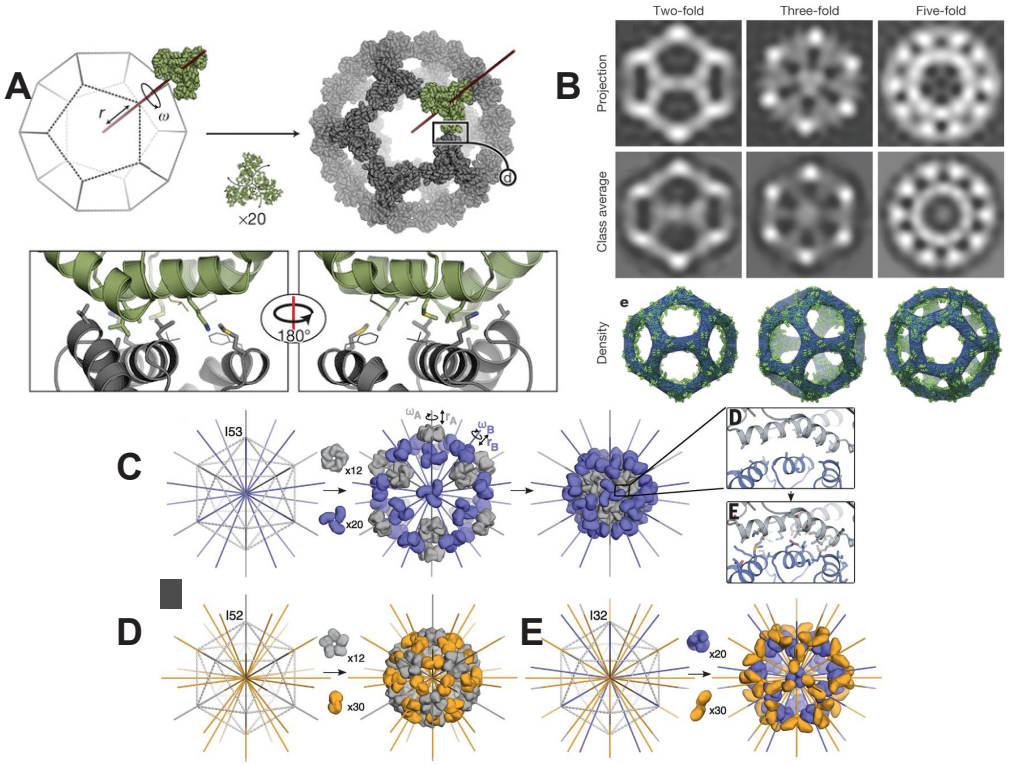


Figure 1.9. Hsia *et al.* designed large icosahedral self-assemblies of small repeating proteins. A) Icosahedral three-fold axis in red and aligned building block in green. Optimization of r and ω yields closely opposed interfaces between subunits. (Bottom) Designed low-energy interfaces; in the I3-01 case, composed of five designed residues and two native residues. B) Table showing (top) back-projections of I3-01 from the design model. (middle) Cryo-EM class averages closely match the design projections along all three symmetry axes. (bottom) The calculated initial, unrefined density (blue, 3.22σ) closely matches the design model (green). C) Bale *et al.* expanded this work to include multiple subunits within an icosahedron. The I53 architecture comprises 12 pentamers (gray) and 20 trimers (purple) aligned along the fivefold and threefold icosahedral symmetry axes. D) The I52 architecture comprises 12 pentamers (gray) and 30 dimers (orange) aligned along the fivefold and twofold icosahedral symmetry axes. E) The I32 architecture comprises 20 trimers (blue) and 30 dimers (orange) aligned along the threefold and twofold icosahedral symmetry axes. A & B adapted by permission from Macmillan Publishers Ltd: Nature, Ref. [46], copyright 2016. C-E adapted with permission from Ref. [47]. Copyright 2016 the American Academy for the Advancement of Science.

then run energy minimization calculations to optimize the protein-protein interacting face. To illustrate the efficacy of this design strategy they created assemblies with tetrahedral and octahedral point group symmetry which required the design of a specific number of proteins to be oriented, 12 and 24 copies respectively. To satisfy to symmetrical requirements of both point-group designs, they elected to use C3 symmetric molecules and ran docking studies on 271 naturally occurring trimeric proteins. Through iterative rotations they were able to isolate a series of docked proteins that exhibited no backbone or beta-carbon clashes. Energy minimization calculations were run on the best of the bunch. The structures resulting from Rosetta energy minimization calculations (8 tetrahedral & 33 octahedral) were expressed. Figure 1.8A-B and C-D show the representative designed structures from the octahedral and tetrahedral structures respectively. This design methodology of docking followed by energy minimization has gained much recognition of the past 5 years. The Baker lab has successfully applied this methodology to significantly more complex materials, like the largest of the platonic solid, an icosahedron.[46] Hsia *et al.* implemented this approach to create a 60 sub-unit icosahedral structure that has great implications as protein-based cargo carriers (Figure 1.9A and B). To say that these assemblies are stable is a bit of a minimization given they are stable at 80 ^{o}C and in 6.7 M GuHCl. This design was further expanded by Baker and coworkers, based on 120-mer icosahedrons which incorporated two assembling units into the cage.[47] Here, they computationally designed structures with molecular weights ranging from ~2-3 MDa and with cage diameters ranging from 24-40 nm (Figure 1.9C-E).

1.5. **Why repeat proteins? Engineering advantages for materials**

While the previous discussion has pointed to some useful features that proteins can impart on functional materials, genetically engineering proteins is not a trivial venture. Many of the above strategies relied on modified natural proteins that exhibit fold-symmetry or known multimerization, as a driving force for assembly. Of these proteins, many are globular. Globular proteins are a class of proteins whose structure, albeit complex, is stabilized through interactions between residues far apart in the primary sequence (Figure 1.10A). These long-range interactions limit a

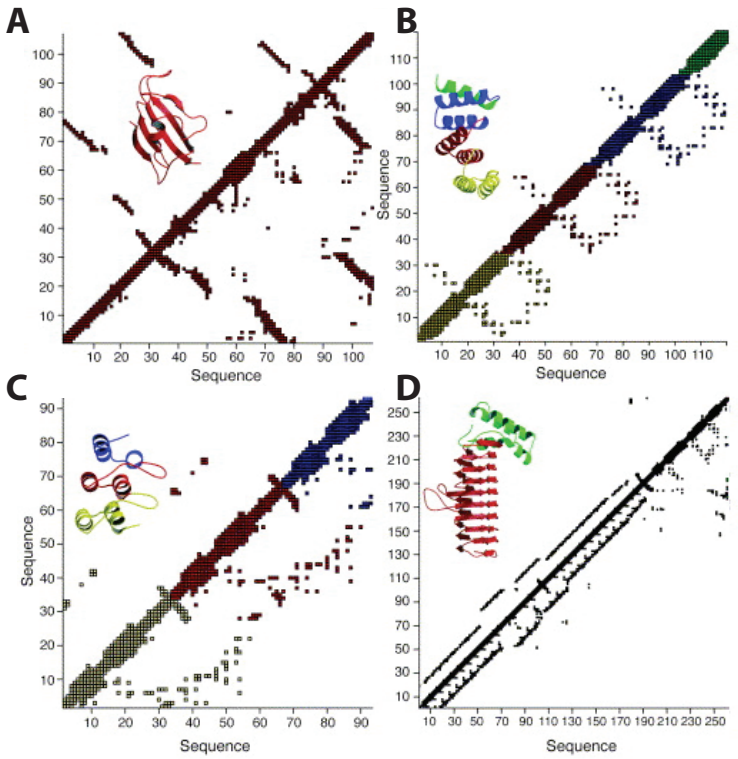


Figure 1.10. A comparison of the contact maps of multiple proteins. The line y = x indicates purely inter-residue interactions. Any point deviating from y = x indicates interactions between different residues in the primary sequence. A) FKBP12 – a globular protein of 107 amino acids, B) CTPR3 - a designed repeat protein containing 3.5 consensus tetratricopeptide (TPR) motifs, C) 3ANK – a designed repeat protein containing 3 consensus ankyrin motifs and D)N-acetylglucosamine acyltransferase – a protein that contains approximately 9 hexapeptide repeats (HPRs). Reproduced with permission from Ref. [48]. Copyright 2003 Elsevier.

globular protein's designability because small point mutations may destabilize the fold. Herein lies the benefit of another class of proteins, repeat proteins. Repeats comprise multiple short motifs, 20-50 amino acids long. They are aptly named for this repetitive motif that creates strings of tandem secondary structural elements. These structural motifs consist of two-linked secondary structural elements such as: helix-turn-helix (TPR and HEAT repeats), helix-turn-beta strand (leucine-rich repeats, LRR), (beta strand-turn-beta strand)$_3$ in HPRs along with many other architectures. These structures fold nearly identically to neighboring repeats, and coalesce forming an extended open domain with a continuous surface.[48] There are examples of closed repeat proteins as well.[49]

Repeat primary sequences are often degenerate; meaning the primary sequence of repeats can often be varied while maintaining its structure. This provides modularity for engineers looking to

incorporate specific binding domains[50] or reactive handles for material crosslinking[51]. Structurally, tandem repeats are stabilized by only inter-repeat, local interactions (Figure 1.10B-D). This makes their structure modular, making it possible to add/remove repeats while still producing a stable, folded protein. Taking this a step further, Zerbe and coworkers showed it was possible to express a split version (N- and C-terminus) of the consensus armadillo repeat, ArmRP. The C-terminal domain expressed as a folded protein, whereas the N-terminal domain was unstructured. However, these subunits were able non-covalently assemble in solution, with the C-terminal module serving as a template, with the resulting assembly resembling the crystal structure.[52]

While the work comprising this thesis focuses on one protein, the consensus tetratricopep-

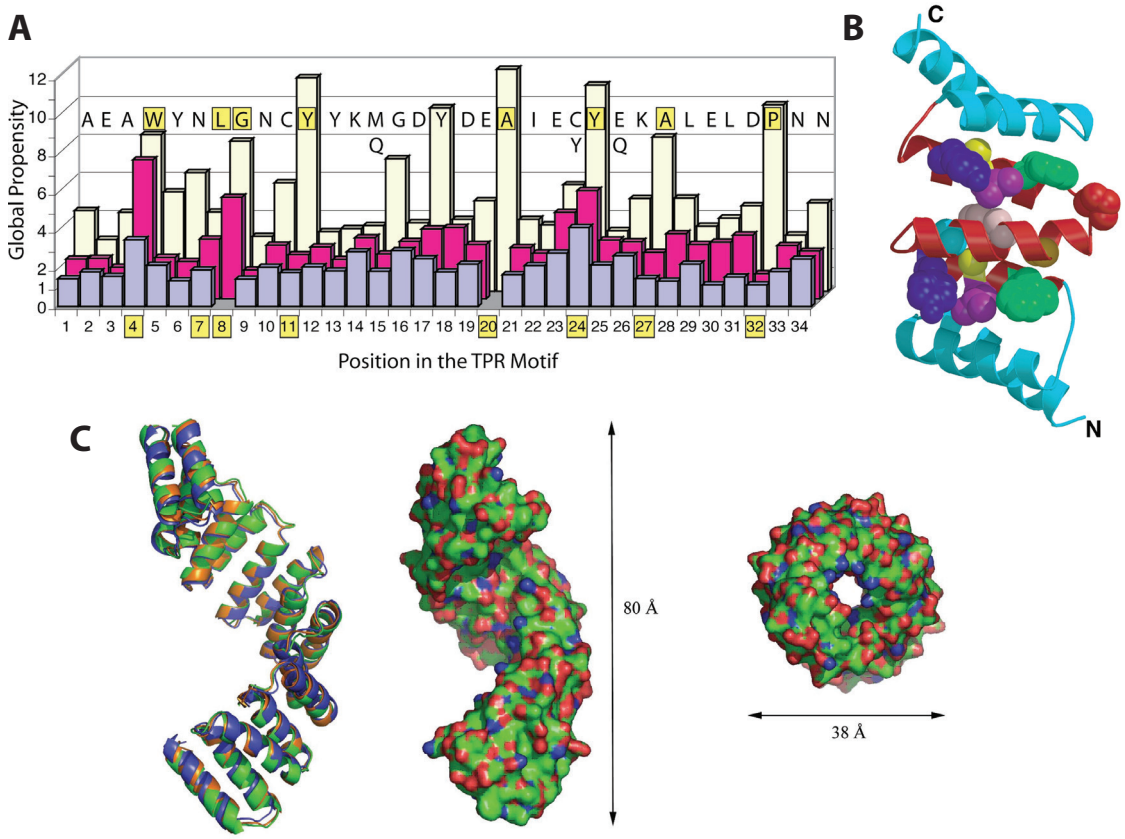


Figure 1.11. A) Histogram showing the global propensity of the three most prevalent amino acids (if above a threshold value of 1) at each of the 34 positions of the TPR motif. The numbers of the highly conserved "TPR signature residues" are boxed in yellow. B) Ribbon representation of the structure of CTPR3. The consensus residues are shown in space fill. C) Alignment of the CTPR8 eight-repeat fragment from different crystal forms. Middle, A moleclar-surface of CTPR8 superhelix with the dimensions shown. A & B reproduced with permission from Ref. [54]. Copyright 2003 Elsevier. C reproduced through open access license from Ref. [56].

tide repeat (CTPR) protein, it is prudent to briefly discuss some design strategies and technologies that are available to design and modify these proteins. The sections below are meant to introduce strategies that have proved useful in the design of repeat proteins: consensus design and *de novo* computational design.

1.5.1 Semi-rational, consensus design of repeat proteins

As the name indicates CTPR18 is the consensus sequence of the TPR family of repeat proteins that were first discovered in 1990.[53] TPR domains are 34 amino acids long and appear between 3-16 repeats in native proteins. They serve to mediate protein-protein interactions in a variety of complexes in nature. Initial consensus design of TPRs was presented by Main *et al.*[54]

Consensus protein design is a semi-rational form of genetic engineering that utilizes the evolutionary history of a desired protein domain to evaluate the conservation of residues in the

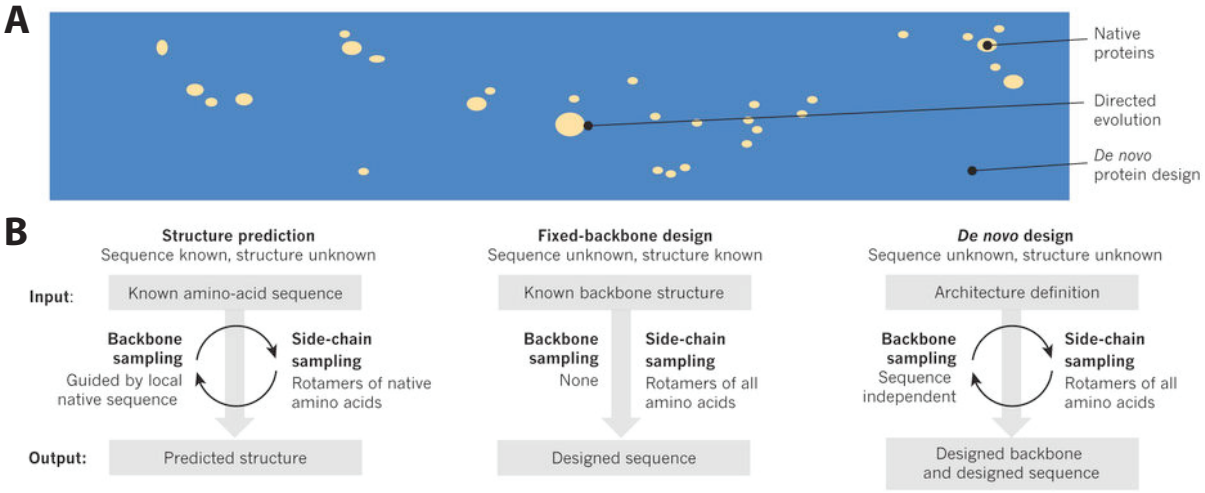


Figure 1.12. A) A schematic of the protein sequence space. Evolution has sampled only a tiny fraction of the total possible sequence space (blue), and the incremental nature of evolution results in tightly clustered families of native proteins (beige), which are analogous to archipelagoes in a vast sea of unexplored territory. Directed evolution is restricted to the region of sequence space that surrounds native proteins, whereas *de novo* protein design can explore the whole space. B) Structure prediction, fixed-backbone design and *de novo* protein design are global optimization problems with the same energy function but different degrees of freedom. In structure prediction, the sequence is fixed and the backbone structure is unknown; in fixed backbone protein design, the sequence is unknown but the structure is fixed; and in *de novo* protein design, neither is known. Reprinted by permission from Macmillan Publishers Ltd: Nature, Ref. [58], copyright 2016.

primary sequence. This design process can be broken up into four steps that include: 1) identifying a target protein, domain or sequence, 2) acquiring homologous sequences from a database, 3) analysis of a multiple sequence alignment (MSA), and 4) a determination of sequence conservancy determined by statistical analysis of amino acid count at each position. It must be noted that there is a relatively high amount of variability in this process. Interested readers should refer to Porebski and Buckle's recent review for more in depth discussion.[55]

Using the consensus design strategy above, Main and coworkers developed a consensus sequence that serves as the basis for the CTPR18 used in this thesis.[54] The consensus sequence showed that 8 of the 34 amino acids were most important for the folding stability. These hydrophobic residues sit on the interacting face of the helix-turn-helix motif (Figure 1.11). They expressed CTPR1,2 and 3. Kajander *et al.* later published the crystal structures of the CTPR8 and CTPR20.[56] CTPR assumes a superhelical structure with one full twist comprising 8-repeats that is of 8.0 x 3.8 nm. CTPR8 was crystalized in trigonal ($P3_121$), orthorhombic ($P2_12_12_1$) and tetragonal ($P4_12_12$) crystals while CTPR20 only assembled in a tetragonal ($P4_12_12$). These crystals show CTPR assembles head-to-tail into continuous superhelices while side-to-side stacking suggests CTPRs are offset by 1 repeat. These stacking modalities are the same that promote CTPR18 self-assembly discussed in Chapters 3-5.

1.5.2 De novo design of repeat proteins

While not directly applicable to the work in this thesis, the advances made in the past five years, it would be remiss to not include *de novo* design given the remarkable advances made in the past 5 years. While indirectly addressed earlier in this chapter, *de novo* designed proteins, or those designed from first principles, where the sequence and backbone structure are unknown, have made a huge impact in the literature. One of the many advantages to this design methodology is that it explores area in the sequence space, not accessible through conventional homology methods (Figure 1.12). Even though, there are multiple groups with varying computational suites attempting this,[57] Baker and coworkers have seemingly corned the market with their Rosetta suite of software. While the discussion here is only meant to skim the surface, we would like to refer interested readers to a

recent review by Huang, Boyken and Baker[58] and more specifically to their work with repeat proteins: 1) a-helical toroids,[59] 2) TIM barrel,[60] 3) variable twist and topologies in tandem repeats,[61] & 4) multi-dimensional assembly of symmetrically designed repeat protein arrays.[45, 47, 62]

1.6. **References**

[1] a) G. Karsenty, H. M. Kronenberg and C. Settembre, *Annu. Rev. Cell Dev. Biol.* **2009**, *25*, 629; b) V. Achal, A. Mukherjee, D. Kumari and Q. Z. Zhang, *Earth-Sci Rev*. **2015**, *148*, 1.
[2] a) K. Michielsen and D. G. Stavenga, *J R Soc Interface*. **2008**, *5*, 85; b) F. P. Barrows and M. H. Bart, *Nanomater Nanotechno*. **2014**, *4*; c) P. Vukusic and J. R. Sambles, *Nature*. **2003**, *424*, 852.
[3] a) B. D. Wilts, K. Michielsen, H. De Raedt and D. G. Stavenga, *J R Soc Interface*. **2012**, *9*, 1609; b) J. W. Galusha, L. R. Richey, J. S. Gardner, J. N. Cha and M. H. Bartl, *Phys Rev E*. **2008**, *77*; c) X. Wu, A. Erbe, D. Raabe and H. O. Fabritius, *Adv. Funct. Mater.* **2013**, *23*, 3615.
[4] A. R. Parker, R. C. McPhedran, D. R. McKenzie, L. C. Botten and N. A. P. Nicorovici, *Nature*. **2001**, *409*, 36.
[5] J. Teyssier, S. V. Saenko, D. van der Marel and M. C. Milinkovitch, *Nat. Commun.* **2015**, *6*.
[6] K. Autumn, Y. A. Liang, S. T. Hsieh, W. Zesch, W. P. Chan, T. W. Kenny, R. Fearing and R. J. Full, *Nature*. **2000**, *405*, 681.
[7] a) A. Blagodatski, A. Sergeev, M. Kryuchkov, Y. Lopatina and V. L. Katanaev, *Proc Natl Acad Sci U S A*. **2015**, *112*, 10750; b) T. Darmanin and F. Guittard, *Mater Today*. **2015**, *18*, 273.
[8] a) J. R. McDaniel, J. A. MacKay, F. G. Quiroz and A. Chilkoti, *Biomacromolecules*. **2010**, *11*, 944; b) D. E. Meyer and A. Chilkoti, *Biomacromolecules*. **2002**, *3*, 357.
[9] a) N. K. Li, F. G. Quiroz, C. K. Hall, A. Chilkoti and Y. G. Yingling, *Biomacromolecules*. **2014**, *15*, 3522; b) T. Christensen, W. Hassouneh, K. Trabbic-Carlson and A. Chilkoti, *Biomacromolecules*. **2013**, *14*, 1514; c) F. G. Quiroz and A. Chilkoti, *Nat Mater*. **2015**, *14*, 1164.
[10] a) W. B. Zhang, F. Sun, D. A. Tirrell and F. H. Arnold, *J Am Chem Soc*. **2013**, *135*, 13988; b) F. Sun, W. B. Zhang, A. Mahdavi, F. H. Arnold and D. A. Tirrell, *Proc Natl Acad Sci U S A*. **2014**, *111*, 11269.
[11] a) W. Hassouneh, T. Christensen and A. Chilkoti, *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]*. **2010**, *CHAPTER*, Unit; b) S. R. MacEwan, W. Hassouneh and A. Chilkoti, *JoVE*. **2014**, 10.3791/51583, 51583; c) D. E. Meyer and A. Chilkoti, *Nat biotechnol*. **1999**, *17*, 1112.
[12] W. A. Hendrickson, A. Pahler, J. L. Smith, Y. Satow, E. A. Merritt and R. P. Phizackerley, *Proc Natl Acad Sci U S A*. **1989**, *86*, 2190.
[13] a) B. Zakeri, J. O. Fierer, E. Celik, E. C. Chittock, U. Schwarz-Linek, V. T. Moy and M. Howarth, *Proc Natl Acad Sci U S A*. **2012**, *109*, E690; b) B. Zakeri and M. Howarth, *J Am Chem Soc*. **2010**, *132*, 4526.
[14] a) G. Veggiani, T. Nakamura, M. D. Brenner, R. V. Gayet, J. Yan, C. V. Robinson and M. Howarth, *Proc Natl Acad Sci U S A*. **2016**, *113*, 1202; b) X. W. Wang and W. B. Zhang, *Angew Chem Int Ed Engl*. **2016**, *55*, 3442; c) X. Gao, J. Fang, B. Xue, L. Fu and H. Li, *Biomacromolecules*. **2016**, *17*, 2812.
[15] R. Wang, Z. Yang, J. Luo, I. M. Hsing and F. Sun, *Proc Natl Acad Sci U S A*. **2017**, *114*, 5912.
[16] X. Gao, J. Fang, B. Xue, L. Fu and H. Li, *Biomacromolecules*. **2016**, *17*, 2812.

[17] C. Gilbert, M. Howarth, C. R. Harwood and T. Ellis, *ACS Synth Biol*. **2017**, 6, 957.
[18] a) T. Z. Grove, C. O. Osuji, J. D. Forster, E. R. Dufresne and L. Regan, *J Am Chem Soc.* **2010**, *132*, 14024; b) T. Z. Grove, J. Forster, G. Pimienta, E. Dufresne and L. Regan, *Biopolymers*. **2012**, *97*, 508.
[19] B. D. Olsen, J. A. Kornfield and D. A. Tirrell, *Macromolecules*. **2010**, *43*, 9094.
[20] P. B. Rapp, A. K. Omar, J. J. Shen, M. E. Buck, Z. G. Wang and D. A. Tirrell, *J Am Chem Soc*. **2017**, *139*, 3796.
[21] a) K. Maruyama, *FASEB J.* **1997**, *11*, 341; b) L. Tskhovrebova and J. Trinick, *Nat Rev Mol Cell Biol*. **2003**, *4*, 679.
[22] H. Li, W. A. Linke, A. F. Oberhauser, M. Carrion-Vazquez, J. G. Kerkvliet, H. Lu, P. E. Marszalek and J. M. Fernandez, *Nature*. **2002**, *418*, 998.
[23] a) S. Lv, D. M. Dudek, Y. Cao, M. M. Balamurali, J. Gosline and H. Li, *Nature*. **2010**, *465*, 69; b) J. Fang, A. Mehlich, N. Koga, J. Huang, R. Koga, X. Gao, C. Hu, C. Jin, M. Rief, J. Kast, D. Baker and H. Li, *Nat Commun*. **2013**, *4*, 2974.
[24] M. A. Gonzalez, J. R. Simon, A. Ghoorchian, Z. Scholl, S. Lin, M. Rubinstein, P. Marszalek, A. Chilkoti, G. P. Lopez and X. Zhao, *Adv. Mater.* **2017**, *29*.
[25] M. Torculas, J. Medina, W. Xue and X. Hu, *ACS Biomaterials Science & Engineering*. **2016**, *2*, 1211.
[26] a) T. E. DeCoursey, *J Physiol*. **2008**, *586*, 5305; b) C. A. Wraight, *Biochim Biophys Acta*. **2006**, *1757*, 886.
[27] a) D. D. Ordinario, L. Phan, W. G. t. Walkup, J. M. Jocson, E. Karshalev, N. Husken and A. A. Gorodetsky, *Nat Chem*. **2014**, *6*, 596; b) D. D. Ordinario, L. Phan, Y. Van Dyke, T. Nguyen, A. G. Smith, M. Nguyen, N. M. Mofid, M. K. Dao and A. A. Gorodetsky, *Chem Mater*. **2016**, *28*, 3703; c) N. Amdursky, X. Wang, P. Meredith, D. D. Bradley and M. M. Stevens, *Adv. Mater.* **2016**, *28*, 2692; d) O. Silberbush, M. Amit, S. Roy and N. Ashkenasy, *Adv. Funct. Mater.* **2017**, *27*, 1604624.
[28] K. L. Naughton, L. Phan, E. M. Leung, R. Kautz, Q. Lin, Y. Van Dyke, B. Marmiroli, B. Sartori, A. Arvai, S. Li, M. E. Pique, M. Naeim, J. P. Kerr, M. J. Aquino, V. A. Roberts, E. D. Getzoff, C. Zhu, S. Bernstorff and A. A. Gorodetsky, *Adv. Mater.* **2016**, *28*, 8405.
[29] C. Zhong, Y. Deng, A. F. Roudsari, A. Kapetanovic, M. P. Anantram and M. Rolandi, *Nat Commun*. **2011**, *2*, 476.
[30] a) D. Ivnitski, M. Amit, O. Silberbush, Y. Atsmon-Raz, J. Nanda, R. Cohen-Luria, Y. Miller, G. Ashkenasy and N. Ashkenasy, *Angew Chem Int Ed Engl*. **2016**, *55*, 9988; b) M. Amit, G. Cheng, I. W. Hamley and N. Ashkenasy, *Soft Matter*. **2012**, *8*, 8690.
[31] M. A. Kostiainen, O. Kasyutich, J. J. L. M. Cornelissen and R. J. M. Nolte, *Nat Chem*. **2010**, *2*, 394.
[32] M. A. Kostiainen, P. Hiekkataipale, A. Laiho, V. Lemieux, J. Seitsonen, J. Ruokolainen and P. Ceci, *Nat Nano*. **2013**, *8*, 52.
[33] V. Liljeström, J. Seitsonen and M. A. Kostiainen, *ACS Nano*. **2015**, *9*, 11278.
[34] J. Mikkilä, E. Anaya-Plaza, V. Liljeström, J. R. Caston, T. Torres, A. de la Escosura and M. A. Kostiainen, *ACS Nano*. **2016**, *10*, 1565.
[35] M. V. Golynskiy, T. C. Davis, J. D. Helmann and S. M. Cohen, *Biochemistry*. **2005**, *44*, 3380.
[36] H. Lee, N. F. Scherer and P. B. Messersmith, *Proc Natl Acad Sci U S A*. **2006**, *103*, 12999.
[37] K. J. Waldron and N. J. Robinson, *Nat Rev Micro*. **2009**, *7*, 25.
[38] E. N. Salgado, R. J. Radford and F. A. Tezcan, *Acc Chem Res*. **2010**, *43*, 661.
[39] G. Yang, L. Wu, G. Chen and M. Jiang, *Chem Commun (Camb)*. **2016**, *52*, 10595.

[40] Q. Luo, C. Hou, Y. Bai, R. Wang and J. Liu, *Chem Rev*. **2016**, 116, 13571.
[41] a) E. N. Salgado, J. Faraone-Mennella and F. A. Tezcan, *J Am Chem Soc*. **2007**, *129*, 13374; b) J. D. Brodin, A. Medina-Morales, T. Ni, E. N. Salgado, X. I. Ambroggio and F. A. Tezcan, *J Am Chem Soc*. **2010**, *132*, 8610.
[42] J. D. Brodin, S. J. Smith, J. R. Carr and F. A. Tezcan, *J Am Chem Soc.* **2015**, *137*, 10468.
[43] W. Zhang, Q. Luo, L. Miao, C. Hou, Y. Bai, Z. Dong, J. Xu and J. Liu, *Nanoscale*. **2012**, *4*, 5847.
[44] Y. S. Bai, Q. Luo, W. Zhang, L. Miao, J. Y. Xu, H. B. Li and J. Q. Liu, *J Am Chem Soc.* **2013**, *135*, 10966.
[45] N. P. King, W. Sheffler, M. R. Sawaya, B. S. Vollmar, J. P. Sumida, I. Andre, T. Gonen, T. O. Yeates and D. Baker, *Science*. **2012**, *336*, 1171.
[46] Y. Hsia, J. B. Bale, S. Gonen, D. Shi, W. Sheffler, K. K. Fong, U. Nattermann, C. Xu, P. S. Huang, R. Ravichandran, S. Yi, T. N. Davis, T. Gonen, N. P. King and D. Baker, *Nature*. **2016**, *540*, 150.
[47] J. B. Bale, S. Gonen, Y. Liu, W. Sheffler, D. Ellis, C. Thomas, D. Cascio, T. O. Yeates, T. Gonen, N. P. King and D. Baker, *Science*. **2016**, *353*, 389.
[48] E. R. Main, S. E. Jackson and L. Regan, *Curr Opin Struc Biol*. **2003**, *13*, 482.
[49] P. S. Huang, K. Feldmeier, F. Parmeggiani, D. A. Fernandez Velasco, B. Hocker and D. Baker, *Nat Chem Biol*. **2016**, *12*, 29.
[50] a) Y. L. Boersma and A. Pluckthun, *Curr Opin Biotechnol*. **2011**, *22*, 849; b) P. Ernst and A. Pluckthun, *Biol. Chem*. **2017**, *398*, 23; c) M. E. Jackrel, R. Valverde and L. Regan, *Protein Sci*. **2009**, *18*, 762.
[51] S. H. Mejias, B. Sot, R. Guantes and A. L. Cortajarena, *Nanoscale*. **2014**, *6*, 10982.
[52] R. P. Watson, M. T. Christen, C. Ewald, F. Bumbak, C. Reichen, M. Mihajlovic, E. Schmidt, P. Güntert, A. Caflisch, A. Plückthun and O. Zerbe, Structure. 2014, 22, 985.
[53] a) T. Hirano, N. Kinoshita, K. Morikawa and M. Yanagida, *Cell*. **1990**, *60*, 319; b) R. S. Sikorski, M. S. Boguski, M. Goebl and P. Hieter, *Cell*. **1990**, *60*, 307.
[54] E. R. Main, Y. Xiong, M. J. Cocco, L. D'Andrea and L. Regan, *Structure*. **2003**, *11*, 497.
[55] B. T. Porebski and A. M. Buckle, *Protein Eng Des Sel*. **2016**, *29*, 245.
[56] T. Kajander, A. L. Cortajarena, S. Mochrie and L. Regan, *Acta Crystallogr D Biol Crystallogr*. **2007**, *63*, 800.
[57] a) C. W. Wood, M. Bruning, A. A. Ibarra, G. J. Bartlett, A. R. Thomson, R. B. Sessions, R. L. Brady and D. N. Woolfson, *Bioinformatics*. **2014**, *30*, 3029; b) C. Negron and A. E. Keating, *Methods Enzymol*. **2013**, *523*, 171.
[58] P. S. Huang, S. E. Boyken and D. Baker, *Nature*. **2016**, *537*, 320.
[59] L. Doyle, J. Hallinan, J. Bolduc, F. Parmeggiani, D. Baker, B. L. Stoddard and P. Bradley, *Nature*. **2015**, *528*, 585.
[60] P. S. Huang, K. Feldmeier, F. Parmeggiani, D. A. F. Velasco, B. Hocker and D. Baker, *Nat Chem Biol*. **2016**, *12*, 29.
[61] T. J. Brunette, F. Parmeggiani, P. S. Huang, G. Bhabha, D. C. Ekiert, S. E. Tsutakawa, G. L. Hura, J. A. Tainer and D. Baker, *Nature*. **2015**, *528*, 580.
[62] a) Y. Mou, P. S. Huang, F. C. Hsu, S. J. Huang and S. L. Mayo, *Proc Natl Acad Sci U S A*. **2015**, *112*, 10714; b) N. P. King, J. B. Bale, W. Sheffler, D. E. McNamara, S. Gonen, T. Gonen, T. O. Yeates and D. Baker, *Nature*. **2014**, *510*, 103; c) S. Gonen, F. DiMaio, T. Gonen and D. Baker, *Science*. **2015**, *348*, 1365; d) Y. Hsia, J. B. Bale, S. Gonen, D. Shi, W. Sheffler, K. K. Fong, U. Nattermann, C.

Xu, P. S. Huang, R. Ravichandran, S. Yi, T. N. Davis, T. Gonen, N. P. King and D. Baker, *Nature*. **2016**, *535*, 136

Chapter 2. Design of Self-assembling Protein-polymer Conjugates

N. A. Carter, X. Geng and T. Z. Grove, in *Protein-based Engineered Nanostructures. Advances in Experimental Medicine and Biology*, vol 940. (Eds: A. L. Cortajarena, T. Z. Grove), Springer International, Switzerland **2016**, pp. 179-214.

Attributions: NAC and XG were the primary writers of the manuscript with equal contributions. TZG assisted in writing and editing the chapter.

2.1. Introduction to Protein-polymer Materials

Protein-polymer conjugates are of particular interest for nanobiotechnology applications because of the various and complementary roles that each component may play in composite hybrid-materials..[1] Proteins provide a degree of complexity in structure and resulting function[2] that, as it stands, we cannot mimic with synthetic materials. For example, the tertiary structure of protein enzymes allows them to carry out complex reaction cascades all the while creating enantiopure products.[2-3] This structure can also provide complex and specific molecular recognition as in antibodies, Avidin/biotin, etc.[4] While we have a good understanding of the interactions that drive secondary structure formation and can thus make reliable predictions, the complexity of tertiary structures have made it exceedingly difficult to create de novo sequences with predictable structure and function.[5] Simply put, the creation of novel protein functionality from scratch is still beyond our reach.

This limitation of only being able to use existing or slightly modified proteins is a key driving force behind the design of protein-polymer conjugates. Polymers, which have a seemingly endless supply of chemical functionalities, can add a great number of beneficial properties to the material, like increased thermal stability[5j], altered solubility[6], added stimuli responsiveness[7], etc. Often, these modifications can be quite simple. In fact, the most prevalent conjugated polymer to date is the hydrophilic polymer, poly(ethylene glycol), PEG. PEGylation was first used in 1970s through

conjugation to one or more sites on the protein to increase protein solubility. PEGylation has since established itself as a solid method for reducing the immunogenicity and increasing protein stability, thus increasing the half-life of protein drugs in vivo. While still of great interest and significance, the scientific community has recognized the large potential of protein-polymer conjugate materials beyond simple PEGylation.

To create the next generation of protein-polymer conjugate materials, researchers have begun to use polymers with greater structural and chemical diversity to create protein-polymer hybrids that self-assemble and form hierarchical structures with defined and predictable structure-property relationships.[7-8] Such hierarchical structures give rise to useful patterning, mechanical and transport properties that may help realize new, more efficient materials for energy generation, catalysis, micropropulsion, etc. While this review focuses on the design principles and applications of self-assembling protein-polymer conjugates, we will address the 1) general design methodology from both a synthetic and genetic perspective, 2) conjugation strategies, 3) protein vs. polymer driven self-assembly 4) conjugate topology and finally 5) emerging applications for conjugate materials.

2.2. Synthetic approaches toward protein-polymer conjugates

2.2.0 Peptide/protein synthesis and controlled radical polymerization (CRP)

Ring opening polymerization of N-carboxyanhyrides (NCA) derivatives is routinely utilized to prepare homopolypeptides or random polypeptides that lack sequence specificity and monodispersity.[9] Solid-phase peptide synthesis (SPPS) is a good alternative method that provides an efficient avenue to monodisperse sequence-specific products after purification.[10] While SPPS is limited to around 50 amino acid long peptides, a complementary technique, native chemical ligation (NCL), may be employed to fuse two or more peptide fragments to generate molecules up to 200 amino acid residues.[11] In view of the large size and complex structure, recombinant DNA technology is a potent tool to express proteins containing a sequence-specific folded structure.[12]

Controlled radical polymerization (CRP) has become one of the most prolific research areas in

polymer science, as attested by the rapid growth in the number of publications. CRP has striking features like controlled chain length, narrow molecular weight distribution (i.e. low polydispersity) and high end-group fidelity. With the aid of CRP, remarkable progress has also been made in the preparation of protein-polymer bioconjugates with high complexity and versatile functionality. Overall, CRPs can be classified into three categories: nitroxide mediated radical polymerization (NMP), atom transfer radical polymerization (ATRP) and reversible addition fragmentation chain transfer (RAFT). The fourth type is still debated single-electron transfer living radical polymerization (SET-LRP).[13]

2.2.1 General conjugation methodology

A large number of chemistries are available to conjugate small molecules and polymers to biomolecules, such as proteins, peptides, nucleic acids and sugars. Many of these commonly exploited reactions are chemoselective, and are too great in number to comprehensively cover in this review. While there are a large variety of ever emerging chemistries used for bio-conjugation, we limit our discussion to chemistries that have been used to conjugate protein and polymers along with a few

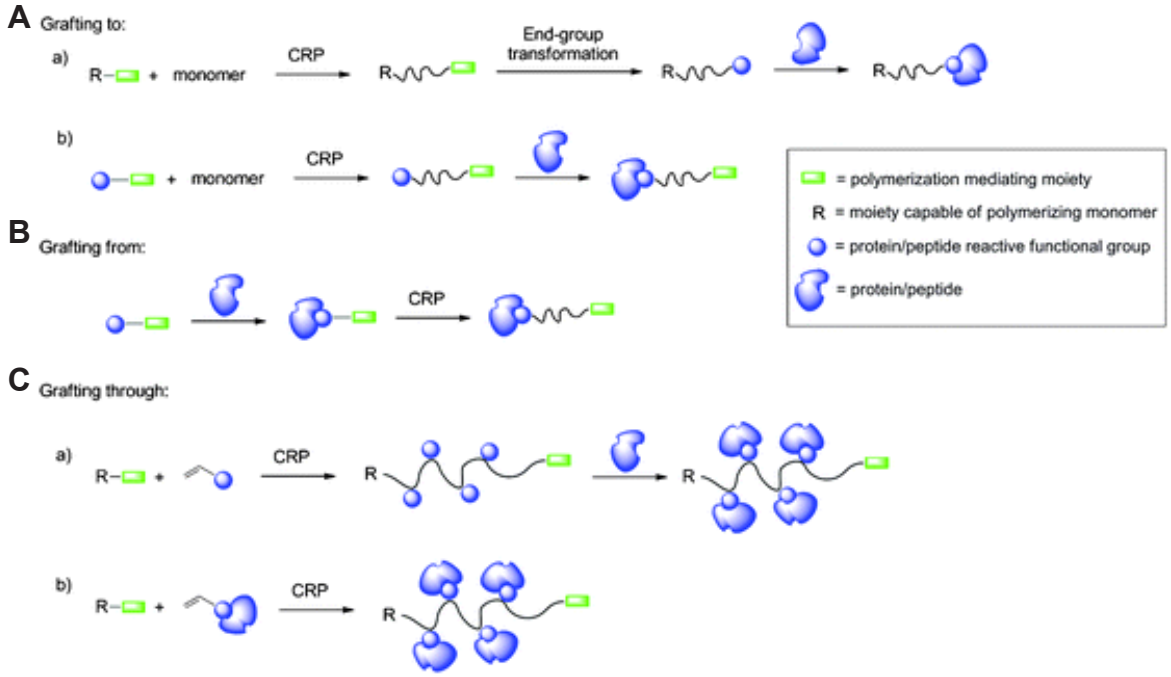


Figure 2.1. Graft to/from/through strategies or the synthesis of protein-polymer conjugates. Reproduced from Ref. [39] with permission from The Royal Society of Chemistry.

newer promising chemistries. We would like to direct interested readers to Hackenberger and Schwarzer's 2008, Jung and Theato's 2013, and Boutureira and Bernardes's 2015, reviews which provide a more comprehensive look into available bio-conjugation chemistries.[14],[15],[16]

When designing protein-polymer conjugates the first thought must focus on how to attach the two materials together. Figure 2.1 below shows the three general methodologies that have been used to produce synthetic grafted co-polymers: 1) grafting-to, 2) grafting-from, and 3) grafting-through, although to our knowledge only the grafting-to and grafting-from methods have been used to design protein-polymer conjugates. Grafting-to (Figure 2.1A) refers to a post-synthetic conjugation method in which two separate macromolecules, i.e. a protein and polymer, are coupled together through some reactive handle (e.g. amine and EDC activated carboxylic acid). Figure 2.1Aa schematically shows this strategy where the polymer is synthesized using a controlled radical polymerization (CRP) and the terminal functional group of the polymer is modified to be reactive with proteins. An alternate method (Figure 2.1Ab) shows the same process only using a hetero-difunctional initiator molecule, where one functional group is CRP reactive and the other is protein reactive. There are a wide variety of chemistries available, both covalent and non-covalent, to conjugate proteins to existing polymer chain-ends or pendant groups. While these will be discussed later in the chapter it is important to note that one must carefully select reactive handles on the monomer and hetero-functional initiator to eliminate the possibility of a bi-directional polymerization. The advantage of grafting-to method is that polymerization steps are isolated from the biomolecule conjugation step allowing for a greater variety of reaction conditions for synthetic polymerization schemes, which may be incompatible with proteins. Likely issues with this strategy are the appearance of side reactions and poor polymer solubility, which may result in aggregated or denatured protein. One limitation of the grafting-to method is in the fabrication of a bottle-brush architecture. Diffusion and steric issues inherent to large molecules largely affect the efficiency of multivalent reaction, leading to irregularity in regards to the density and position of the attached pendant macromolecules.

In contrast, the grafting-from method (Figure 2.1B) exploits a reactive handle on the protein,

using it for in situ polymerization from the protein modified with a chain-transfer agent (CTA), also called a macro-initiator. This strategy allows for polymerization of a suitable monomer directly from the protein. The approach provides a more regular, densely packed structure because of the direct polymerization off of the protein. Additionally, the graft-from method has advantages in terms of purification, attachment specificity, and yield. However, immobilization of a RAFT agent onto proteins/peptides may alter their conformation and thus bioactivity. Additionally, the presence of bulky biomolecules may affect the chain-transfer kinetics and consequently give rise to a relatively broad molecular weight distribution. In contrast, the graft to approach is superior with respect to bioactivity and polydispersity, but it may necessitate multistep purification and suffer from the relatively low yield.

2.2.2 Chemoselective synthetic approaches using natural amino acids

The early work in bio-conjugation was focused on PEGylating proteins and small molecule drugs to increase the circulation time and thus bioavailability.[17]. This work focused on chemical ligation to reactive groups that are inherent to a protein's primary structure. Of the 20 canonical or natural amino acids present in proteins, conjugation chemistries exist for 11 of them. The most widely used are amines (Figure 2.2A) and sulfhydryls (Figure 2.2B), the reactive moieties in lysine and cysteine respectively. Figure 2.2C highlights some of the more common reactions utilizing natural amino acid side chains in red (including amine, thiol, phenol, tyrosyl and carboxyl) and non-natural amino acid side-chains in blue (including aniline, azide, and alkyne). While there are a number of reactive chemistries that are not shown, we will discuss strategies that will first incorporate residue-specific chemistries, going from the more promiscuous Lys to more specific chemistries in Tyr conjugation, and where possible highlight methods to provide site-specific conjugation as well. From a materials design standpoint, precise control over both the conjugation reaction and site of conjugation is necessary for predictable and controllable self-assembly.

A good rule of thumb is the more reactive an amino acid's side-chain is, the less control one

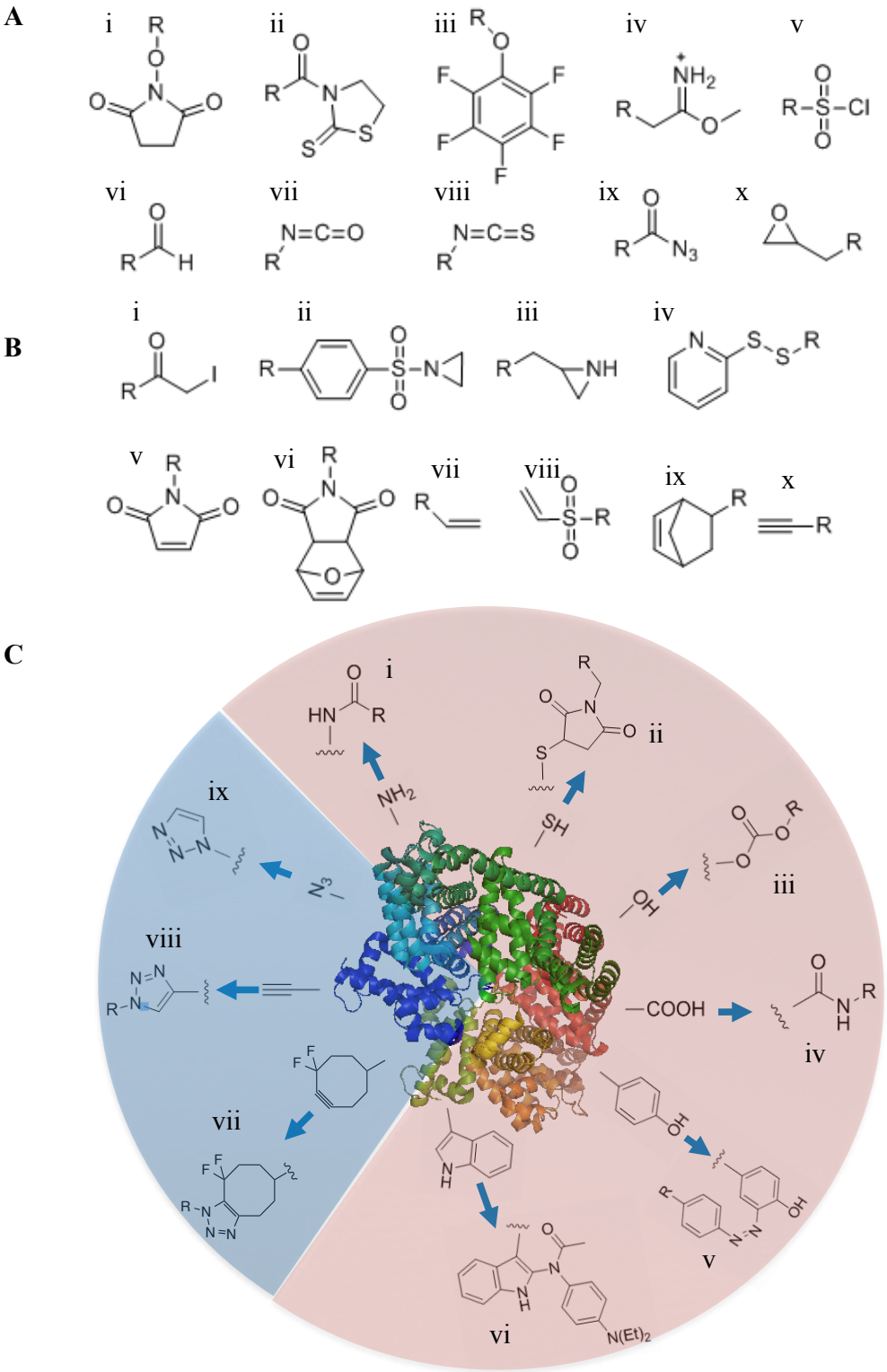


Figure 2.2. A) amine reactive functionalities B) thiol-reactive functionalities, Ref. [11] B) commonly used chemi-cal handles for bio-conjugation, red are natural, noncanonical are blue, Ref. [11]

will have over conjugation. However, other key points can play into and significantly affect the optimal design of a conjugation strategy using canonical amino acids. One key aspect is the relative abundance of an amino acid. For example, Lys is more frequently used in the protein sequence than Trp. Thus, that number of available chemistries for Lys makes it very simple to pick a compatible reaction with good yields, but this will likely come at the sacrifice of site-specificity. Whereas, using Trp will likely increase site-specific conjugation, sacrificing the efficiency because of tryptophans are seldom sovent accessible. Solvent accessibility is a second important consideration one must account for. In the discussion below, we will show examples where solvent accessibility affords site-specificity where the conjugation chemistry otherwise would not.

2.2.2.1 Lysine conjugation (including N-terminus)

The primary amine groups on the lysine residues and the N-terminus of a protein provide prime nucleophilic sites for chemical conjugation. In general, amines are a promiscuous handles that are able to react with a multitude of functional groups, some of which are shown in Figure 2.2A: such as activated esters (i-iv), thionyl chlorides (v), aldehydes (vi), isocyanates/isothiocyanates (vii-viii), acyl azides (ix) and epoxides (x).[16] An activated N-hydroxysuccinimide (NHS)-ester (Figure 2.2Ai) terminated CTA is widely utilized for the synthesis of polymers bearing a terminal NHS moiety. For example, a thermoresponsive protein-polymer conjugate was synthesized through the coupling reaction between NHS functionalized poly(N-isopropyl acrylamide) (PNIPAAM) with amines in lysozyme.[18] In another example, a novel thiazolidine-2-thione (Figure 2.2Aii) functionalized CTA was used to prepare the well-defined telechelic poly(N-(2-hydroxypropyl)methacrylamide) (PHPMA).[19] The number of polymeric chains conjugated to the lysozyme was dependent both on the pH of the reaction mixtures and molecular weight of the polymer. At the pH below the pK_a of lysine, nucleophilicity of Lys is significantly reduced thus resulting in lower reactivity towards conjugation.[19]

Utilizing a free amine for bio-conjugation is the idiomatic 'low-hanging fruit'. This is due to the relatively high frequency with which lysine appears in proteins. The omnipresence of Lys

combined with it's hydrophilic side-chain, that is often solvent accessible, makes it an easy target for functionalization. While this increases the probability of a successful conjugation reaction, it does little for positional control. Thus, multiple groups have sought to develop functionalization strategies that are specific for the free amine at the N-terminus of a protein. Their rational lies in the inherent difference in pKa of the α-amino group, which is generally involved in a peptide bond, compared to the ε-amino group on Lys side-chain. A recently published promising strategy utilizes a functionalized ketene to achieve greater than 99:1 (α-amino:ε-amino) conjugation selectivity in over half of the tested peptides. While short peptides with phenylalanine and lysine at the N-terminus yielded modest selectivity of 6:1 and 29:1, proteins with these terminal functionalities insulin (F), lysozyme (K), and RNAseA (K) showed no evidence of γ-amino functionalization, contrary to functionalization using NHS esters.[20]

The Francis group has formulated multiple strategies to specific conjugation at the N-terminal amine. Initially, pyridoxal-5-phosphate (PLP) was used to create oxime/hydrazine linkages with N-terminal Gly, Val, Lys and Met in a series of proteins. While this reaction has good conversions, between 30-80% for the aforementioned amino acids, this technique would likely not work for proteins with terminal Ser, Thr, Cys and Trp due to their reactivity towards aldehydes.[21] A later attempt where 2-pyridinecarboxyaldehydes (2PCA) was reacted with the terminal amine proved much more successful. This single step process yielded good to excellent (43-96%) conjugation in mild conditions at physiological pH for all proteins tested with the exception of a protein with Pro adjacent to the N-terminus and one that is acylated at the N-terminus. These are modest limitations given the fact that the 2PCA system gave no interference and subsequent functionalization at the ε-amino position of a terminal lysine.[22]

2.2.2.2 Cysteine conjugation

When comparing to amino- and carboxyl- modifications, sulfhydryls (or thiols) are more desirable target residues for functionalization. First, their relative frequency, i.e. that of Cys, is much lower when compared to Lys, Glu and Asp, limiting the number of possible conjugation events. Its

infrequent appearance has proven to be advantageous in the design of certain proteins. The second advantage of Cys is that many sulfhydryls reside on the interior of proteins forming disulfide bridges, further limiting the sites for functionalization. While this limits the accessibility of the moieties shown in Figure 2.2B, there are a few clever methods that have exploited the proximity of vicinal or through-space Cys neighbors for site-specific conjugation. These will be discussed a little further in this section. Lastly, at physicological pH (~7.0) thiols are much more reactive than the protonated amine groups of Lys giving an overall milder reaction scheme.

Thiol-based reactions that take advantage of the reduced sulfhydryl state, such as thiol-halogeno, thiol-parafluoro, thiol-ene, thiol-yne, thiol-vinylsulfone, thiol–maleimide, thiol–bisulfone, and thiol–pyridyl disulfide (PDS) that have all been exploited for connecting polymers with natural building blocks (Figure 2.2B).[23] Haddleton et al. first reported a furan protected maleimide functionalized initiator used to PEGylate BSA.[24] Maynard and coworkers further extended this method through (RAFT) polymerization of a furan-protected maleimide-CTA. A Retro-Diels-Alder reaction deprotected the maleimide-end-group, allowing for conjugation between the polymer and mutated lysozyme in PBS buffer.[25] In addition to the maleimide-CTA, a pyridyl disulfide (PDS) functionality can be readily introduced into the RAFT agent for the meditated polymerization of acrylate and acrylamide monomers.[26], [27], [28] Since the PDS-thiol coupling reaction is very selective and efficient, it permits facile conjugation of polymers with BSA, glutathione or designer proteins

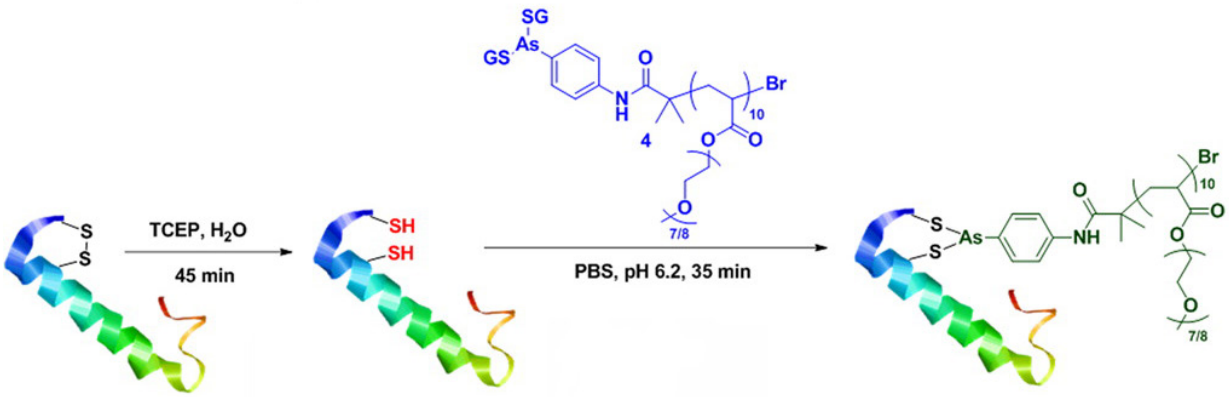


Figure 2.3. Davis and coworkers showed that organic arsenicals are specific for and efficient at bridging the gap between short disulfides. This reaction proves useful for site-specific conjugation of polymers to proteins (Reprinted with permission from Ref. [54]. Copyright 2015 American Chemical Society

containing cysteine residues. Furthermore, the formed disulfide bond could be reversibly cleaved by adding tris (2-Carboxyethyl) phosphine hydrochloride (TCEP) or dithiothreitol (DTT) for in vitro release of the protein/peptide.

In addition to conjugation of reduced, solvent accessible cysteines, there are examples of site-selective strategies that conjugate across naturally occurring disulfide bridges. Bis-alkylation modifications take advantage of solvent accessible disulfide bridges and as a result are both chemo- and site-specific. In a proof-of-principle study Shaunak *et al.* used a monosulfone functionalized PEG to bis-alkylate across a solvent accessible disulfide bridge in Interferon a-2b. This approach is versatile in that when stoichiometric amounts of the functionalized PEG and protein are used, the PEG-monosulfone bridges the S-S gap forming two disulfides that maintained the integrity of the proteins tertiary structure. Disulfides could be disrupted by doubling the amount of PEG-monosulfone, where each of the Cys involved in the disulfide had now formed a disulfide with its own functionalized PEG.[29] The Davis group (Figure 2.3) used arsenical radicals to bridge a newly reduced disulfide. This reaction is highly specific for reduced disulfides rather than lone sulfhydryls because two vicinal or through space neighbors are needed to coordinate As. The conjugation was as efficient as the 'benchmark' molecule dibromomaleimide. Furthermore, a functionalized arsenical radical was used as an initiator for aqueous SET-LRP polymerization of a poly(poly[ethylene glycol] methyl ether acrylate) macromonomer that was then grafted-to salmon calcitonin (sCT). This bridging ligand was releasable in response to ethanedithiol (EDT), which in high concentration outcompetes the organic arsenical and releases it. While this system does incorporate arsenic, it may still be a medically relevant method due to the significantly diminished toxicity of organic arsenicals when compared to their inorganic counterparts.[30] Lastly, a promising new strategy by Griebnow *et al.* shows the ability to bridge disulfides in a single-pot reaction light-initiated thiol-yne 'click' reaction. They originally demonstrated this ability with small peptides in 2015, but have since expanded their system to incorporate 6-heptynoic acid into the interchain disulfide bond in a 46 kDa Fab-fragment M14-G07.[31] While only carried out with small molecule bridges, this method will likely to be useful for polymer bioconjugation in future research given the simplis-

tic synthetic setup and efficiency of 'click' reactions.

2.2.2.3 Tyrosine Conjugation

Multiple strategies to functionalize tyrosine have been developed in the recent years. Tyrosine, like Cys is not very abundant in nature and lends itself to more site-selective chemistries than Lys. Tyrosine itself is susceptible to oxidation, generating a hydroxyl radical. In Nature these hydroxyl radicals form dityrosine cross-links in oxidatively stressed proteins.[32] Recent work using transition metals successfully mimics this reaction on solvent accessible tyrosines. Finn *et al.* showed the ability to oxidatively cross-link two neighboring Tyr residues using a nickel(II) catalyst in the presence of a GGH peptide and hydrogen peroxide.[33] A similar method utilizes an N-terminal tyrosine to react with superoxide produced in situ forming a Tyr-hydroxide/hydroperoxide complex which then participates in a Michael-type addition of glutathione.[34] However, a significant limitation to this reaction is possible side-reaction with neighboring cysteines since thiols are also susceptible to these modifications. A different transition-metal mediated conjugation of Tyr was proposed by the Francis lab. Here the Trost allylic-alkylation mediated by a palladium catalyst was used to specifically functionalize Tyr.[35]

Tyr can also be functionalized using a three-component Mannich-type reaction coupling phenol, aniline, and an aldehyde to create the Mannich adduct. Surface accessible tyrosines can be conjugated under mild conditions with a variety of aldehydes, at protein concentrations as low as 20 uM.[36] While effective, this conjugation strategy is susceptible to side reactions with Trp and reduced disulfides. Another chemistry for conjugating to Tyr is to react aniline-functionalized macromolecules in the presence of formaldehyde. This strategy was successfully used to attach aniline-functionalized peptides to protein Tyr.[37] While this conjugation strategy is very selective, it is slow and may not be amenable for scaled-up materials. For this reason, Haddleton and coworkers utilized a diazonium mediated conjugation of Tyr - another of the techniques originally developed by the Francis lab - to PEGylate salmon calcitonin (sCT), a 32 amino acid peptide, used to treat bone diseases like osteoporosis and hypercalcimia. Haddleton and coworkers were able to show

efficient conjugation without sacrificing bioactivity.[38] The main advantages of this reaction are relative speed, reaction is catalyst free and it is already applied in pharmaceutical industry setting. sCT was selected because of its small size, simple structure and because it has a single Tyr which in a non-conserved position. While this will not directly transferable to all systems of interest, this type of reaction remains a more viable option than the Mannich reaction or Trost-alkylation.

A more recent and now commercially available tyrosine specific conjugation strategy, based on a 'click-like' reaction of tyrosine with triazolinediones (TAD), was introduced by the late Carlos Barbas III.[39] While this reaction is an electrophilic aromatic substitution (EAS), which is typically very slow, the authors found that is was significantly sped up in aqueous buffer. The reaction of 4-phenyl-3H-1,2,4-triazoline-3,5(4H)-diones (PTAD) with Tyr further proved to be highly specific when tested at physiological pH against all possible PTAD reactive side-chains: Trp, Ser, Glu, Lys, Arg, and His.[40] This reaction is compatible with a multiple labeling scheme using 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and NHS-ester chemistries to label a Lys and Cys respectively, and that is useful to PEGylate proteins at multiple Tyr sites.[40] Recently Novartis used this chemistry for efficient post-translational synthetic glycosylation when they developed a glycosylated pilus protein from Streptococcus agalactiae, designed to help treat common infections in pregnant women and fetuses.[41] This strategy may have broad implications on the production of therapeutic proteins due to the inherent limitations of producing glycosylated proteins in E. coli.[42]

2.2.2.4 Tryptophan Conjugation

Tryptophan is the least abundant amino acid found in proteins.[43] While its relatively low abundance may make Trp conjugation a less attractive approach for a universal method, it provides a better site selectivity. Tam and coworkers utilized the Picter-Spengler reaction for N-terminal Trp where in the acidic conditions and and in the presence of aldehyde, a condensation reaction occurs to form the 6-membered heterocycle.[44] The mechanism is also the suggested route for conjugation of terminal His and Trp in the previously discussed paper where Francis and coworkers utilized PLP to label the N-terminus of proteins.[21b]

A more accessible method for Trp conjugation is using a pathway catalyzed by metallocarbenoids.[45] Lately, it has been shown that N-(tert-butyl)hydroxylamine allows for solvent accessible Trp modification over a wide range of different pH buffers.[46] This chemical modification strategy was implemented with coiled-coil peptides where assembly of coiled-coil motifs is used to promote a proximity induced, dirhodium catalyzed modification of aromatic side-chains in Trp, Tyr and for the first time Phe.[47]

2.2.3 Chemoselective synthetic approaches using non-natural amino acids

While there are many elegant strategies to chemoselectively modify the 20 natural amino acids, specific conjugation is difficult because each amino acid generally appears more than once in a protein sequence. Thus, conjugation becomes a balancing act between stoichiometry, solvent accessibility, specificity, etc. An alternative method (that was first reported almost 60 years ago!) utilizes genetically modified E. coli bacteria to insert non-canonical amino acid analogs. In the seminal publication, Cowie and Cohen reported a substitution of methionine with a selenomethionene in bacterial protein synthesis. Substitution is achieved in minimal media using a Met auxotroph, or deficient, bacteria.[48] This first expansion of the genetic code has significantly aided studying protein structure/function relationships through crystallography, and was the first publication that pioneered the idea of inserting unnatural or non-canonical amino acids (ncAAs) into protein structure. In the past two decades or so the Schultz, Tirrell and Bertozzi's groups, among others, have made significant contributions to expand the library of ncAAs that are available. Schultz's group has even engineered a bacterium that can express a '21st' amino acid. While we will specifically address a few of the prominent chemistries that are used or hold great promise for protein-polymer conjugation, we would like to refer interested readers to some great reviews on the topic for a more exhaustive look.[49]

2.2.3.1 Conjugation reactions with azides

2.2.3.1.1 Azide/Alkyne 'Click' Chemisty

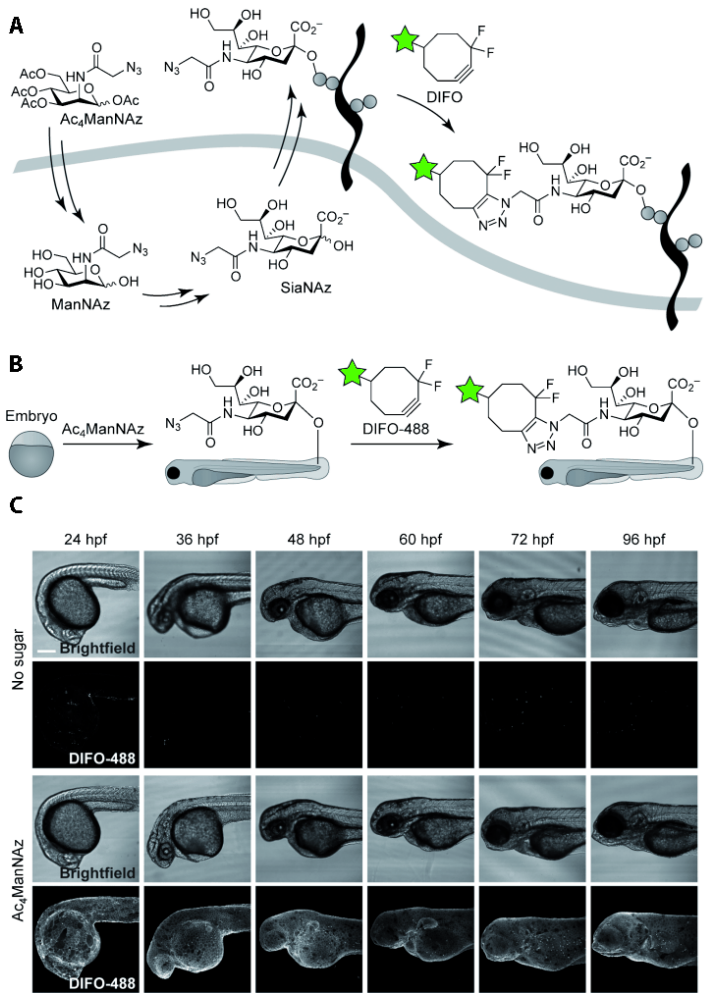


Figure 2.4. Metabolic labeling strategy and visualization of sialylated glycans in developing zebrafish. A) Scheme for metabolic labeling of sialylated glycans. Cell-permeable, peracetylated N-azidoacetylmannosamine (Ac4ManNAz) enters the cell, and once its acetyl groups are cleaved, ManNAz traverses the steps of the sialic acid biosynthetic pathway and is converted to azido sialic acid (SiaNAz). After conversion to the nucleotide sugar CMP-SiaNAz, SiaNAz is incorporated into cell-surface glycoconjugates by sialyltransferase enzymes. At the cell surface, SiaNAz is visualized by reaction with fluorophore-conjugated difluorinated cyclooctyne (DIFO) reagents. B) Scheme for visualizing sialylated glycans in zebrafish embryos. Embryos are incubated in solution containing Ac4ManNAz, which provides metabolic labeling of cell-surface sialylated glycans over the course of development. In a second step, reaction with DIFO-488 enables visualization of the labeled glycans. C) Treatment with Ac4ManNAz followed by reaction with DIFO-488 provides labeling of sialic acids in the enveloping layer during the first four days of development. Embryos were incubated in a solution containing Ac4ManNAz (or no sugar) beginning at 4 hpf and then were reacted with DIFO-488 at the indicated times. Shown are Z projection images of DIFO-488 fluorescence and corresponding brightfield images of embryos, viewed laterally. Scale bar: 200 μm. Reprinted with permission from Ref. 85. Copyright 2012 John Wiley & Sons, Inc.

The Sharpless group coined the term 'Click Chemistry' in 2001 to describe their highly efficient modification of the 1,3-dipolar Huigsen cycloaddition, the copper-catalyzed azide/alkyne cycloaddition (CuAAC). The name 'Click' seemingly refers to the ease with which the reaction progresses and its tolerance to multiple reaction environments. 'Click' has been extensively adopted for the bio-conjugation of alkyne and azide functionalized building blocks. Although the click reaction could be accomplished in an efficient manner, there are great challenges in introducing unnatural functionality into biomolecules. Van Hest and coworkers synthesized clickable elastin-like peptides by replacing methionine with unnatural amino-acids containing azido or propargyl groups. [50] In another study, they designed a pathway to convert amine groups into azide moieties using a diazotransfer agent.[51] While these reactions are amazingly tolerant to a multitude of reaction conditions, the presence of metal catalysts always raises concerns when considering biomedical applications like in vivo imaging. This concern motivated a seminal work from the Bertozzi group that introduced a copper-free method utilizing the strain in a cyclooctyne to promote the azide-alkyne 1,3-dipolar cycloaddition. This strain-promoted 'click' reaction was used to immobilize diagnostic peptides onto polymer coatings for an enzyme-linked immunosorbent assay (ELISA). [52] The method has since been extended to other strained rings such as oxanorbornadienes[53] and dibenzocyclooctynes.[54] The technology has emerged as an invaluable tool for in vivo imaging as in the monitoring the development of the Zebrafish neural networks by fluorescently labeling the sialylated glycans (Figure 2.4).[55]

2.2.3.1.2 Staudinger Ligation

Bertozzi and coworkers additionally revitalized the Staudinger reaction, where an amide forms between an azide and phosphine, to functionalize azide modified glycan on cell surfaces with a fluorogenic phosphine reagent.[56] While this reaction was specific, near quantitative, and produced no reactive byproducts, its aza-ylide product would rapidly hydrolyze to a primary amine and a phosphine oxide (Figure 2.5A). In the original work this hydrolysis was mitigated by attaching an intramolecular electrophilic-trap (Figure 2.5B) which created a stable amide bond by cyclizing the

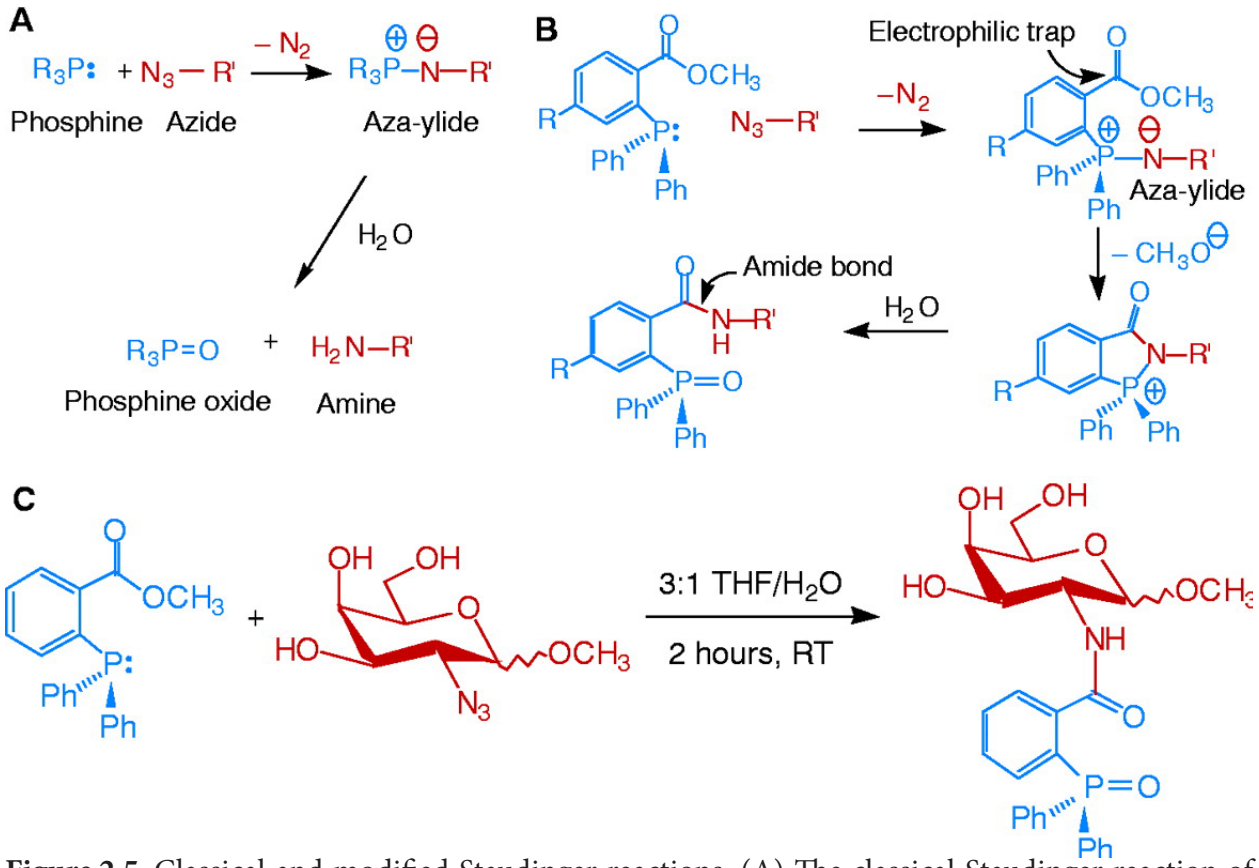


Figure 2.5. Classical and modified Staudinger reactions. (A) The classical Staudinger reaction of phosphines and azides. Hydrolysis of the aza-ylide produces an amine and a phosphine oxide. (B) A modified Staudinger reaction that produces a stable covalent adduct by amide bond formation, even in the presence of water as solvent. (C) A model reaction that produces a single amide-linked product. The limited water solubility of the phosphine necessitated an organic cosolvent (THF). Reprinted with permission from Ref. [86]

aza-ylide. This reactions is referred to as the 'nontraceless' Staudinger Ligation. The follow up on 'nontraceless' Staudinger Ligation developed in parallel by the Bertozzi and Raines groups, each with a slightly different approach, is the 'traceless' Staudinger Ligation. The 'traceless' Staudinger Ligation eliminates the phosphine oxide from the final product while creating a stable amide. Raines method accomplished this by creating a phosphinothiol that reacts with the azide to form iminophosphorane. The nucleophilic nitrogen in this aza-ylide attacks the carbonyl group cleaving the thioester. Final hydrolysis of the rearranged product produces a native amide and liberates the phosphine oxide.[57] Among the suitable phosphines the Raines ligation reagent – dipenylphisphinemetanethiol - has found the widespread application due to it's excellent reactivity profile.

2.2.4 Oxime/hydrazone chemistry

Incorporation of aminooxy or carbonyl groups into the protein offers an alternate conjugation route based on the oxime/hydrazone formation. Original work by the Haddleton and Maynard groups demonstrated the utility of aldehyde or aminooxy functionalized ATRP initiators for the polymerization of methacrylate monomer.[58], [59] Analogous methodology was used to prepare protected aminooxy CTA for the RAFT polymerization of NIPAM. After deprotection in TFA, the α-terminal aminooxy PNIPAAM was conjugated to levulinic functionalized BSA. The resulting bioconjugate exhibited thermo-responsive properties conferred by the PNIPAAM component. [60]

2.2.5 Enzyme-mediated conjugation

There is an entire host of different enzymatic reactions that are carried out to functionalize proteins. Sortases, transferases, transglutaminase and ligases have all been utilized to modify proteins for conjugation or simply carry out the conjugation themselves (e.g. ligases). To tie-in a previous example, Ting's group utilized a biotin ligase to incorporate an azide into a biotin acceptor protein that allowed then to FLAG-tag it through a Staudinger Ligation.[61] The Distefano group showed the ability to carry out a one-pot site-specific labeling of two proteins, Green Fluorescent

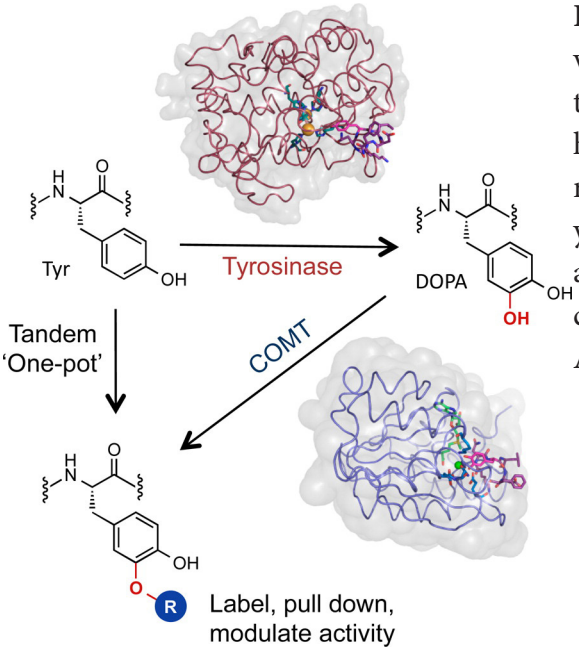


Figure 2.6. A general strategy for regioselective derivatization of tyrosine residues in peptides and proteins by a tandem enzymatic alkoxylation. Tyrosinase hydroxylates tyrosine to give an intermediate l-DOPA residue, which is then alkylated by catechol-O-methyltransferase (COMT) with AdoMet or AdoMet analogues. Reproduced from Ref. [95] under ACS creative commons (CC-BY) license. Published by American Chemical Society.

Protein (GFP) and Red Fluorescent Protein (RFP), using two sequence specific prenyltransferases. Additionally GFP and RFP were coupled in the same one-pot reaction by labeling each with a azide-containing and alkyne-containing isoprenoid analogue respectively.[62] More recently, a two enzyme cascade for hydrolysis and subsequent alkylation of Tyr was reported. Cascade consists of tyrosinase and catechol-O-methyltransferase (COMT) that hydrolyzes tyrosine into L-DOPA and subsequently alkylates the tyrosine with analogues of S-adenosyl methionine (AdoMet), Figure 2.6.[63] While this reaction occurs faster at the termini, it provides a good proof of principle and green alternative to some of the tyrosine chemistries discussed above. We would like to point interested readers to a review written by the Distefano in 2013 in Bioconjugate Chemistry for an in-depth discussion of enzymatic strategies for bioconjugation.[64]

2.2.6 Noncovalent conjugation through protein-ligand interactions

Biotin, also known as vitamin H or B7, is a cofactor produced in the metabolism of fatty acids and amino acids. The complexation between biotin and avidin, with a dissociation constant of $K_d = 10^{15}$, represents the strongest noncovalent protein-ligand interaction.[16] A number of mono- or hetero-functional biotinylated CTAs have been synthesized via esterification of a carboxylic acid terminated RAFT agent with a biotinylated alcohol.[65], [66] Carbohydrates which play essential roles in many biological processes are known for binding lectins with high affinity and specificity through multivalent interactions.[14] In a related work, a series of amphiphilic glycolpolymers were synthesized for the selective reorganization of Concanavalin A lectin.[66] Folate, another biological ligand, has also been reported to selectively target human cancer cells. In a study by Sumerlin et al., PNIPAAM-poly(N,N-dimethylacrylamide) (PDMA) micelles were functionalized with folate for potential therapeutic and diagnostic applications.[67] Another highly specific method for protein-polymer conjugation utilizes the well-described complementarity of DNA base-pairs. The Das lab utilized alkyne functionalized complementary DNA strands for direct 'click' conjugation to an azide functionalized green fluorescent protein (GFP) and poly(oligo(ethylene glycol)methyl ether methacrylate) (pOEOMA).[68] This method was further used to create clickable DNA-bottlebrush

polyer (DBBP) as nano-tag scaffolds. These scaffolds showed the ability to densely pack intercalating fluorescent dyes such as YOYO-1 (~1800/brush), as well concurrently incorporate secondary antibodies without hindering their binding capacity.[69] This versatile method can be adjusted to incorporate a number of different targeting moieties and dyes by simply altering the ratio of DNA sequences in the bottlebrush functionalization reaction. Overall, these non-covalent interactions present very specific and selective avenue for connecting two macromolecules.

2.3. **Rational design of conjugate architecture and topology**

The self-assembly behaviors of protein-polymer conjugates are closely associated with their specific architecture. In practice, the apparent starting point for controlling architecture, also topology, is to modify the number and position of reactive handles on either the polymer or the protein or both. While methods for controlling the topology of polymers are well established in the literature (multi-arm initiators or monomers, grafting strategies, etc.); controlling the number of reactive handles can be quite imprecise. Protein and peptide macromolecules on the other hand afford site-specific control over conjugation utilizing the available genetic engineering methods. While precise control over amino acid placement in protein and peptide materials is invaluable for controlling topology, it must be noted that there are possible issues with mutant stability and scalability in production.

The best control in synthetic polymerization comes from end-functionality derived from the polymerization; for example, creating a mono-functional PEG by using methanol as an initiator or creating the diol with ethylene oxide. Alternatively, end-group modification can be efficient by introducing a post-polymerization synthetic step, but only when working in excess. Some groups have had success producing hetero-di-functional materials, however separation of the homo-di-functional from the hetero-di-functional and non-functionalized polymer can prove to be quite problematic. Another method to create reactive handles on polymers is to have functional pendant groups in the backbone (i.e. grafting sites, Figures 2.1C and 2.7). Direct polymerization of pendant functional groups is very efficient but gives little control over the percent functionality.

Post-polymerization modification provides some control over the percent functionality and there are a large variety of functional group modifications, both covalent and non-covalent, temperature modifications, etc. that can be exploited to create a random distribution of modifications that can be used as reactive handles to create new conjugate materials.

Proteins on the other hand have an added advantage of unique, complex primary sequences that even the most complex copolymers do not. While there is a large variety of conjugation strategies (previously discussed) that take advantage of natural amino acids, these methods, more times than not, are only relatively chemo-specific rather than positionally specific. A few exceptions to this are the selective functionalization of solvent accessible disulfides, termini modifications and some enzymatic transformations. Conversely, site or residue-specific introduction of non-canonical amino acids offers precise control of reactive handle placement. This method has been used for direct polymerization from proteins to create cross-linked protein networks[70] and AB and ABA block co-polymers where non-natural amino acids have served as ATRP initiators[71] as well as di-

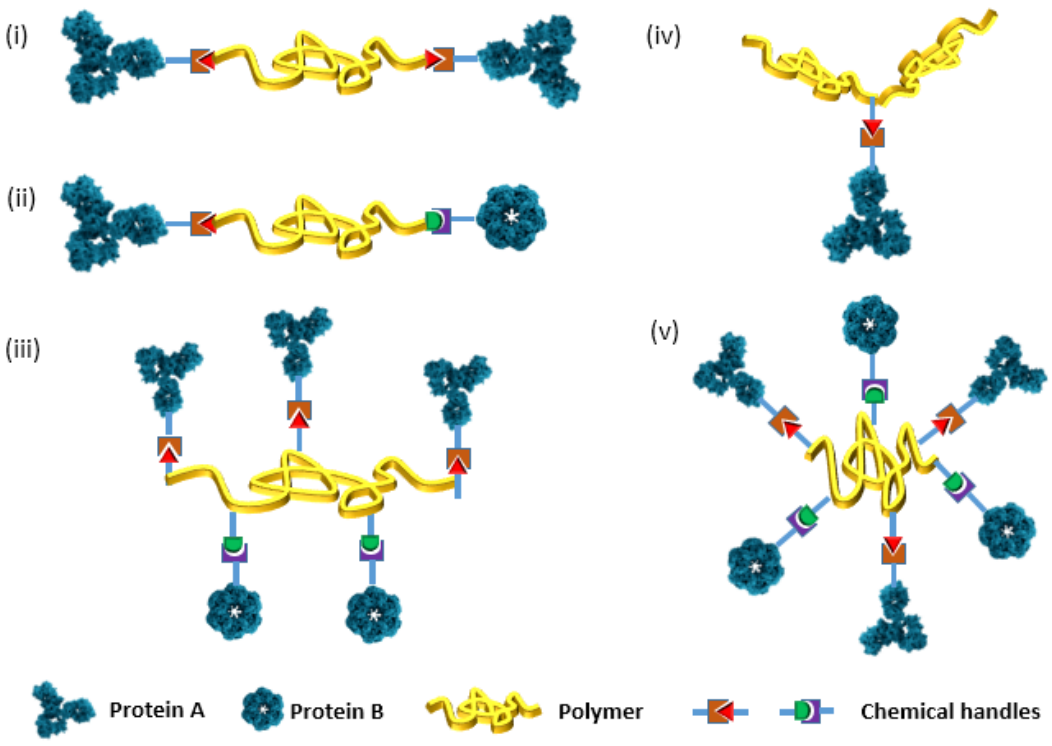


Figure 2.7. Schematic representation of the topologies of (i) homodimeric (ii) heterodimetric (iii) comb-like (iv)Y-junction (v) star-like protein-polymer conjugates.

rected placement of RAFT CTAs.[7]

Since the advent of CRP, a few novel architectures have been constructed to impart appealing attributes to polymer-based bioconjugates. For instance, hyper-branched or dendritic globular bioconjugates exhibit superior solubility as compared to their linear counterparts. In addition, multi-armed bioconjugate ligands are apt to show enhanced binding affinity for specific receptors.

2.3.1 Homodimeric and heterodimeric conjugate

Maynard et al. proposed an efficient route to prepare α, ω-dual functional oxime or PDS-CTAs which were exploited for the polymerization of various monomers with high functionality and low dispersity.[72] The same group also reported the synthesis of a heterodimeric bioconjugate where the central PNIPAAM block was flanked by heterogeneous proteins. To pursue site-specific orthogonal conjugation, biotin and protected-maleimide were incorporated into the telechelic polymer through sequential esterification and radical cross-coupling.[73] A similar strategy was used by Boyer et al. to prepare a series of α-azide, ω-PDS bis-functional polymers. Two consecutive coupling reactions were carried out to conjugate the functional polymer with pargyl-biotin and glutathione or BSA.[65]

2.3.2 Y- junction/branched conjugate

Poly(N-(2-hydroxylproyl) methacrylamide), PHPMA, is a hydrophilic, nonimmunogenic polymer that has been extensively studied as drug delivery carrier due to the prominent stealth effect and pronounced stability in the blood circulation. The Davis group synthesized a Y-junction PHPMA-BSA conjugate using RAFT polymerization techniques, in which mid-chain bis-trithiocarbonate CTA was prepared for the sulfhydryl-reactive coupling of the BSA. This straightforward methodology for the synthesis of branched mid-functionalized polymers for protein conjugation has the potential to enhance the properties of polymer-protein conjugates by improving biomolecule stability and enhancing circulation time. (Figure 2.8A).[74],[75]

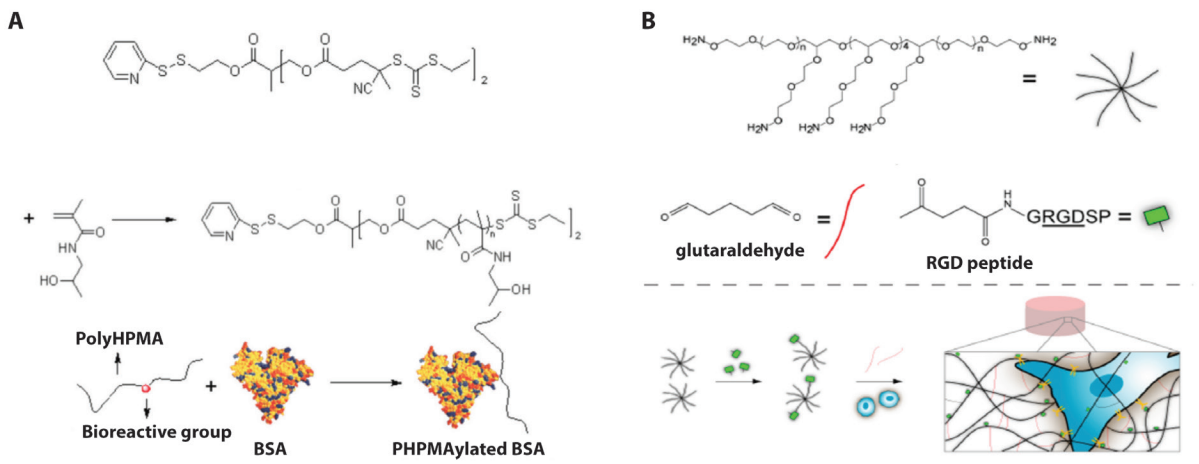


Figure 2.8. (a) Synthesis of branched PHPMA-BSA conjugate1, 2 and (b) encapsulation of MSC cells within RGD-functionalized oxime-cross-linked PEG hydrogels.. 2.8A adapted with permission from Ref. [107]. Copyright 2009 American Chemical Society. 2.8B adapted with permission from Ref. [110]. Copyright 2012 American Chemical Society.

2.3.3 Star-like multimeric conjugate

In nature, proteins such as enzymes and antibodies are prone to oligomerize spontaneously to achieve higher bioactivity. Maynard's group took advantage of the star-like PNIPAAM scaffold to produce multimeric protein-polymer conjugates for enhanced binding affinity. In this work, a tetrameric CTA was prepared followed by the RAFT polymerization of four-armed PNIPAAM. A radical coupling reaction was then conducted to attach maleimide onto each arm for the ensuing malemide-thiol conjugation.[76] In a recent study, 8-armed aminoxy PEG was used to ligate ketone-terminated RGD peptide via oxime chemistry (Figure 2.8B). The synthesized conjugate was further crosslinked by glutaraldehyde to generate a hydrogel with high cell viability and tunable mechanical properties.[77] Likewise, Grove et al. also provided a paradigm for the synthesis of stimuli-responsive smart gel in which modular designed repeat protein was non-covalently crosslinked with a star-like building block containing both PEG backbone and terminal peptide ligands. It has been further demonstrated that the encapsulated protein could be conveniently released under the high ionic strength condition and their rheological properties are well-applicable to potential biomedical applications.[78] Using the similar strategy of non-covalent cross-linking through specific protein-peptide ligand interactions, biomaterials for tissue regeneration.[79]

2.3.4 Comb-like conjugates

In addition to branching from a single node, polymers can impart a brush- or comb-like structure to the conjugate using grafting-to strategy to add a biomolecule the polymer backbone.[69, 80] Petkau-Milroy and coworkers recently used a grafting-to method using the bioorthogonal SNAP-tag to attach a yellow and cyan fluorescent protein.[81] This system incorporated the SNAP-active benzylguanine moiety onto a monovalent discotic supramolecular polymer. The dynamic backbone afforded control over the density of multiple incorporated proteins by simply changing the ratio of functionalized and non-functionalized backbone discotics in solution (Figure 2.9).

2.3.5 Controlling topology through protein design

While controlling topology in synthetic polymers holds advantages in scalability and sometimes simplicity, the ever growing number of simplified procedures and commercially available kits has made it easier than ever to produce sequence-controlled, genetically engineered proteins for a variety of different applications. With this in mind in vitro produced proteins hold key advantages over polymers in that they: 1) are monodisperse, 2) have significantly more complex structures/functions, 3) in many cases, can tolerate insertion of natural and non-natural moieties in a sequence-specific manner, which can be site-specific when the crystal structure is known. These advantages afford the ability to create a number of protein scaffolds diverse in structure and function.

2.3.5.1 Precision Polypeptide Copolymers

One of the key limitations of synthetic polymers is the detrimental affect that polydispersity, inherent to the polymerization process, has on the resulting materials properties (i.e. assembly, mechanical, optical, transport). Early work by the Whitesides group leveraged the monodisperse size of proteins to create sequence defined, single molecular weight polyamide backbones through chemical denaturation of commercially available proteins.[82] With the large variety of

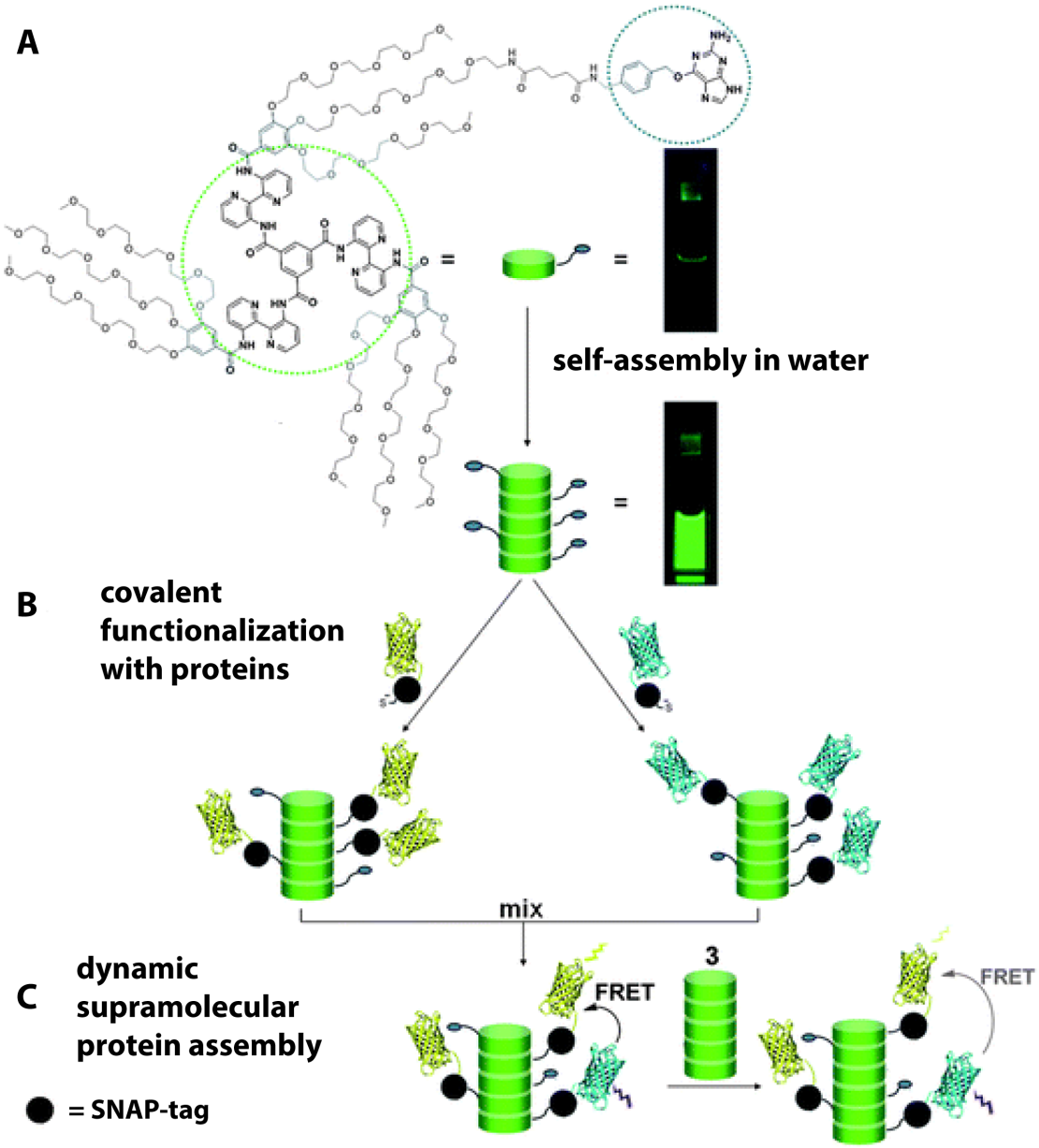


Figure 2.9. (a) Structure of the O-benzylguanine (BG) discotic 2 carrying a single moiety for conjugation to SNAP-tag fusion proteins. In water these discotics self-assemble to form auto-fluorescent columnar stacks, displaying moieties for conjugation at their periphery. The photographs show 10 mM solutions of 2 in CH2Cl2 (top) and in water (bottom) excited with UV light (lex . 350 nm). (b) Site-selective covalent functionalization of the supramolecular polymer with a SNAP-tag fused to cyan and yellow fluorescent protein (CFP and YFP). (c) Dynamic intermixing of supramolecular protein assemblies results in efficient Forster resonance energy transfer (FRET) between CFP and YFP. Intercalation of inert-discotic 3 allows tuning of the distance between the fluorescent proteins. Reproduced from Ref. [118] under Creative Commons Attribution 3.0 Unported License. Published by The Royal Society of Chemistry.

commercially available, inexpensive proteins this method can be applied to create a host of, what is termed, precision-polypeptide copolymers (PcPs), which incorporate a polyamide backbone functionalized with pendant polymer groups, Figure 2.10. These materials may be tailored for different applications based on proteins with the desired size, charge and amino acid content.[83] While useful, using wild-type PcPs lacks the ability to tune the frequency and distribution of reactive handles within the polyamide backbone. In search of better topological control over protein-polymer materials, many groups have turned to genetically encoded recombinant polypeptide materials that incorporate modular domains.

2.3.5.2 Modular Polypeptide Materials

Key examples of modular recombinant polypeptides are the elastin-like polypeptides (ELPs), introduced by the Ury group and further studied by Chilkoti and coworkers,[84] and the silk-elastin like polypeptides (SELPs), extensively studies by the Kaplan group.[85] ELPs are a family of polypeptides which exhibit a lower critical transition temperature (LCST) and consist of pentapeptide repeats $(VPGXG)_n$ where X can be any amino acid other than proline.[84a, b] SELPs incorporate additional silk domains (GAGAGS) which impart tightly packed, insoluble β-sheets that contribute to phase separation and mechanical stability.[85a] Overall, these materials represent a larger class of materials that which use recombinant DNA to create single block co-polymer biomolecules from a[86] toolbox of modular peptide motifs (individual blocks) which impart useable material properties like: an LCST, structure, stimuli responsiveness and crosslinking.[87]

One modular motif that has proved very useful in tailoring backbone architecture is the SpyCatcher/SpyTag protein-peptide pair. In the search to create a more robust covalent protein tag, Howarth and coworkers found the Streptococcus pyogenes (Spy) CnaB2 protein, one know to form isopeptides spontaneously.[88] They split SpyCnaB2 into a large N-terminal domain (Catcher) and a small C-terminal (Tag) b-strand, splitting the autocatalytic triad of $Lys^{31}/Asp^{117}/Glu^{77}$ (Figure 2.11A), all key for isopeptide formation. These fragmented domains, SpyCatcher and SpyTag, were able to reassemble in solution to create the isopeptide between corresponding Lys^{31} and Asp^{117}. This robust reaction occurs across a variety of reaction conditions.[89]

Arnold *et al.* developed the strategy for recombinant incorporation within modular ELP blocks to create a large variety of cleavable backbone architectures. Topologies they can produce are cyclic, tadpole, multi-arm, H-shaped, all of which afford the same control as the previously discussed polymeric systems (Figure 2.11C).[90] More recently Wang and Zhang incorporated a module containing the p53 dimerization domain, which promotes intertwining ELP chains thus

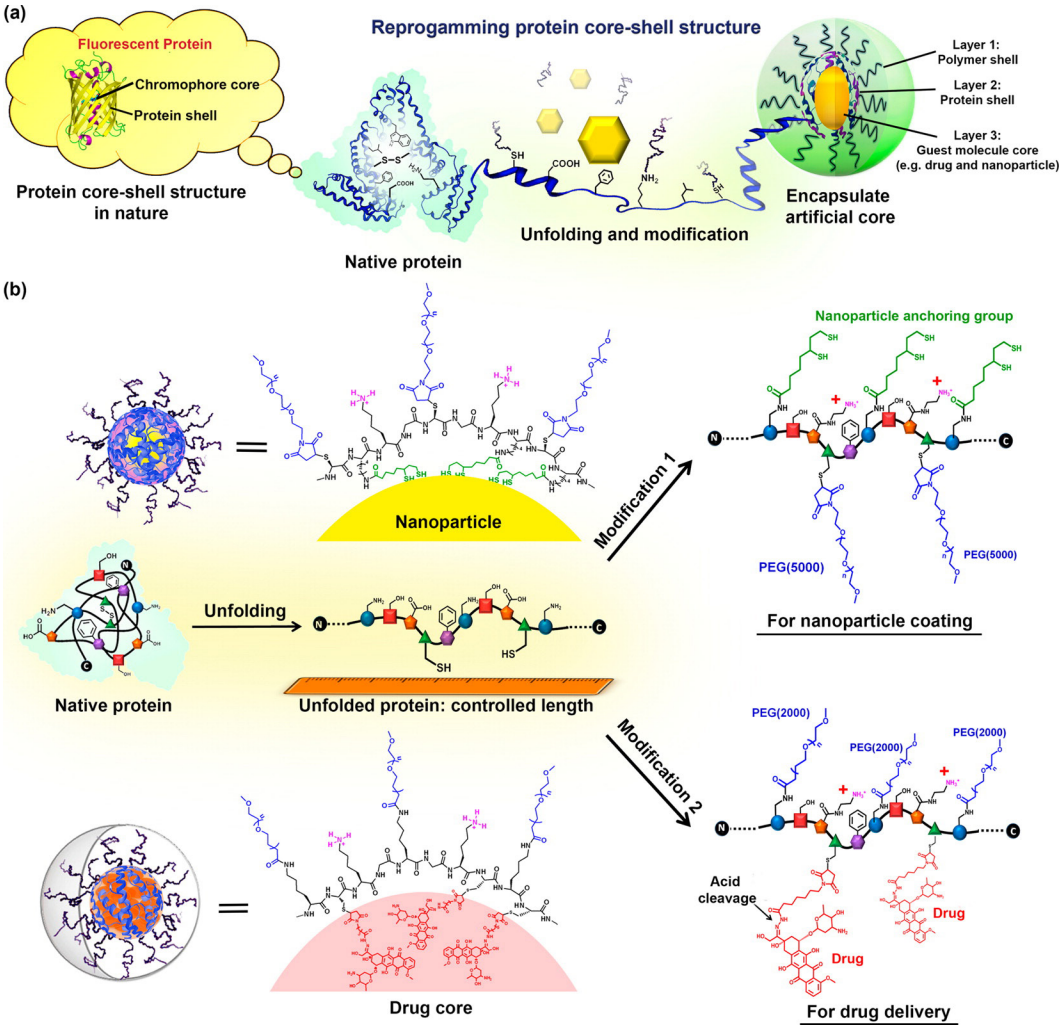


Figure 2.10. (a) Chemically induced unfolding of a globular protein into denatured protein–PEO hybrid and backfolding of globular structure induced by nanoparticles and hydrophobic guest molecules. The term of "backfolding" does not imply refolding of the secondary or tertiary structure of the native protein, but it refers to the spontaneous assembly of the polypeptide chain into a globular architecture. (b) Chemical programming of denatured protein polypeptide using orthogonal modifications on different sets of amino acids. Adapted with permission from Ref. 122. Copyright 2014 American Chemical Society.

giving a method to produce polypeptide cantenanes and psuedorotaxanes.[91]

In addition to SpyCatcher/SpyTag there are a number of technologies that have been used to create varied protein topologies: split-intein,[92] sortase-mediated[93] and native chemical ligation. [94] Again, these systems hold advantages over polymeric architectures because of their monodispersity and sequence specificity. This readily available topological diversity combined with the advent of efficient lab-scale bioprocessing,[95] scalable combinatorial gene production,[96] new cloning techniques[86] and tunable material properties[97] all have positioned these modular proteins materials as competitive alternatives to synthetic polymers.

2.3.5.3 Native Protein Copolymer Conjugates

The benefit of using native proteins in conjugate materials is the significantly more complex structure/function relationships one can exploit. For wild-type proteins the topology is dictated by the specificity of the conjugation chemistry and solvent accessibility of the target side-chains. While effective, this often lends poor control over the density and position of polymer placement.

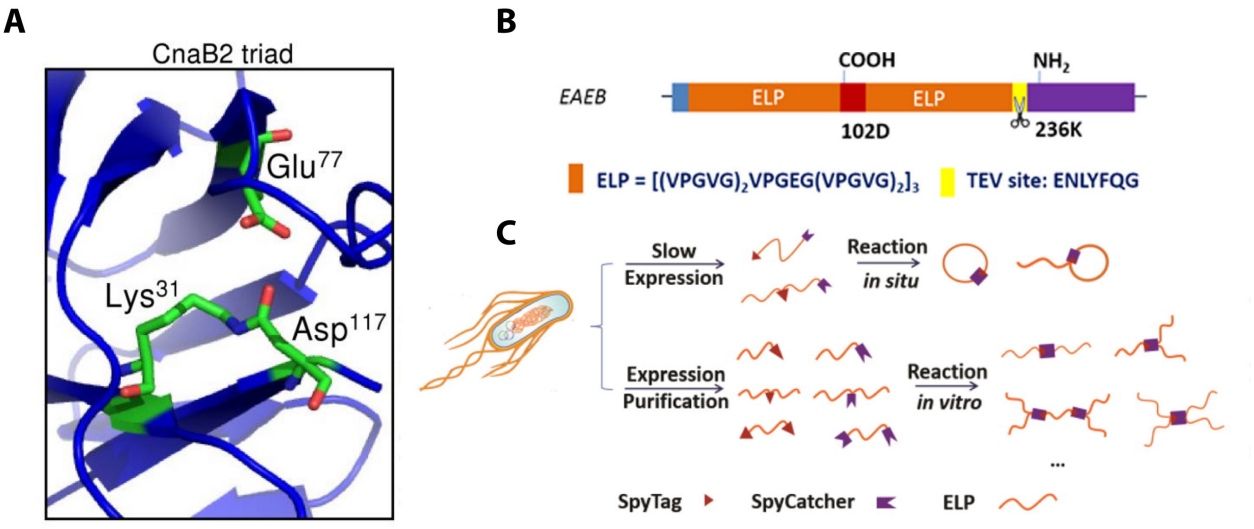


Figure 2.11. (a) Crystal structure of the autocatalytic triad of SpyCnaB2 which forms an isopeptide between Lys31 and Asp117. (b) A representative gene structure that forms a tadpole architecture. The Asp and Lys are noted below. (c) Using modular peptide blocks, like ELPs allows for facile design of multiple architectures including cyclic, tadpole, multi-arm and H-shaped backbones. (a adapted with permission from Ref. [134]. b and c adapted with permission from Ref. [135]. Copyright 2013 American Chemical Society)

We discussed strategies that afford site-specific control in wild-type proteins, such as N-terminal modification[20-21, 44] and disulfide bridging,[29-30, 31b, 98] but these severely limit the available topologies.

One method to supplement the available conjugation handles in native proteins is through genetic incorporation of new sites for grafting-to and grafting-from, both canonical and non-canonical. There are a variety of techniques for non-cannonical amino acid incorporation such as nonsense suppression, where tRNA gene is mutated where the anticodon is mutated to recognize a stop codon (nonsense) thus making it a sense codon. Amber (TAG) and ochre (TAA) are the most common of these. We would like to direct interested readers to further explore these topics by reading Chin's review in Annual Reviews Biochemistry, which contains a comprehensive overview of these techniques as well as an additional discussion of the expanded quadruplet codon.[99]

While current methods have significantly reduced the time and effort into genetically modifying proteins, maintaining the structure and function may not be trivial. If the crystal structure is known, it makes structural modifications easier to predict but not perfect. Globular proteins present a problem because of long-range interactions between distant residues in the primary sequence, which make it difficult to ascertain their global affect on the protein's structure. For this reason, many groups have studied a class of polymer-like proteins, called repeat proteins. These proteins comprise small repeating motifs with secondary structural elements, only local contacts and relatively low conservation. Regan *et al.* used a concensus sequence tetratricopeptide repeat (CTPR) protein to create stimuli-responsive hydrogels, which incorporated a modular peptide-binding pocket into the protein with control over the spacing and pitch of binding sites.[4a, 78b] TPRs have been further modified to incorporate terminal cysteines and polymerized into fibers.[100]

Combining the methods above with self-assembled protein structures, both supremolecular and de novo structures, can further increase the topological complexity of conjugate materials. We will cover protein directed self-assembly later on, but we would be remiss to not point out the implications for affecting topology. While there are only a few examples in the literature, tobacco

mosaic virus,[101] viral capsids,[102] and vault nanocapsules,[103] there are many promising protein assemblies that could prove useful in future material design. Some promising candidates are the proliferating cell nuclear antigen (PCNA), which has been used to scaffold enzyme cascades,[104] designed protein cages,[105] crysalins, which form 1D, 2D and 3D assemblies based on the rotational symmetry of assembling units,[106] template designed motifs[107] and de novo proteins[108]

2.4. Self-assembly of protein-polymer conjugates

Self-assembly has emerged as a powerful and cost-efficient approach to generate well-defined nanostructures containing both polymer and protein/peptide building blocks. The self-assembled structures over multiple length scales may find potential applications in gas separation, high-density media storage, drug delivery and biosensing.[109] In order to achieve well-aligned structures with tunable domain sizes, it is necessary to gain a deeper understanding of the intra/intermolecular interactions among all the components. In the following section, recent advances in the patterning of peptide/protein-polymer conjugate will be briefly discussed, providing insight into the influence of key factors upon the phase behavior and morphology of the self-assembled structure.

2.4.1 Block-co-Polymer (BCP) Template-assisted patterning of protein arrays

Although some proteins/peptides have been discovered to spontaneously assemble into fibrils, capsules or rod-like structures, a large number of proteins/peptides are not inclined to form hierarchical structures over a macroscopic scale. Hence, two bottom-up techniques have been developed so far for non-lithographic patterning of proteins/peptides.

In the template-assisted approach, highly ordered substrates, in particular strongly segregated BCPs are employed to selectively partition biomolecules into micro-phase separated domains. Patterning biomolecules has been shown to enhance activity and stability exhibiting fine control over the micro-environment of the protein in the solid state. Early research done by Kumar et al. demonstrated the utility of polystyrene-b-poly(methyl methacrylate) (PS-b-PMMA) and PS-

b-poly(vinyl pyridine) (PVP) BCPs for the deposition of well-organized protein arrays.[110] For instance, spin-coating a toluene solution of PS-b-PMMA onto a silicon wafer generated a thin-film with upright cylindrical PS micro-domains. This micro-phase separation phenomenon could be reasonably explained by the preferential wetting of the PMMA on the Si substrate. In the following step, IgG protein was adsorbed onto the PS domain due to the favorable hydrophobic interaction between protein and PS.[110] The author further expanded their system to template-patterned functional enzymes where not only the precise spatial control was attained, but the conformational integrity and bioactivity were also maintained.[111],[112] While non-specific physiosorption works, the main concern for using this method is lack of orientational control since the adsorption is stochastic in nature. This leads to a distribution of orientations where the enzyme active sight may be blocked reducing the overall efficiency. Many groups have adopted the strategy of covalently or non-covalently conjugating proteins using a reactive handle at a designed site, especially helpful in the case of enzymes where one can select a position away from the active site, ensuring an outward orientation. Kadir *et al.* created a nitrolotriacetic acid functionalized polystyrene in order to capture the His_6-tag of GFP. While they didn't use these aggregates to directly purify the protein, it might be possible depending on their size. Instead of using physiosorption, covalent linkage of an azide tagged protein to an alkyne functionalized PS-b-PHMA BCP was presented in a recent study. The hexgonally packed PS cylindrical microphase could accommodate individual protein molecules with appropriate domain sizes, which was confirmed by both atomic force microscopy (AFM) and immunoassays.[113]

Electrostatic interactions are also employed for building up bioconjugates with highly ordered structures. The zwitterionic nature of proteins permits the convenient control of their surface charge by simply adjusting pH. For example, RAFT was employed to synthesize thermo and pH sensitive PNIPAAM-b-PDMAEA, in which the PDMAEA block could conjugate with a fluorescent protein, mCherry, by complexation in a pH dependent manner. UV-visible spectroscopy (UV-vis), Dynamic Light Scattering (DLS) and Zeta potential measurements illustrated the reversible protein segregation driven by the electrostatic interaction.[114]

An original method based on the electrochemical redox reaction was described by Schwarzacher and coworkers in pursuit of controllable adsorption of ferritin onto the organometallic BCP substrate. In this work, PS-b-PFS thin films were chosen as nanopatterned templates due to the redox activity of PFS block in the bulk state. To suppress nonspecific protein adsorption, the pH value was adjusted above the isoelectric point of ferritin. Selective adsorption of protein on vertically aligned cylindrical PFS domains was triggered in-situ by applying positive potentials.[115]

2.4.2 Direct self-assembly into well-defined nanostructures

2.4.2.1 Self-assembly in solution

Over the past decade, great efforts have been devoted to the supramolecular assembly of protein/peptide-based BCPs in solution. Pioneering work done by Nolte's group first demonstrated the self-assembly behavior of giant amphiphiles, namely protein-polymer conjugates in aqueous solution.[116],[117],[118] In one study, horseradish peroxidase (HRP) or myoglobin (Mb) enzyme was covalently attached onto the PS block spanned by a hydrophilic spacer. After incubating function-

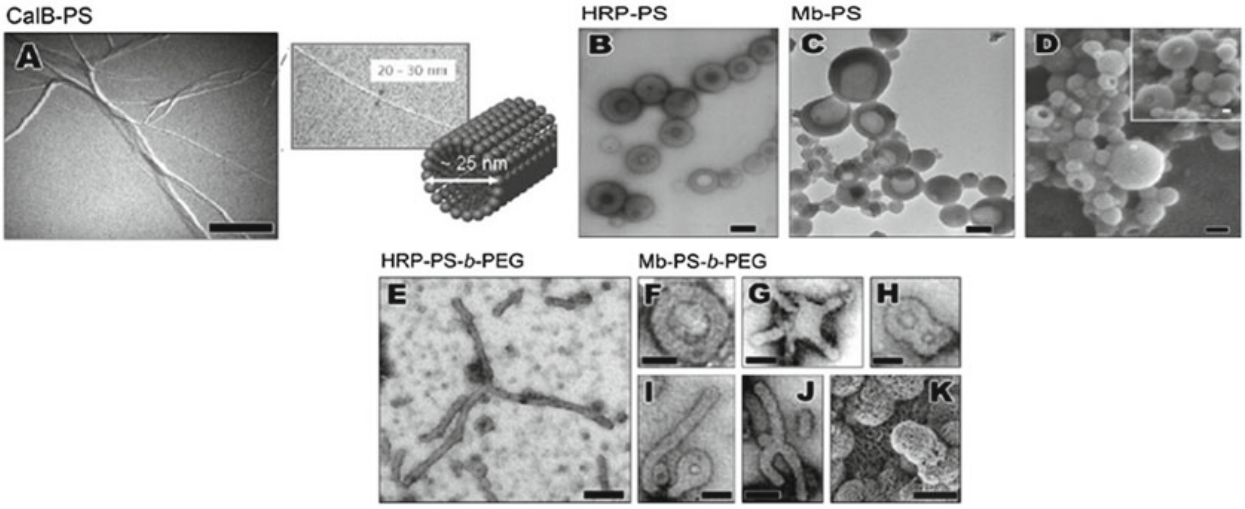


Figure 2.12. TEM and SEM micrographs of self-assembled protein-polymer giant amphiphiles in aqueous media. (A) Fibers of CalB–PS40 (B) Vesicles of HRP–PS90 (C, D) Spherical aggregates of Mb–PS90 (E) Assemblies of HRP–PS144-b-PEG113 (F–J) Unusual architectures of Mb–PS144-b-PEG113, K) Lamellae-containing spheres of Mb–PS48-b-PEG113. Bars represent 500nm (A, K), 200nm (B, E), 100nm (C, D, F–I). Figure 2.12A and 2.12E-K adapted with permission from Ref. [175] & [176]. Copyright 2002 & 2007 American Chemical Society. Figure 2.12B-D adapted with permission from Ref. [174]. Copyright 2002 John Wiley & Sons, Inc.

alized-PS with the enzyme for a certain time, vesicular aggregates were observed using Transmission Electron Microscopy (TEM) (Figure 2.12B, C, D).[116] An analogous giant amphiphile was synthesized through the conjugation between maleimide-functionalized PS and cysteine-terminated Lipse B enzyme. It was noted that the slow addition of water into the THF solution of the bioconjugate gave rise to nanofibrils, which were made up of monodispersed cylindrical micellar bundles (Figure 2.12A). Unfortunately, the self-assembly process was accompanied by a dramatic loss in bioactivity, which was probably caused by the conformational change in the presence of hydrophobic PS.[117] In general, the amphiphilic protein-polymer conjugate exhibits very similar supramolecular morphologies as conventional amphiphilic BCP in solution. Interestingly, a variety of unique morphologies including toroids, octopus and figure eight were also identified for self-assembly of ABC triblock copolymer synthesized through the reconstitution of HRP or Mb enzyme with heme-PS-b-PEG (Figure 2.12E-K).[118]

2.4.2.2 Self-assembly into well-defined solid-state nanostructures

Despite the fact that template-assisted self-assembly is able to regulate the nanoscale alignment of proteins/peptides with high precision, it is generally restricted to 2D patterning due to a lack of control over the protein orientations. Unlike the coil-coil BCP system where the phase separation could be precisely predicted relying on the classical Flory-Huggins treatment, the rod-coil BCP can self-assemble into intriguing nanophase separated domains such as tetragonally perforated lamellae (TPL), smectic C and Fddd.[119] The introduction of a rigid block will cause geometrical asymmetry and meanwhile appreciably influence the phase behavior through the additional interactions between the rod segments during the rod-rod packing process. As a consequence, the rod-coil BCPs possess reinforced phase-segregation abilities which are attributed to the preferential alignment of rod block components in the constrained nanospace. Conversely, the confinement effect of phase-separation is postulated to stabilize the ordered phase of the rod block and expedite the formation of ordered liquid crystalline structures. Thus, the ultimate morphology of rod-coil BCPs are determined by the Flory parameter, the geometry factor and the conformation of the rod-shaped

segment. In general, polypeptide-based BCPs have been examined to show a strong preference for lamellar structures with alternating polymer and peptide sheets. For instance, poly(Z-L-lysine)-PS BCPs were found to exhibit hexagonal-in-zigzag lamellar morphologies due to the fact that the α helical peptide acted as a rod-like mesogen. The α helical chains were arranged in a periodic hexagonal array while the zigzag structure was formed when the planar lamellae were disrupted by the kinks of hexagonally packed helical peptide components.[120]

Recently, Olsen's group published a series of papers on the self-assembly of BCP containing mCherry protein and PNIPAAM.[121],[122],[123] The globular mCherry was mutated to contain one terminal cysteine for coupling with a low PDI α-maleimide PNIPAAM. The bulk self-assembly was achieved by solvent evaporation under protein-selective ($T=40^{o}C > LCST_{PNIPAAM}$) or nonselective conditions ($T=25\,^{o}C$). Small-angle X-ray Scattering (SAXS) and TEM measurements revealed that hexagonally perforated lamellae (HPL) were formed under the former condition due to the presence of kinetically trapped micelle phases. In the latter case, a similar metastable HPL morphology was observed with improved ordering.[121] In a subsequent research, the microphase separation was also examined by solvent evaporation using polymer-favorable conditions ($pH=7.5>PI_{mCherry}$) (Figure 2.13A). It was concluded that the PNIPAAM component dominated the phase behavior of the BCP regardless of the composition of the conjugate. By varying the volume fraction of mCherry, hexagonal cylinders, HPL, lamellae, or hexagonal phases were obtained, resembling the morphologies formed by phase segregation of conventional coil-coil BCPs (Figure 2.13B).[122]

For BCPs, thermal annealing is generally performed above T_g to facilitate the evolution of an equilibrium morphology. In the case of protein-polymer conjugates, solvent annealing was preferentially adopted due to the vulnerability of proteins at elevated temperature. It was found that annealing in a protein-selective solvent is not advantageous for structural rearrangement, whereas annealing in nonselective solvents promoted nanostructure evolution with improved ordering. Moreover, Olsen et al. investigated the order-order transition (OOTs) and ordered-disordered transition of PNIPAAM-mCherry conjugate in concentrated solutions, and in turn develop a phase diagram to elucidate the phase behavior as a function of temperature and concentration (Figure

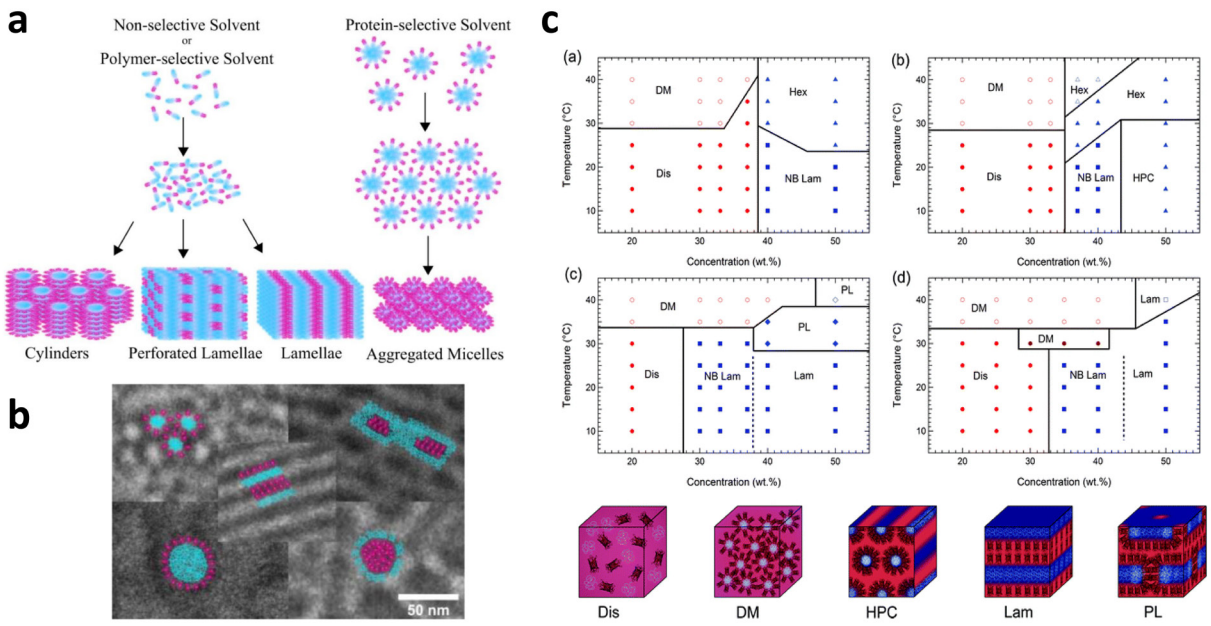


Figure 2.13. (A) Schematic illustration of two pathways toward the self-assembly (B) TEM images of self-assembly morphologies of mCherry–PNIPAAM (C) and phase diagrams of (a) mChP8 (b) mChP17 (c) mChP30 (d) mChP57 as a function of temperature and concentration. The various phases are denoted as disordered (Dis), disordered micellar (DM), nonbirefringent lamellar (NB Lam), lamellar (Lam), nonbirefringent hexagonal (Hex), hexagonally packed cylinders (HPC), and perforated lamellar (PL). Open symbols represent regions where macrophase separation between a conjugate rich ordered phase and a water-rich phase are observed. Figure 2.13A and 2.13 B adapted with permission from Ref. [179] & [180] respectively. Copyright 2011 & 2012 American Chemical Society. Figure 2.13C adapted with permission from Ref. [174]. Copyright 2012 The Royal Society of Chemistry.

2.13C).[123] In summary, the self-assembly of the protein-polymer conjugate in the solid state involves a complex thermodynamic and kinetic process, which can be finely tailored by the composition as well as diverse aspects during the casting, evaporation and annealing process.

2.4.3 Protein-driven supramolecular self-assembly

The previously discussed conjugate systems rely on phase separation between polymeric and protein blocks. These types of materials heavily populate the literature due to the relatively easy design associated with simplistic structures such as micellar, hexagonal and lamellar phases. Their advantage is that phase separation has been well characterized for a variety of systems and

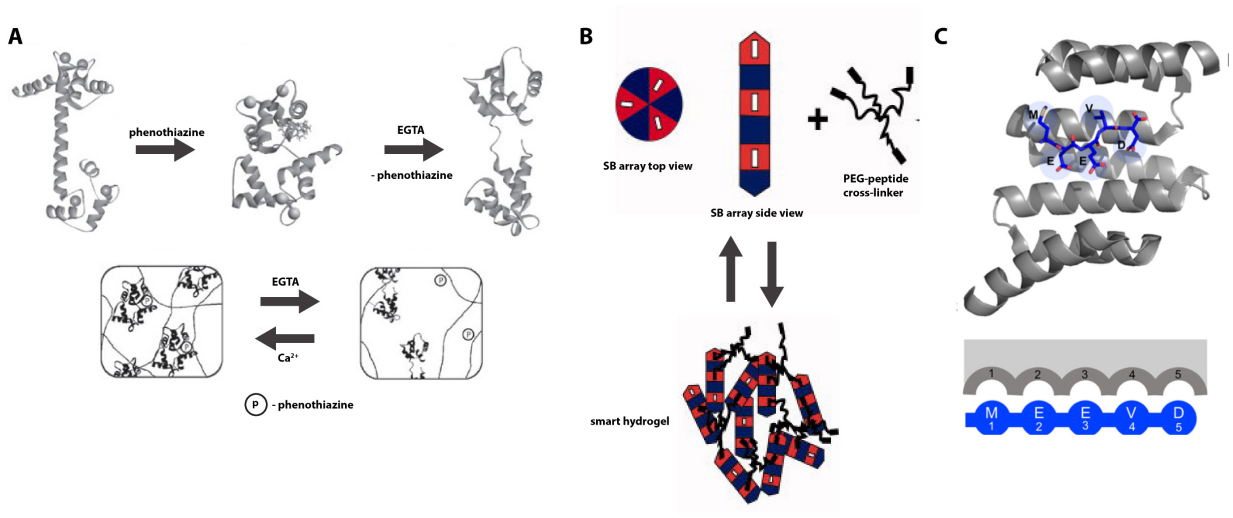


Figure 2.14. Examples of two-components hydrogels where assembly is dictated by protein interactions. A) Conformational changes in calmodulin in response to EGTA (swell) and Ca2+ (de-swell) cause modulate the hydrogels size and mechanical properties. B) Peptide binding domains incorporated into CTPR allow to tunable mechanical properties by changing the spacing and orientation of DESVD peptide binding pockets around the protein. C) Further work expanded the available peptides that can bind the modified TPR binding motif. . Figure 2.14A adapted with permission from Ref. [194]. Copyright 2005 Nature Publishing Group. Figure 2.14B adapted with permission from Ref. [112]. Copyright 2012 John Wiley & Sons, Inc. Figure 2.14C adapted with permission from Ref. [196]. Copyright 2015 American Chemical Society.

is relatively predictable. The same cannot be said for designer protein-based assembly, where we have a limited understanding of complex protein folds. Although, much work is being done on de novo protein sequences,[108] and designed supramolecular assemblies.[105-107, 124] Many self-assembled bio-hybrids incorporate small secondary structure elements attached to polymers to drive assembly in hydrogels. There are a multitude of examples for ELP/SELP,[125] beta-sheet[126] and coiled-coil[127] cross-linked hydrogels, along with many good reviews that cover a variety of others.[128]

Another subclass of protein-linked hydrogels, incorporate proteins that undergo conformational changes in response to a specific ligand. Figure 2.14A shows a key example where Daunert *et al.* polymerized a phenothiazine functionalized, acrylamide (AAm), and N,N'-methylenebis(acrylamide) (MBAA) copolymer off of the C-terminus of calmodulin (CaM). This hydrogel would reversibly swell/de-swell in response to Ca^{2+} in response to a two-step conformational change of

CaM. When Ca^{2+} was present the phenothiazine could bind, creating a non-covalent cross-link within the gel. Removing Ca^{2+} had two effects on the protein, first releasing the phenothiazine cross-link and then causing the conformational change in CaM that allowed further water uptake. [129] Daunert *et al.* also introduced a protein-binding responsive hydrogel that would actuate in response to binding glucose. This was accomplished by incorporating the accordion-like glucose/galactose binding protein (GBP) that collapses when bound to glucose and opens when free.[130]

Figure 2.14B shows a modular genetically engineered hydrogel system based around the consensus tetratricopeptide repeat (CTPR) protein. Here, Regan and coworkers spaced a series of three tandem repeats peptide-binding pockets into CTPR18. CTPRs superhelical structure allowed for easy control over the 3D cross-linking of the hydrogel, by spacing cross-link sites length wise as well as along the screw axis. The authors were able to tune the mechanical properties by altering the topology of the peptide functionalized PEG cross-linker.[4a, 78b] In a subsequent work Regan *et al.* modified the peptide-binding pocket to create three non-natural protein/peptide (Tetratricopeptide Repeat Affinity Protein, TRAP) binding pairs (Figure 2.14C), highlighting the ability to tune the cross-linking interactions in future materials.[131] Repeat proteins are known to drive complex self-assembly both in nature[132] and the lab. [133] They are further being designed from the bottom up, to create novel architectures[134] which will prove useful in designing the next generation of protein-driven hybrid self-assemblies.

Another class of supramolucular protein materials are multimeric cages and particles. A major portion within this class of materials are the virus-like particles (VLPs). VLPs are formed using virus-derived coat proteins from viruses like the tobacco mosic virus (TMV), Hepatitis B (HBc), bacteriophage M13 (M13), cowpea chlorotic mottle virus (CCMV), cowpea mosaic virus (CPMV), among many others. The particles take advantage of the specific self-assembly of the viral coat proteins while removing the infectious viral nucleic acids. These robust assemblies have shown promise in biomedicine[135] as drug carriers, vaccines, and imaging agents as well as in catalysis as nanoreactors,[136] Internal polymer conjugation can increase loading capacity,[137] whereas external conjugation can contribute to their stability and increase their solubility.[138] For example,

the Douglas group modified a P22 coat protein monomer with a S39C point mutation to introduce an interior, accessible Cys. This served as a reactive handle to attach a maleimide functionalized ATRP initiator to polymerize the free-amine rich 2-aminoethyl methacrylate. This internal templated polymerization exploited the monodisperse size of the VLP container while increasing the loading capacity of Gd-diethylenetriaminepentacetate, the MRI contrast agent Magnevist, by 28-fold.[137] In addition to the well studied VLP materials, new developments in designer coat proteins,[139] coiled-coil cages,[140] as well as novel self-assembling proteins[5h, 141] provide many new materials for designer protein-polymer conjugates that possess tailorable architectures suitable for an array of applications.

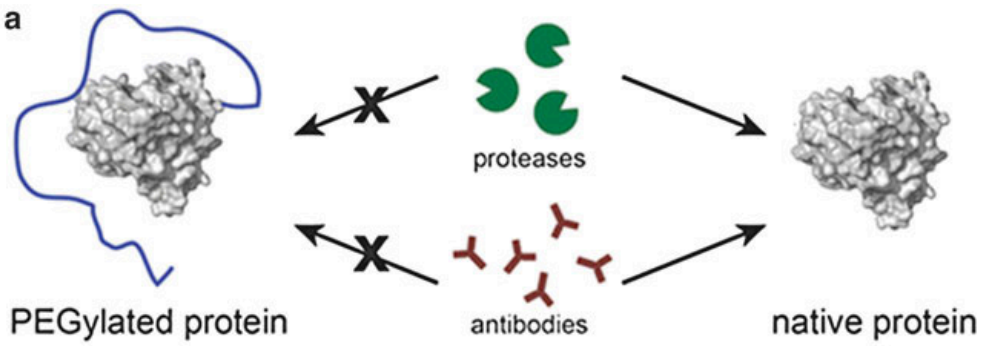
2.5. Emerging and potential applications

2.5.1 Biomedical applications and drug delivery

A polymer corona, as a protecting layer can significantly increase the blood circulation time and circumvent non-specific interactions. Polymer functionalization can also affect the needed dose, drug release time, and targeting location. A tremendous amount of research has been conducted on the inert PEG or PHPMA functionalized protein-polymer conjugates because of the enhanced stability and improved pharmacokinetics.[74] PEGylated proteins like Adagen and On-caspar have been approved by the FDA for therapeutic or diagnostic purposes, resulting in the growth of a multi-billion dollar business.[142] Aside from that, smart bioconjugates containing stimuli-responsive polymers such as PNIPAAM and poly(2-(dimethylamino)ethyl methacrylate) were employed to tailor enzyme activity as well as for environmentally-triggered drug release.[143] Increasing attention has also been paid to targeted gene therapy, which could be fulfilled by the rational design of cationic polyelectrolyte-based conjugates for DNA complexation and receptor recognition.

2.5.2 Non-biomedical applications

In addition to biomedical applications, protein/peptide-polymer conjugates with well-ordered hierarchical structures derived from self-assembly may hold promise for other diverse ap-



b

trade name	protein	PEG size	indication	company
adagen	adenosine deaminase	5 kDa	SCID	Enzon
oncaspar	asparaginase	5 kDa	ALL	Enzon
somavert	GH antagonist	5 kDa	excessive growth	Pfizer
PEG-intron	interferon α2b	12 kDa	hepatitis C	Schering-Plough
pegasys	interferon α2a	40 kDa	hepatitis C	Roche
neulasta	granulocyte CSF	20 kDa	neutropenia	Amgen
mircera	EPO	30 kDa	anemia	Roche
cimzia	anti-TNFα Fab'	40 kDa	rheumatoid arthritis	Nektar/UCB
krystexxa	uricase	10 kDa	chronic gout	Savient

Figure 2.15. Therapeutic protein–polymer conjugates. (a) Schematic of some of the biological effects of protein PEGylation. (b) Table of protein-PEG therapeutics with FDA approval (GH, growth hormone; CSF, colony stimulating factor; EPO, erythropoietin; TNFα, tumor necrosis factor α; SCID, severe combined immunodeficiency disease; ALL, acute lymphatic leukemia). Adapted with permission from Ref. [214]. Copyright 2015 American Chemical Society.

plications including gas separation, optoelectronic devices and catalysis. For example, Xu et al. reported the fabrication of a sub-nanometer porous thin-film membranes by synergistic co-assembly of cyclic peptide-polymer conjugate and BCP.[144] The growth of cyclic peptide nanotubes in the self-assembled BCP framework provides well-aligned channels normal to the surface, exhibiting selective gas transportation properties. Thiophene-based peptide-polymer conjugates with fibrillar organization represent a novel class of bio-inspired materials, where the macroscopic alignment of the superstructure is governed by the interplay between different intermolecular forces in the peptide and thiophene components. The obtained noodle-like hydrogel displayed anisotropic pho-

tophysical and electrical responses.[145] Diez et al. also reported the use of highly ordered oligopeptide-PEG biohybrids for template-directed nucleation and growth of silver NP arrays. Unique optical and electronic properties of these nanoparticle (NP) arrays with high densities and short interparticle spacings make them good candidates for light trapping.[146]

In the realm of catalysis an of area of special interest, that will benefit nanostructured protein-polymer conjugates is the growing series reaction pathways catalyzed by in vitro enzymatic systems. Patterning enzymes involved in in vitro cascade reactions has been shown to give a variety of benefits over solution based systems: increased stability,[147] enhanced activity,[148] reusability[147a] and metabolic channeling[149] to name a few. Nolte *et al.* encapsulated GOx in HRP-PS and HRP-PMMA amphiphile vesicles for use as high-performance bio-nanoreactors for the cascade oxidation of glucose and ABTS.[150] Städler *et al.* utilized a mutli-compartment encapsulation strategy to functionalize polydopamine-coated capsosomes with three enzyme to carry out two separate enzymatic reaction pathways.[151] The field of compartmentalized nanoreactors has been recently reviewed.[152] Beyond simple encapsulation strategies lies the more complex creation of artificial enzyme scaffolds. By controlling the number of enzymes and their positioning with nanometer precision, groups have shown increases in reaction efficiency from 16-fold[153] to 77-fold.[154]

Experimentally in vitro enzyme cascade systems show great promise as biorefineries for the conversion biomass to energy sources such as biofuels and in bio-batteries[155] as well as commercially relevant products such as sweeteners[156] and other materials.[157] A main disadvantage for these systems is the long-term stability of their enzymes, which adds significantly to the cost. The inherent stability increase as well as proposed benefits to reaction rate and efficiency of surface immobilization or patterning as bio-hybrids may help assuage these concerns. But as the world's need for energy and food continue to grow the market for biomass-derived fuels and food products will become more competitive, especially if the cost of non-renewable fuel sources such as petroleum gasoline and their derived products remain high.

2.6. **Conclusion and perspectives**

The concept that a protein's primary sequence affects its structure, which further affects its function, is well established and accepted in the literature. While significant strides are being made to further expand our understanding of this relationship through the modeling and designing new protein architectures, there is still much work to be done before we can design more complex functions like catalysis. Conversely, polymeric materials have a much greater diversity in chemical structure and can be easily designed to impart structure through phase separation or by the myriad available post-polymerization processing techniques. By marrying proteins and polymers into conjugated bio-hybrid materials, materials scientists, chemists, and biologists alike, have at their fingertips a vast toolkit for material design. While polymers still hold the advantage in time invested and scalability, protein materials, especially modular recombinant polypeptides, are becoming more competitive with the advent of scalable fermentation processes and kit-based genetic modification strategies. As these conjugation strategies continue to mature, interest in new conjugates will grow. With growth in demand for materials, cost will decrease, opening up a large number of opportunities that may be currently deemed cost-prohibitive.

2.7. **References**

[1] a) S. F. Oliveira, G. Bisker, N. A. Bakh, S. L. Gibbs, M. P. Landry and M. S. Strano, *Carbon*. **2015**, *95*, 767; b) M. A. Ali, P. R. Solanki, S. Srivastava, S. Singh, V. V. Agrawal, R. John and B. D. Malhotra, *ACS Appl Mater Inter*. **2015**, *7*, 5837; c) S. Bhattacharyya, C. Sinturel, J. P. Salvetat and M. L. Saboungi, *Appl Phys Lett*. **2005**, *86*; d) K. Makyla, C. Muller, S. Lorcher, T. Winkler, M. G. Nussbaumer, M. Eder and N. Bruns, *Adv Mater.* **2013**, *25*, 2701; e) G. W. M. Vandermeulen and H. A. Klok, *Macromol Biosci*. **2004**, *4*, 383; f) X. Hu, P. Cebe, A. S. Weiss, F. Omenetto and D. L. Kaplan, *Mater Today*. **2012**, *15*, 208.

[2] D. E. Koshland, Jr., *Science*. **1963**, *142*, 1533.

[3] N. Furnham, I. Sillitoe, G. L. Holliday, A. L. Cuff, R. A. Laskowski, C. A. Orengo and J. M. Thornton, *PLoS Comput Biol*. **2012**, *8*.

[4] a) T. Z. Grove, C. O. Osuji, J. D. Forster, E. R. Dufresne and L. Regan, *J Am Chem Soc.*. **2010**, *132*, 14024; b) R. Parker, A. Mercedes-Camacho and T. Z. Grove, *Protein Sci*. **2014**, *23*, 790; c) S. Cho, Q. Wang, C. P. Swaminathan, D. Hesek, M. Lee, G. J. Boons, S. Mobashery and R. A. Mariuzza, *Proc Natl Acad Sci U.S.A*. **2007**, *104*, 8761.

[5] a) D. Baker, *Biochem. Soc. Trans.* **2014**, *42*, 225; b) C. A. Rohl, C. E. M. Strauss, K. M. S. Misura and D. Baker, *Method Enzymol*. **2004**, *383*, 66; c) J. Meiler and D. Baker, *Proc Natl Acad Sci U.S.A*. **2003**, *100*, 12105; d) M. R. Lee, J. Tsai, D. Baker and P. A. Kollman, *J. Mol. Biol.* **2001**, *313*, 417; e) D. Baker and A. Sali, *Science*. **2001**, *294*, 93; f) R. Bonneau and D. Baker, *Annu Rev Bioph Biom*. **2001**, *30*, 173; g) G. A. Khoury, A. Liwo, F. Khatib, H. Y. Zhou, G. Chopra, J. Bacardit, L. O. Bortot,

R. A. Faccioli, X. Deng, Y. He, P. Krupa, J. L. Li, M. A. Mozolewska, A. K. Sieradzan, J. Smadbeck, T. Wirecki, S. Cooper, J. Flatten, K. F. Xu, D. Baker, J. L. Cheng, A. C. B. Delbem, C. A. Floudas, C. Keasar, M. Levitt, Z. Popovic, H. A. Scheraga, J. Skolnick, S. N. Crivelli and F. Players, *Proteins*. **2014**, *82*, 1850; h) N. P. King, J. B. Bale, W. Sheffler, D. E. McNamara, S. Gonen, T. Gonen, T. O. Yeates and D. Baker, *Nature*. **2014**, *510*, 103; i) N. Preiswerk, T. Beck, J. D. Schulz, P. Milovnik, C. Mayer, J. B. Siegel, D. Baker and D. Hilvert, *Proc Natl Acad Sci U.S.A*. **2014**, *111*, 8013; j) Y. C. Cui, F. Liu, X. Li, L. Wang, H. W. Wang, G. J. Chen, L. Yuan, J. L. Brash and H. Chen, *ACS Appl Mater Inter*. **2015**, *7*, 21913.

[6] a) T. Matsudo, K. Ogawa and E. Kokufuta, *Biomacromolecules*. **2003**, *4*, 1794; b) R. M. Kramer, V. R. Shende, N. Motl, C. N. Pace and J. M. Scholtz, *Biophys. J.* **2012**, *102*, 1907.

[7] D. Moatsou, J. Li, A. Ranji, A. Pitto-Barry, I. Ntai, M. C. Jewett and R. K. O'Reilly, *Bioconjugate Chem*. **2015**, *26*, 1890.

[8] a) M. M. Lorenzo, C. G. Decker, M. U. Kahveci, S. J. Paluck and H. D. Maynard, *Macromolecules*. **2016**, *49*, 30; b) D. Russo, M. Plazanet, J. Teixeira, M. Moulin, M. Hartlein, F. R. Wurm and T. Steinbach, *Biomacromolecules*. **2016**, *17*, 141; c) Q. Zhang, M. X. Li, C. Y. Zhu, G. Nurumbetov, Z. D. Li, P. Wilson, K. Kempe and D. M. Haddleton, *J Am Chem Soc.*. **2015**, *137*, 9344.

[9] T. J. Deming, *Adv Mater.* **1997**, *9*, 299.

[10] G. B. Fields and R. L. Noble, *Int J Pept Prot Res*. **1990**, *35*, 161.

[11] P. E. Dawson and S. B. Kent, *Annual review of biochemistry*. **2000**, *69*, 923.

[12] A. L. Demain and P. Vaishnav, *Biotechnol adv*. **2009**, *27*, 297.

[13] a) M. W. Jones, M. I. Gibson, G. Mantovani and D. M. Haddleton, *Polym Chem-Uk*. **2011**, *2*, 572; b) A. Anastasaki, V. Nikolaou, G. Nurumbetov, P. Wilson, K. Kempe, J. F. Quinn, T. P. Davis, M. R. Whittaker and D. M. Haddleton, *Chem Rev*. **2016**, *116*, 835; c) D. Konkolewicz, Y. Wang, P. Krys, M. Zhong, A. A. Isse, A. Gennaro and K. Matyjaszewski, *Polym Chem-Uk*. **2014**, *5*, 4396; d) P. Krys, T. G. Ribelli, K. Matyjaszewski and A. Gennaro, *Macromolecules*. **2016**, 10.1021/acs.macromol.6b00058, 10.1021/acs.macromol.6b00058; e) C. Boyer, N. A. Corrigan, K. Jung, D. Nguyen, T. K. Nguyen, N. N. Adnan, S. Oliver, S. Shanmugam and J. Yeow, *Chem Rev*. **2016**, *116*, 1803.

[14] C. Boyer, V. Bulmus, T. P. Davis, V. Ladmiral, J. Liu and S. Perrier, *Chem Rev*. **2009**, *109*, 5402.

[15] B. Le Droumaguet and J. Nicolas, *Polym Chem-Uk*. **2010**, *1*, 563.

[16] R. M. Broyer, G. N. Grover and H. D. Maynard, *Chem. Commun.* **2011**, *47*, 2212.

[17] I. W. Hamley, *Biomacromolecules*. **2014**, *15*, 1543.

[18] H. Li, A. P. Bapat, M. Li and B. S. Sumerlin, *Polym Chem-Uk*. **2011**, *2*, 323.

[19] L. Tao, J. Liu, J. Xu and T. P. Davis, *Org Biomol Chem*. **2009**, *7*, 3481.

[20] A. O. Y. Chan, C. M. Ho, H. C. Chong, Y. C. Leung, J. S. Huang, M. K. Wong and C. M. Che, *J Am Chem Soc.*. **2012**, *134*, 2589.

[21] a) J. M. Gilmore, R. A. Scheck, A. P. Esser-Kahn, N. S. Joshi and M. B. Francis, *Angew Chem Int Edit*. **2006**, *45*, 5307; b) R. A. Scheck, M. T. Dedeo, A. T. Iavarone and M. B. Francis, *J Am Chem Soc*. **2008**, *130*, 11762.

[22] J. I. MacDonald, H. K. Munch, T. Moore and M. B. Francis, *Nat Chem Biol*. **2015**, *11*, 326.

[23] M. H. Stenzel, *ACS Macro Letters*. **2012**, *2*, 14.

[24] G. Mantovani, F. Lecolley, L. Tao, D. M. Haddleton, J. Clerx, J. J. Cornelissen and K. Velonia, *J Am Chem Soc*. **2005**, *127*, 2966.

[25] E. Bays, L. Tao, C. W. Chang and H. D. Maynard, *Biomacromolecules*. **2009**, *10*, 1777.

[26] D. Bontempo, K. L. Heredia, B. A. Fish and H. D. Maynard, *J Am Chem Soc*. **2004**, *126*, 15372.

[27] V. Vazquez-Dorbatt, Z. P. Tolstyka, C. W. Chang and H. D. Maynard, *Biomacromolecules*.

2009, *10*, 2207.
[28] C. W. Chang, T. H. Nguyen and H. D. Maynard, *Macromol Rapid Comm*. **2010**, *31*, 1691.
[29] S. Shaunak, A. Godwin, J. W. Choi, S. Balan, E. Pedone, D. Vijayarangam, S. Heidelberger, I. Teo, M. Zloh and S. Brocchini, *Nat Chem Biol*. **2006**, *2*, 312.
[30] P. Wilson, A. Anastasaki, M. R. Owen, K. Kempe, D. M. Haddleton, S. K. Mann, A. P. R. Johnston, J. F. Quinn, M. R. Whittaker, P. J. Hogg and T. P. Davis, *J Am Chem Soc*. **2015**, *137*, 4215.
[31] a) N. Griebenow, S. Bräse and A. M. Dilmac, *RSC Advances*. **2015**, *5*, 54301; b) N. Griebenow, A. M. Dilmac, S. Greven and S. Brase, *Bioconjugate Chem*. **2016**, 10.1021/acs.bioconjchem.5b00682.
[32] a) C. A. Foerder and B. M. Shapiro, *Proc Natl Acad Sci U.S.A.* **1977**, *74*, 4214; b) J. W. Heinecke, *Toxicology*. **2002**, *177*, 11.
[33] S. Meunier, E. Strable and M. G. Finn, *Chem Biol*. **2004**, *11*, 319.
[34] P. Nagy, T. P. Lechte, A. B. Das and C. C. Winterbourn, *J. Biol. Chem.* **2012**, *287*, 26068.
[35] S. D. Tilley and M. B. Francis, *J Am Chem Soc*. **2006**, *128*, 1080.
[36] N. S. Joshi, L. R. Whitaker and M. B. Francis, *J Am Chem Soc.*. **2004**, *126*, 15942.
[37] D. W. Romanini and M. B. Francis, *Bioconjugate Chem.* **2008**, *19*, 153.
[38] M. W. Jones, G. Mantovani, C. A. Blindauer, S. M. Ryan, X. X. Wang, D. J. Brayden and D. M. Haddleton, *J Am Chem Soc.*. **2012**, *134*, 7406.
[39] H. Ban, J. Gavrilyuk and C. F. Barbas, 3rd, *J Am Chem Soc*. **2010**, *132*, 1523.
[40] H. Ban, M. Nagano, J. Gavrilyuk, W. Hakamata, T. Inokuma and C. F. Barbas, 3rd, *Bioconjugate Chem*. **2013**, *24*, 520.
[41] Q. Y. Hu, M. Allan, R. Adamo, D. Quinn, H. L. Zhai, G. X. Wu, K. Clark, J. Zhou, S. Ortiz, B. Wang, E. Danieli, S. Crotti, M. Tontini, G. Brogioni and F. Berti, *Chem Sci*. **2013**, *4*, 3827.
[42] a) H. Nothaft and C. M. Szymanski, *J Biol Chem.* **2013**, *10*, 6912; b) H. Nothaft and C. M. Szymanski, *Nat Rev Microbiol*. **2010**, *8*, 765.
[43] A. Moura, M. A. Savageau and R. Alves, *PLoS One*. **2013**, *8*, e77319.
[44] X. Lia, L. Zhanga, S. E. Halla and J. P. Tam, *Tetrahedron Lett*. **2000**, *41*, 4069.
[45] J. M. Antos and M. B. Francis, *J Am Chem Soc*. **2004**, *126*, 10256.
[46] J. M. Antos, J. M. McFarland, A. T. Iavarone and M. B. Francis, *J Am Chem Soc*. **2009**, *131*, 6301.
[47] B. V. Popp and Z. T. Ball, *J Am Chem Soc*. **2010**, *132*, 6660.
[48] D. B. Cowie and G. N. Cohen, *Biochim. Biophys. Acta*. **1957**, *26*, 252.
[49] a) J. T. Ngo and D. A. Tirrell, *Acc. Chem. Res.* **2011**, *44*, 677; b) D. A. Dougherty and E. B. Van Arnam, *Chembiochem*. **2014**, *15*, 1710; c) E. M. Sletten and C. R. Bertozzi, *Acc. Chem. Res.* **2011**, *44*, 666; d) C. H. Kim, J. Y. Axup and P. G. Schultz, *Current Opinion in Chemical Biology*. **2013**, *17*, 412.
[50] R. L. M. Teeuwen, S. S. van Berkel, T. H. H. van Dulmen, S. Schoffelen, S. A. Meeuwissen, H. Zuilhof, F. A. de Wolf and J. C. M. van Hest, *Chem. Commun.* **2009**, 10.1039/b903903a, 4022.
[51] S. Schoffelen, M. B. van Eldijk, B. Rooijakkers, R. Raijmakers, A. J. R. Heck and J. C. M. van Hest, *Chem Sci*. **2011**, *2*, 701.
[52] L. A. Canalle, T. Vong, P. H. Adams, F. L. van Delft, J. M. Raats, R. G. Chirivi and J. C. van Hest, *Biomacromolecules*. **2011**, *12*, 3692.
[53] S. S. van Berkel, A. T. J. Dirks, M. F. Debets, F. L. van Delft, J. J. L. M. Cornelissen, R. J. M. Nolte and F. P. J. T. Rutjes, *Chembiochem*. **2007**, *8*, 1504.
[54] X. H. Ning, J. Guo, M. A. Wolfert and G. J. Boons, *Angew Chem Int Edit*. **2008**, *47*, 2253.
[55] K. W. Dehnert, J. M. Baskin, S. T. Laughlin, B. J. Beahm, N. N. Naidu, S. L. Amacher and C. R. Bertozzi, *Chembiochem*. **2012**, *13*, 353.

[56] E. Saxon and C. R. Bertozzi, *Science*. **2000**, *287*, 2007.
[57] a) B. L. Nilsson, L. L. Kiessling and R. T. Raines, *Org Lett*. **2000**, *2*, 1939; b) B. L. Nilsson, L. L. Kiessling and R. T. Raines, *Org Lett*. **2001**, *3*, 9; c) E. Saxon, J. I. Armstrong and C. R. Bertozzi, *Org Lett*. **2000**, *2*, 2141.
[58] L. Tao, G. Mantovani, F. Lecolley and D. M. Haddleton, *J Am Chem Soc*. **2004**, *126*, 13220.
[59] K. L. Heredia, Z. P. Tolstyka and H. D. Maynard, *Macromolecules*. **2007**, *40*, 4772.
[60] V. Vázquez-Dorbatt, Z. P. Tolstyka and H. D. Maynard, *Macromolecules*. **2009**, *42*, 7650.
[61] S. A. Slavoff, I. Chen, Y. A. Choi and A. A. Y. Ting, *J Am Chem Soc.*. **2008**, *130*, 1160.
[62] Y. Zhang, M. J. Blanden, C. Sudheer, S. A. Gangopadhyay, M. Rashidian, J. L. Hougland and M. D. Distefano, *Bioconjugate Chem.* **2015**, *26*, 2542.
[63] A.-W. Struck, M. R. Bennett, S. A. Shepherd, B. J. C. Law, Y. Zhuo, L. S. Wong and J. Micklefield, *J Am Chem Soc*. **2016**, *138*, 3038.
[64] M. Rashidian, J. K. Dozier and M. D. Distefano, *Bioconjugate Chem.* **2013**, *24*, 1277.
[65] C. Boyer, J. Liu, V. Bulmus, T. P. Davis, C. Barner-Kowollik and M. H. Stenzel, *Macromolecules*. **2008**, *41*, 5641.
[66] K. L. Heredia, L. Tao, G. N. Grover and H. D. Maynard, *J Polym Sci A1*. **2010**, *1*, 168.
[67] P. De, S. R. Gondi and B. S. Sumerlin, *Biomacromolecules*. **2008**, *9*, 1064.
[68] S. E. Averick, E. Paredes, D. Grahacharya, B. F. Woodman, S. J. Miyake-Stoner, R. A. Mehl, K. Matyjaszewski and S. R. Das, *Langmuir*. **2012**, *28*, 1954.
[69] M. F. Fouz, K. Mukumoto, S. Averick, O. Molinar, B. M. McCartney, K. Matyjaszewski, B. A. Armitage and S. R. Das, *ACS Central Science*. **2015**, *1*, 431.
[70] C. Albayrak and J. R. Swartz, *ACS Synth. Biol.* **2014**, *3*, 353.
[71] a) S. E. Averick, C. G. Bazewicz, B. F. Woodman, A. Simakova, R. A. Mehl and K. Matyjaszewski, *Eur Polym J*. **2013**, *49*, 2919; b) J. C. Peeler, B. F. Woodman, S. Averick, S. J. Miyake-Stoner, A. L. Stokes, K. R. Hess, K. Matyjaszewski and R. A. Mehl, *J Am Chem Soc.*. **2010**, *132*, 13575.
[72] G. N. Grover, S. N. Alconcel, N. M. Matsumoto and H. D. Maynard, *Macromolecules*. **2009**, *42*, 7657.
[73] K. L. Heredia, G. N. Grover, L. Tao and H. D. Maynard, *Macromolecules*. **2009**, *42*, 2360.
[74] L. Tao, J. Liu and T. P. Davis, *Biomacromolecules*. **2009**, *10*, 2847.
[75] Y. Nojima, Y. Suzuki, K. Yoshida, F. Abe, T. Shiga, T. Takeuchi, A. Sugiyama, H. Shimizu and A. Sato, *Pharm Res*. **2009**, *26*, 2125.
[76] L. Tao, C. S. Kaddis, R. R. Loo, G. N. Grover, J. A. Loo and H. D. Maynard, *Macromolecules*. **2009**, *42*, 8028.
[77] G. N. Grover, J. Lam, T. H. Nguyen, T. Segura and H. D. Maynard, *Biomacromolecules*. **2012**, *13*, 3013.
[78] a) T. Z. Grove, C. O. Osuji, J. D. Forster, E. R. Dufresne and L. Regan, *J Am Chem Soc*. **2010**, *132*, 14024; b) T. Z. Grove, J. Forster, G. Pimienta, E. Dufresne and L. Regan, *Biopolymers*. **2012**, *97*, 508.
[79] a) K. Vulic and M. S. Shoichet, *Biomacromolecules*. **2014**, *15*, 3867; b) K. Vulic and M. S. Shoichet, *J Am Chem Soc.*. **2012**, *134*, 882; c) C. T. S. W. P. Foo, J. S. Lee, W. Mulyasasmita, A. Parisi-Amon and S. C. Heilshorn, *Proc Natl Acad Sci U.S.A.* **2009**, *106*, 22067; d) L. Cai, R. E. Dewi and S. C. Heilshorn, *Adv Funct Mater.* **2015**, *25*, 1344.
[80] a) M. Danial, M. J. Root and H. A. Klok, *Biomacromolecules*. **2012**, *13*, 1438; b) N. K. Singha, M. I. Gibson, B. P. Koiry, M. Danial and H. A. Klok, *Biomacromolecules*. **2011**, *12*, 2908.
[81] K. Petkau-Milroy, D. A. Uhlenheuer, A. J. H. Spiering, J. A. J. M. Vekemansb and L. Brunsveld,

Chem Sci. **2013**, *4*, 2886.
[82] J. Yang, I. Gitlin, V. M. Krishnamurthy, J. A. Vazquez, C. E. Costello and G. M. Whitesides, *J Am Chem Soc.*. **2003**, *125*, 12392.
[83] a) S. L. Kuan, Y. Z. Wu and T. Weil, *Macromol Rapid Comm.* **2013**, *34*, 380; b) D. Y. W. Ng, Y. Z. Wu, S. L. Kuan and T. Weil, *Acc. Chem. Res.* **2014**, *47*, 3471; c) Y. Z. Wu, G. Pramanik, K. Eisele and T. Weil, *Biomacromolecules*. **2012**, *13*, 1890; d) Y. Z. Wu, C. Li, F. Boldt, Y. R. Wang, S. L. Kuan, T. T. Tran, V. Mikhalevich, C. Fortsch, H. Barth, Z. Q. Yang, D. S. Liu and T. Weil, *Chem. Commun.* **2014**, *50*, 14620; e) Y. Z. Wu, S. Ihme, M. Feuring-Buske, S. L. Kuan, K. Eisele, M. Lamla, Y. R. Wang, C. Buske and T. Weil, *Adv Healthc Mater*. **2013**, *2*, 884.
[84] a) D. E. Meyer and A. Chilkoti, *Biomacromolecules*. **2004**, *5*, 846; b) S. R. MacEwan and A. Chilkoti, *Biopolymers*. **2010**, *94*, 60; c) D. W. Urry, *J. Phys. Chem. B*. **1997**, *101*, 11007.
[85] a) X. X. Xia, Q. B. Xu, X. Hu, G. K. Qin and D. L. Kaplan, *Biomacromolecules*. **2011**, *12*, 3844; b) X. X. Xia, M. Wang, Y. A. Lin, Q. B. Xu and D. L. Kaplan, *Biomacromolecules*. **2014**, *15*, 908; c) Z. Y. Sun, G. K. Qin, X. X. Xia, M. Cronin-Golomb, F. G. Omenetto and D. L. Kaplan, *J Am Chem Soc.*. **2013**, *135*, 3675.
[86] W. Ott, T. Nicolaus, H. E. Gaub and M. A. Nash, *Biomacromolecules*. **2016**, 10.1021/acs.biomac.5b01726.
[87] C.-Y. Lin and J. C. Liu, *Curr. Opin. Biotechnol.* **2016**, *40*, 56.
[88] B. Zakeri and M. Howarth, *J Am Chem Soc*. **2010**, *132*, 4526.
[89] B. Zakeri, J. O. Fierer, E. Celik, E. C. Chittock, U. Schwarz-Linek, V. T. Moy and M. Howarth, *Proc Natl Acad Sci U S A*. **2012**, *109*, E690.
[90] W. B. Zhang, F. Sun, D. A. Tirrell and F. H. Arnold, *J Am Chem Soc*. **2013**, *135*, 13988.
[91] X. W. Wang and W. B. Zhang, *Angew Chem Int Ed Engl*. **2016**, *55*, 3442.
[92] J. K. Bocker, K. Friedel, J. C. Matern, A. L. Bachmann and H. D. Mootz, *Angew Chem Int Ed Engl*. **2015**, *54*, 2116.
[93] a) J. M. Antos, M. W. Popp, R. Ernst, G. L. Chew, E. Spooner and H. L. Ploegh, *J Biol Chem*. **2009**, *284*, 16028; b) J. Hu, W. Zhao, Y. Gao, M. Sun, Y. Wei, H. Deng and W. Gao, *Biomaterials*. **2015**, *47*, 13.
[94] J. A. Camarero, J. Pavel and T. W. Muir, *Angew Chem Int Ed Engl*. **1998**, *37*, 347.
[95] T. Collins, M. Barroca, F. Branca, J. Padrao, R. Machado and M. Casal, *Biomacromolecules*. **2014**, *15*, 2701.
[96] N. C. Tang and A. Chilkoti, *Nat Mater*. **2016**, *15*, 419.
[97] a) M. B. van Eldijk, F. C. Smits, N. Vermue, M. F. Debets, S. Schoffelen and J. C. van Hest, *Biomacromolecules*. **2014**, *15*, 2751; b) Y. N. Zhang, R. K. Avery, Q. Vallmajo-Martin, A. Assmann, A. Vegh, A. Memic, B. D. Olsen, N. Annabi and A. Khademhosseini, *Adv Funct Mater*. **2015**, *25*, 4814; c) M. V. Imakaev, K. M. Tchourine, S. K. Nechaev and L. A. Mirny, *Soft Matter*. **2015**, *11*, 665.
[98] a) H. Khalili, A. Godwin, J. W. Choi, R. Lever and S. Brocchini, *Bioconjugate Chem.* **2012**, *23*, 2262; b) M. Zloh, S. Shaunak, S. Balan and S. Brocchini, *Nat Protoc*. **2007**, *2*, 1070; c) S. Brocchini, S. Balan, A. Godwin, J. W. Choi, M. Zloh and S. Shaunak, *Nat Protoc*. **2006**, *1*, 2241.
[99] J. W. Chin, *Annu Rev Biochem, Vol 83*. **2014**, *83*, 379.
[100] S. H. Mejias, B. Sot, R. Guantes and A. L. Cortajarena, *Nanoscale*. **2014**, *6*, 10982.
[101] P. G. Holder, D. T. Finley, N. Stephanopoulos, R. Walton, D. S. Clark and M. B. Francis, *Langmuir*. **2010**, *26*, 17383.
[102] a) Y. Hu, R. Zandi, A. Anavitarte, C. M. Knobler and W. M. Gelbart, *Biophys J*. **2008**, *94*, 1428; b) M. Comellas-Aragonès, A. d. l. Escosura, A. T. J. Dirks, A. v. d. Ham, A. Fusté-Cuñé, J. J. L. M.

Cornelissen and R. J. M. Nolte, *Biomacromolecules*. **2009**, *10*, 3141; c) J. Lucon, S. Qazi, M. Uchida, G. J. Bedwell, B. LaFrance, P. E. P. Jr and T. Douglas, *Nat Chem*. **2012**, *4*, 781.
[103] a) N. M. Matsumoto, P. Prabhakaran, L. H. Rome and H. D. Maynard, *ACS Nano*. **2013**, *7*, 867; b) N. M. Matsumoto, G. W. Buchman, L. H. Rome and H. D. Maynard, *Eur Polym J*. **2015**, *69*, 532.
[104] T. Haga, H. Hirakawa and T. Nagamune, *PLoS One*. **2013**, *8*, e75114.
[105] a) Y. T. Lai, E. Reading, G. L. Hura, K. L. Tsai, A. Laganowsky, F. J. Asturias, J. A. Tainer, C. V. Robinson and T. O. Yeates, *Nat Chem*. **2014**, *6*, 1065; b) N. P. King, J. B. Bale, W. Sheffler, D. E. McNamara, S. Gonen, T. Gonen, T. O. Yeates and D. Baker, *Nature*. **2014**, *510*, 103; c) Y. T. Lai, D. Cascio and T. O. Yeates, *Science*. **2012**, *336*, 1129.
[106] J. C. Sinclair, K. M. Davies, C. Venien-Bryan and M. E. Noble, *Nat Nanotechnol*. **2011**, *6*, 558.
[107] A. R. Voet, H. Noguchi, C. Addy, D. Simoncini, D. Terada, S. Unzai, S. Y. Park, K. Y. Zhang and J. R. Tame, *Proc Natl Acad Sci U S A*. **2014**, *111*, 15102.
[108] a) P. S. Huang, K. Feldmeier, F. Parmeggiani, D. A. Fernandez Velasco, B. Hocker and D. Baker, *Nat Chem Biol*. **2016**, *12*, 29; b) D. N. Woolfson, G. J. Bartlett, A. J. Burton, J. W. Heal, A. Niitsu, A. R. Thomson and C. W. Wood, *Curr Opin Struc Biol*. **2015**, *33*, 16; c) N. H. Joh, T. Wang, M. P. Bhate, R. Acharya, Y. Wu, M. Grabe, M. Hong, G. Grigoryan and W. F. DeGrado, *Science*. **2014**, *346*, 1520.
[109] B. D. Olsen, *Macromol Chem Phys*. **2013**, *214*, 1659.
[110] N. Kumar and J. I. Hahm, *Langmuir*. **2005**, *21*, 6652.
[111] N. Kumar, O. Parajuli, A. Dorfman, D. Kipp and J. I. Hahm, *Langmuir*. **2007**, *23*, 7416.
[112] N. Kumar, O. Parajuli and J. I. Hahm, *J Phys Chem B*. **2007**, *111*, 4581.
[113] L. Shen, A. Garland, Y. Wang, Z. Li, C. W. Bielawski, A. Guo and X. Y. Zhu, *Small*. **2012**, *8*, 3169.
[114] B. Kim, C. N. Lam and B. D. Olsen, *Macromolecules*. **2012**, *45*, 4572.
[115] J.-C. Eloi, S. E. W. Jones, V. Poór, M. Okuda, J. Gwyther and W. Schwarzacher, *Adv Funct Mater.* **2012**, *22*, 3273.
[116] M. J. Boerakker, J. M. Hannink, P. H. Bomans, P. M. Frederik, R. J. Nolte, E. M. Meijer and N. A. Sommerdijk, *Angew Chem Int Ed Engl*. **2002**, *41*, 4239.
[117] K. Velonia, A. E. Rowan and R. J. Nolte, *J Am Chem Soc*. **2002**, *124*, 4224.
[118] I. C. Reynhout, J. J. Cornelissen and R. J. Nolte, *J Am Chem Soc*. **2007**, *129*, 2327.
[119] L.-Y. Shi, Y. Zhou, X.-H. Fan and Z. Shen, *Macromolecules*. **2013**, *46*, 5308.
[120] H. Schlaad, B. Smarsly and M. Losik, *Macromolecules*. **2004**, *37*, 2210.
[121] C. S. Thomas, M. J. Glassman and B. D. Olsen, *ACS Nano*. **2011**, *5*, 5697.
[122] C. S. Thomas, L. Xu and B. D. Olsen, *Biomacromolecules*. **2012**, *13*, 2781.
[123] C. N. Lam and B. D. Olsen, *Soft Matter*. **2013**, *9*, 2393.
[124] a) R. Matsunaga, S. Yanaka, S. Nagatoishi and K. Tsumoto, *Nat Commun*. **2013**, *4*, 2211; b) J. D. Brodin, X. I. Ambroggio, C. Tang, K. N. Parent, T. S. Baker and F. A. Tezcan, *Nat Chem*. **2012**, *4*, 375.
[125] a) A. A. Dinerman, J. Cappello, H. Ghandehari and S. W. Hoag, *Biomaterials*. **2002**, *23*, 4203; b) R. Dandu, A. Von Cresce, R. Briber, P. Dowell, J. Cappello and H. Ghandehari, *Polymer*. **2009**, *50*, 366; c) K. Greish, K. Araki, D. Q. Li, B. W. O'Malley, R. Dandu, J. Frandsen, J. Cappello and H. Ghandehari, *Biomacromolecules*. **2009**, *10*, 2183.
[126] N. Tzokova, C. M. Fernyhouph, P. D. Topham, N. Sandon, D. J. Adams, M. F. Butler, S. P. Armes and A. J. Ryan, *Langmuir*. **2009**, *25*, 2479.

[127] a) C. Wang, R. J. Stewart and J. Kopecek, *Nature*. **1999**, *397*, 417; b) C. Wang, J. Kopecek and R. J. Stewart, *Biomacromolecules*. **2001**, *2*, 912; c) J. Y. Yang, C. Y. Xu, C. Wang and J. Kopecek, *Biomacromolecules*. **2006**, *7*, 1187; d) P. Jing, J. S. Rudra, A. B. Herr and J. H. Collier, *Biomacromolecules*. **2008**, *9*, 2438.
[128] a) A. M. Jonker, D. W. P. M. Lowik and J. C. M. van Hest, *Chem Mater*. **2012**, *24*, 759; b) A. Dasgupta, J. H. Mondal and D. Das, *RSC Advances*. **2013**, *3*, 9117.
[129] J. D. Ehrick, S. K. Deo, T. W. Browning, L. G. Bachas, M. J. Madou and S. Daunert, *Nat. Mater.* **2005**, *4*, 298.
[130] J. D. Ehrick, M. R. Luckett, S. Khatwani, Y. N. Wei, S. K. Deo, L. G. Bachas and S. Daunert, *Macromol Biosci*. **2009**, *9*, 864.
[131] E. B. Speltz, A. Nathan and L. Regan, *ACS Chem. Biol.* **2015**, *10*, 2108.
[132] J. A. Ybe, F. M. Brodsky, K. Hofmann, K. Lin, S.-H. Liu, L. Chen, T. N. Earnest, R. J. Fletterick and P. K. Hwang, *Nature*. **1999**, *399*, 371.
[133] a) K. L. Gurunatha, A. C. Fournier, A. Urvoas, M. Valerio-Lepiniec, V. Marchi, P. Minard and E. Dujardin, *ACS Nano*. **2016**, *10*, 3176; b) R. P. Watson, M. T. Christen, C. Ewald, F. Bumbak, C. Reichen, M. Mihajlovic, E. Schmidt, P. Güntert, A. Caflisch, A. Plückthun and O. Zerbe, *Structure*. **2014**, *22*, 985; c) N. A. Carter and T. Z. Grove, *Biomacromolecules*. **2015**, *16*, 706.
[134] a) L. Doyle, J. Hallinan, J. Bolduc, F. Parmeggiani, D. Baker, B. L. Stoddard and P. Bradley, *Nature*. **2015**, *528*, 585; b) T. J. Brunette, F. Parmeggiani, P. S. Huang, G. Bhabha, D. C. Ekiert, S. E. Tsutakawa, G. L. Hura, J. A. Tainer and D. Baker, *Nature*. **2015**, *528*, 580.
[135] I. Yildiz, S. Shukla and N. F. Steinmetz, *Curr. Opin. Biotechnol.* **2011**, *22*, 901.
[136] a) D. P. Patterson, P. E. Prevelige and T. Douglas, *ACS Nano*. **2012**, *6*, 5000; b) D. P. Patterson, B. Schwarz, R. S. Waters, T. Gedeon and T. Douglas, *ACS Chem. Biol.* **2014**, *9*, 359.
[137] J. Lucon, S. Qazi, M. Uchida, G. J. Bedwell, B. LaFrance, P. E. Prevelige and T. Douglas, *Nat. Chem.* **2012**, *4*, 781.
[138] T. L. Schlick, Z. B. Ding, E. W. Kovacs and M. B. Francis, *J Am Chem Soc.*. **2005**, *127*, 3718.
[139] A. Hernandez-Garcia, D. J. Kraft, A. F. J. Janssen, P. H. H. Bomans, N. A. J. M. Sommerdijk, D. M. E. Thies-Weesie, M. E. Favretto, R. Brock, F. A. de Wolf, M. W. T. Werten, P. van der Schoot, M. C. Stuart and R. de Vries, *Nat. Nanotechnol.* **2014**, *9*, 698.
[140] a) J. M. Fletcher, R. L. Harniman, F. R. H. Barnes, A. L. Boyle, A. Collins, J. Mantell, T. H. Sharp, M. Antognozzi, P. J. Booth, N. Linden, M. J. Miles, R. B. Sessions, P. Verkade and D. N. Woolfson, *Science*. **2013**, *340*, 595; b) H. Gradisar, S. Bozic, T. Doles, D. Vengust, I. Hafner-Bratkovic, A. Mertelj, B. Webb, A. Sali, S. Klavzar and R. Jerala, *Nat Chem Biol*. **2013**, *9*, 362.
[141] a) J. C. Sinclair, *Curr Opin Chem Biol*. **2013**, *17*, 946; b) Y. T. Lai, E. Reading, G. L. Hura, K. L. Tsai, A. Laganowsky, F. J. Asturias, J. A. Tainer, C. V. Robinson and T. O. Yeates, *Nat. Chem.* **2014**, *6*, 1065.
[142] a) S. N. S. Alconcel, A. S. Baas and H. D. Maynard, *Polym Chem-Uk*. **2011**, *2*, 1442; b) A. C. Obermeyer and B. D. Olsen, *ACS Macro Lett.* **2015**, *4*, 101.
[143] H. Murata, C. S. Cummings, R. R. Koepsel and A. J. Russell, *Biomacromolecules*. **2013**, *14*, 1919.
[144] T. Xu, N. Zhao, F. Ren, R. Hourani, M. T. Lee, J. Y. Shu, S. Mao and B. A. Helms, *ACS Nano*. **2011**, *5*, 1376.
[145] S. Zhang, M. A. Greenfield, A. Mata, L. C. Palmer, R. Bitton, J. R. Mantei, C. Aparicio, M. O. de la Cruz and S. I. Stupp, *Nat Mater*. **2010**, *9*, 594.
[146] I. Diez, H. Hahn, O. Ikkala, H. G. Borner and R. H. A. Ras, *Soft Matter*. **2010**, *6*, 3160.

[147] a) C. Mateo, V. Grazu, J. M. Palomo, F. Lopez-Gallego, R. Fernandez-Lafuente and J. M. Guisan, *Nat Protoc*. **2007**, *2*, 1022; b) J. Pedroche, M. D. Yust, C. Mateo, R. Fernandez-Lafuente, J. Giron-Calle, M. Alaiz, J. Vioque, J. M. Guisan and F. Millan, *Enzyme Microb Tech*. **2007**, *40*, 1160.
[148] Y. F. Zhang, J. Ge and Z. Liu, *ACS Catal.* **2015**, *5*, 4503.
[149] a) Y. H. P. Zhang, *Biotechnol. Adv.* **2011**, *29*, 715; b) O. Idan and H. Hess, *ACS Nano*. **2013**, *7*, 8658; c) I. Wheeldon, S. D. Minteer, S. Banta, S. C. Barton, P. Atanassov and M. Sigman, *Nat. Chem.* **2016**, *8*, 299.
[150] G. Delaittre, I. C. Reynhout, J. J. Cornelissen and R. J. Nolte, *Chemistry*. **2009**, *15*, 12600.
[151] L. Hosta-Rigau, M. J. York-Duran, Y. Zhang, K. N. Goldie and B. Stadler, *ACS Appl Mater Inter*. **2014**, *6*, 12771.
[152] L. Schoonen and J. C. M. van Hest, *Adv Mater.* **2016**, *28*, 1109.
[153] O. I. Wilner, Y. Weizmann, R. Gill, O. Lioubashevski, R. Freeman and I. Willner, *Nat. Nanotechnol.* **2009**, *4*, 249.
[154] J. E. Dueber, G. C. Wu, G. R. Malmirchegini, T. S. Moon, C. J. Petzold, A. V. Ullal, K. L. J. Prather and J. D. Keasling, *Nat. Biotechnol.* **2009**, *27*, 753.
[155] a) A. Zebda, C. Gondran, A. Le Goff, M. Holzinger, P. Cinquin and S. Cosnier, *Nat. Commun.* **2011**, *2*; b) Y. H. Kim, E. Campbell, J. Yu, S. D. Minteer and S. Banta, *Angew Chem Int Edit*. **2013**, *52*, 1437.
[156] a) Y. R. Wang, W. D. Huang, N. Sathitsuksanoh, Z. G. Zhu and Y. H. P. Zhang, *Chem Biol*. **2011**, *18*, 372; b) A. Moradian and S. A. Benner, *J Am Chem Soc.*. **1992**, *114*, 6980.
[157] C. You, H. G. Chen, S. Myung, N. Sathitsuksanoh, H. Ma, X. Z. Zhang, J. Y. Li and Y. H. P. Zhang, *Proc Natl Acad Sci U.S.A*. **2013**, *110*, 7182.

Chapter 3. Repeat-Protein Films Exhibit Hierarchical Anisotropic Mechanical Properties

N. A. Carter and T. Z. Grove, Biomacromolecules. 2015, 16, 706.

Attributions: NC conceived all experiments and performed all unless otherwise noted. Greg Fahs performed SAXS/WAXD experiments. NC and TZG wrote and edited the manuscript.

3.1. **Abstract**

Complex hierarchical structures provide beneficial structure-property relationships that can be exploited for a variety of applications in engineering and biomedical fields. Here we report on molecular organization and resulting mechanical properties of self-assembled designed repeat-protein films. Wide-angle x-ray diffraction indicates the designed 18-repeat consensus tetratricopeptide repeat protein (CTPR18) orients normal to the casting surface while small-angle measurements and electron microscopy show a through-plane transversely aligned laminar sheet-like morphology. Self-assembly is driven by the combination of CTPRs head-to-tail stacking and weak dipole-dipole interactions. We highlight the effect that this hierarchical structure has on the material's mechanical properties. We use nanoindentation and dynamic mechanical analysis to test the mechanical properties over multiple length scales, from the molecular level to the bulk. We find that morphology predictably affects the film's mechanics from the nano- to the macro-scale, with the axial modulus values ranging from 2-5 GPa. The predictable nature of the structure-property relationship of CTPR proteins and their assemblies proves them a promising platform for material engineering.

3.2. **Introduction**

Designing materials with defined, hierarchical structures is a key area of focus in both the synthetic and bio-materials fields. Hierarchical structures manifest interesting structure-property

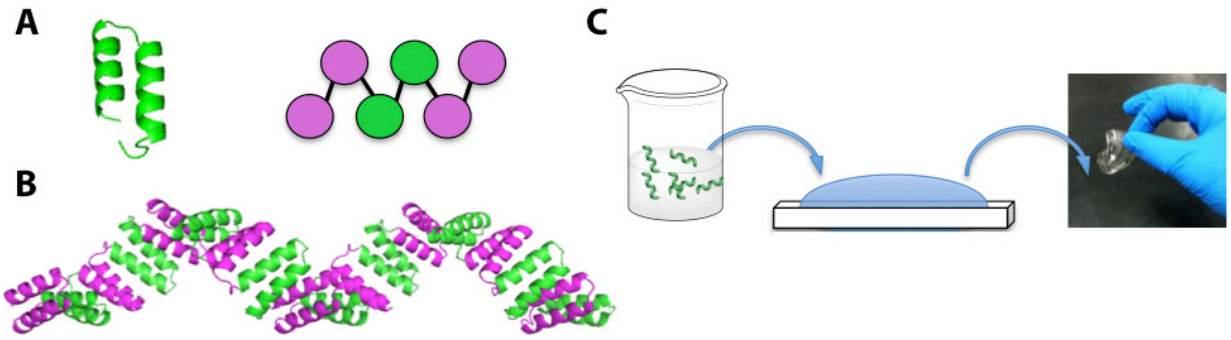


Figure 3.1. A) A single 34 amino acid helix-turn-helix motif within CTPRn. The ball and stick model to the right illustrates the predominant role that local interactions have on determining CTPR stability. The purple and green balls represent helices within neighboring tandem repeats that interact only with near neighbors. B) The cartoon representation of CTPR18 illustrates how the predominance of local interactions results in modular, extended structure of CTPRn. Image was prepared from X-ray coordinates using MacPyMol (www.pymol.org).[14e] When in tandem, eight repeats form a full super-helical turn with a pitch of 7.2 nm and width of 3.5 nm. CTPR18 is 16.2 nm in length. C) Schematic diagram of film casting. CTPR18 is solvent cast on a PTFE sheet and allowed to dry at room temperature. The resulting films are 90% transparent and uniformly birefringent.

relationships that can be tuned for a variety of applications in tissue engineering,[1] electron and ion transport,[2] or as mechanical actuators.[3] However, full realization of the properties of hierarchical architectures for applications and devices requires a good understanding of the structure-property relationships of each level in the hierarchy.

Synthetic block-co-polymers (BCPs) rely on top-down fabrication methods, where surface templating methods like chemo- and graphoepitaxy, electrospinning, or exposure to magnetic or flow fields direct long-range order.[4] Bottom-up design is a more attractive approach because it leverages the inherent ability of molecular subunits to self-assemble into larger structures removing the need for sometimes complex and costly processing methods. In recent years, self-assembly driven by specific biomolecular interactions has gained attention as a powerful approach for bottom-up fabrication and templating of nano-structured materials.[5] There are a number of examples where polymers[6], peptides[7], nucleic acids[8] and proteins[9] have all been exploited to create self-assembled nanostructures, yet there remain few instances where structural hierarchy propagates across the nano-meso gap.[9b, 10]

Commonly utilized self-assembled structures of polypeptides and proteins are based on the amyloid-like structures that can be formed from naturally occurring and designed sequences. The downside of amyloid-like assemblies is that they have limited tunability since amyloids are a generic structure that provides only one type of architecture no matter the starting material.[11] The sequence-to-structure relationship is well defined for alpha helices and helical bundles, thus allowing for a more rational approach to designing de novo alpha helical peptides that rely on the association of amphiphatic helices in the alpha-helical coiled-coil motif.[12] Assemblies of naturally occurring proteins, such as resilins and reflectins, exhibit interesting and useful functions, albeit their structures are not well understood thus precluding rational design approaches to device engineering.[13]

Most recently, designed repeat-proteins were exploited for the bottom-up assembly of ordered and functional protein films.[10] The structure of repeat proteins is characterized by a distinctive, regular pattern of contacts between adjacent residues in the primary sequence (Figure 3.1A). As a consequence, repeat proteins tend to be linear, modular, extended structures where there is a clear delineation of structurally and functionally important residues.[14] The modular architecture and tunable stability and ligand binding properties make this class of proteins an especially attractive platform for functional material design.[10, 15]

Here we use the 34 amino acid, helix-turn-helix concensus tetratricopeptide repeat (CTPR) (Figure 3.1A).[16] The simplest repeat protein system is one in which all the repeats within a protein are identical, such as the CTPRn protein array, where 'n' stands for a number of tandem repeats. Such CTPRn proteins can therefore be treated as simple homopolymeric molecules where the monomer is a single-repeat, helix-turn-helix motif, 34 amino acids in length. CTPRn arrays form a superhelical structure, where one full turn of the superhelix comprises eight repeats (n=8) (Figure 3.1B). Additionally, only eight of these 34 amino acids are conserved, allowing for facile engineering of proteins and incorporation of non-canonical functionalities.[14f, 15b, 17] CTPR proteins have been previously used to assemble hydrogels, ordered protein films, and nanometer-scale fibers, revealing their inherent self-assembling properties.[10, 15a, 18]

In the present study we examine the structure and mechanical properties of macro-scale films (centimeters in diameter) that form from the self-assembly of CTPR with 18 tandem repeats (CTPR18). The high aspect-ratio and rigid, superhelical structure play key roles in the evaporation induced self-assembly that produces macro-sized films with order that spans from protein secondary structure all the way to larger-scale, global morphologies. The resultant anisotropic morphology correlates with the mechanical properties across multiple length scales. Since morphology and mechanical properties dictate a material's usefulness in a variety of applications, such as proton or ion-transport, separation membranes, or actuators, this study is meant to provide valuable insight into the interplay between morphology and emergent collective properties.

We observe the bulk morphology of CTPR18 films using polarized light microscopy (PLM) and scanning electron microscopy (SEM) of a freeze-fracture surface. Orientation of CTPR18 molecules within the material was determined using small-angle x-ray scattering (SAXS) and wide-angle x-ray diffraction (WAXD). We tie morphology to the film's mechanical properties by testing all relevant length scales, ranging from a single molecule to the bulk. We accomplish this using nanoindentation and dynamic mechanical analysis (DMA).

3.3. **Materials and methods**

3.3.1 Protein synthesis.

Protein was expressed and purified based on previously published protocols for His_6-tagged TPR proteins.[19] The concentration of CTPR18 was determined from UV absorbance at 280 nm using extinction coefficient of 252,470 M^{-1} cm^{-1}, calculated from the amino acid composition. The protein concentration is reported as mg mL^{-1} and the MW of CTPR18 is 77068.9 g mol^{-1} as estimated using ProtParam: http://web.expasy.org/protparam/.

3.3.2 Film preparation.

Concentrated solutions of CTPR18 (~35 mg mL^{-1}) were prepared in 10 mM NaCl, 10 mM Na_2HPO4 buffer at pH 7.2. Films were cast on a PTFE mold and allowed to dry at ambient conditions.

3.3.3 Optical birefringence.

Images were obtained using a Nikon LV100N polarizing microscope with a mounted DS-Fi2 in-line camera. Films were illuminated through plane and imaged between two perpendicular plane-polarizers. Camera settings were held consistent to enable direct comparison of images.

3.3.4 Small-angle x-ray scattering.

SAXS experiments were performed using a Rigaku S-Max 3000 3 pinhole SAXS system, equipped with a rotating anode emitting X-ray with a wavelength of 0.154 nm (Cu Kα). The sample-to-detector distance was 1,603 mm. Two-dimensional SAXS patterns were obtained using a fully integrated 2D multi-wire, proportional counting, gas-filled detector with an exposure time of 16 hours. Three orthogonal directions were measured and are denoted:

Z – measured through-plane, where the incident beam is normal to the casting surface

X- measured in-plane, laid down so that the incident beam is 15^{o} from parallel with the casting surface

Y – measured in-plane, a 90^{o} rotation from X

Intensity vs. pixel and azimuthal data was obtained from 2D scattering patterns using the SAXSGUI software package. In order to correlate pixels to the momentum transfer vector q the scattering data was calibrated against a silver behenate standard. All scattering profiles were corrected for transmission and background scattering and are expressed in arbitrary, relative intensity units as a function of the scattering vector q, expressed by the following relationship:

$$\mathbf{q = (4\pi/\lambda)\ \sin(\theta)} \tag{1}$$

where, θ is half of the scattering angle (2θ) and is the wavelength (0.154 nm). Peaks greater than d = 85 nm are hidden by the beam stop.

2D scattering patterns were integrated azimuthally over a range of scattering vectors near the beam stop, from q = 0.008 to 0.03 A^{-1}. We fit a liner baseline to correct for the beam stop in order to calculate a Hermann's order parameter (f), Equation 2. The Hermann's orientation function was calculated for each of the azimuthal peaks and averaged.

$$f = \frac{3\langle \cos^2 \ \phi \rangle - 1}{2} \tag{2}$$

$$\langle \cos^2 \ \phi \rangle = \frac{\int_{x_1}^{x_2} I(\phi) \cos^2 \ \phi \ \sin \ \phi \mathrm{d}\phi}{\int_{x_1}^{x_2} I(\phi) \sin \ \phi \mathrm{d}\phi} \tag{3}$$

3.3.5 Scanning electron microscopy.

Images were taken using a Zeiss LEO 1550 SEM with a field-emission source, Everhart-Thornley detector and in-lens detector. Images were acquired using an accelerating voltage of 5 keV using a variety of magnifications and sample orientations. Freeze fractured samples were immersed in liquid nitrogen for 30 minutes and broken across a long surface. Tensile fractured samples were imaged directly after yield. All samples were sputter coated with a Au/Pd (60:40) for 30 seconds to avoid charging.

3.3.6 Nanoindentation.

Samples ranged in thickness from 80-130 μm. The through-plane Z-axis was mounted directly on the stage, while samples that were indented in-plane (X and Y) were embedded in epoxy and cut using a RMC CR-X Cryosectioning System equipped with a Diatome 35° cryo diamond knife. The in-plane samples were indented on the exposed section of the block-face of the embedded sample. The indented block-faces were examined with a HIROX KH-7700 3D digital microscope and SEM to check for brittle fracture.

Indentation was performed on a Hysitron TI 900 Triboindenter®. This instrument has a load resolution of <1 nN, displacement resolution of 0.0004 nm and a displacement drift of <0.05

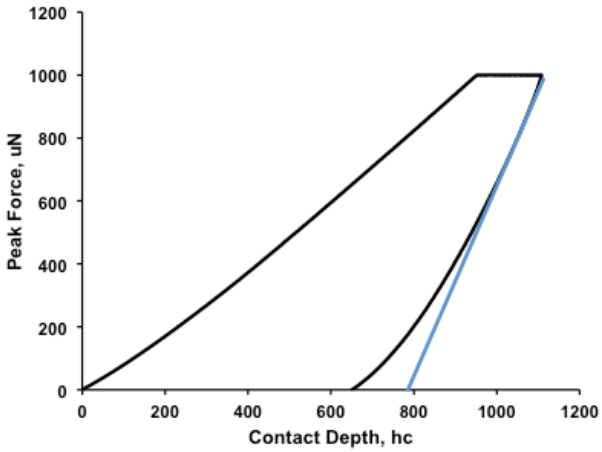


Figure 3.2. Representative load-displacement curve of CTPR18 films. Blue line indicates where S = dP/dh.

nm s^{-1}. Using a 1 μm in diameter cono-spherical tip (60°) and a trapezoidal load function, peak loads of 5-500 μN (n≥4 per data point) were applied to the films with a loading rate of 5 μN s^{-1}, hold for 10s and unloaded at 5 μN s^{-1}. While the contact depth (h_c) depends on the peak load (P_{max}) applied, the deepest indents were limited to ~1 μm. Figure 3.2 shows a characteristic load curve of the indents with no excursions that might indicate material fracture. To determine the reduced elastic modulus (E_r), we employ the Oliver and Pharr method for plastically deforming solids.[20] We assume the Poisson ratio of the film, $\upsilon_{film} = 0.3$. This method estimates E_r using Equation 4, where S = dP/dh of the unloading portion of the curve Figure 3.2, and A is the contact area. The instrument was calibrated to account for the non-linear relationship between contact area and contact depth when using a cono-spherical tip.

$$E_r = \frac{1}{2 \times 1.034} \sqrt{\pi} \frac{S}{\sqrt{A}} \tag{4}$$

$$\frac{1}{E_r} = \frac{1 - \upsilon_{film}{}^2}{E_{film}} + \frac{1 - \upsilon_{indenter}{}^2}{E_{indenter}} \tag{5}$$

The E_r is a value that encompasses the elastic modulus of both the CTPR18 film and the diamond indenter tip, Equation 5. Since the modulus of the diamond tip is so much larger, 1140 GPa, the $E_{indenter}$ is essentially zero and it is assumed that $E_r = E_{film}$.[21]

3.3.7 Dynamic mechanical analysis.

Orthogonal film sections (8 mm x 30 mm) were cut from circularly cast films, with thicknesses ranging from 80-130 μm, for testing on a TA Instruments Q800 equipped with a standard tensile geometry, with and without a relative humidity (RH) controller. Strain sweeps were run with a frequency of 1 Hz from 0.00% until failure at 0.002% increments.

3.3.8 Protein dipole calculation.

Protein dipoles were calculated from PDB files using the Weizmann Institute's Protein Dipole Moments Server. This server calculates a protein's overall dipole moment through summation of all of the dipole moments associated with the protein backbone. http://dipole.weizmann.ac.il/

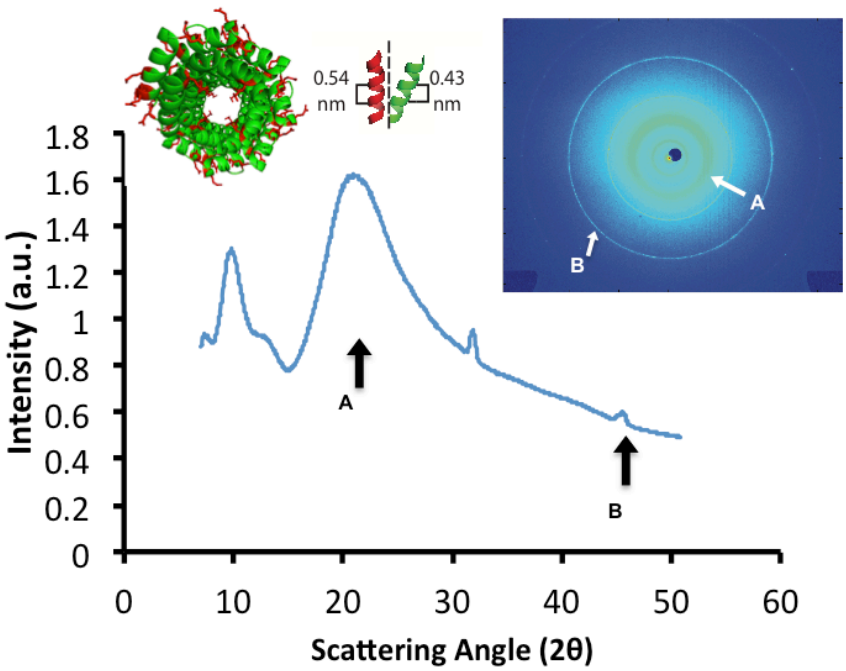


Figure 3.3. Through-plane 1D and 2D WAXD data indicates that the CTPR18 super-helical axis is oriented normal to the casting surface. Arrow A points to the characteristic 0.43 nm signal that has been associated with the alpha-helical pitch of tandem repeats that reside in the super-helix.16 Arrow B points to the 0.19 nm diffraction that results from the D and E side chains (red side chains, looking down helical axis) that are positioned on the internal and external face of the helix.

3.4. **Results and discussion**

Our strategy for the design of self-assembled CTPR films is shown in Figure 3.1C. Briefly, CTPR18 protein in 10 mM NaCl and 10 mM Na_2HPO_4 at pH = 7.2 is deposited into a Teflon mold and allowed to dry at ambient conditions. CTPR18 consists of 18 tandem repeats and assumes a superhelical structure with 2.25 turns of the super-helix. CTPR18 is an acidic protein with an isoelectric point of pI= 4.2, therefore, under casting conditions CTPR18 has an overall negative charge. To minimize protein-casting surface interactions, we chose to cast our films on Teflon (PTFE). Teflon's fluorinated surface minimizes protein-surface interactions and promotes protein-protein interactions, which aids self-assembly. The resulting films are 90% transparent (Figure S3.1) and uniformly birefringent (Figure S3.2) under full λ-plate, an indication of high-order alignment.

Previously, we reported that CTPR18 proteins self-assemble into films where individual CTPR18 molecules are aligned along a common axis.[10] Observation of uniform molecular ori-

entation in combination with birefringence indicates multi-scale order within the films. Here we hypothesize that this multi-scale order will result in multi-scale mechanical properties, thus we characterized hierarchical morphology and mechanical properties over all relevant length scales.

4.1 Defining the underlying morphology in macro-CTPR18 films.

Controlling orientation of molecules within films has great implications for the emergence of collective properties such as mechanical strength and for processes such as ion or proton transport. Here we use a combination of WAXD, SAXS, and SEM to describe the alignment of individual CTPR18 molecules and the resulting morphology of these self-assembled films.

WAXD data of CTPR18 films, Figure 3.3, was measured with the incident beam perpendicular to the casting surface, Z-axis. The most intense signal shows a d-spacing of 0.43 nm (Figure 3.3A) corresponds to the alpha-helical pitch in tandem repeats that are tilted 37° off of the superhelical axis, as observed in the high-resolution crystal structure when the helical axis is parallel with the incident beam.[14d] This suggests that the superhelical axis within the films is parallel to the incident beam and thus aligned along the Z-axis. Crystalline rings at 0.19 nm (Figure 3.3B) correspond to the distance between oxygens in the carboxylate group of aspartic (D) and glutamic (E) acid side-chains. Because of the repeated sequence and structure of CTPRs, D and E side-chains are periodically distributed along the internal and external faces of the super-helix (colored red in Figure 3.3 inset) resulting in isotropic diffraction when the CTPR18 helical axis is parallel to the incident beam. When films are rotated off-axis, the characteristic arc pattern appears, thus confirming axial alignment.[10]

Spontaneous assembly of rod-like particles in a vertical orientation in respect to the casting surface has been observed in a variety of different inorganic systems[22] but is uncommon in soft or bio-materials.[23] Baker *et al.* previously showed the importance of minimizing particle-surface interactions to promote vertical orientation of CdS nanorods.[22a] Here we note that CTPR18 films indeed are self-assembled under conditions that minimize molecule-surface interactions.

While WAXD provides good information on the orientation of individual CTPR18 proteins relative to the film, it is limited in scale to 10 nm. We use SAXS to probe the film's larger scale

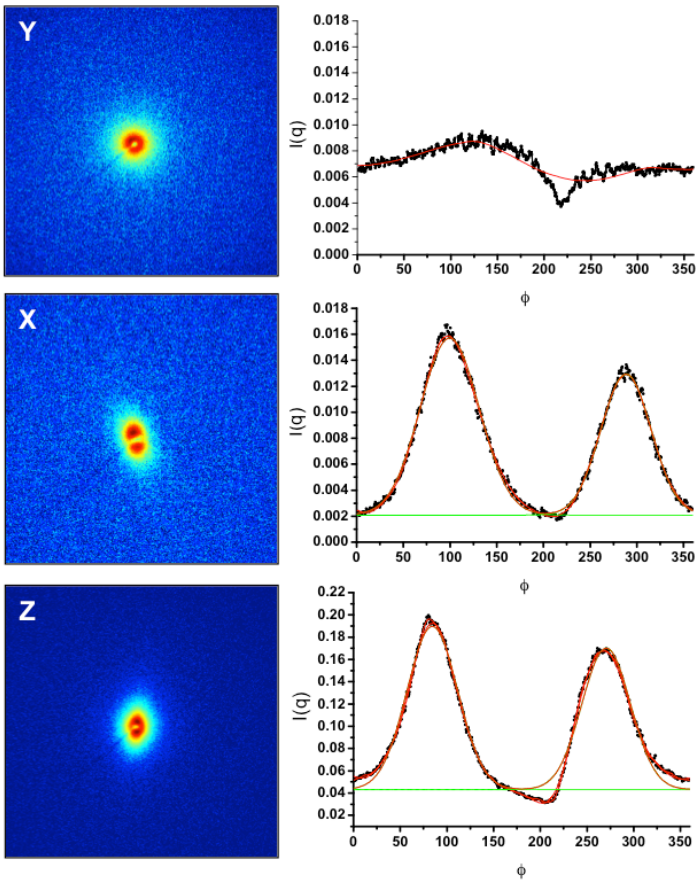


Figure 3.4. Left) 2D scattering patterns for three orthogonal axes. The circular shape of Y is isotropic while the elliptical signals of both X and Z indicate anisotropy. Right) azimuthal integration (black line) from q =0.008 to 0.03 A-1 highlights the elliptical shape of the spectra. X and Z were fit with a Gaussian function (red line) and flat baseline (green line) to correct for the dip in intensity at Φ = 220 that occurs because of the beam-stop.

morphology between 10-85 nm. The left panel in Figure 3.4 shows the 2D scattering patterns obtained from the through-plane (Z) and two orthogonal in-plane directions (X,Y). Z was collected through-plane with the incident beam perpendicular to the film, while X and Y directions were collected in-plane where the incident beam was 15° off parallel to the film's surface, rotated 90° with respect to each other. These axis designations remain constant throughout each experiment and correspond to optical birefringence images seen in Figure S3.2A-C respectively.

By linearly integrating the 2D scattering profile we can obtain a plot of intensity, I(q), vs. momentum transfer vector, q. Figure S3.3 shows a peak corresponding to 79 nm. Correcting for the 15° sample tilt gives a d-spacing of 81 nm which coincides with the length of five CTPR18s

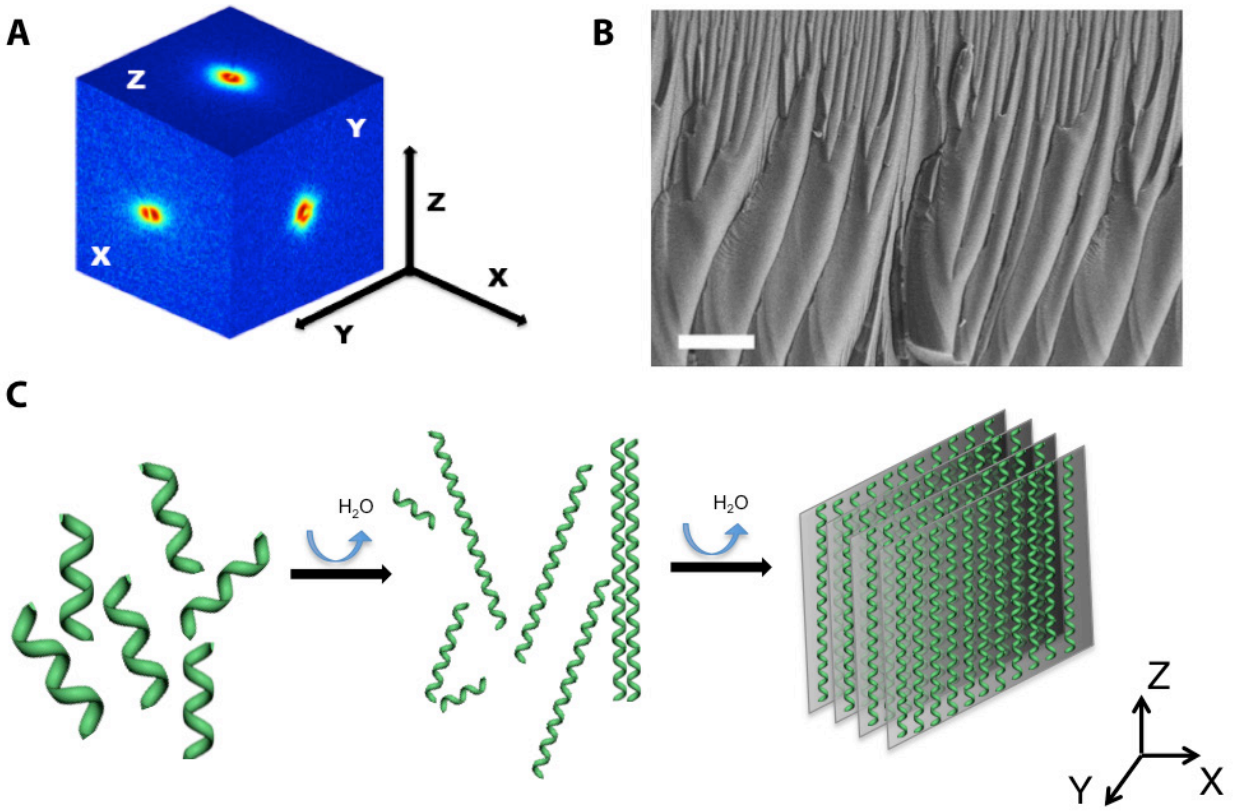


Figure 3.5. A) Cubic overlay of 2D SAXS patterns illustrate directions within the film. The through-plane (Z) and in-plane (X) are anisotropic while the in-plane (Y) is isotropic. Order on two axes while the third remains isotropic is characteristic of a lamellar or laminar morphology. Radial integration gives a d-spacing on the order of 81 nm. B) Secondary electron SEM image of a brittle freeze fracture of a CTPR18 film. The fracture surface is oriented slightly off of the X-direction so the sheet-like structure can be easily seen. Scale bar is 10 um. C) A schematic representation of the proposed lamellar-like structure formed by CTPR18 when cast in films. SAXS indicates that 5 CTPR18s stack head-to-tail to form these sheets. This is consistent with CTPR fibers in which 5 molecules are stacked head-to-tail.[15a]

stacked head-to-tail. Head-to-tail stacking is favorable in CTPR proteins and has been observed in solution, films, and fibers.[10, 14d, 15a, 18] Recently Mejias *et al.* reported on polymerized CTPR nanowires, where the average degree of polymerization was five CTPR20 molecules stacked in head-to-tail configuration.[15a]

Figure 3.4 shows the 2D SAXS profiles of all three axes. The morphology is anisotropic along the Z and X axes, as indicated by the extended ellipse pattern, while the film is isotropic along the X-axis, as indicated by the circular pattern. Order on two axes accompanied by isotropy on the

third suggests a lamellar or laminar sheet-like structure. The shapes of the 2D patterns are highlighted by the accompanying plots of I(q) vs. azimuthal angle , ϕ, which show two peaks along the X and Z axes when integrated from 0.008 to 0.03 A^{-1}. Using these data we calculated the Herman's orientation factor (f) for the X, Y, and Z planes which are -0.20 and -0.02 and -0.17 respectively. Values near -0.5 indicate the axis of orientaion is perpendicular to the reference axis. Figure 3.5A shows a cubic overlay of the 2D profiles in order to reference the SAXS measurements with respect to film's orientation.

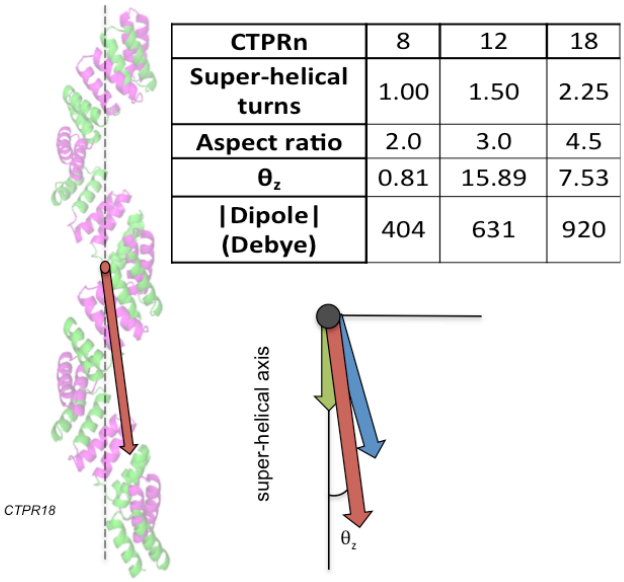
In a single crystal, CTPRn molecules are most commonly observed in a hexagonal arrangement but are also known to orient in an orthorhombic geometry.[24] In an orthorhombic lattice CTPR molecules are packed differently along the lateral coordinates perpendicular to the helical axis. This geometric arrangement might promote the laminar morphology seen in SAXS. SEM of a cryo-fractured brittle surface, along the Y-axis (Figure 3.5B), shows a repeated structure consistent with laminar sheets broken near perpendicular to their alignment axis, thus validating our interpretation of SAXS profiles. With this collection of data we propose that CTPR18 self-assembles into laminar-like sheets that are formed by head-to-tail and lateral stacking of individual super-helices. A cartoon representation of this process is depicted in Figure 3.5C.

4.2 Driving force for orientation in CTPR films.

Evaporation driven assembly is an inherently complex process because of its deviation from equilibrium.[25] Assembly is affected by a large number of factors including intermolecular interactions, molecule-surface interactions, changes in local solvent density, etc. As previously mentioned we minimize CTPR18-surface interactions by using a PTFE casting surface. This promotes assembly that is driven through intermolecular interactions between neighboring CTPR18 molecules. These interactions occur through previously described head-to-tail stacking, analogous to intra-repeat interactions that stabilize the CTPRn fold.[10, 14d, 15a, 18] The Gibbs free energy (ΔG) for this association has been approximated to be -5 kcal mol^{-1}. This value was obtained both experimentally through kinetic studies of fiber assembly[15a] and through theoretical treatments using free energy calculations accounting for burial of hydrophobic residues[15a] and using a linear 1D-Ising

model to describe CTPR's stability.[14b, e] The energy associated with head-to-tail stacking is on the same order of magnitude as hydrogen bonding. It is driven by the entropy gained through the dissociation of water from the hydrophobic surface on the exposed first and last repeats in a CTPRn molecule. Though small individually, the additive stabilization effect from multiple stacking events is significant. However, head-to-tail stacking alone doesn't explain the apparent uniform alignent of CTPR18 on a common axis.

Previous studies established that CTPRn proteins behave as rigid rods in solution, i.e. their molecular dimensions in solution are identical to those observed in high-resolution crystal structures.[14c, 19] For CTPR18 molecules with aspect ratio of 4.5, head-to-tail interactions increase the aspect ratio by a factor of a = 4.5*n, where n is the number of stacked units. Onsager's theory states that at a critical volume fraction, the entropy associated with mixing, which favors order, will predominate to compensate for the decreased excluded volume and resulting loss in orientational entropy.[26] In CTPR18 films, the existing head-to-tail stacking combined with the loss of solvent



CTPRn	8	12	18
Super-helical turns	1.00	1.50	2.25
Aspect ratio	2.0	3.0	4.5
θ_z	0.81	15.89	7.53
|Dipole| (Debye)	404	631	920

Figure 3.6. Molecular dipole (Debyes) of CTPRn change significantly in both magnitude and orientation as the CTPR super-helix is elongated. The figure shows the dipole calculated from the Protein Dipole Moment server that are overlaid at the estimated center of mass. Each dipole vector has been translated to the same plane for comparison. θz is the angle that the dipole vector deviates from the super-helical axis. The table additionally relates the number of super-helical turns to the vector of CTPRn's dipole.

due to evaporation promotes alignment of the CPTR18 molecules. While the propensity for head-to-tail stacking is a consequence of protein sequence and structure inherent to CTPR proteins, CTPR8 and CTPR12 do not form birefringent films under the same conditions. This indicates an additional parameter charactistic to CTPR structure that favors self-assembly for CTPR18 but not shorter proteins.

We propose that 3-dimensional assembly is driven by a synergistic effect of entropically favorable packing of high aspect-ratio rods and weak dipole-dipole interactions between superhelices. For anisotropic molecules, such as CTPR18, parallel orientation leads to favorable dipole-dipole interactions because it minimizes the distance between mass moments.[27] Stabilization of side-to-side packing within protein assemblies using weak dipole-dipole interactions have been well established for peptides and proteins that form alpha-helical bundles.[28]

CTPRs assume a superhelical structure, where 8 tandem repeats constitute a full turn of the superhelix. Using the Weizmann Protein Dipole Moment Server we calculated the dipole moments for CTPR8, 12, and 18 (Figure 3.6).[29] This server calculates a protein's overall dipole moment from the summation of the dipoles associated with the peptide backbone excluding side-chains. While inclusion of side chains will invariably affect the proteins dipole we accept these data as a close approximation. More importantly, given the repeat nature and superhelical structure of CTPR we suggest that the greatest effector of dipole moment is the number of complete turns of the super-helix. The table in Figure 3.6 relates the dipole moment to the number of complete superhelical turns (# = n/8). CTPRn proteins, where n/8 = integer, have a complete superhelical turn where the resulting dipole will be parallel to the superhelical axis. In the CTPRn proteins we used, the magnitude of the dipole is proportional to the aspect ratio and number of superhelical turns. CTPR18 has 2.25 complete turns, thus head-to-tail stacking will result in each molecule rotated 90° around the superhelical axis in relation to its nearest stacking neighbor. When 5 CPTR18s are stacked head-to-tail, as interpreted from SAXS, the summation of individual CTPR18 molecular dipoles gives a collective dipole 7.53° off the superhelical axis. As CTPR18 stacks assemble into sheets, a fully anti-parallel orientation of superhelices would result in a bulk cancellation of dipole moments,

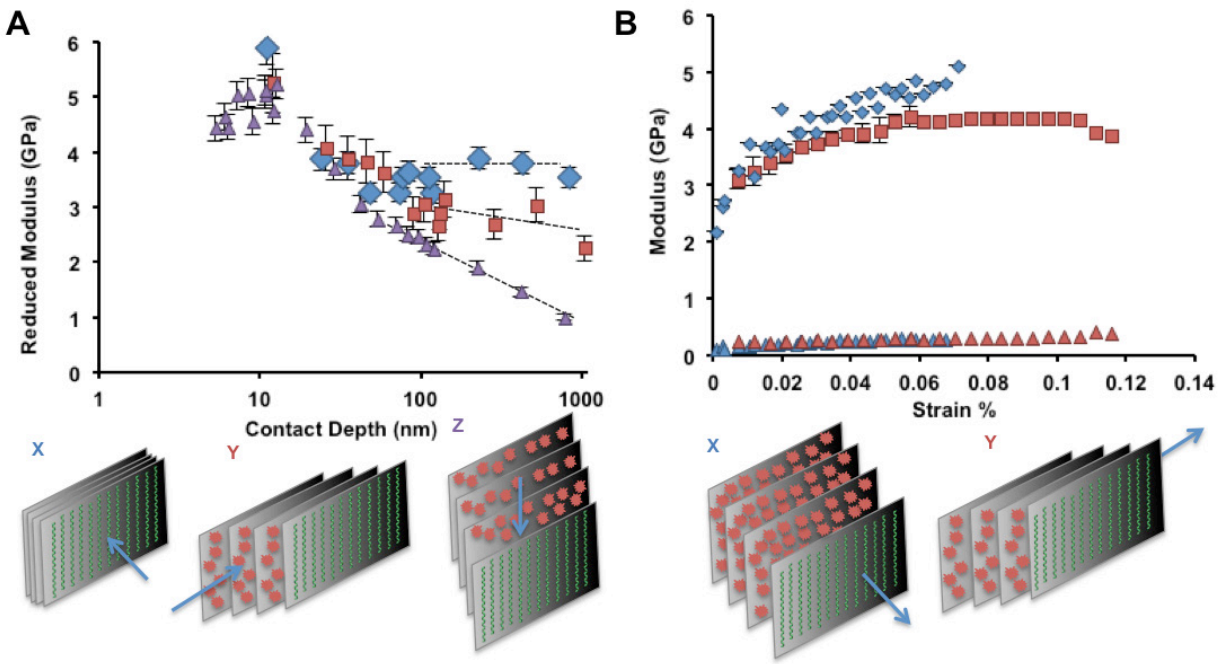


Figure 3.7. A) Nano-indentation of CTPR18 films compressed along X-, Y-, and Z-axes (blue, red, purple respectively). Major differences in modulus values and trends appear at contact depths greater than 70 nm. B) DMA (under tension) shows a marked difference in the storage moduli of the X- and Y- axes (blue and red respectively). Interestingly the loss moduli show no difference between the two axes (triangles of same color). The cartoon representations show the direction of applied compression/tension in relation to the film's underlying structure. The red 'blobs' represent the exposed surface between sheets as a result of plastic deformation.

whereas a parallel orientation of superhelices would give the film a net dipole resulting in non-linear optical properties, as observed for certain cellulose assemblies.[30] While Figure 3.5 illustrates a schematic representation of how CTPR18 stacks and assembles into sheets, further investigation into the relative orientation of dipole moments is needed.

4.3 Morphology affects mechanical properties across multiple length scales.

Intuitively yet not trivially, Dietz and coworkers showed that within single protein molecules, the secondary and tertiary structures influence the deformation response in a direction dependent manner.[31] Likewise, mechanical properties have been shown to be dependent on protein secondary structure in self-assembled amyloid fibers.[32] Although analyzing the mechanics of individual proteins and fibers gives many insights into the nano-scale phenomenon related to secondary structure, our contribution builds upon the collection of studies on bone, proteins, and other

biomaterials that show functional anisotropy is a result of the multi-scale hierarchical structures. [33] There are four discrete levels within the structural hierarchy in CTPR18 films: 1) highly regular helical secondary structure of individual CTPR repeats 2) superhelical structure of tandem repeats 3) head-to-tail stacks of CTPR18 molecules 4) laminar-like sheets. The question remains, to what degree does this structural hierarchies affect the film's properties?

In order to probe the mechanical properties of CTPR films on nano/micro scale we employed nanoindentation with the use of a Hysitron Triboindenter, equipped with a 1 μm cono-spherical tip. A triboindenter rather than an atomic force microscope is a better choice for axial compression in this case because it provides a compressive z-resolution on the order of angstroms and a large enough tip diameter to encompass all levels of the structural hierarchy orthogonal to the direction of indentation. Since CTPR18 films are brittle at room temperature we were able use a standard Oliver-Pharr model to calculate the reduced modulus, $E_{r,}$ (Figure 3.2 shows characteristic load curve). We tested the E_r as a function of contact depth, h_c (Figure 3.7A). This was accomplished using a trapezoidal load function with +/- 5 μN s^{-1} and a hold time of 10s to limit relaxation artifacts commonly seen in biomaterials.

Nanoindentation data for compression along three orthogonal directions are shown in Figure 3.7A with schematics illustrating the direction of indentation. Two regimes are visibly apparent in the semi-log plot of reduced modulus (E_r) vs. contact depth (h_c). The first regime ranges from contact depths of 5.0-65.0 nm. The Z-axis peaks at h_c=11.2 nm and then non-linearly decreases while the X- and Y-axis both non-linearly decrease in this region. When considering a single CTPR18 molecule, the simplest model would calculate the compression modulus along the superhelical axis as a function of the Hookian spring-like compression of the super-helix (7.2 nm pitch). This would be a reasonable approximation given the repeat structure of CTPRs. Considering the bulk, where a perfectly aligned system has all CTPR18s oriented perpendicular to the casting surface, the Z-axis modulus should reflect the Hookian spring-like compression, but would need to account for the compressibility of the material below, analogous to a stack of compressible springs. The modulus for compression along the X and Y-axes are more complex in that these directions

have multiple modes for plastic relaxation.

The second regime corresponds to contact depths larger than the 70 nm, close to the 81 nm d-spacing observed with SAXS. In this regime we see markedly different trends in the E_r vs. h_c of all three axes. The X-axis modulus levels off at high h_c with a magnitude of about 3.8 GPa. Whereas the Z-axis, and to a much lesser extent, the Y-axis modulus values monotonically decrease with contact depth. At these contact depths, we see a synergistic effect on the mechanical properties, mainly a function of the laminar morphology, explained in the next paragraph. We limited contact depths to around 1 μm because of the appearance of fractures in the materials surface. Indenting to high contact depths generally creates artifacts due to build-up and pile-in, therefore we explored the bulk mechanical properties along X and Y directions with dynamic mechanical analysis (DMA, Figure 3.7B) placing the films under tension. CTPR18 films were tested to failure on both the X and Y-axis (Figure 3.7B). Both the X and Y-axes are quite brittle, only withstanding 0.08 and 0.12% strain respectively.

The mechanical properties measured with both compression and tension can be rationalized by considering three modes of deformation: 1) spring-like compression of superhelical pitch, 2) slip of lamina sheets, and 3) sheet compression. The Z-axis (Figure 3.7) can deform through

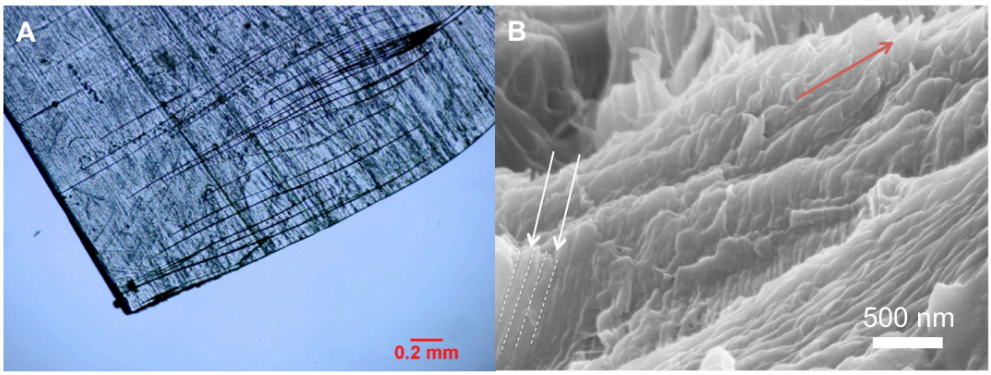


Figure 3.8. A) Optical micrograph of fracture lines in film put under to tension in the X-direction to failure. Cracks are straight and uniform in direction, and propagate along the laminar alignment axis. This fracture mechanism compares to that of oriented fibers.70 B) SEM of the fracture surface of a film put under tension in the Y-direction (red arrow indicates Z-axis). The surface shows a mostly brittle fracture with very little crazing. Aside from the apparent sheet structure the white arrows point to a repeated ~100nm structure which may be the CTPR18 stacks.

modes 1 and 2, where the combination of the elastic spring-like compression and plastic sheet-slip account for the lowest modulus that monotonically decreases with contact depth. Lamina slippage minimizes the enthalpic loss of local Van der Waals interactions where lost interactions are replaced by near neighbors. The Y-axis has the next lowest modulus due to its ability to deform through mode 2, or sheet slip, but no longer helical compression. The X-axis slightly differs between compression and tension. Compression pushes sheets together giving no mode of deformation other than buildup and brittle fracture, which results in the leveling off of the reduced modulus before fracture. Tensile stress applied along the Y-axis again gives larger modulus values across various strains to failure resulting from the greater enthalpic loss associated with pulling sheets apart rather than slipping. The inability to slip is the reason that the X-axis breaks first, where the Y-axis yields but can withstand roughly twice the strain (DMA data, Figure 3.7B).

It is interesting to note that the modulus values for both nanoindentation and DMA are in the GPa range, quite high for a non-entangled system. For comparison, modulus values of CTPR18 films are three orders of magnitude higher than covalently cross-linked resilin with a Young's modulus ranging from 0.1-3 MPa and on the same order of magnitude as physically cross-linked collagen fibers.[9b, 34] These high values suggest strong protein-protein interactions on every length scale.

Lastly we observed the fracture surfaces and crack propagation of films broken under tension to probe the effect that alignment has on the micro and bulk mechanical properties. Figure 3.8A shows an optical micrograph taken of a film tested to failure in the X-direction. While qualitative, these cracks exhibit the characteristics of a highly aligned material.[33c, 35] The linear fracture lines propagate down the lamina sheet, orthogonal to the stretching axis much as they would in reinforced polymer materials.[36] Figure 3.8B shows the SEM of the fracture surface of films tested to failure in the Y-direction (red arrows indicate Z-axis). The highly regular fracture surface shows a largely brittle fracture with little or no crazing. It also shows the laminar-like structure broken perpendicular to the alignment axis. The uniformity in crack propagation in the X-direction and regular tensile fracture in the Y-direction further points the great effect that morphology has on the properties of CTPR18 films.

3.5. **Conclusions**

The modular repeat protein CTPR18 self-assembles into highly ordered, transparent and birefringent, macroscopic films. WAXD shows that superhelical axis of CTPR18 is oriented normal to the casting surface, whereas SAXS indicates laminar morphology with d-spacing of 81 nm. CTPR18 self-assembly is driven by the propensity of CTPR for head-to-tail stacking that effectively increases aspect ratio of assembling molecules during the solvent evaporation. Self-assembly is additionally aided by favorable dipole-dipole interactions. Ergo as water evaporates and solvation decreases these weak interactions grow in effect culminating in the high, axial, mechanical strength of the dry, non-entangled CTPR18 films. Finally, we show how collective properties of CTPR18 films develop across length-scales that mirror morphological features. The multi-scale order in CTPR18 films is reflected in anisotropic mechanical properties across the nano-meso and micro/macro gap, with orientation dependent moduli ranging from 2-5 GPa.

3.6. **Acknowledgments**

This material is partially based upon work supported by the Institute of Critical Technology and Applied Science: JFP Grant 119106. This material is also partially based upon work supported by the National Science Foundation under Grant No. DMR-0923107. Nanoindentation and SEM experiments were completed at the ICTAS Nano-scale Characterization and Fabrication Laboratory (NCFL). We would like to thank Gregory Fahs for experimental assistance with WAXD and SAXS, Dr. Robert Moore and Dr. Bruce Orler for valued discussion and use of the SAXS/WAXD, DSC and DMA. We also acknowledge members of the Grove lab for careful reading of the manuscript.

Abbreviations

CTPR, concensus tetratricopeptide repeat; WAXD, wide-angle x-ray diffraction; SAXS, small-angle x-ray scattering; SEM, scanning electron microscopy.

3.7. **References**

[1] a) G. C. Engelmayr, Jr., M. Cheng, C. J. Bettinger, J. T. Borenstein, R. Langer and L. E. Freed, *Nat Mater.* **2008**, 7, 1003; b) T. Courtney, M. S. Sacks, J. Stankus, J. Guan and W. R. Wagner, *Biomaterials.* **2006**, 27, 3631; c) Y. Q. Xia, Y. F. Gu, X. Zhou, H. Xu, X. B. Zhao, M. Yaseen and J. R. Lu, *Biomacromolecules.* **2012**, 13, 2299; d) T. G. Kirn, S. H. Park, H. J. Chung, D. Y. Yang and T. G. Park, *Adv Funct Mater.* **2010**, 20, 2303.

[2] a) M. Tiitu, N. Volk, M. Torkkeli, R. Serimaa, G. ten Brinke and O. Ikkala, *Macromolecules*. **2004**, 37, 7364; b) O. Ikkala and G. ten Brinke, *Chem Comm*. **2004**, 0, 2131; c) B. G. Choi, Y. S. Huh, W. H. Hong, H. J. Kim and H. S. Park, *Nanoscale*. **2012**, 4, 5394.

[3] a) A. Buguin, M. H. Li, P. Silberzan, B. Ladoux and P. Keller, *J Am Chem Soc*. **2006**, 128, 1088; b) Y. X. Zhou, N. Sharma, P. Deshmukh, R. K. Lakhman, M. Jain and R. M. Kasi, *J Am Chem Soc.* **2012**, 134, 1630; c) M. M. Ma, L. Guo, D. G. Anderson and R. Langer, *Science*. **2013**, 339, 186.

[4] a) I. Bita, J. K. W. Yang, Y. S. Jung, C. A. Ross, E. L. Thomas and K. K. Berggren, *Science*. **2008**, 321, 939; b) S. Park, D. H. Lee, J. Xu, B. Kim, S. W. Hong, U. Jeong, T. Xu and T. P. Russell, *Science*. **2009**, 323, 1030; c) C. B. Tang, E. M. Lennon, G. H. Fredrickson, E. J. Kramer and C. J. Hawker, *Science*. **2008**, 322, 429; d) M. R. Badrossamay, K. Balachandran, A. K. Capulli, H. M. Golecki, A. Agarwal, J. A. Goss, H. Kim, K. Shin and K. K. Parker, *Biomaterials*. **2014**, 35, 3188.

[5] a) T. He, G. Abbineni, B. R. Cao and C. B. Mao, Small. 2010, 6, 2230; b) N. C. Seeman, *Nature*. **2003**, 421, 427.

[6] K. Watanabe, K. Suda and K. Akagi, *J Mater Chem C*. **2013**, 1, 2797.

[7] H. Wackerbarth, A. P. Tofteng, K. J. Jensen, I. Chorkendorff and J. Ulstrup, *Langmuir*. **2006**, 22, 6661.

[8] a) Y. He, T. Ye, M. Su, C. Zhang, A. E. Ribbe, W. Jiang and C. D. Mao, *Nature*. **2008**, 452, 198; b) P. W. K. Rothemund, *Nature*. **2006**, 440, 297.

[9] a) R. Kawamura, A. Kakugo, K. Shikinaka, Y. Osada and J. P. Gong, *Biomacromolecules*. **2008**, 9, 2277; b) T. P. J. Knowles, T. W. Oppenheim, A. K. Buell, D. Y. Chirgadze and M. E. Welland, *Nat Nanotechnol*. **2010**, 5, 204.

[10] T. Z. Grove, L. Regan and A. L. Cortajarena, J. R. Soc. *Interface*. **2013**, 10, 20130051.

[11] A. J. Baldwin, R. Bader, J. Christodoulou, C. E. MacPhee, C. M. Dobson and P. D. Barker, *J Am Chem Soc*. **2006**, 128, 2162.

[12] a) M. G. Ryadnov, D. Papapostolou and D. N. Woolfson, *Methods Mol Biol*. **2008**, 474, 35; b) C. T. Armstrong, A. L. Boyle, E. H. C. Bromley, Z. N. Mahmoud, L. Smith, A. R. Thomson and D. N. Woolfson, *Faraday Discuss*. **2009**, 143, 305.

[13] a) L. Phan, W. G. t. Walkup, D. D. Ordinario, E. Karshalev, J. M. Jocson, A. M. Burke and A. A. Gorodetsky, *Adv Mater*. **2013**, 25, 5621; b) D. D. Ordinario, L. Phan, W. G. Walkup Iv, J. M. Jocson, E. Karshalev, N. Husken and A. A. Gorodetsky, *Nat Chem*. **2014**, 6, 596; c) S. Lv, D. M. Dudek, Y. Cao, M. M. Balamurali, J. Gosline and H. Li, *Nature*. **2010**, 465, 69.

[14] a) A. L. Cortajarena and L. Regan, *Protein Sci.* **2011**, 20, 336; b) A. L. Cortajarena, S. G. Mochrie and L. Regan, *Protein Sci.* **2011**, 20, 1042; c) A. L. Cortajarena, G. Lois, E. Sherman, C. S. O'Hern, L. Regan and G. Haran, *J Mol Biol*. **2008**, 382, 203; d) T. Kajander, A. L. Cortajarena, S. Mochrie and L. Regan, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, 63, 800; e) T. Kajander, A. L. Cortajarena, E. R. Main, S. G. Mochrie and L. Regan, *J Am Chem Soc*. **2005**, 127, 10188; f) A. L. Cortajarena, T. Kajander, W. L. Pan, M. J. Cocco and L. Regan, *Protein Eng Des Sel*. **2004**, 17, 399.

[15] a) S. H. Mejias, B. Sot, R. Guantes and A. L. Cortajarena, *Nanoscale*. **2014**, 6, 10982; b) T. Z. Grove, C. O. Osuji, J. D. Forster, E. R. Dufresne and L. Regan, *J Am Chem Soc*. **2010**, 132, 14024.

[16] E. R. Main, S. E. Jackson and L. Regan, *Curr Opin Struc Biol*. **2003**, 13, 482.

[17] a) J. T. Ngo and D. A. Tirrell, *Acc Chem Res*. **2011**, 44, 677; b) M. E. Jackrel, A. L. Cortajarena, T. Y. Liu and L. Regan, *ACS Chem Biol.* **2010**, 5, 553.

[18] J. J. Phillips, C. Millership and E. R. Main, *Angew Chem Int Ed*. **2012**, 51, 13132.

[19] T. Z. Grove, J. Forster, G. Pimienta, E. Dufresne and L. Regan, *Biopolymers*. **2012**, 97, 508.

[20] W. C. Oliver and G. M. Pharr, *J Mater Res*. **1992**, 7, 1564.

[21] Y. F. Yin, L. Berglund and L. Salmen, Biomacromolecules. **2011**, 12, 194.

[22] a) J. L. Baker, A. Widmer-Cooper, M. F. Toney, P. L. Geissler and A. P. Alivisatos, *Nano Lett.* **2010**, 10, 195; b) B. D. Smith, D. J. Kirby and C. D. Keating, *Small*. **2011**, 7, 781.

[23] a) T. Jiang, C. F. Xu, Y. Liu, Z. Liu, J. S. Wall, X. B. Zuo, T. Q. Lian, K. Salaita, C. Y. Ni, D. Pochan

and V. P. Conticello, *J Am Chem Soc*. **2014**, 136, 4300; b) T. Jiang, C. Xu, X. Zuo and V. P. Conticello, *Angew Chem Int Ed*. **2014**, 53, 8367.

[24] S. Forster, A. Timmann, C. Schellbach, A. Fromsdorf, A. Kornowski, H. Weller, S. V. Roth and P. Lindner, *Nat Mater*. **2007**, 6, 888.

[25] E. Rabani, D. R. Reichman, P. L. Geissler and L. E. Brus, *Nature*. **2003**, 426, 271.

[26] H. N. W. Lekkerkerker, P. Coulon, R. Vanderhaegen and R. Deblieck, *J Chem Phys.* **1984**, 80, 3427.

[27] J. N. Israelachvili, *Intermolecular and surface forces*. **1991**, Academic Press London.

[28] a) L. Shahied, E. H. Braswell, W. M. LeStourgeon and A. M. Krezel, *J Mol Biol*. **2001**, 305, 817; b) L. Regan and W. F. DeGrado, *Science*. **1988**, 241, 976.

[29] C. E. Felder, J. Prilusky, I. Silman and J. L. Sussman, *Nucleic Acids Res*. **2007**, 35, W512.

[30] M. Zimmerley, R. Younger, T. Valenton, D. C. Oertel, J. L. Ward and E. O. Potma, *J Phys Chem B*. **2010**, 114, 10200.

[31] H. Dietz, F. Berkemeier, M. Bertz and M. Rief, *Proc Natl Acad Sc USA*. **2006**, 103, 12724.

[32] J. Adamcik, C. Lara, I. Usov, J. S. Jeong, F. S. Ruggeri, G. Dietler, H. A. Lashuel, I. W. Hamley and R. Mezzenga, *Nanoscale*. **2012**, 4, 4426.

[33] a) R. M. Kramer, W. J. Crookes-Goodson and R. R. Naik, *Nat Mater*. **2007**, 6, 533; b) G. K. Qin, P. B. Dennis, Y. J. Zhang, X. Hu, J. E. Bressner, Z. Y. Sun, W. J. Crookes-Goodson, R. R. Naik, F. G. Omenetto and D. L. Kaplan, *J Polym Sci Pol Phys*. **2013**, 51, 254; c) P. Fratzl, *Nat Mater*. **2008**, 7, 610; d) V. I. Ternovsky, Y. Okada and R. Z. Sabirov, *FEBS Lett*. **2004**, 576, 433; e) Y. Y. Wang, A. L. Agapov, F. Fan, K. L. Hong, X. Yu, J. Mays and A. P. Sokolov, *Phys Rev Lett*. **2012**, 108; f) P. W. Majewski, M. Gopinadhan and C. O. Osuji, *Soft Matter*. **2013**, 9, 7106; g) M. J. Park and N. P. Balsara, *Macromolecules*. **2010**, 43, 292; h) M. B. van Eldijk, C. L. McGann, K. L. Kiick and J. C. M. van Hest, *Top Curr Chem*. **2012**, 310, 71; i) C. H. Ye, G. Singh, M. L. Wadley, A. Karim, K. A. Cavicchi and B. D. Vogt, *Macromolecules*. **2013**, 46, 8608.

[34] G. K. Qin, X. Hu, P. Cebe and D. L. Kaplan, *Nat Commun*. **2012**, 3, Doi 10.1038/Ncomms2004.

[35] a) M. Mirkhalaf, A. K. Dastjerdi and F. Barthelat, *Nat Commun*. **2014**, 5, Doi 10.1038/Ncom-

ms4166; b) L. Han, L. F. Wang, J. H. Song, M. C. Boyce and C. Ortiz, *Nano Lett.* **2011**, 11, 3868.

[36] A. K. Subramaniyan and C. T. Sun, *Compos Part A - Appl Sci.* **2007**, 38, 34.

3.8. **Supplemental information**

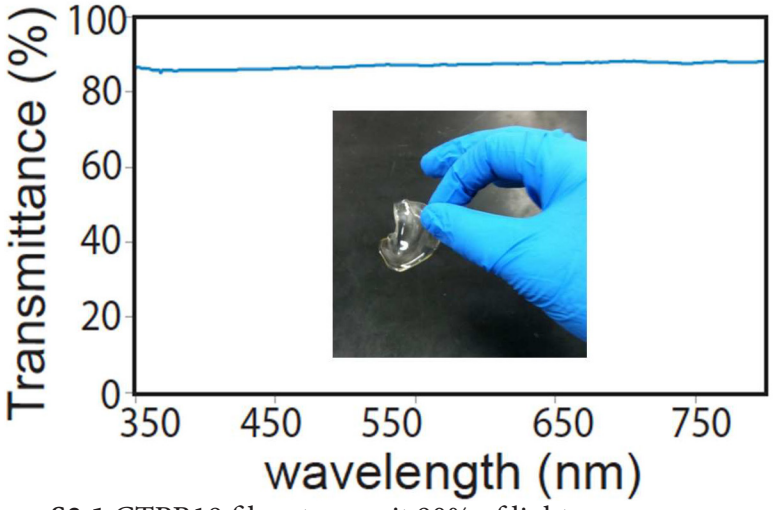


Figure S3.1 CTPR18 films transmit 90% of light.

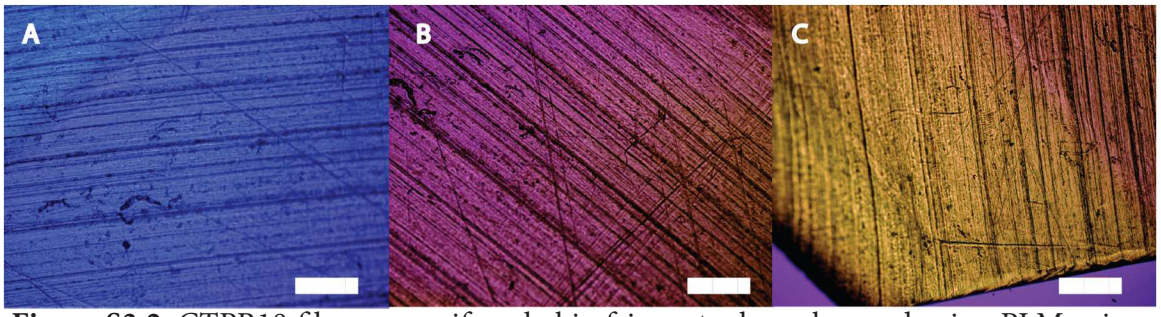


Figure S3.2 CTPR18 films are uniformly birefringent when observed using PLM using a full-lambda plate. Panels A-C show 45° rotations of the film. The scale bar is 0.2 mm.

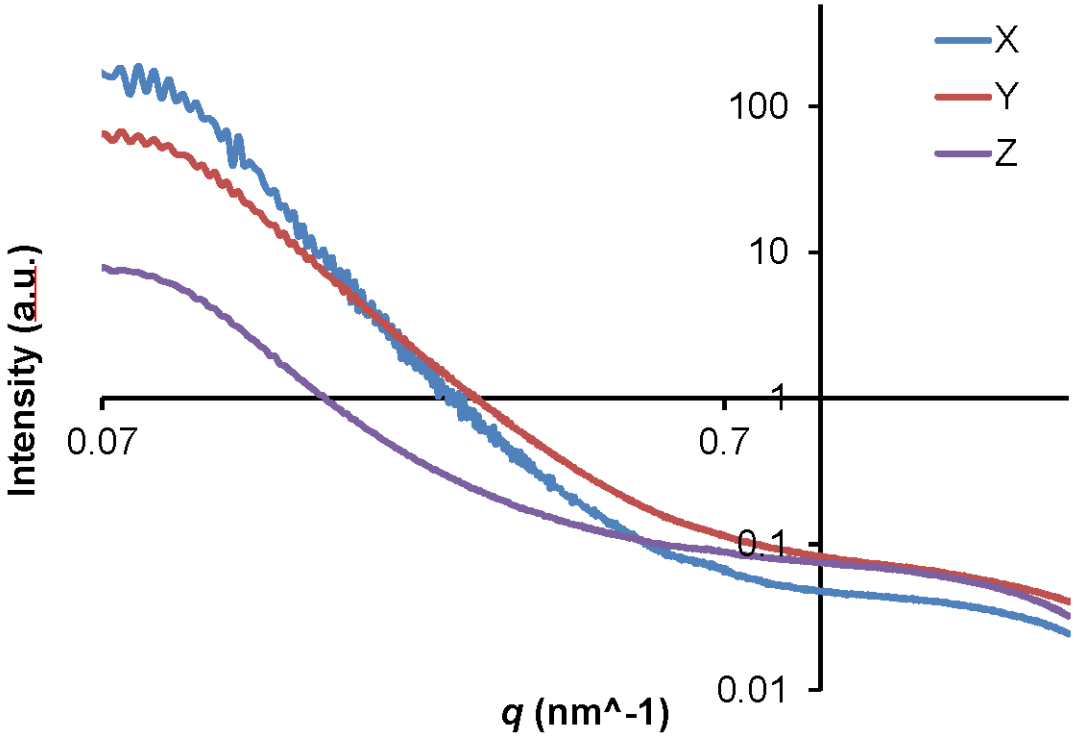


Figure S3.3 Linear integration along elliptical axis of 2D SAXS profiles all three planes.

Chapter 4. Self-assembled gradients of ionic proteins make dual-functional actuator that can bend or twist.

Nathan A. Carter and Tijana Z. Grove

Attributions: NAC and TZG conceived project. NAC performed all experiments unless noted. Dr. E. Margaretta ran TGA-SA and Dr. C. Jangu performed in-plane EIS measurements. NAC and TZG wrote and edited the manuscript.

4.1. **Abstract**

Herein, we report all-protein, humidity- and electro-active, large-displacement, fast, multi-directional and thermally stable actuators. Actuators were fabricated from the engineered repeat protein Consensus Tetratricopeptide Repeat protein (CTPR18). The actuation of CTPR18 materials is tunable using a through-thickness gradient of its underlying axial aligned morphology. This gradient structure provides the ability to select between faster, highly water-sensitive actuation or vastly increased mechanical strength by changing the morphology through-thickness penetration. Tuning of the mode of motion, e.g. bending, twisting and folding, is achieved by changing the morphological director within the actuator. Furthermore, we show that CTPR18 materials exhibit one of the highest proton conductivities seen in protein materials to date (7.1×10^{-2} S cm^{-1}). We were able to exploit this property to drive a low-voltage, ionically driven actuation and mechanically driven generation/discharge of voltage. The actuators and fabrication methods reported here provide an opportunity toward development of new materials requiring multifunctional and tunable stimuli responsiveness for healthcare, wearable devices, and microrobotics.

4.2. **Introduction**

Materials that exhibit adaptive locomotion and reversible shape change in response to specific stimuli are key components for the development of soft robotics, sensors, small-scale power

generators, or 'smart' textiles.[1] Synthetic and bio-polymer based actuators have been developed to respond to changes in electric voltage,[2] light,[3] temperature,[4] or humidity.[5] While electric fields can be generated with compact setups and are thus especially amenable to miniaturization, humidity-driven actuators are drawing special interest for the ease of access, cost, and environmentally benign nature of water vapor as the stimulus.[6]

More recently, there is a growing need for biodegradable and biocompatible responsive materials for medical applications. To that end, biopolymers such as polysaccharides, DNA, and proteins present a natural choice. However, precise control over the dynamics and modes of locomotion in hygroresponsive biopolymeric actuators still presents a considerable challenge.[7]

Biology has been a rich source of inspiration for design principles towards hygroresponsive actuation and locomotion. While biological structures exhibit a level of complexity that is extremely difficult to design, the underlying mechanisms of using water vapor to facilitate motion can be reduced to relatively simple, yet elegant, processes like buckling of perpendicular cellulose fibers, seen in the hygromorphic opening of pine cones[8] and wheat awns,[9] or the change in chiral-handedness of confined tendrils.[3a] In almost all cases locomotion is directed by hierarchical biological structures.

In the realm of synthetic materials, cross-linked liquid crystals[5c, 10] and polymeric gels[11] have been used for the fabrication of materials responsive to a variety of stimuli, including water. Cross-linked liquid crystal polymer (CLCP) materials have gained popularity as actuators due to the easily manipulated alignment orientation. For example, photoresponsive actuators were fabricated from cross-linking LC with an azobenzene moiety,[12] whereas hygroresponsive actuators were prepared from carboxylic acid containing polymers.[3b] However, tuning the rate and mode of actuation in such systems requires changing the material components.

Non-structured hygroscopic materials undergo either non-directional volumetric changes in response to changes in humidity, i.e. shrink-swell, or can produce asymmetric motion by selectively exposing one side of the material to water. [5a, 13] For instance, photocrosslinked hydrogels of poly(ethylene glycol) diacrylate (PEG-DA) could be actuated by asymmetrically applying a humid-

ity gradient.[6] More consistent anisotropic mechanical deformation has been achieved with bilayer architectures, where two layers have different responses to water and/or with a different stiffness.[14] An advantage of bilayer architecture is that the rate of locomotion can be tuned by simply changing layer thickness, whereas, dual-responsiveness can be achieved by the addition of a different responsive material as the second layer, causing deformation in the opposite direction.[14-15]. For example, hygromorhic, electrically controllable actuators were prepared by spin coating Poly(3,4-ethylenedioxythiophene):polystyrene sulfonate (PEDOT:PSS) films as an active layer onto poly(dimethysiloxane) (PDMS) as a passive layer.[14] In addition to passive bending in response to changes in relative humidity, an electrically driven gripper fabricated of this material was able to grasp and lift objects with weights comparable to its own weight.

Recently, Naumov and colleagues reported composite hygroresponsive actuators in which high directionality and complex modes of motion were achieved by incorporating glass fibers as directing and reinforcing materials into agarose gels.[7] However, a significant limitation of multilayered and composite systems lays at the layer-layer and material interfaces, the proverbial 'weak link', where delamination occurs with prolonged use.[16] Thus, materials and fabrication strategies that afford consistent macroscopic deformation in a predefined direction in a single layer actuator are preferred.[12, 17]

In this manuscript, we combined positive aspects of these two aforementioned design strategies using an aligned material and bilayer-like architecture to impart predetermined 3D motion to humidity responsive protein actuator. To that end, we have fabricated self-standing films of engineered repeat protein Consensus Tetratricopeptide Repeat protein (CTPR18). The underlying through-thickness lamellar morphology of these films facilitates hygroresponsive, large-displacement, fast, and multi-directional actuation. We additionally show that these films are highly conductive (7.1×10^{-2} S cm^{-1} @ 60 °C and 60% RH). This property has been further exploited to drive electro-responsive actuation as well as the basis for the mechanically driven generation/discharge of voltage. To our knowledge, this is the first fully protein-based actuator of this kind. Furthermore, fabrication strategies for tuning the rate and modes of locomotion are generally applicable for other

self-assembled materials. The actuators and fabrication methods described here will contribute to further development of materials requiring multifunctional and tunable stimuli responsiveness for healthcare, wearable devices, and microrobotics.

4.3. **Materials and methods**

4.3.1 Materials

All chemicals were purchased from Sigma-Aldrich or Fisher Scientific and used as received. CTPR18 was expressed as previously described.[33] Briefly, the protein was expressed in BL21 E. coli, which were grown at 37 °C to an OD 0.6-0.8. Protein expression was induced with 1mM Isopropyl β-D-1-thiogalactopyranoside for 16-18 hours at 17 °C. The his-tagged protein was purified using a Ni-NTA columns and size-exclusion chromatography. Films were cast from 10 mM sodium phosphate and 10 mM NaCl buffer into a Teflon mold. One set of films were placed in a humidity-controlled chamber (~43% RH over saturated potassium carbonate) and allowed to dry. The second set were placed in the same chamber and allowed to sit overnight. Air was then slowly purged through the chamber to speed up the drying process. After casting films were stored at ambient conditions in a petri dish until testing. Film birefringence and cross-section were qualitatively checked using a Leica polarized light, optical microscope.

4.3.2 Thermal gravimetric analysis

All samples were analyzed using a TA Instruments Q500. Non-sorption samples were run from 25 °C – 600 °C under an inert nitrogen atmosphere. Samples were heated 5 °C min^{-1}. Sorption analysis was performed using on a Q500 with the water sorption attachment at an isotherm of 30 °C (the closest stable condition to room temperature). Samples were initially equilibrated at 0% humidity until the mass was stable. Humidity was then ramped from 0% to 95% in 5% increments. Each step was allowed to equilibrate with respect to mass gained, before moving to the next humidity step.

4.3.3 Dynamic mechanical analysis

Films were tested on a TA Instruments Q800 equipped with a standard tensile geometry

and relative humidity controller. Films were stretched with a frequency of 1 Hz and 5 μm amplitude for the duration of the test. Humidity was increased at a rate of 2% min^{-1}.

4.3.4 Nanoindentation

All samples were tested using a trapezoidal load function. The loading rate and unloading rate was 5 uN min^{-1}, with a 10 second hold in between. E_r calculations were made using the Oliver and Pharr method. We assume a Poisson ratio of: $\upsilon_{film} = 0.3$. This method estimates E_r using Equation 1, where $S = dP/dh$ (unloading portion of the curve) and A is the contact area. The instrument was calibrated to account for the nonlinear relationship between contact area and contact depth when using a cono-spherical tip. The E_r encompasses the elastic modulus of both the CTPR18 film and the diamond indenter tip, Equation 2. However, since $E_{indenter} >>>> E_{film}$ (1140 GPa for a diamond tip), the second term is negligible and it is assumed that $E_r = E_{film}$ (μm).

$$E_r = \frac{1}{2 \times 1.034} \sqrt{\pi} \frac{S}{\sqrt{A}} \tag{1}$$

$$\frac{1}{E_r} = \frac{1 - \upsilon_{film}{}^2}{E_{film}} + \frac{1 - \upsilon_{indenter}{}^2}{E_{indenter}} \tag{2}$$

4.3.5 Scanning electron microscopy

All images were acquired using a Zeiss LEO 1550 SEM with a field-emission source, Everhart–Thornley and in-lens detectors. Images were acquired using an accelerating voltage between 2-5 keV. Samples were immersed in liquid nitrogen for 30 min before fracture and then sputter coated with ~3 nm of carbon prior to imaging.

4.3.6 Humidity response

All films that were used measured 60 ± 8 μm in thickness and were cut to length with a scalpel. The cut films were exposed to variable sources of humidity: 1) higher humidity testing was accomplished clamping them above a warm water bath while, 2) lower humidity gradients were produced by simple holding a finger close to the film and allowing them to actuate freely. Videos were captured using Canon DSLRs or an iPhone 6.

4.3.7 Blocking force measurement

The measurement of the blocking force of the actuators was performed using an Aurora Scientific Dual-Mode lever system. The tip of the force sensor was located 7 mm laterally from the actuator base. The sensor was operated in zero-displacement mode giving the blocking force of the actuator. The voltage readout from the instrument was calibrated against know masses to create a linear standard curve. Only data for the SLOW samples are shown, FAST films were too flexible and bent around the cantilever.

4.3.8 Electrical impedance spectroscopy

Experiments were performed on an Autolab PGSTAT 302N with a four-point electrode sample cell. AC potential was applied at 0.2 V with frequencies ranging from 0.1 Hz to 1 MHz. All measurements were taken in a controlled humidity chamber where samples were allowed to equilibrate for 120 minutes prior to measurement. Resistance (R) is determined from the x-intercept of the Nyquist plots and further used to calculate conductivity (σ) using Equation 3 below where L is the channel length and A is the cross-sectional area:

$$\sigma = \mathbf{L/A} * \mathbf{R} \quad (3)$$

4.3.9 Electromechanical testing

An Agilent 1260 wave-form generator was used to apply square potentials Amp = 200 mV to the films while an Agilent 34420A 7.5 Digit Nano Volt / Micro Ohm Meter was used to measure the voltage generated by the films.

4.4. **Results and discussion**

The CTPR18 protein is a highly-charged, rod-like biopolymer that forms macro-scale, highly ordered, lamellar films when solvent cast (Figure 4.1A). Consequently, this hierarchical, lamellar morphology has been shown to directly affect the film's mechanical behavior in a directional manner.[18] Films can be cast in a range of thicknesses, ranging from 10-150 μm, are thermally stable to 150 °C (Figure S4.1) and produce a sustained, swelling-induced actuation when placed in humid environments (Figure 4.1B). Swelling is driven by the absorption of (61 wt% water measured by TGA, Figure 4.1C), and plasticization by water (1 GPa, dry and 0.5 GPa, wet measured by DMA

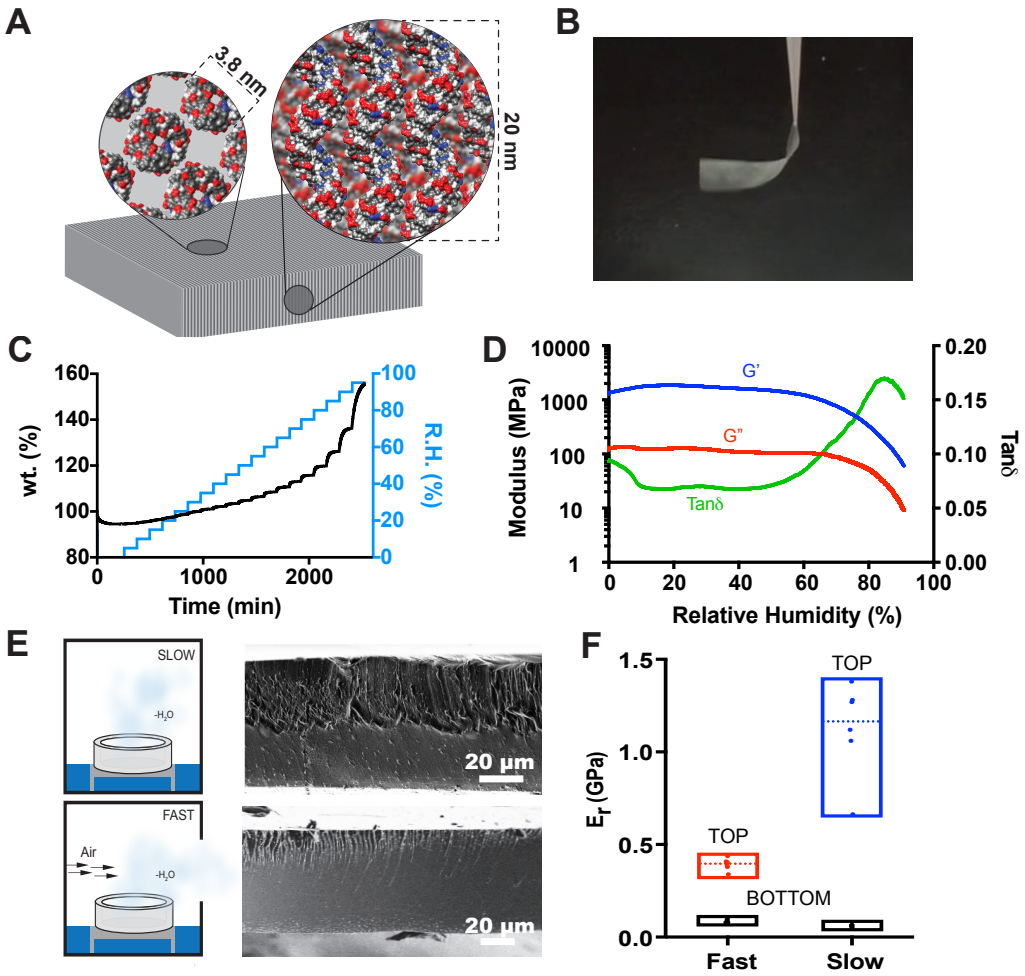


Figure 4.1. A) Schematic representation of CTPR18 lamellar orientation in films. Dimensions represent the length and width of a CTPR18 superhelix. Residues are colored by: acidic (red), basic (blue), polar (dark gray) and non-polar (light gray). B) Persisent bending of CTPR18 films when exposed to a humid environment. C) Ramping R.H.% (blue line) at an isotherm of 30°C in TGA shows CTPR18 films can absorb 61% by wt. water (black line). D) Dynamic mechanical analysis shows the films are mechanically stable across a broad range of humidities, with an order of magnitude decrease in storage modulus near the dew point. E) Schematic of the casting process in which a high-wall Teflon mold is placed in a humidity controlled chamber. Airflow is altered to increase the evaporation rate. SEM images show representative freeze-fracture, cross-sections of the resulting films. They illustrate the role of kinetic control over lamellar structure where the SLOW film has ~60% penetration through the film's thickness, in the FAST film, the hastened evaporation process limits lamellar penetration to ~30%. F) Nanoindentation measurements on the top (air-water) and bottom (casting surface) of the films show a marked difference in the resulting reduced modulus (E_r) values. Differences arise in the "Fast' vs 'Slow' top E_r because of matrix effects on the indenter tip, with average E_r values of 0.39 ± 0.3 GPa and 1.13 ± 0.23 GPa respectively. Since the 'Fast' actuator's gradient is much thinner, the indenter tip feels the lower modulus material below.

(Figure 4.1D) on the side exposed to humidity. While similar behavior was observed in nanocellulose thin films,[13] CTPR18 films reported here are thicker and actuate faster.

With this material in hand, we realized that by simply changing the evaporation rate we can alter the kinetics of self-assembly and impose a morphological gradient on the materials (Figure 4.1E). This morphological gradient endows bilayer-like properties to CTPR18 films. In this architecture, one side of film is highly aligned, whereas the opposite side is amorphous, further translating to different mechanical properties on the opposing sides (Figure 4.1F). This architecture is somewhat analogous to agarose gels reinforced by glass fibers, but in a single material, single layer actuator.

For films of fixed thickness, the rate of actuation is a function of degree to which aligned morphology penetrates into amorphous region whereas the direction of 3D motion, e.g. bending, twisting, curling, is predefined by the relative orientation of morphological features within actuator. Thus, a morphological gradient allows for simultaneous modulation of the rate and mode of actuation. This fabrication strategy further minimizes the 'weak-link' effect seen at the interfaces in bilayer and composite materials.

4.4.1. Morphology affects humidity-induced motion in CTPR18 films

CTPR18 films were drop-cast from aqueous buffer (10 mM NaCl, NaH_2PO_4, pH 7.2).[18] For a fixed size mold, thickness is controlled through CTPR18 concentration in the casting solution. Films used here are 60 ± 8 μm in thickness, as measured with a drop micrometer. Since solvent casting is a non-equilibrium process, the resulting morphology can be tuned by changing the evaporation kinetics.[19] For CTPR18 materials, we found that the evaporation rate is dependent on environmental humidity. Accordingly, two sets of actuators were assembled to illustrate the effect of gradient thickness on the actuator's properties.

Films were solvent cast in an in-house made humidity chamber at 43.16% RH (saturated potassium carbonate). The first set, were allowed to dry for 48 hrs (Figure 4.1E, SLOW), whereas the second set were purged with air after a 10 hour incubation, for a total casting time of 18 hrs (Figure 4.1E, FAST). Overnight incubation allows the initial stages of assembly to start, where a

subsequent air-purge hinders further assembly. Note that a high-wall Teflon mold is used to eliminate possible shear alignment resulting from airflow across the drying film. The air purge step has a similar effect to quenching systems below the glass transition, thus kinetically trapping structures.[20]

Analogous to bilayer architectures, a structural gradient produces defined sides (Figure 4.1E), with varying mechanical properties within the same film (Figure 4.1F).[21] Nanoindentation of the two actuators show marked differences in their surface mechanics. Indentation was performed using a 1 μm diameter cono-spherical tip (60°) and a trapezoidal load function, with a peak load of 500 μN ($n \geq 3$), a loading/unloading rate of 5 μN/s and a 10 s hold. E_r calculations were made using the Oliver and Pharr method as previously described.[18, 22]

SLOW films, ~60% lamellae though-thickness, have a top-down reduced modulus (E_r) of 1.13 ± 0.23 GPa, consistent with previously published values for aligned CTPR18 films.[18] The top-down E_r for FAST films, ~30% lamellae through-thickness, was 0.39 ± 0.3 GPa. We attribute this difference in E_r to matrix effects within the FAST film. Here, the indenter tip feels softer material below the thin gradient. In both cases, the bottom (amorphous) side of the film has a significantly reduced E_r, 0.07 ± 0.01 GPa, up to an order of magnitude softer that the aligned side.

The resulting force generated through the humidity-induced actuation for the SLOW, stiffer actuator, is significant. When tested in constant displacement mode, the actuator produced an average blocking force of 63.9 mN or 1.6×10^4 N kg^{-1}, generating force 10 times its weight (Figure 4.2A). For comparison, the per weight force generated is larger than human arm muscle (180 N kg^{-1})[17] and previously discussed PEDOT:PSS/PDMS hygromorphs from Mattoli and coworkers (118 N kg^{-1}) that were able to lift things three times their weight.[14] Similar per weight forces were generated by reduced graphene oxide gradient actuators that produced 7.5×10^5 N kg^{-1} and were able to lift objects 26 times their weight.[23] The periodicity seen in the first few cycles of the loading curve is attributed to the water absorption and evaporation as previously described.[13] As water content becomes more uniform through the film's structure, force fluctuations disappear and the swelling force increases.

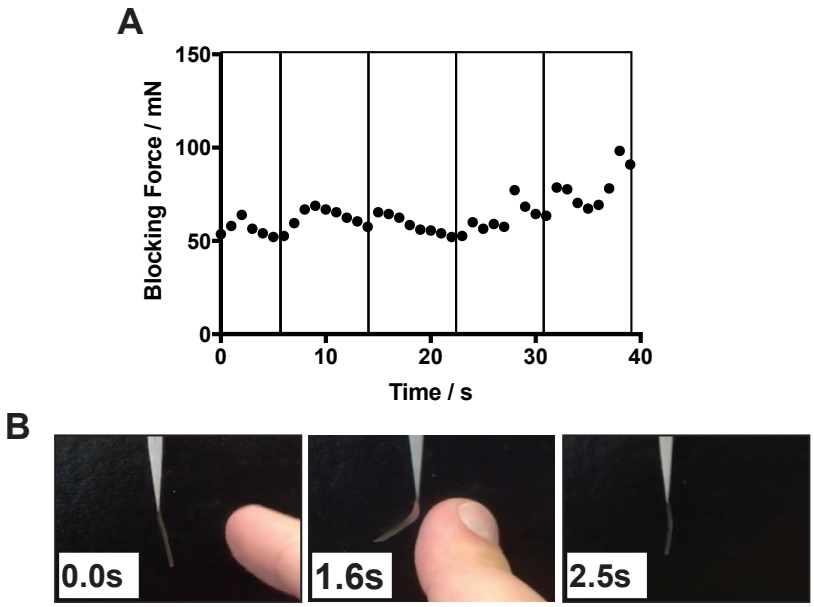


Figure 4.2. A) Blocking force measurements on the more mechanically stiff 'Slow' actuators can generate a force = $1.6x10^4$ N kg^{-1}, a force which is 10x its weight where B) the softer 'Fast' actuators show their hight humidity sensitivity, actuating in response to RH fluctuations caused by a finger.

The lower modulus, FAST actuators readily deflected around the cantilever sensor in the same experimental setup, due to their lower rigidity. However these softer films showed greatly increased sensitivity to humidity changes, compared to the SLOW actuators. Indeed, small humidity fluctuations coming from the bare skin of a finger cause the films to deflect away, but recover their original shape within two seconds (Figure 4.2B, Movie S1).

As seen in other patterned systems the actuation direction is dependent on the alignment of oriented structures within the films.[4, 24] This is also true for CTPR18 films fabricated into high aspect-ratio actuators. Actuators exhibit three distinct modes of actuation when cut 45° relative to each other from the same film: 1) Long bend (Figure 4.3A); 2) Twist (Figure 4.3B); and 3) Short bend (Figure S4.2). The films are denoted by the angle of the lamellar director relative to the long-axis of the film: 90°, 45° and 0° respectively. Bending on the short-axis illustrates that morphology, not the film shape determines the actuation mode.[7, 25]

Figures 3A and B show schematics of the long bending and twisting actuations where the inset shows a representative freeze-fracture (the viewers perspective looking down the film's long-axis). When comparing the SLOW (Figure 4.3Aii-v) and FAST (Figure 4.3Avii-x) actuators,

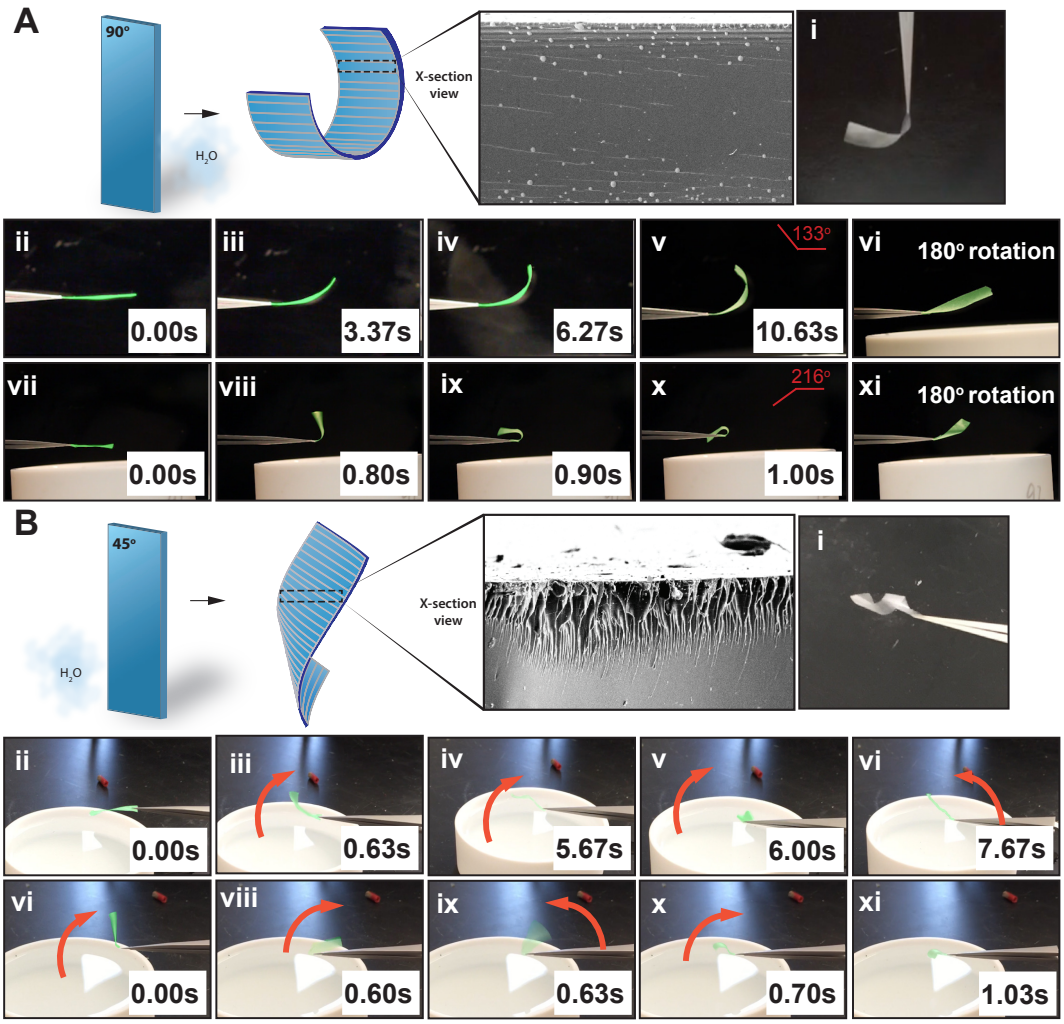


Figure 4.3. A) Actuators cut where their long-axis is perpindicular to the lamellar-axis (90°) exhibit bending, A(i). The inset shows a cross-sectional view of a freeze fractured surface, where the viewer's perspective is looking down the long axis of the film. A(ii-v) show the actuation stroke of the 'Slow' gradient where A(vii-x) show the actuation stroke of the 'Fast' gradient. A(vi & xi) show that when actuators are rotated 180° around their long-axis to expose the opposite side, they no longer exhibit the same actuation. This behavior mimics that of bilayer materials.[7] B) Actuators cut where their long-axis is 45° diagonal to the lamellar-axis exhibit twisting, B(i). B(ii-vi) show the actuation stroke of the 'Slow' gradient where B(vii-xi) show the actuation stroke of the 'Fast' gradient. Green images are false-colored for ease of viewing.

we see meaningful differences in the stroke length and speed in both orientations. The trajectory of the SLOW actuator ends with a peak tip deflection of 133° (Figure 4.3Av) where the FAST actuator has a peak deflection of 216° (Figure 4.3Ax). Moreover the deflection of the SLOW actuator is 10x slower than that of the FAST, which takes only 1.0 s (Figure 4.3Av & x).

Actuation preferentially occurs when the structured side (air-water when cast) is exposed to humidity. This is most obvious in the motion of the long bending actuator (Figure 4.3A). When films were rotated 180° around the long-axis to expose the softer side of the actuators to humidity, both SLOW and FAST actuators show significantly diminished actuation strokes (Figure 4.3Avi & xi, Movie S2 & S3). This behavior is similar to bilayer systems, which are designed to selectively absorb water on one side.[7, 14] However, here it derives from the limited ability of the softer side to compress the higher modulus portion of the film. This is further exacerbated by the plasticization and softening that occurs as the low modulus material absorbs water.

Orienting the lamellar axis diagonally (45°) to the long axis of the actuator creates the twisting motion (Figure 4.3B). Water absorption triggers a swelling-induced decrease in anisotropy between lamellar stacks, leading to contraction parallel to lamellar alignment and expansion in all other directions. Like the long bending actuators, significant differences are apparent in the rate of actuation strokes when comparing the SLOW (Figure 4.3Bii-vi) and FAST (Figure 4.3Bvi-xi) actuators.

The FAST actuator is more sensitive to changes in local humidity, seen in the first frame (Figure 4.3Bvi) as the film deflects before it is fully over the water. This is not observed for the SLOW actuator in the same position (Figure 4.3Bii). On average, the FAST actuator coils in less than 0.3 s, where the SLOW actuator coiling time ranges between 0.7-1.5 s (Movie S4). Unlike the limited motion of the long bending actuators, when the 45° films are rotated 180° the coil changed handedness (Movie S4).

One common feature between the actuation of the two films is the relatively chaotic nature of their movement. As both films move and coil, the ratio of exposed surface (top:bottom) is constantly changing. The difference in water sorption of the two sides makes the actuation stroke inherently more sporadic. The gradation of water sorption produces an actuation similar to the chaotic motion seen in light-triggered liquid crystal actuators produced by Schenning and coworkers.[3c] In this case, the chaotic motion was in part attributed to differential, local photo-softening when exposed to blue and green light. This is analogous to the differential swelling stiffness observed

within CTPR18s structural gradient.

4.4.2. Protonic conductivity gives rise to secondary functionality

We've illustrated that a through-thickness morphology gradient in CTPR18 films provides an easy avenue to manipulate their humidity sensitivity, force generation, stroke length and actuation speed. Given the highly ionic nature of CTPR18 protein we endeavored to find whether or not these materials would exhibit the proton conductivity. Ordinario and coworkers recently showed that proton conductivity of reflectin materials, 2.6×10^{-3} S cm^{-1} at 65 °C, is dependent on the content of acidic residues Glu (D) and Asp (E) in reflectin's primary sequence.[26] Interestingly, reflectin is not an acidic protein. It has an isoelectric point, pI = 8.86, with its primary sequence comprising only 9.7% acidic residues. More recently, Amdursky *et al*. showed that electrospun mats of the slightly anionic globular protein BSA (pI = 5.82, 16.3% Glu+Asp) also exhibited protonic conductivity which peaked at 4.9×10^{-5} S cm^{-1}.[27] CTPR18 is similar in molecular weight and acidic residue composition to BSA, ~69 kDa and 16.0% respectively. Yet, unlike either of the prior examples, it exhibits complex hierarchical morphologies that span nano to micro scales (Figure 4.4A). Analogous to continuous morphological phases that have been shown to greatly increase conductivity in synthetic proton conducting membranes[28], long-range alignment in CTPR18 films may contribute to conductivity of CTPR18 materials.

We probed the humidity dependent conductivity of the films using electrochemical impedance spectroscopy (EIS). The Nyquist plots were semicircular in the high-frequency region with an inclined Warburg spur in the low-frequency region. This is consistent with bulk proton impedance (semicircle) with capacitive buildup at the electrode surface (Warburg spur). Figure 4.4B shows the conductivity of CTPR18 films (thickness = 8 μm) as a function of temperature at various humidities (n=5). The observed peak conductivity was $7.1 \times 10^{-2} \pm 0.0003$ S cm^{-1} at 60 °C and 60% humidity, a full order of magnitude higher than reflectin at 90% humidity. Figure 4.4C shows the activation energies obtained from an Arrhenius-type conductivity plot (Figure 4.4B). The low activation energy of 0.22 ± 0.021 eV at 60% RH is in agreement with a Grotthuss-type conduction mechanism.[29] Similar activation energies have been measured for bulk proton conductivity in reflectin and other

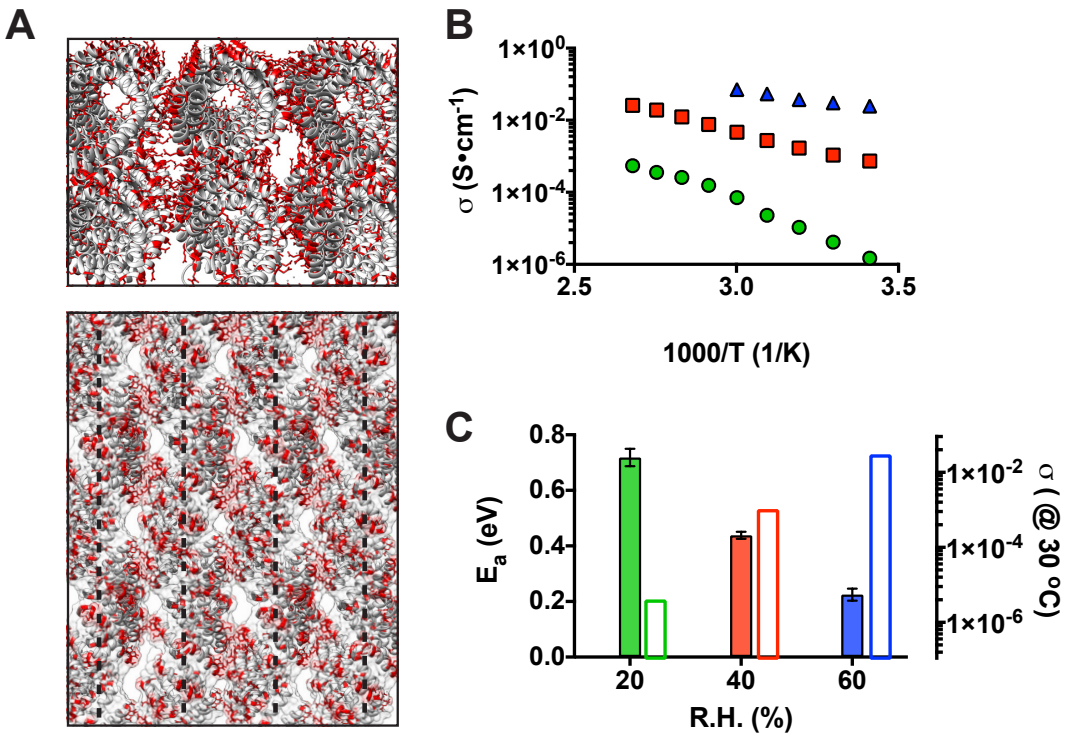


Figure 4.4. A) Cartoon schematic of aligned CTPR18 with a top-down (top panel) and side-view (bottom panel, dashed line indicates helical axis) showing the distribution of acidic residues (colored red) in CTPR18 films. This alignment provides inter- and intra-helix 'channels' of negatively charged residues (Glu and Asp) that aid in proton transduction. B) EIS measurements show the conductivity as a function of temperature at 20 (green), 40 (red) and 60 (blue) percent humidity for n=5 films. The peak observed conductivity of $7.1 \times 10^{-2} \pm 0.0003$ S cm^{-1} occured at 60% RH and 60 °C. Error bars are smaller that the points C) Activation energy (solid bars) varies inversely with humidity while conductivity increases (hollow bars). $E_a = 0.22 \pm 0.021$ eV at 60% RH is is in agreement with Grotthus-type proton conduction.[29]

protein materials.[26]

Consequently, CTPR18 films also actuate in response to applied electrical potential. Actuation was driven by a square-wave current with a 200 mV amplitude. However, actuation only occurred when current was accompanied by humidity. Initially discouraging, because this behavior was relatively indistinguishable from the previous swelling induced actuation, we found that rotating the films 180° about the long axis induced actuation in the opposite direction (i.e. towards the water), Figure 4.5A. Figure 4.5C shows the marked speed increase (t = 0.3 s) that can be imparted on the SLOW films when current is applied. Additionally the resulting tip deflection is exaggerat-

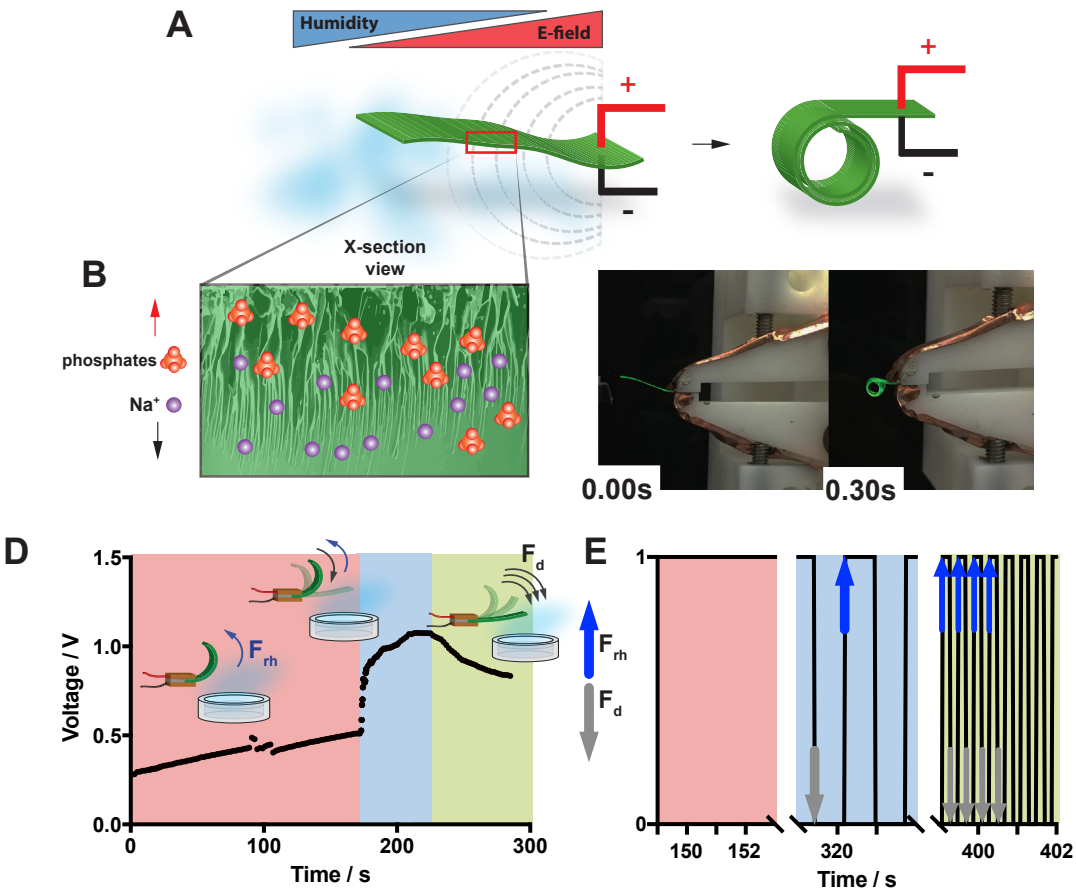


Figure 4.5. A) Cartoon schematic that shows how films exposed to humidity can be directed to curl in the opposite direction of the applied humidity. B) This actuation is driven by water plasticized ion transport across the membrane. The relative ion-size difference allow the cathode side to swell. C) Image of the current induced curling. Film is false-colored green for better visualization. D) Complementary to voltage induced motion, actuators can generate their own voltage in response to mechanical stresses (F_{rh} - swelling force & F_d - mechanically pushing down). E) A pulse-train representation of the applied forces.

ed, a full 360°; much larger than the 133° tip deflection seen in response to humidity alone.

The electro-active actuation of CTPR18 films is in large part supported by the highly ionic character of CTPR18 protein. The theoretical ion-exchange capacity of CTPR18 is 1.49 mmol g^{-1}, comparable to the gold-standard, ion conductor, Nafion® (1 mmol g^{-1}).[30] Additionally, in both polymers ion transport is related to negatively charged groups, carboxylates in CTPR18 and perfluorsulfonates in Nafion®. Correspondingly, the ion-transport mechanisms that drive actuation in CTPR18 films are likely similar to Nafion®, where both mobile ions and water molecules in their

hydration shells quickly move through films when electric potential is applied.[31]

Other work on anionic actuators have shown increased actuator performance by limiting the mobility of cations within the actuator matrix.[32] Given the complex ionic environment within the CTPR18 films, we posit that the anionic protein matrix limits the mobility of cations (e.g. Na^+), allowing larger anions (phosphates) to swell at the anode (positive electrode) when the higher modulus material is exposed to water, causing further swelling and a greater overall tip deflection. Selectively limiting cation/anion mobility opens up possibilities for future generations of actuators, actuation of which can be tuned by modulating the strength of the ionic bonds within the network, simply by changing the casting buffer salts.

Complementary to the voltage driven motion, CTPR18 also showed a capacitive effect where it generated potential upon application of mechanical stress. Swelling stress that results from absorbed water generates an upward force (F_{RH}), while stress in the opposite direction is applied by mechanically pushing down on the films (F_D). Monitoring voltage as a function of time and applied force we were able to show that these films behave like 'mechanical batteries' in that they are able to generate, briefly hold and discharge their own charge in response to opposite mechanical forces.

Since charging occurs through ionic motion, we tested the effect that the frequency of applied forces had on the charge/discharge profile. Three regimes are visible in Figure 4.5D. Initially, a constant swelling force pushes on the film generating a linear increase in voltage (t = 0-171s). Since humidity cannot be quickly cycled, we applied low frequency FD to load and release the film (t = 172-270s). The acceleration upwards was instantaneous and caused a steep increase in the stored voltage, peaking around 1.15V. To counteract the effect of FRH we increased the frequency of the applied FD four-fold (t = 271-end) in an attempt to make the sum force applied in the downward direction. The high-frequency of the FD changed the current and discharged the actuator as seen in the exponential decay in the third regime. Figure 4.5E shows the frequency as a pulse-train of applied forces.

4.5. **Conclusion**

In conclusion, we introduce the first protein actuator of its kind that predictably actuate in response to humidity, actuates through applied electric voltage, and can be used as a mechanical battery that can generate, hold and discharge its own potential. We show that the actuation of CTPR18 films is tunable by imparting a through-film gradient to the underlying axially aligned morphology. The gradient architecture provides the ability to select between faster, highly water-sensitive actuation or vastly increased mechanical strength (1.6 $x10^4$ N kg^{-1}), by changing the alignment through-thickness penetration. Further tuning arises from the ability to choose the mode of motion, e.g. bending, twisting and folding, by changing the morphological director within the film. This further combines with gradient structures for an additive effect. Taking into account these general design principles that exploit the kinetic nature of self-assembly, other ordered systems may benefit from through-thickness gradient morphologies as a mechanism to tune the actuator properties.

4.6. **Acknowledgments**

We would like to thank MII and Prof. T. Long's lab for the use of instruments, specifically Dr. E. Margaretta for running TGA-SA on films and Dr. C. Jangu for performing in-plane EIS measurements. Nanoindentation and SEM experiments were completed at the ICTAS Nanoscale Characterization and Fabrication Laboratory (NCFL).

4.7. **References**

[1] a) L. Hines, K. Petersen, G. Z. Lum and M. Sitti, Adv Mater. 2016, 10.1002/adma.201603483 b) Q. M. Zhang and M. J. Serpe, ChemPhysChem. 2017, 10.1002/cphc.201601187
[2] a) Y. Yan, T. Santaniello, L. G. Bettini, C. Minnai, A. Bellacicca, R. Porotti, I. Denti, G. Faraone, M. Merlini, C. Lenardi and P. Milani, Adv Mater. 2017, 10.1002/adma.201606109, 1606109; b) L. Kong and W. Chen, Adv Mater. 2014, 26, 1025.
[3] a) M. Wang, B.-P. Lin and H. Yang, Nat Commun. 2016, 7, 13981; b) L. T. de Haan, V. Gimenez-Pinto, A. Konya, T. S. Nguyen, J. M. N. Verjans, C. Sanchez-Somolinos, J. V. Selinger, R. L. B. Selinger, D. J. Broer and A. P. H. J. Schenning, Adv Funct Mater. 2014, 24, 1251; c) K. Kumar, C. Knie, D. Bleger, M. A. Peletier, H. Friedrich, S. Hecht, D. J. Broer, M. G. Debije and A. P. H. J. Schenning, Nat Commun. 2016, 7; d) S. Iamsaard, S. J. Aßhoff, B. Matt, T. Kudernac, J. L. M. CornelissenJeroen, S. P. Fletcher and N. Katsonis, Nat Chem. 2014, 6, 229; e) L. T. de Haan, C. Sanchez-Somolinos, C. M. W. Bastiaansen, A. P. H. J. Schenning and D. J. Broer, Angew Chem Int Edit. 2012, 51, 12469.

[4] J. Deng, J. F. Li, P. N. Chen, X. Fang, X. M. Sun, Y. S. Jiang, W. Weng, B. J. Wang and H. S. Peng, J Am Chem Soc. 2016, 138, 225.
[5] a) M. M. Ma, L. Guo, D. G. Anderson and R. Langer, Science. 2013, 339, 186; b) H. Arazoe, D. Miyajima, K. Akaike, F. Araoka, E. Sato, T. Hikima, M. Kawamoto and T. Aida, Nat Mater. 2016, 15, 1084; c) L. T. de Haan, J. M. N. Verjans, D. J. Broer, C. W. M. Bastiaansen and A. P. H. J. Schenning, J Am Chem Soc. 2014, 136, 10585.
[6] C. Lv, H. Xia, Q. Shi, G. Wang, Y.-S. Wang, Q.-D. Chen, Y.-L. Zhang, L.-Q. Liu and H.-B. Sun, Adv Mater Inter. 2017, 10.1002/admi.201601002, 1601002.
[7] L. Zhang, S. Chizhik, Y. Wen and P. Naumov, Adv Funct Mater. 2016, 26, 1040.
[8] E. Reyssat and L. Mahadevan, J. R. Soc. Interface. 2009.
[9] R. Elbaum, L. Zaltzman, I. Burgert and P. Fratzl, Science. 2007, 316, 884.
[10] R. R. Kohlmeyer and J. Chen, Angew Chem Int Edit. 2013, 52, 9234.
[11] a) Y. W. Hu, J. S. Kahn, W. W. Guo, F. J. Huang, M. Fadeev, D. Harries and I. Willner, J Am Chem Soc. 2016, 138, 16112; b) H. Zhang, A. Mourran and M. Moller, Nano Lett. 2017, 17, 2010; c) E. Borre, J. F. Stumbe, S. Bellemin-Laponnaz and M. Mauro, Angew Chem Int Edit. 2016, 55, 1313.
[12] Y. Liu, B. Xu, S. Sun, J. Wei, L. Wu and Y. Yu, Adv Mater. 2017, 29, 1604792.
[13] M. Wang, X. Tian, R. H. A. Ras and O. Ikkala, Advanced Materials Interfaces. 2015, 2, 1500080.
[14] S. Taccola, F. Greco, E. Sinibaldi, A. Mondini, B. Mazzolai and V. Mattoli, Adv Mater. 2015, 27, 1668.
[15] a) T. Lan, Y. Hu, G. Wu, X. Tao and W. Chen, J. Mater. Chem. C. 2015, 3, 1888; b) X. Chen, L. Mahadevan, A. Driks and O. Sahin, Nat Nano. 2014, 9, 137.
[16] a) H. Annepu and J. Sarkar, Phys Rev E. 2012, 86; b) S. Yang, K. Khare and P. C. Lin, Adv Funct Mater. 2010, 20, 2550.
[17] J. Mu, C. Hou, B. Zhu, H. Wang, Y. Li and Q. Zhang, Sci Rep-Uk. 2015, 5, 9503.
[18] N. A. Carter and T. Z. Grove, Biomacromolecules. 2015, 16, 706.
[19] K. Roger, M. Liebi, J. Heimdal, Q. D. Pham and E. Sparr, Proc Natl Acad Sci. U.S.A. 2016, 113, 10275.
[20] J. Kunzelman, B. R. Crenshaw and C. Weder, J Mater Chem. 2007, 17, 2989.
[21] D. D. Han, Y. L. Zhang, Y. Liu, Y. Q. Liu, H. B. Jiang, B. Han, X. Y. Fu, H. Ding, H. L. Xu and H. B. Sun, Adv Funct Mater. 2015, 25, 4548.
[22] a) Y. F. Yin, L. Berglund and L. Salmen, Biomacromolecules. 2011, 12, 194; b) W. C. Oliver and G. M. Pharr, J Mater Res. 1992, 7, 1564.
[23] J. K. Mu, C. Y. Hou, B. J. Zhu, H. Z. Wang, Y. G. Li and Q. H. Zhang, Sci Rep-Uk. 2015, 5.
[24] X. G. Gu, Q. X. Fan, F. Yang, L. Cai, N. Zhang, W. B. Zhou, W. Y. Zhou and S. S. Xie, Nanoscale. 2016, 8, 17881.
[25] a) G. Stoychev, S. Zakharchenko, S. Turcaud, J. W. C. Dunlop and L. Ionov, ACS Nano. 2012, 6, 3925; b) I. S. Chun, A. Challa, B. Derickson, K. J. Hsia and X. L. Li, Nano Lett. 2010, 10, 3927; c) P. Cendula, S. Kiravittaya, I. Monch, J. Schumann and O. G. Schmidt, Nano Lett. 2011, 11, 236.
[26] D. D. Ordinario, L. Phan, W. G. Walkup, J. M. Jocson, E. Karshalev, N. Husken and A. A. Gorodetsky, Nat Chem. 2014, 6, 597.
[27] N. Amdursky, X. Wang, P. Meredith, D. D. Bradley and M. M. Stevens, Adv Mater. 2016, 28, 2692.
[28] J. Li, J. K. Park, R. B. Moore and L. A. Madsen, Nat Mater. 2011, 10, 507.
[29] N. Agmon, Chem Phys Lett. 1995, 244, 456.
[30] J. K. Park, J. Li, G. M. Divoux, L. A. Madsen and R. B. Moore, Macromolecules. 2011, 44, 5701.

[31] J. K. Park, P. J. Jones, C. Sahagun, K. A. Page, D. S. Hussey, D. L. Jacobson, S. E. Morgan and R. B. Moore, Soft Matter. 2010, 6, 1444.
[32] O. Kim, H. Kim, U. H. Choi and M. J. Park, Nat Commun. 2016, 7.
[33] A. L. Cortajarena, T. Kajander, W. L. Pan, M. J. Cocco and L. Regan, Protein Eng Des Sel. 2004, 17, 399.

4.8. **Supplemental Information**

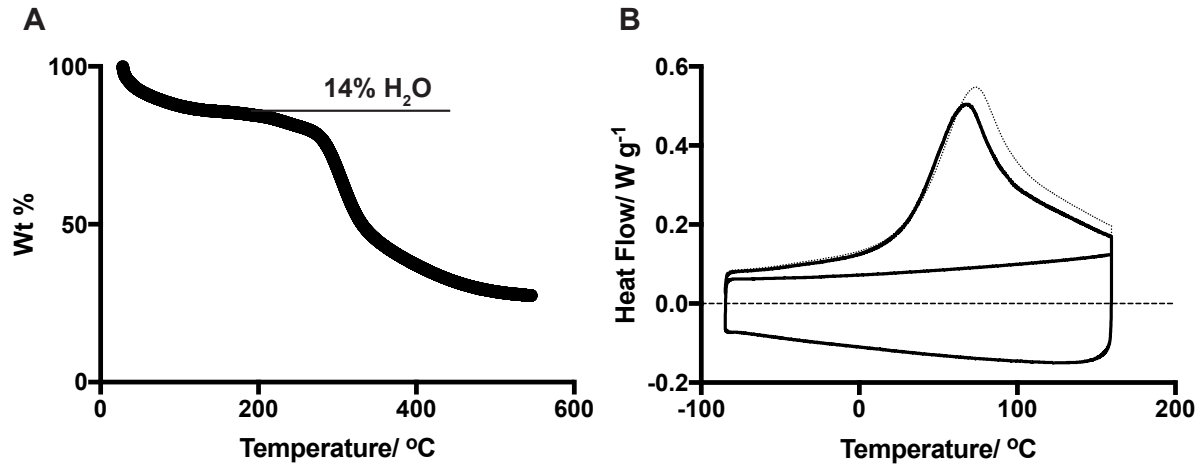


Figure S4.1 - A) TGA shows 14% mass loss up to 200 ºC. This is attributed to the loss of free and bound water. B) The first heat in DSC shows an endotherm resulting from the loss of water. This peak is recoverable by setting the film out in environmental conditions for an hour. No other transitions can be seen, and the films remain birefringent after heat treatment. This shows the films are thermally stable to at least 150 ºC.

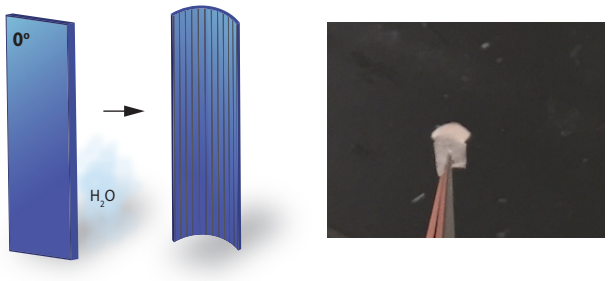


Figure S4.2 - Short-axis bending

Chapter 5. Shape control of 2D protein nanosheets using monovalent salt modifiers

Nathan A. Carter, Christopher Winkler, and Tijana Z. Grove

(Manuscript in preparation)

Attributions: NAC is the primary author of the manuscript who initiated the project and performed all experiments. CW assisted with TEM characterization, specifically SAED.

5.1 Abstract

Natural self-assembled structures are some of the most complex and diverse functional materials in existence. Scientists have long since explored the underlying mechanisms behind these assembly processes with the goal to rationally control the resulting structures. In recent efforts, researchers have begun to explore the self-assembly of protein-based two-dimensional (2D) nanomaterials as a natural extension of 0D fibers and 1D tubes. 2D materials have shown much promise in the realm of biosensing, catalysis, and drug delivery. While there are many examples of synthetic materials, to date there are few examples of self-supported 2D assemblies of proteins. Of these examples, many focus on protein interfaces that are designed around symmetrical building blocks. Here, nanomaterial shape is dictated by symmetric binding interactions where the result is: one morphology from one building block. While efficient at times, this strategy creates single-use molecules, where redesign is necessary for each new structure. Moreover, computational interface design may exclude a number of functional proteins that are impractical to genetically engineer.

Here we present a more universal, physical approach to tuning interactions at protein interfaces, thus imparting control over assembly. By simply varying the ionic content of monovalent cations we illustrate the ability to modify protein self-assembly at both hydrophobic and electrostatic driven interfaces. We use the modular 18-repeat, consensus tetratricopeptide repeat protein (CTPR18) as a model protein. CTPR18 self-assembles into anisotropic nanotapes driven by two perpendicular processes: 1) head-to-tail hydrophobic packing and 2) side-to-side electrostatic interactions. Herein, we showed the ability to promote directional growth, i.e. selecting

hydrophobic vs. electrostatic, by changing the ionic components in solution. From a single building block we illustrated the ability of monovalent cations to create different morphologies: nanotapes, nanowires, 'footballs' and short aspect-ratio tubes.

5.2 Introduction

Biological assemblies are the most complex structures in nature. From a relatively small set of building blocks, proteins fold and assemble into larger scale structures, which impart vastly different and complex functionalities. Nature assembles these immensely complex hierarchical materials, which carry out every function of life, in three-dimensions with good speed and high fidelity. For example the tobacco mosaic virus is a linear structure that comprises 2130 identical coat proteins,[1] or the cantenane HK97 which comprises 60 hexamers and 12 pentamers to assemble a much larger encapsulated structure.[2] These assemblies are driven by specific interfacial interactions that scientists have long sought to mimic and control. While the precise control of hierarchical structures in the third dimension remains an elusive target and the ultimate goal, interest in self-assembled two-dimensional (2D) materials has risen in the recent past due to their potential applications in biosensing, catalysis, energy harvesting, drug delivery, and cell culture. Additionally, 2D nanomaterials bring a wide variety of functionality in areas such as nanoelectronics and drug delivery because of their efficiency in loading and patterning due to their high relative surface area. As such, well-defined self-assembling materials are of special interest because modifications to the assembling subunits may afford control over material properties.

There are many examples of simple biomolecules self-assemble. DNA origami affords great structural control, but is severely limited by the high negative charge and lack of reactive chemistries to readily modify DNA.[3] Peptides as small as two and three amino acids long have been shown to make a variety of nanomaterials and gels.[4] Self-assembling larger peptides often fit into a two categories: 1) alpha-helical and amyloid-like, beta-sheet assemblies. The Woolfson lab and others have demonstrated the ability to rationally design *de novo* peptide sequences to form fibers, cages and other assemblies.[5] Amyloids have especially proven accessible to researchers, because beta-

strands tend to aggregate creating filamentous structures. In one of the only examples of designed protein materials to show hierarchical order Knowles and coworkers utilized amyloid-like peptides to create macro-scale birefringent films.[6] While established in the literature, these peptide structures are relatively simplistic and limited in function when compared to the possibilities of self-assembled protein materials.

Toward creating regular, controllable protein assemblies, groups have utilized cross-linking strategies, based on metal chelation,[7] disulfide-bond formation[8] or on ligand interactions.[9] Recent work has focused on computational design of protein-protein interfaces of symmetric molecules. In a tour de force example showing the advantages of computational protein design Gonen *et al.* designed more than 60 proteins to promote lateral assembly of circularly symmetric protein oligomers. They illustrated this design strategy could quickly and reliably produce assemblies with a variety of architectures, exhibiting wallpaper groups: p321, $p42_12$, and p6.[10] However, while these materials readily assemble, researchers have yet to exhibit any kind of shape control over the resulting assemblies. This is a far cry from the control exhibited by extended, linear, TMV assemblies. One contrary example, recently published by Brodin *et al*, shows the ability to modulate the diameter of self-assembled nanotubes comprising the D_2 symmetric protein cytochrome cb_{562} by changing the solvent conditions and timing of Zn^{2+} additions.[11] With assemblies based on designed interface a general rule of thumb is one design, one morphology. Thus a more universal physical method to modulate protein interfacial interactions, akin to crystal modifiers used in inorganic crystallization, processes would be advantageous.[12]

Herein, we use the model protein, consensus tetratricopeptide repeat protein (CTPR) as the modular template to explore interfacial interactions in self-assembly. Repeat-proteins carry multiple advantages over globular proteins seen in many of the designs above. Their structure comprises covalently linked 'monomer' motifs of secondary structural elements. These motifs are stabilized by only local interactions; that is to say, interactions occurring only between residues close in the primary sequence. The result is that each 'monomer' motif may be individually modified without affecting the global architecture. CTPR has a 34 amino acid repeating sequence in which only 8 are

structurally conserved, making it an ideal candidate for bioengineering.[13] CTPR has been shown to assemble into 2D and 3D arrays, with higher order structures such as those seen in hierarchical lamellar films,[14] nanofibers,[15] monolayers,[16] and most recently nanotubes.[17]

In these works, assembly is mainly driven by the same interactions seen in CTPR crystallization processes. Head-to-tail (HTT) stacking between proteins is driven by hydrophobic interactions akin to the inter-repeat stacking, which occurs between neighboring tandem repeats in CPTR. Side-to-side stacking (STS) is mediated by electrostatic interactions between neighboring proteins. In the most recent example, Sanchez-deAlcazar *et al.* elegantly illustrated the ability to rationally design nanotubes by introducing new side-to-side interfaces, exhibiting three types of interactions: hydrophobic, electrostatic and π-π interactions.[17] While the ability to genetically modify protein interfaces has been shown to be an effective way to direct protein self-assembly, this strategy may not be amenable to all proteins. As a method to explore a more universal method to affect protein assembly, we hypothesized that changing solvent conditions might be able to promote

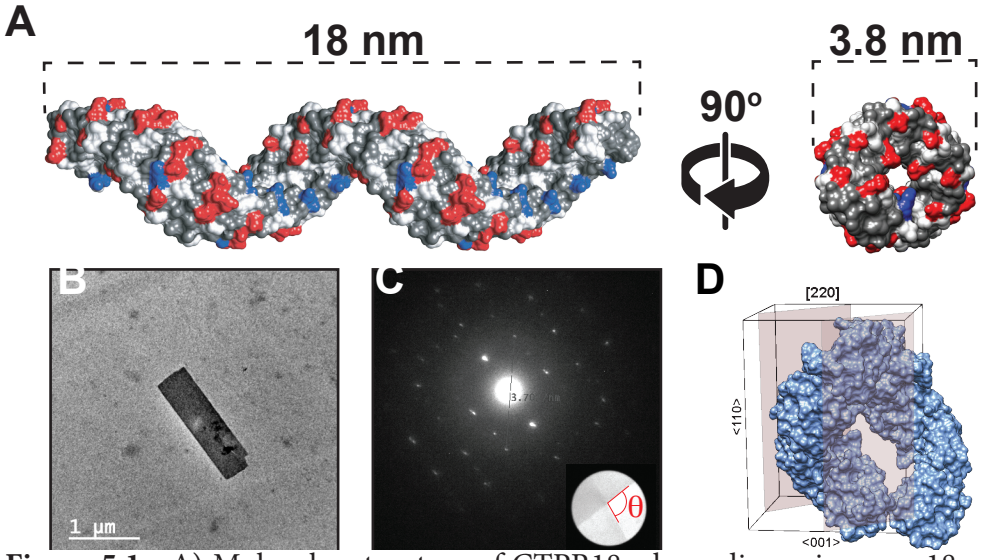


Figure 5.1 - A) Molecular structure of CTPR18 whose dimensions are 18 x 3.8 nm. The surface is colored to show acidic residues (red), basic residues (blue), polar uncharged (dark gray) and non-polar (light grey). B) TEM image of slef-assembled CTPR18 nanotape. D) Single-area electron diffraction of two nanotapes overlayed at an angle, θ = 97°. The diffraction patterns confirm nanotapes are single crystals with two-fold symmetry, Pmm2. D) This may correspond with the two-fold axis of the tetragonal $P4_12_12$ crystals (PDB: 2AVP) seen in CTPR20 and CTPR8.[9]

either HTT or STS growth relative to the other.

As proteins are relatively intolerant toward organic solvents we elected to use aqueous solutions in which we varied the salt content. We ignored multivalent ions as they have been well utilized in protein self-assembly to effectively create coordination cross-links between neighboring proteins. Rather, we were more interested the effects of cations, because CTPR18 has a high density of acidic residues on its surface. At pH 7.2 CTPRs 99 acidic residues (Figure 5.1A, red) and 27 basic residues (Figure 5.1A, blue) give the protein an ideal isoelectric point of 3.96 (calculated from using the ExPasy ProtParam tool), making it highly negatively charged. Additionally, cations have been shown to have a lesser effect on the stability of folded proteins, when compared to their anionic counterparts, and may thus be suited for fine-tuning interfacial interactions that drive self-assembly without affecting the protein monomer's structure.[18] Since CTPR's epitaxial assembly is driven by two differing mechanisms, hydrophobic HTT and electrostatic STS, we hypothesized that selecting salts of varying size, charge character, carboxylate binding strength and aromatic binding strength would effectively change the rate with which CTPR assembled in each of the two directions.

5.3 Material and methods

5.3.1 Materials

All chemicals were purchased from Sigma Aldrich and used as received unless otherwise noted. Before use, PEG-10, 000, and all salts were dried in a vacuum oven until no mass change was recorded. Tetramethylammonium chloride was produced with a 1.1x molar titration of tetramethylammonium hydroxide with hydrochloric acid.

5.3.2 Protein-protein interface characterization

All structural analysis on TPRs was carried out using UCSF Chimera, https://www.cgl.ucsf.edu/chimera/. CTPR18 models were made by looking at multiple unit cells using the tetragonal form of CTPR8 as the building block, PDB-ID: 2AVP. While the PDB2PQR server and ABPS were used to calculate the electrostatic and isopotential surfaces of CTPR molecules, the figures are false

colored to highlight acidic residues (red), basic residues (blue), polar uncharged (dark gray) and non-polar in (light gray). The primary sequences of both 2AVP and the CTPR18 used here are shown in the supplemental information section.

5.3.3 Preparation of CTPR18 nano-assemblies

CTPR18 was expressed in *E. coli* (BL21) as previously described.[14a] Briefly, the CTPR18 gene cloned into pPro-EX-HTam vector was transformed into E. coli (BL21) and a single colony was used to inoculate an overnight culture of fresh LB medium containing 100 μg/mL of ampicillin and then grown overnight at 37°C in a shaking incubator. 5 mL of overnight culture was used to inoculate per1 L of growth medium, again containing 100 μg/mL of ampicillin. When cultures reached an optical density of 0.6–0.8 at 600 nm, protein expression was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) (final concentration 1.0 mM). Cells were pelleted via centrifugation and re-suspended in lysis buffer (standard Ni-NTA purification methods). ~10 mg of lysozyme was added to each tube and allowed to sit on ice for 30 min. The cells were then lysed using an immersion sonicator tip. Finally, DNAse was added and allowed to sit on ice for 30 min. The cellular debris was separated from the supernatant via centrifugation. The resulting supernatant was purified using a Ni-NTA column to bind the genetically encoded His_6-tag on CTPR18. After elution, the protein was concentrated using a centrifugal filter (MW cutoff 10,000 Da) to an approximate concentration of 10-15 mg/mL (measure at 280 nm using UV/Vis spectroscopy) and subsequently purified using size exclusion chromatography (FPLC buffer, 150 mM sodium chloride and 50 mM sodium phosphate, pH 8.0) and collected using a fractionating collector. After FPLC, the proteins were concentrated again to ~13 mg/mL. This protein stock was dialyzed into 10 mM sodium chloride and 10 mM sodium phosphate at pH 7.2 using SnakeSkin dialysis tubing. Subsequent dialysis to various salt solutions was then carried out using SnakeSkin Dialysis tubing with three exchanges at 4 °C. Finally PEG-10k was added to a final concentration of 2 wt%. Samples were allowed to sit at 4°C for ~24 before prepping for the sample for electron microscopy.

5.3.4 Transmission electron microscopy

Samples for electron microscopy were prepared by spotting 2 uL of the dialyzed sample onto a carbon-coated copper TEM grid (Electron Microscopy Sciences, 400 mesh). Samples were allowed to sit undisturbed for 20 seconds, and were then washed twice by blotting the surface on a water droplet. For stained samples, the grid was blotted on a droplet of 3.75% uranyl formate (purchased from Electron Microscopy Sciences), blotted dry and blotted again on the uranyl formate solution for and additional 20 seconds. Grids were then washed 3 times using water droplets in the same manner as before, dried under vacuum and stored in ambient conditions until use. All images and single area electron diffraction patterns were captured on a JEOL 2100 equipped with a lanthanum hexaboride filament in Virginia Tech's ICTAS Nanoscale Characterization and Fabrication Laboratory.

5.3.5 Scanning electron microscopy, SEM

Images were taken using a Zeiss LEO 1550 SEM equipped with a field-emission source and an in-lens secondary electron detector in Virginia Tech's ICTAS Nanoscale Characterization and Fabrication Laboratory. Images were acquired using an accelerating voltage of 5 keV using a variety of magnifications and sample orientations. All mounted samples were coated with ~3 nm of carbon before imaging.

5.4 Results and Discussion

5.4.1 Self-assembled, two-dimensional arrays of CTPR18.

At dilute concentrations CTPR remains a monomer in solution. However, a solid formed from solution directly after dialysis of the concentrated, SEC-purified protein (150 mM sodium chloride and sodium phosphate, pH 7.2 to 10 mM sodium chloride and sodium phosphate buffer pH 7.2). This occurred in samples with concentrations of CTPR18 ranging from 150 to 200 μM as determined by UV-Vis spectroscopy. From dialysis tubes with solid present, we collected samples for transmission electron microscopy (TEM). In these samples, rectangular nanosheets termed 'nanotapes' regularly appeared, each with a consistent aspect ratio, Figure 5.1B. Samples collected immediately after dialysis had the highest population of nanotapes. It is likely that over time, CTPR

disassociates into its monomeric soluble form, similar to the observations of Sanchez-deAlcazar and coworkers when examining CTPR nanotube self-assembly.[17] Dialysis and sample storage at 4 ^{o}C prolonged the usable lifetime of samples by at least a few hours.

Much to our surprise, the nanotapes diffracted extremely well under an electron beam using a TEM in non-cryogenic conditions. Unsurprisingly, this diffraction was short lived as the samples would become amorphous with prolonged exposure to the electron beam. The collected images exhibit single-crystalline, single-area electron diffraction (SAED) patterns with defined spots rather than poly-crystalline rings or diffuse amorphous halos. Figure 5.1C shows an SAED pattern collected where two nanotapes overlapped. The inset image shows the area where the SAED pattern was collected. The two overlapping nanotapes are rotated with respect to each other at an angle, $\theta = 97^{o}$. This is consistent with the SAED image showing two identical patterns superimposed with the same rotation. These diffraction patterns exhibit Pmm2 symmetry with a two-fold symmetrical axis and perpendicular mirror planes (Figure 5.1C). The Pmm2 layer group is one of the 17 two-dimensional layer groups and one of seven with a rectangular shape crystal.[19] While the CTPR18 structure has not been directly solved using single-crystal x-ray diffraction, the Pmm2 symmetry exhibited by nanotapes correlates well with the crystal coordinates of other CTPRs. Both CTPR8 and CTPR20 exhibit tetragonal crystals, with a $P4_12_12$ point-group.[20] As such, the Pmm2 wallpaper group represents the two-fold axis of the $P4_12_12$ point-group, which is shown in Figure 5.1D with crystal planes and directions labeled with respect to the tetragonal crystal form along the [220] plane. In other words, longitudinal nanotape growth, HTT, occurs in the <001> direction whereas, latitudinal growth, STS occurs in the <110> direction. As previously mentioned, the nanotapes show a consistent morphology and aspect ratio. This aspect ratio is similar to that of CTPR18, which suggest that nanotape growth occurs at a similar rate in both perpendicular epitaxial directions <001> and <110> along the [220] plane, Figure 5.1D.

In order to systematically test the effect that alkali metal cations would have on nanotape growth we needed to increase the stability and yield of nanotapes. To increase yield, we added 2 wt% of 10,000 M_w poly(ethylene) glycol (PEG). PEG acts as a depletant or molecular crowding agent and

aids in assembly by increasing the osmotic pressure on CTPR18 molecules, pushing them closer together (Figure S5.1). This leads to an overlap in the CTPR18 excluded volumes, and increases the entropy of mixing for the smaller PEG by increasing the volume it has to explore. This augments the already established rigid-rod-like assembly exhibited by CTPR18 in solvent cast films.[14a] Briefly, rigid-rod molecules, like CTPR18, tend to align, orienting their mass moments along a common axis. This further minimizes the void volume or the space occupied by the protein superhelix. This molecular crowding phenomenon is well established in the literature with discussion on protein systems mainly from the perspective of solubility vs. aggregation.[21] Nanotapes produced with PEG-10,000 were present in higher quantities and significantly stabilized, being able to persist for multiple days (Figure S5.2). Further study is needed to determine the long-term stability of these materials.

5.4.2 Monovalent salts affect epitaxial growth

Within a protein there are many different types of ligands that can interact with ions in solution. In the case of alkali metal cations, the complexing properties of N-donor ligands, such as amine side-chains (e.g. Lys), are very weak and neglected from further discussion. Rather, we target carboxylate and aromatic side-chains by systematically varying alkali cations. Figure 5.2 shows the HTT (A) and STS (B) interfaces, with the sequences shown in Figure S5.7. Mainly aromatic residues reside at the HTT interface whereas the majority of the STS interface comprises charged acidic residues, with one basic side-chain present.

To determine the effect that monovalent salts would have on epitaxial growth of CTPR arrays we tested the chloride salts of the following cations: $N^{+}(CH_3)_4$ > Cs^{+} > Na^{+} > Li^{+}, listed in order of ionic radius. We additionally tested the bromide salts of lithium and sodium but the samples in each case precipitated with very few to no higher-order structures observed. Thus, these salts will not be addressed further. Concentrations of the salts were selected to compliment the number of carboxylate side chains in CTPR18. 0.02M solutions provide an approximately 1:1 ratio of salt to acidic residues whereas 0.20M solutions were made to provide a 10-fold excess of salt to acidic residues. Figure 5.3 shows the TEM images of the characteristic morphologies that were

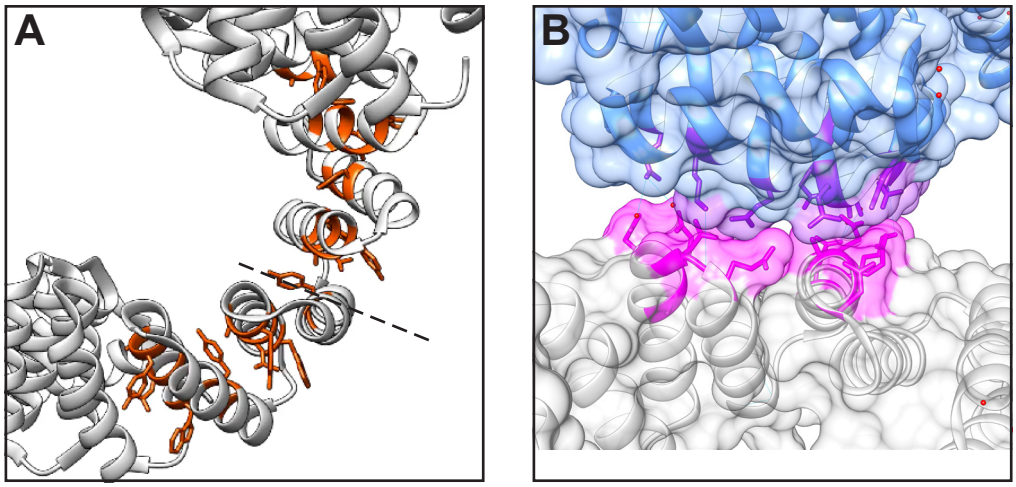


Figure 5.2 - A) Illustration showing the head-to-tail hydrophobic stacking interface of two CTPR18 molecules. The indistinguishable side-chain positions illustrate how the inter-moleculer interactions are the same as the intra-repeat interactions that drive folding. Highlighted are the 8 conserved residues in the terminal 2-repeats of each superhelix. The dashed line cuts through this interface. B) Side-to-side packing is aided by electrostatic interactions between neighboring CTPR18s. Residues at the interface are shown and colored purple and magenta to help differentiate between the anti-parallel CTPR18s. This interface involves 4-repeats from each protein with two hydrogen bonds forming between K26 and G15 of the 2nd and 3rd repeat respectively of each CTPR. These models were adapted symmetrically from PDB: 2AVP.

observed. Table S5.1 shows a representative picture of the majority population seen in each of the tested samples.

5.4.2.1 Nanofootballs

In samples with 0.02M tetramethylammonium-Cl ($N^{+}(CH_3)_4$ or TMA^{+}) a new morphology, oblate discs, or so called 'nanofootball' appeared, Figures 5.3 & S3. It is possible that the oblate disc shape is an intermediate to the rectangular nanotape. The appearance of the oblate intermediate can be explained by faster STS growth in the <110> direction when compared to the HTT growth in the <001> direction. These assemblies also stack together like nanotapes, but rather than stacking as a deck of cards, they produced flower-like assemblies (Figure S5.3).

Differences between these two morphologies arise with closer examination of the boundaries, suggesting that the nanofootball is not simply an intermediate assembly. TEM images suggest that CTPR18 nanotapes assemble through a non-classical pathway, Figure 5.4A-B. A non-classical

crystallization process involves a multi-step assembly that progresses through monomer-oligomer and oligomer-crystal. A classical process on the other hand involves monomer deposition directly on the growing crystal face.[12] Intuitively, one can visualize a non-classical pathway for nanotape assembly because the crystallization process combines multiple steps: HTT super-helix elongation and STS lateral packing. Figure 5.4 A-B shows step-like inclinations on the short edges of nanotapes. This is consistent with a non-classic crystallization process where oligomers, in this case, extended HTT superhelical stacks, assemble first and add to the growing crystal face. Comparatively, Figure 5.4C-E shows the assembly process aided by TMA^+ fits more with monomer-by-monomer addition

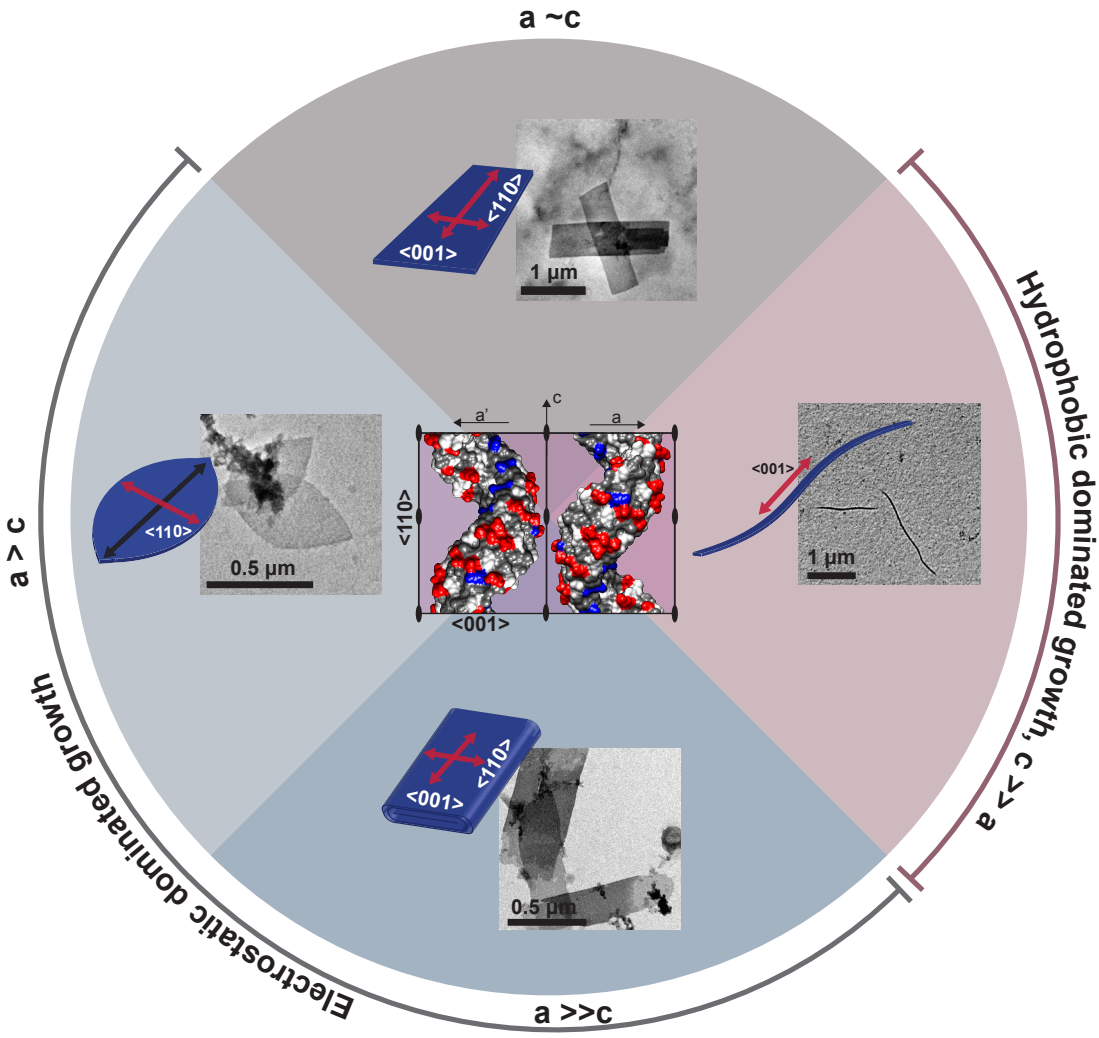


Figure 5.3 - Schematic illustration showing the different morphologies assembled by CTPR18 in different salt solutions. The red arrow indicates the preferred growth axis.

to the growing crystal surface. The smooth and uniform shape seen in Figure 5.4C suggests a crystal growing from a single nucleus. Figure 5.4D-E show what we interpret as protein monomer surrounding the growing [220] plane. This supposition is further supported by the high-resolution image in panel E where crystalline materials can be observed with a d-spacing corresponding to the inter-helical pitch of CTPR.

Mechanistically the difference between these two cations, Na^+ and TMA^+, arises from their ability to bind and shield hydrophobic regions, i.e. affecting growth on the HTT interface. Vlachy and coworkers found that TMA^+ is quite good at salting-in hydrophobic molecules.[22] They suggest

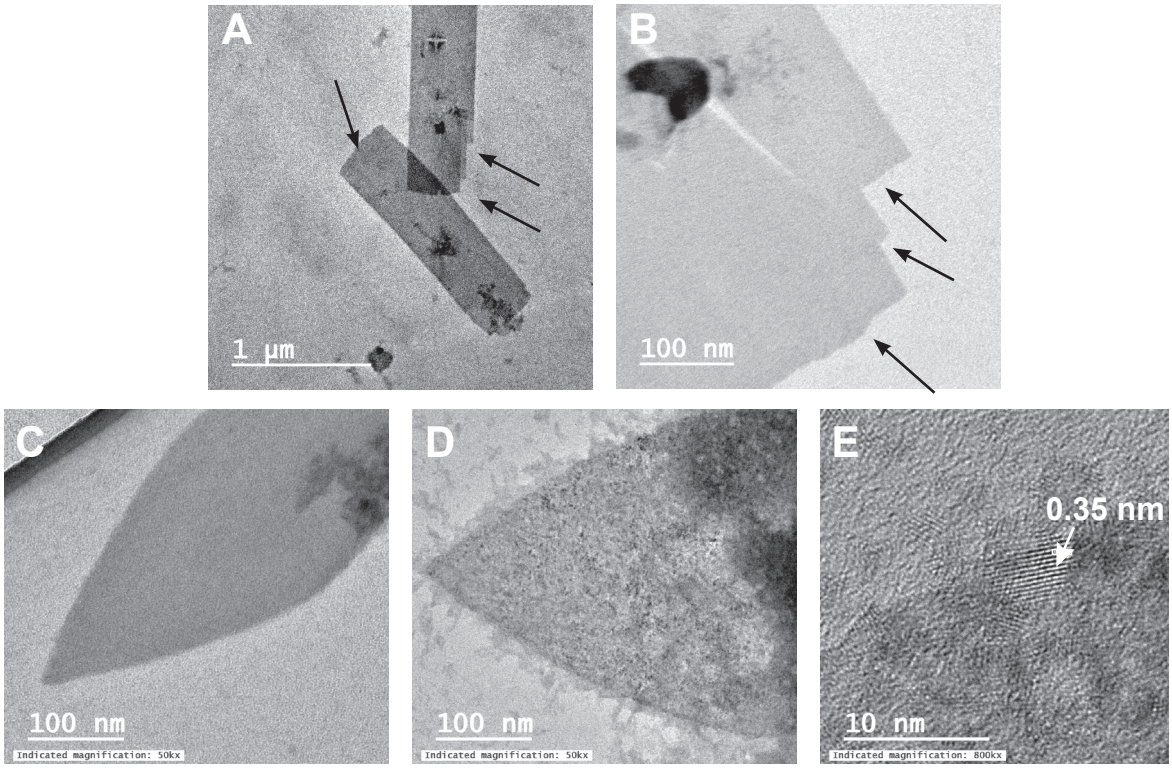


Figure 5.4 - A-B) Disclinations on the edge of CTPR18 nanotapes suggest a non-classical crystallization method where crystal growth occurs through the attachment of oligomeric subunits. The disclinations here occur from HTT stacks that are offset. C-E) The assembly process seen with TMA-Cl that yields nanofootballs is more characteristic of a classical, monomer-by-monomer assembly process. Here the crystal grows from a central nucleation site and add similar to bricks laid by a mason. D) TEM image of a snapshot in the assembly process. The monomer (CTPR18) is adding to all edges simultaneously. E) Stained image show crystalline striations

a 'receptacle model' where water molecules highly orient in the first water shell of TMA^+. This dipolar alignment orients water in the second shell, essentially making cavities for hydrophobic molecules to sit in.[22] We propose the TMA^+ is able to populate the HTT interface at the terminal CTPR repeats and temporarily slow HTT stacking. As the crystallite nucleates each CTPR18 added to the growing face satisfies both the electrostatic and hydrophobic interactions with its neighbors and displaces the TMA^+.

While there is quantitative data on the gas phase interactions between alkali cations and amino acids, peptides and other similar ligands, studies in aqueous environments are sparse.[23] In the gas phase, the strength of alkali cation interactions with carboxylates varies inversely with ionic radius. For free amino acids the bond-dissociation energies (BDE) decrease down the periodic table: $Li^+ > Na^+ > K^+ > Rb^+ > Cs^+$. The ionic radii of these cations are respectively: 0.70, 0.98, 1.33, 1.49, and 1.69 Å.[24] The inverse relationship between BDE and ionic radius is caused by the increase in the ion-amino acid bond distance, which reduces the overall electrostatic interactions.

This trend also holds true for cation-pi interactions in the gas phase. However, the trend in aqueous solution is essentially flipped. In aqueous solutions, this trend reverses where $K^+ > Rb^+ >> Na^+ > Li^+$, resulting from the high energetic penalty paid by strongly solvated ions when interacting with aromatic molecules.[25] To our knowledge, there is no published data for TMA^+ in aqueous environments. However the binding energy of the closely related NH_4^+ to Trp is 3.1 kcal mol^{-1} with that of ethyl- and methylammonium (NH_3R^+) ions reportedly higher.[26] These binding energies are similar to the calculated binding energies of the CTPR18 HTT interface discussed below and may thus significantly affect the head-to-tail stacking of CTPRs.

The CTPR HTT inter-molecular interface is essentially identical to the intra-molecular repeat-repeat interface and is driven by the burial of hydrophobic residues at the interface. Mejias *et al.* quantified this HTT interaction using two methods.[15a] First using the 1D-Ising model, which has been used to estimate the thermodynamic stability of repeat proteins.[27] The second method was an estimation based on the transfer energy scales associated with moving each of the interfacial side-chains from hydrophobic to polar media.[28] Both methods were in close agreement

with a binding energies averaging to, -5.5 kcal mol^{-1}. These are in the same order of magnitude as the aqueous interactions measured between aromatic side-chains and ammonium cations. While this is an interesting observation, it is hard to quantitatively compare these data because the solvent dependent estimation of the HTT interface is in pure water, and doesn't account for ionic interactions.

5.4.2.2 Low-aspect ratio tubes

The Cs^+ sample produced self-assembled structures that appear to be short aspect-ratio tubes as can be seen in Figure 5.3 and S4. The structures form tubes with STS growth dominating in the <001> direction. Empirically, it is difficult to make direct comparisons between Cs^+ and the other ionic compounds tested in this report. Cs^+ is the most conspicuous omission from studies regarding aqueous interactions. While Cs^+ is the least hydrated of the alkali metal ions, our observations would suggest that it completely blocks the HTT stacking, even more so than TMA^+, suggesting that it may be even less hydrated.

5.4.2.2 Other shapes

Lastly, at high-salt (0.2M) concentrations with both Cs^+ and Na^+ the same nanowire structure forms. It should be noted that there is reasonable amount of aggregation in these samples. However a significant number of the CTPR18 population is able to assemble into nanowires likely as a result of ionic concentration dependent charge shielding of the STS electrostatic interactions in the <110> direction (Figure 5.3 and S5).

Of the salts tested, not all produced stable assemblies. 0.02M lithium chloride mainly aggregated CTPR18 with occasional misshapen sheet-like structure (Figure S5.6 and S7). This may be attributed to the high level of hydration of Li^+. While highly-hydrated cations bind more tightly to carboxylates, lithium can also pull water from the protein and destabilize its structure.[18b] The 0.2M sample of Li^+ completely precipitated the protein. In this sample all CTPR18 fell out of solution into a solid pellet at the bottom.

5.4 Conclusions

We have shown that CTPR18 readily self-assembles into two-dimensional anisotropic nanotapes from phosphate/saline solution. By adding the depletant, PEG-10,000, we were able to increase the yield as well as the stability of the nanotapes increasing their shelf-life from a few hours to a few days. These self-assembled structures are single-crystalline assemblies with two-fold symmetry (Pmm2), akin to the two-fold axis seen in CTPR8 and CTPR20 tetragonal crystals. Among the effectors on this self-assembly process, there are two complementary epitaxial growth mechanisms. The first, relies on hydrophobic head-to-tail inter-molecular interactions, while the second interface directs side-to-side packing between neighboring molecules through electrostatic interactions. These mechanisms correlate to epitaxial growth on the [220] plane in the <001> and <110> directions respectively. Through systematic variation of the monovalent chloride salts of Li^+, Na^+, Cs^+ and TMA^+, we have shown the ability to alter the rate of epitaxial growth in the two directions. This is seen in the resulting shapes of the new self-assembled structures. Nanotapes produced with Na^+ serve as the relative comparison, where the aspect ratio is similar to the building block, CTP18. This represents relatively equivalent growth on the <001> and <110> directions. TMA^+ and Cs^+ salts were able to produce electrostatically driven morphologies, oblate discs or 'footballs' and low aspect ratio tubes respectively. Li^+ gave little to no control over morphology and produced spherical aggregates, albeit of the same size. We also found that morphology was independent of salt identity at high salt concentrations (0.2 M). In all conditions where self-assembly was observed, high salt promoted hydrophobic dominated growth in the form of fibers. This is likely due to shielding of electrostatic interactions by the excess of salt. Together these data not only present a new modular repeat-protein platform for the design of functional nanomaterials, but also illustrates the general utility that that monovalent salts may provide in tuning protein self-assembled networks.

5.5 Acknowledgements

The authors would like to thank the ICTAS Nanoscale Characterization and Fabrication Laboratory

(NCFL) for the use of TEM, and SEM. This work was in part supported by ICTAS and their JFC grant program.

5.6 References

[1] J. M. Alonso, M. L. Gorzny and A. M. Bittner, *Trends Biotechnol*. **2013**, *31*, 530.
[2] W. R. Wikoff, L. Liljas, R. L. Duda, H. Tsuruta, R. W. Hendrix and J. E. Johnson, *Science*. **2000**, *289*, 2129.
[3] a) H. Dietz, S. M. Douglas and W. M. Shih, *Science*. **2009**, *325*, 725; b) F. Hong, F. Zhang, Y. Liu and H. Yan, *Chemical reviews*. **2017**, *117*, 12584.
[4] a) Y. M. Abul-Haija, G. G. Scott, J. K. Sahoo, T. Tuttle and R. V. Ulijn, *Chem Commun (Camb)*. **2017**, *53*, 9562; b) I. P. Moreira, T. K. Piskorz, J. H. van Esch, T. Tuttle and R. V. Ulijn, *Langmuir*. **2017**, *33*, 4986.
[5] a) J. M. Fletcher, R. L. Harniman, F. R. Barnes, A. L. Boyle, A. Collins, J. Mantell, T. H. Sharp, M. Antognozzi, P. J. Booth, N. Linden, M. J. Miles, R. B. Sessions, P. Verkade and D. N. Woolfson, *Science*. **2013**, *340*, 595; b) A. L. Boyle and D. N. Woolfson, *Chem Soc Rev*. **2011**, *40*, 4295.
[6] T. P. J. Knowles, T. W. Oppenheim, A. K. Buell, D. Y. Chirgadze and M. E. Welland, *Nat. Nanotechnol.* **2010**, *5*, 204.
[7] J. D. Brodin, X. I. Ambroggio, C. Tang, K. N. Parent, T. S. Baker and F. A. Tezcan, *Nat Chem*. **2012**, *4*, 375.
[8] a) Y. Suzuki, G. Cardone, D. Restrepo, P. D. Zavattieri, T. S. Baker and F. A. Tezcan, *Nature*. **2016**, *533*, 369; b) S. H. Mejias, B. Sot, R. Guantes and A. L. Cortajarena, *Nanoscale*. **2014**, *6*, 10982.
[9] P. Ringler and G. E. Schulz, *Science*. **2003**, *302*, 106.
[10] S. Gonen, F. DiMaio, T. Gonen and D. Baker, *Science*. **2015**, *348*, 1365.
[11] J. D. Brodin, S. J. Smith, J. R. Carr and F. A. Tezcan, *J Am Chem Soc*. **2015**, *137*, 10468.
[12] K. N. Olafson, R. Li, B. G. Alamani and J. D. Rimer, *Chem Mater*. **2016**, *28*, 8453.
[13] E. R. Main, Y. Xiong, M. J. Cocco, L. D'Andrea and L. Regan, *Structure*. **2003**, *11*, 497.
[14] a) N. A. Carter and T. Z. Grove, *Biomacromolecules*. **2015**, *16*, 706; b) T. Z. Grove, L. Regan and A. L. Cortajarena, *J. R. Soc. Interface*. **2013**, *10*.
[15] a) S. H. Mejias, B. Sot, R. Guantes and A. L. Cortajarena, *Nanoscale*. **2014**, *6*, 10982; b) J. J. Phillips, C. Millership and E. R. Main, *Angew. Chem. Int. Ed.* **2012**, *51*, 13132.
[16] S. H. Mejias, P. Couleaud, S. Casado, D. Granados, M. A. Garcia, J. M. Abad and A. L. Cortajarena, *Colloids and Surfaces B: Biointerfaces*. **2016**, *141*, 93.
[17] D. Sanchez-deAlcazar, S. H. Mejias, K. Erazo, B. Sot and A. L. Cortajarena, *Journal of structural biology*. **2017**, 10.1016/j.jsb.2017.09.002.
[18] a) Y. J. Zhang and P. S. Cremer, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15249; b) J. Kherb, S. C. Flores and P. S. Cremer, *J. Phys. Chem. B*. **2012**, *116*, 7389.
[19] B. L. Nannenga, M. G. Iadanza, B. S. Vollmar and T. Gonen, *Curr Protoc Protein Sci*. **2013**, *Chapter 17*, Unit17 15.
[20] T. Kajander, A. L. Cortajarena, S. Mochrie and L. Regan, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *63*, 800.
[21] a) T. Li, X. J. Zan, Y. Sun, X. B. Zuo, X. D. Li, A. Senesi, R. E. Winans, Q. Wang and B. Lee, *Langmuir*. **2013**, *29*, 12777; b) M. Boncina, J. Rescic and V. Vlachy, *Biophys. J.* **2008**, *95*, 1285.
[22] B. Hribar-Lee, K. A. Dill and V. Vlachy, *The journal of physical chemistry. B*. **2010**, *114*,

15085.
[23] A. S. Reddy and G. N. Sastry, *J. Phys. Chem. A*. **2005**, *109*, 8893.
[24] F. Crea, C. De Stefano, C. Foti, G. Lando, D. Milea, S. Sammartano. *The Alkali Metal Ions: Their Role for Life. Metal Ions in Life Sciences*, Springer, *16*, **2016**, 133.
[25] R. A. Kumpf and D. A. Dougherty, *Science*. **1993**, *261*, 1708.
[26] Q. Lu, D. X. Oh, Y. Lee, Y. Jho, D. S. Hwang and H. Zeng, *Angew Chem Int Ed Engl*. **2013**, *52*, 3944.
[27] a) A. L. Cortajarena and L. Regan, *Protein Sci.* **2011**, *20*, 336; b) C. Millership, J. J. Phillips and E. R. Main, *J Mol Biol*. **2016**, *428*, 1804.
[28] A. Radzicka and R. Wolfenden, *Biochemistry*. **1988**, *27*, 1664.

5.8 Supplemental information

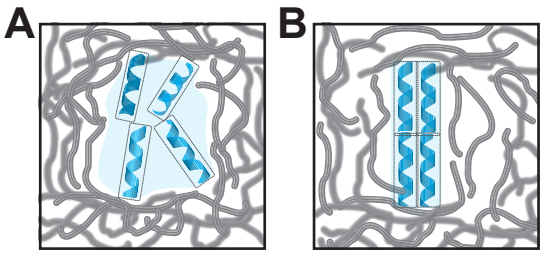


Figure S5.1 - Illustration of the depletion interaction by PEG (gray lines) on CTPR18 (coils). A) Disperse CTPR18 takes up volume in its pre-assembled state. B) PEG exerts pressure on CTPR18 promoting alignment of the superhelical axes. This minimizes the volume taken up by CTPR and allows the PEG to more freely move.

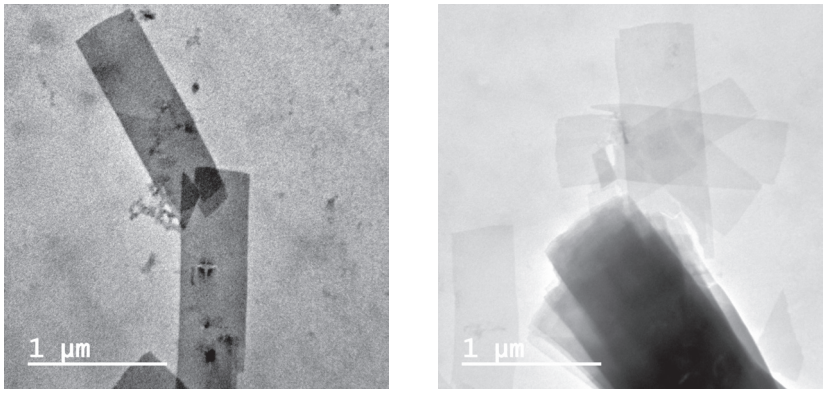


Figure S5.2 - TEM images of nanotapes exhibit the same rectangular morphology, with a consistent aspect ratio that is close to that of CTPR18 .

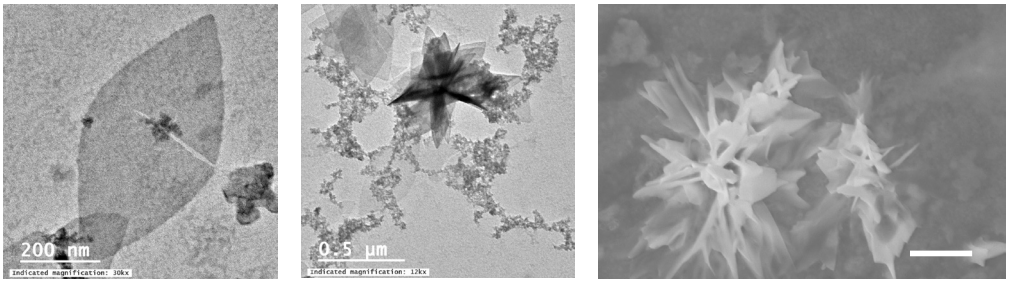


Figure S5.3 - TEM images of oblate discs termed 'nanofootballs' that are assembled in $N^{+}(CH_3)_3$. Far right image is an SEM of a flower-like assembly of nanofootballs similar to the middle image. Scale bar is 1 micrometer.

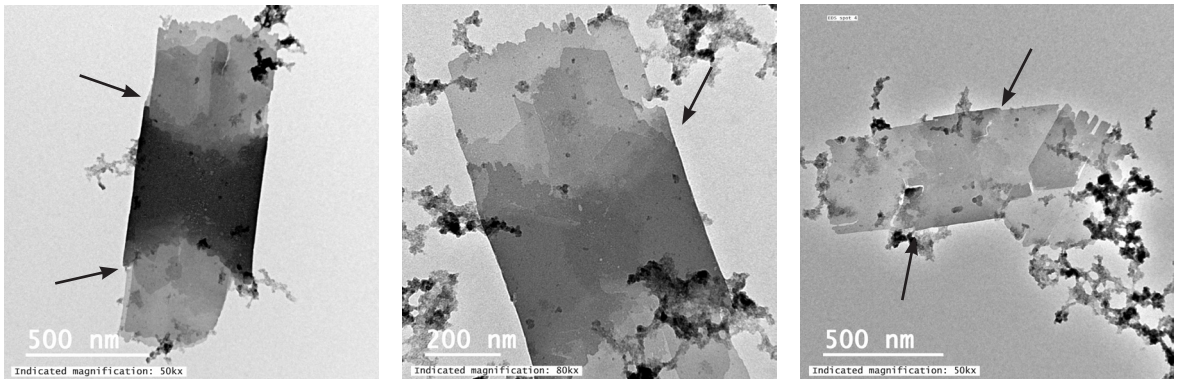


Figure S5.4 - TEM images of short aspect tubes. Arrows point to the curvature.

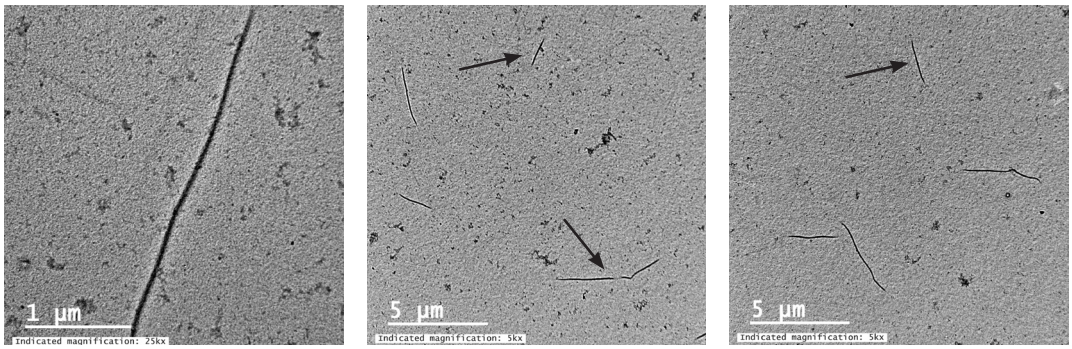


Figure S5.5 - TEM images of fibers assembled at high salt concentrations. Arrows point to some of the nanowires

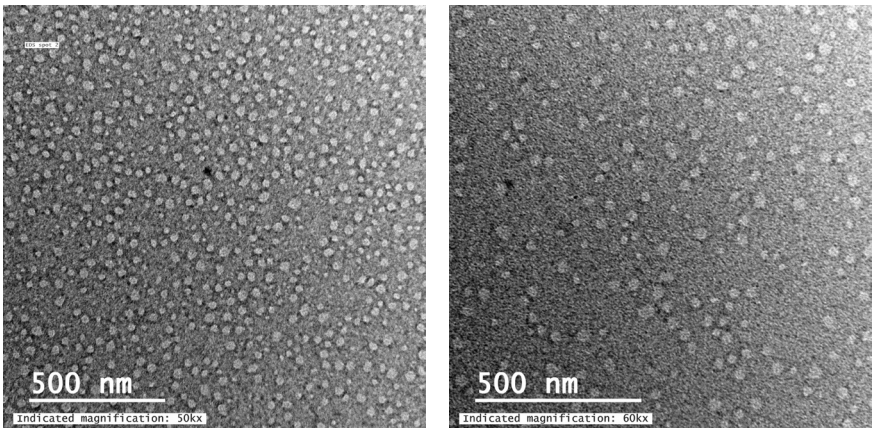


Figure S5.6 - Representative TEM image of aggregation seen in Li samples.

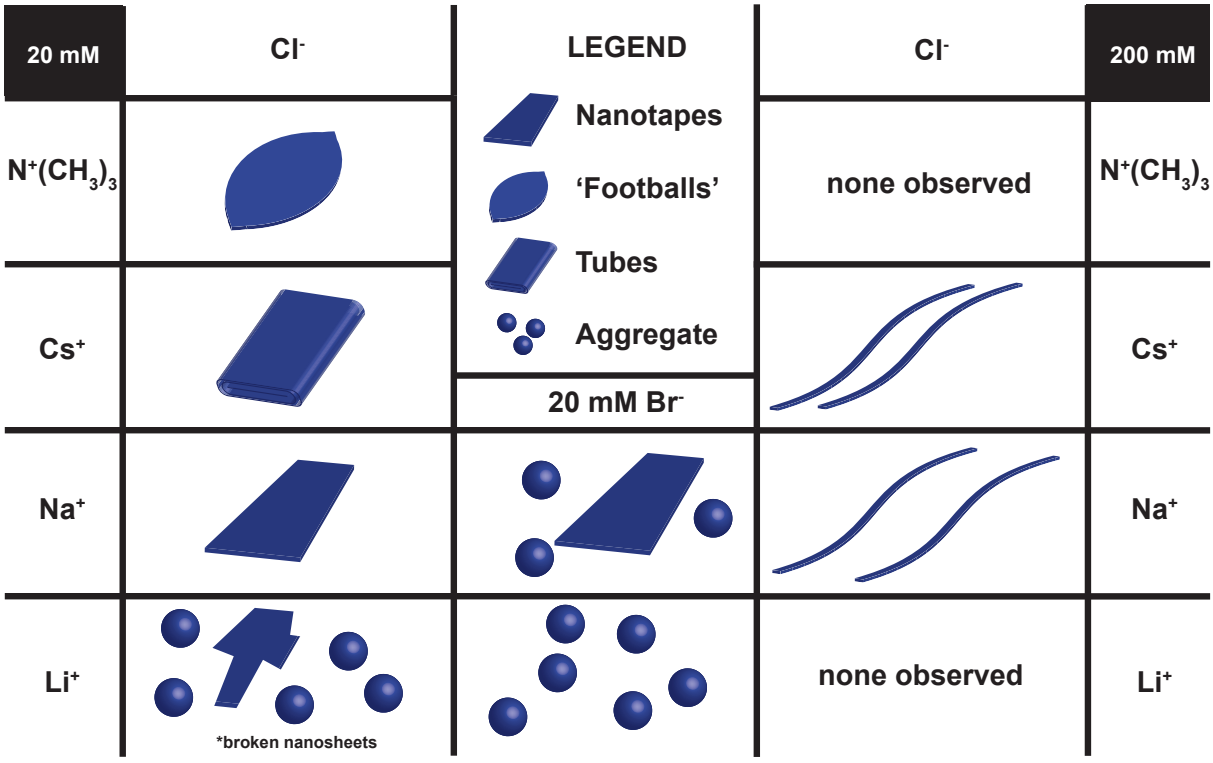


Table S5.1 - Illustrations of the nano-asssembly populations seen in each sample at 20 mM (left) and 200 mM (right). Bromide salts are only tested at 20 mM. CsBr and Tetramethylammonium-Br were not examined.

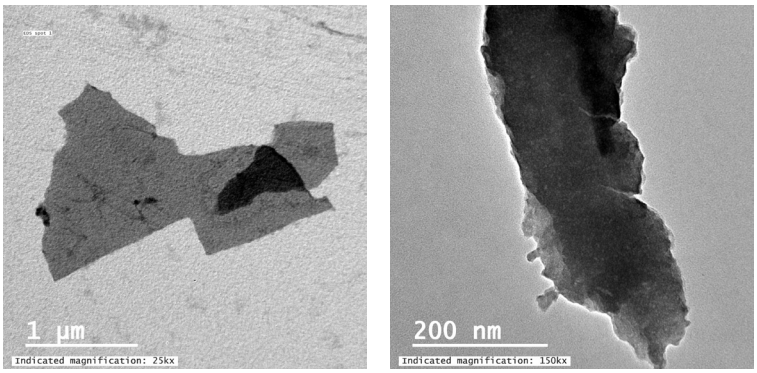


Figure S5.7 - Representative TEM image of misshapen and broken TPR assemblies seen in sample with high aggregation

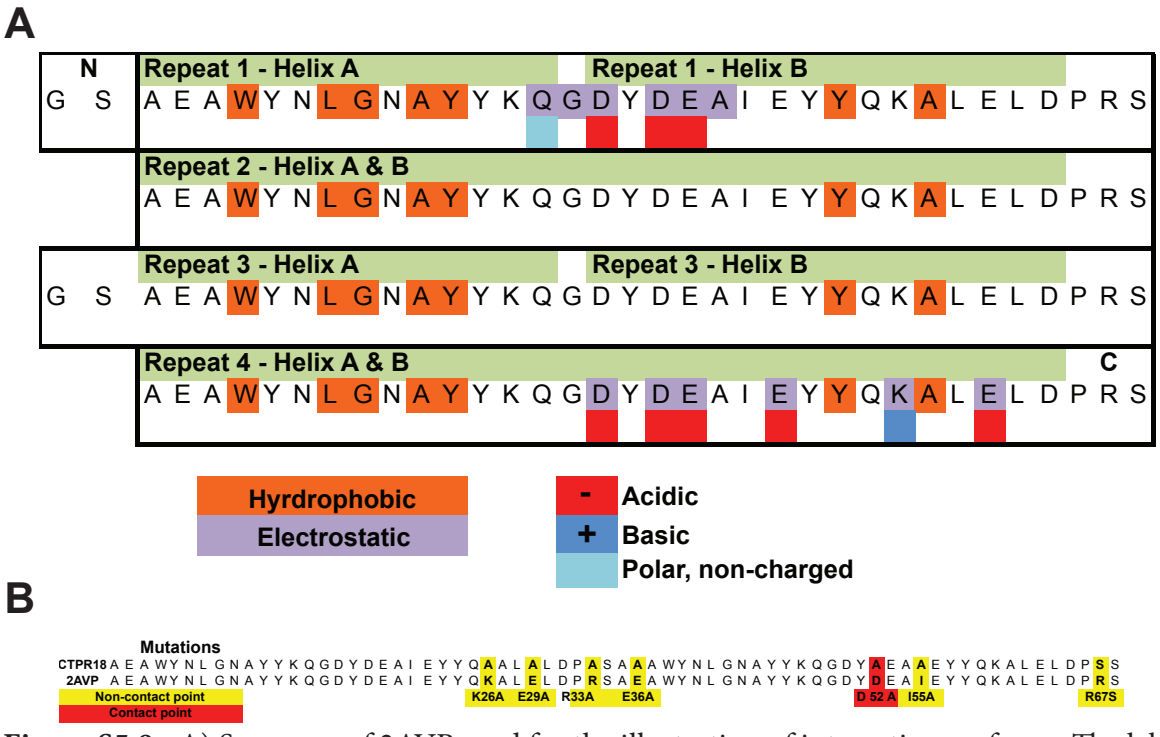


Figure S5.8 - A) Sequence of 2AVP used for the illustration of interacting surfaces. The label, Repeat 1-4 refers the 4-repeats that are involved in the STS packing, not a CTPR4 molecule. Purple highlighted residues sit at the STS interface. The color code below denotes the side-chain chemistry. Orange highlighted residues represent the 8 conserved positions that are important for intra-helix folding and inter-molecular HTT stacking. B) Comparison between CTPR18 sequence and 2AVP show 7 mutations, yet only one sits on an interface.

Chapter 6. Overall Conclusions

The motivation for the diverse work comprising this thesis was not to simply exhaust the possible uses of a single protein but was more of an exploratory examination of the utility and practicality of protein materials in areas traditionally dominated by synthetic polymers. It highlights how the complex structure-property relationships inherent to proteins can add interesting and beneficial properties to materials that are hard to achieve with synthetic materials alone. The caveat to this is that the structural complexity inherent to many proteins limits the ability to rationally modify their structures because of fundamental limitations in our understanding of protein folding. As such, specific interest in this work lies with the sub-class, repeat proteins.

Biologically these proteins are involved in cell signal pathways, protein-protein interactions as well as binding small molecules. These small repeating sequences form structurally independent units that fit together like beads on a string and to analogize, are similar to the repeating monomer unit of a polymer. The resulting assembly of these structurally independent monomer units gives rise to the same fold and relative position, no matter where they are placed in the sequence. This unique structure presents these proteins as prime candidates for bioengineering, both for fundamental studies of protein folding and design as well as well as applied materials with modular structure which are tunable for a variety of applications. This work specifically focuses on the consensus tetratricopeptide repeat protein (CPTR18). TPRs consists of a 34 amino acid helix-turn-helix motif.[1] The consensus TPR (CTPR) sequence derived from the statistical analysis of natural TPR proteins.[2] CTPR is especially amenable to engineering because only 8 of 34 amino acids are conserved. The residues are involved in the intra- and inter-repeat interactions give CTPR its characteristic superhelical fold. It has since been well studied and engineered to maximize stability (with a T_m or 105 ^oC). These successful design efforts as well as the utility shown in this thesis are a testament to possible advantages that consensus design presents in furthering the applicability of designer proteins for functional materials.

It has been well established in the literature that structure begets function in materials. To better understand the structure-property relationships within hierarchical protein assemblies,

more specifically birefringent CTPR18 films, we first explored the effect that hierarchical morphology had on the mechanics of these materials in Chapter 3. CTPR18 films exhibited anisotropic mechanical properties from the nano- to macro-scale. It was determined that these directional mechanical properties resulted from a multi-scale assembly, ultimately forming a highly ordered lamellar phase. This phase had no distinguishable orientational variation using polarized light optical microscopy, an uncommon occurrence in macro-scale films These morphological features provided different mechanisms for mechanical relaxation and failure and exhibited markedly different properties when tested in both tension and compression on each of the Cartesian axes.

Alone, anisotropic mechanical properties may be interesting, but may not be immediately useful. However, hygroscopic materials, such as proteins, generally exhibit swelling when exposed to water. In Chapter 4, we leveraged this swelling to explore the utility of CTPR18 films as hygromorphic actuators for possible applications in soft robotics. The highly oriented morphology provides an avenue for controlling the actuation stroke while morphology thickness affects the kinematics. We communicated a simple strategy for tuning the morphology and the resulting actuation of these films. Changing the humidity in the casting chamber showed a marked effect on the ability to tune the thickness of structural gradients in the material by simply altering the kinetics of assembly with evaporation speed. The lamellar gradients were useful in tuning the actuation stroke and speed of the hygroresponsive actuators. Bending and twisting was selected for by changing the lamellar orientation, perpendicular and diagonal respectively, within cut films. While films with different structural gradients showed divergent actuation speed, strength and RH sensitivity. Lastly we showed that the highly charged character of this protein (pI = 3.8) provides a second property for these actuators, electro-responsiveness. Electro-responsiveness of these films is driven by the inherent proton conductivity of these materials. Excitingly we report one of the highest proton conductivities to date for a biological material, $7.1 \times 10^{-2} \pm 0.0003$ S cm^{-1}. Consequently, directional bending can be induced by applied electrical potential, thus swelling as a result of moving counterions to their respective electrodes. Complementary to this, CTPR18 act as mechanical-batteries, where the direction and frequency of applied bending force drives the capacitive build-up and

discharge of potential on the films with a peak voltage of 1.15V.

In Chapter 5, we took a step back to study the self-assembly process of CTPR18 nanomaterials in solution. We discovered that this protein assembles into two-dimensional, single-crystal, anisotropic sheets. We determined that the growth mechanism matched interactions seen in the single crystal forms of CTPR with hydrophobic head-to-tail growth (HTT) and electrostatic side-to-side stacking (STS). These two different mechanisms provided an opportunity to explore the effect that monovalent salts would have on protein assembly. By systematically varying monovalent salts selected for ionic radius, carboxylate binding energy and aromatic binding energy we were able to rationally change the epitaxial growth rates of HTT:STS. Faster growth on the STS crystal plane gave rise to oblate circles, or footballs, as well as low aspect-ratio tubes, while faster HTT growth yielded nanowires. This is the first example showing the ability to tune the shape and morphology of non-cross-linked, 2D self-assembled proteins without genetically engineering or computationally modeling a new protein interface.. This work is especially exciting because it provides alternate strategies to modify the 2D assemblies of proteins that are not amenable to genetic engineering. This is also immediately applicable to many repeat-proteins that exhibit similar, linear extended structures and has generated significant interest at multiple conferences.

Looking forward, this work lays the foundation for projects that can go a number of ways. Regardless of direction, production scale remains to be a primary concern for protein materials as a whole. On the lab scale, it is hard to match the reaction yields of synthetic polymers. On average, E. coli (BL21) produces CTPR18 at a concentration of ~60 mg per liter of culture. While somewhat scalable using lab-scale fermenters, the limiting step in regards to efficiency remains purification. Metal-affinity chromatography provides a facile route to purify recombinant proteins. However yields are at best ~85% pure CTPR18 protein (estimated from stained polyacrylamide gels), the affinity chromotography matrix is cost-prohibitive and has a limited lifetime. Additionally with the low purity, subsequent size-exclusion columns is necessary to remove impurities. This involved purification strategy means that the production time scales with mass of cells grown.

Alternatively a solution for producing CTPR18 at scale may have serendipitously resulted

from the exhaustive consensus design of CTPR18. Consequently, the highly designed sequence is quite stable with an unfolding temperature (T_m) ~ 105 °C. Borrowing from purification strategies for thermostable enzymes, simply heating the cell lysate above 60 °C will precipitate most of the cellular proteins, leaving CTPR18 folded in solution. Preliminary evidence (data not shown) illustrates that CTPR18 can be purified by heating to 75 °C for as little as 15 min. These early data have are promising because we have seen ~75% purity from simply heating. Its likely that the addition of small molecule osmolytes like PEG-1000 or sucrose will increase purity. Further development will allow for purification of CTPR18 on gram scale, thus alleviating many material supply and timing issues with producing large scale CTPR18 films.

From a protein engineering perspective, many projects will benefit from exploiting the modular nature of CTPR. Specific mutations may be also useful in creating tunable hydrogel materials based on CTPR proteins. With only 8 of 34 amino acids conserved, there is significant room for engineering sites for chemical cross-linking. The easiest would be incorporating Cys into the primary sequence, much like Mejias *et al.* did when making nanowires.[3] An alternative to this would be to incorporate sites for chemical cross-linking using non-canonical amino acids analogues. Consequently, we have genes that would work for both, with mutations at the 26th position of a CTPR2.

Additionally mutation strategies may prove useful in tuning the conductivity of CTPR18 films. Towards this goal, a systematic study of the density, position and residue specific effects on charge carrying amino-acid side-chains in the CTPR superhelix would be useful. The main goal being to optimize the conductivity of CTPR18 films. Preliminary evidence indicates that the conductivity in CTPR18 films in anisotropic, with higher conductivities through the thickness of the film when compared to in-plane measurements. First, more reliable data is needed with measurements taken using the same electrode material with testing cells of similar dimensions. Platinum or gold electrodes are optimal, yet somewhat cost-prohibitive. With such a system, the mutations of interest for CTPR18 would be: 1) mutate all acidic residues to compare Glu vs Asp, 2) mutate out charge carrying residues from the internal surface of the superhelix to test through-plane conduc-

tivity mechanism 3) create a highly positively charged mutant incorporating Lys with high frequency. With careful design, internal modifications to CTPR can yield the proper inter-chain spacings for a proton wire.[4] Additionally of interest in these materials would be to test the conductivity of a variety of ionic liquids. The proposed effect would be multi-fold: 1) conductivity would be water independent, 2) removing water and using a poor solvent ionic liquid may increase the stability of film morphology, 3) additionally this may increase the thermal stability of protein monomers as well, 4) removal of water opens up applications with water reactive metals.

In addition to the possible benefits of removing water for energy applications, replacing water as the charge carrier may aid in actuator performance. For actuation, a systematic study incorporating ionic liquids of varying size may provide: 1) increased stability by minimizing swelling with a good solvent (e.g. water) and 2) enhancement of the ionic induced motion by increasing the size disparity between counterions present in the films.

An additional sets of experiments that would be interesting for the CTPR18 hygromorphic actuators, but may even be more broadly applicable, would be based around the design of a sensor array to monitor humidity at multiple positions in real-time. By collecting this data we could more accurately model the humidity gradient produced around the film. This would more accurately correlate actuator motion, especially the sporadic strokes, to determine if they are a entirely result of morphological features or local variations in the humidity. This would provide a much more quantitative look that is not currently present in the literature.

Finally, there are many exciting opportunities to build upon the two-dimensional assemblies of TPRs. Here, yield and long-term stability are the initial areas to explore. To reach these goals a better fundamental understanding of the assembly process is needed. Using light-scattering, either dynamic light scattering or using Mie scattering, to monitor to appearance or growth of nanomaterials is a much cheaper and quicker route to determine yield when compared with electron microscopy. Additionally these techniques may help in determining the kinetic parameters for further mechanistic studies. Additionally, a systematic study of assembly parameters should be carried out with CTPRs of varying lengths. These experiments would provide set of design rules

for other structurally similar materials that the Grove lab is developing for bio-sensing applications. As a proof-of-concept, nanosheets could be patterened onto an EVD gold electrode array. This configuration would allow CTPR18 (the version with CTPR3-90 incorporated) to operate as 'biosensor' for DESVD functionalized molecules.

In the end, there are innumerable directions this project can be taken. The experiments above, the work currently progressing in the Grove lab and other developments such as the exciting and recent computational work focusing on *de novo* design of protein and protein materials will ensure this is a very exciting field for the future.

References

[1] L. D. D'andrea and L. Regan, *Trends Biochem. Sci.* **2003**, 28, 655.

[2] E. R. Main, S. E. Jackson and L. Regan, *Current opinion in structural biology.* **2003**, 13, 482.

[3] S. H. Mejias, B. Sot, R. Guantes and A. L. Cortajarena, *Nanoscale*. **2014**, *6*, 10982.

[4] R. Pomes and B. Roux, *Biophys J.* **1996**, 71, 19.